Current Research in Systematic Musicology

Alexander Refsum Jensenius *Editor*

Sonic Design

Explorations Between Art and Science

OPEN ACCESS

Springer

Current Research in Systematic Musicology

12

Series Editors

The series covers recent research, hot topics, and trends in Systematic Musicology. Following the highly interdisciplinary nature of the field, the publications connect different views upon musical topics and problems with the field's multiple methodology, theoretical background, and models. It fuses experimental findings, computational models, psychological and neurocognitive research, and ethnic and urban field work into an understanding of music and its features. It also supports a pro-active view on the field, suggesting hard- and software solutions, new musical instruments and instrument controls, content systems, or patents in the field of music. Its aim is to proceed in the over 100 years international and interdisciplinary tradition of Systematic Musicology by presenting current research and new ideas next to review papers and conceptual outlooks. It is open for thematic volumes, monographs, and conference proceedings.

The series therefore covers the core of Systematic Musicology,

- Musical Acoustics, which covers the whole range of instrument building and improvement, Musical Signal Processing and Music Information Retrieval, models of acoustical systems, Sound and Studio Production, Room Acoustics, Soundscapes and Sound Design, Music Production software, and all aspects of music tone production. It also covers applications like the design of synthesizers, tone, rhythm, or timbre models based on sound, gaming, or streaming and distribution of music via global networks.
- Music Psychology, both in its psychoacoustic and neurocognitive as well as in its performance and action sense, which also includes musical gesture research, models and findings in music therapy, forensic music psychology as used in legal cases, neurocognitive modeling and experimental investigations of the auditory pathway, or synaesthetic and multimodal perception. It also covers ideas and basic concepts of perception and music psychology and global models of music and action.
- Music Ethnology in terms of Comparative Musicology, as the search for universals in music by comparing the music of ethnic groups and social structures, including endemic music all over the world, popular music as distributed via global media, art music of ethnic groups, or ethnographic findings in modern urban spaces. Furthermore, the series covers all neighbouring topics of Systematic Musicology.

Alexander Refsum Jensenius
Editor

Sonic Design

Explorations Between Art and Science

Springer

Editor
Alexander Refsum Jensenius
RITMO Centre for Interdisciplinary Studies in Rhythm,
Time, and Motion, Department of Musicology
University of Oslo
Oslo, Norway

ISSN 2196-6966 ISSN 2196-6974 (electronic)
Current Research in Systematic Musicology
ISBN 978-3-031-57891-5 ISBN 978-3-031-57892-2 (eBook)
https://doi.org/10.1007/978-3-031-57892-2

© The Editor(s) (if applicable) and The Author(s) 2024. This book is an open access publication.

Open Access This book is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.
The images or other third party material in this book are included in the book's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

Preface

Beeeeeeep, katchak!!! woooooOOOOooooo .... oh-Ra-ta-ta-oh-Ra-ta-ta-oh-Ra-ta-ta-oh-Ra-ta-ta-oh-Raaa ... ssssssssssssssso! Try to read these words aloud. Try to perform the words while standing. Letters on paper become sounds in the air. It is like a musical score being transformed into sound through the sound-producing actions of a musician interacting with an instrument. This is sonic design, constructing sound events and relationships between them.

In this volume, sonic design is used as a term that includes both artistic and scientific approaches to creating and studying (musical) sound. The term "sound" is typically used to describe vibrations that travel through air, while "sonic" describes anything related to sound. Thus, sound design can be thought of as designing (and producing) sounding sounds, while sonic design could also embrace designing and understanding sonic experiences.

Sonic design includes sound design practices such as recording, editing, and mixing acoustic sounds but also synthesizing entirely new sounds through analog and digital electronic devices. Over the years, these practices have merged with composition, orchestration, and production techniques to create previously unheard-of sonic results. In addition to such creative applications, sonic design also encompasses different sonification strategies aiming to create "objective" representations of data through sound.

Underlying all such practical approaches to the creation of sounds for various purposes are several fundamental research perspectives, in music theory, music perception, embodied cognition, phenomenology, acoustics, cognitive neuroscience, and digital signal processing, to mention just a few. Thus, sonic design can be seen as a meeting point between basic and applied research, "soft" and "hard" approaches, and creative and analytic perspectives.

Such a mix of art and science also summarizes the career of Professor Rolf Inge Godøy, to whom this volume is dedicated. Building on the legacy of the French composer Pierre Schaeffer, Godøy embraced the concept of the sounding object and how composition could be considered a combination of sounding objects in time and space. His background as a performer and composer greatly inspired his theoretical work. Throughout the 1990s, he delved into academic research to understand more about the cognitive foundations for the perception of sound as chunks and how mental imagery—musical imagery—structures how one listens to music, with or without sonic presence. This eventually led to exploring musical gestures, a concept encompassing body motion, sound, and imagery.

I was fortunate to work with Professor Godøy for nearly two decades, first as a student, then as a research fellow, and later as a colleague. I have seen him inspire several generations of students with his innovative and progressive music theoretical thinking. After a nearly 30-year-long career at the University of Oslo, he now has more time to work as a composer, using knowledge from his theoretical work in artistic practice. But his artistic activities also feed back to his scientific inquiries, most recently on the topic of sonic (co)articulation.

This edited volume is based on a selection of contributions at an international seminar organized in May 2022 to celebrate the achievements of Professor Godøy upon his retirement. The 17 chapters cover different approaches to sonic design practice and theory, giving readers historical backdrops and an overview of the current state of both artistic and scientific research in the field. Reflecting the breadth and width of Professor Godøy's activities, the volume will be of interest to students, practitioners, and researchers from the arts and humanities, social and natural sciences, and design and engineering.

I am grateful to the authors for their efforts to turn their seminar presentations into original chapters and review each other's contributions. I would also like to thank the other reviewers who helped improve the manuscripts: Andreas Bergsland, Bilge Serdar Göksülük, Björn Thor Jónsson, Bálint Laczkó, Hugh Alexander von Arnim, Joachim Mossige, Joachim Poutaraud, Kjell Andreas Oddekalv, Laura Bishop, Maham Riaz, Pedro Pablo Lucas Bravo, Qichao Lan, and Remy Martin. Finally, I thank RITMO Centre for Interdisciplinary Studies in Rhythm, Time, and Motion and the Department of Musicology at the University of Oslo for supporting the seminar and this book project and the Research Council of Norway for generous funding over the years.

BalalalalalalalalalalaboM!

August 2023 Alexander Refsum Jensenius

Contents

Theoretical Perspectives

Performance Perspectives

Artistic Exploration

Technological Innovation

Contributors

Rolf Bader Institute of Systematic Musicology, University of Hamburg, Hamburg, Germany

Anna-Maria Christodoulou Department of Musicology, RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo, Oslo, Norway

Gemma L. Crowe Emily Carr University of Art + Design, Vancouver, BC, Canada

John D'Arcy SARC: Centre for Interdisciplinary Research in Sound and Music, Queen's University Belfast, Belfast, Northern Ireland

Çağrı Erdem Department of Musicology, University of Oslo, Oslo, Norway

Kerstin Frödin Luleå University of Technology, School of Music in Piteå, GEMM (Gesture Embodiment and Machines in Music), Piteå, Sweden

Sylvie Gibet IRISA, Campus of Tohannic, Université Bretagne Sud, Vannes, France

Rolf Inge Godøy Department of Musicology, RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo, Oslo, Norway

Mari Romarheim Haugen Department of Musicology, RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo, Oslo, Norway

Ulf A. S. Holbrook Asplan Viak & RITMO Centre for Interdisciplinary Studies in Rhythm, Time, and Motion, Department of Musicology, University of Oslo, Oslo, Norway

Risto Holopainen Oslo, Norway

Patrick Kontopidis Institute of Systematic Musicology, University of Hamburg, Hamburg, Germany

Olivier Lartillot RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, Department of Musicology, University of Oslo, Oslo, Norway

Benjamin Lavastre CIRMMT, IDMIL, DCS, McGill University, Montreal, Canada

Marc Leman Department of Musicology, IPEM, Ghent University, Ghent, Belgium

Annamaria Minafra Conservatorio "N. Piccinni", Bari, Italy

Stefan Östersjö Luleå University of Technology, School of Music in Piteå, GEMM (Gesture Embodiment and Machines in Music), Piteå, Sweden

Albrecht Schneider University of Hamburg, Hamburg, Germany

Rodrigo Schramm Luleå University of Technology, School of Music in Piteå, GEMM (Gesture Embodiment and Machines in Music), Piteå, Sweden

Chris Stover Queensland Conservatorium, Griffith University, Brisbane, Australia

Åsa Unander-Scharin Luleå University of Technology, School of Music in Piteå, GEMM (Gesture Embodiment and Machines in Music), Piteå, Sweden

Georgios Varoutsos SARC: Centre for Interdisciplinary Research in Sound and Music, Queen's University Belfast, Belfast, Northern Ireland

Federico Visi Luleå University of Technology, School of Music in Piteå, GEMM (Gesture Embodiment and Machines in Music), Piteå, Sweden;
Universität der Künste Berlin, Berlin Open Lab, Berlin, Germany

Marcelo M. Wanderley CIRMMT, IDMIL, McGill University, Montreal, Canada

Theoretical Perspectives

Generic Motion Components for Sonic Design

Rolf Inge Godøy(✉)

Department of Musicology, RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo, Oslo, Norway
r.i.godoy@imv.uio.no

Abstract. Sonic design, understood as the activity of intentionally creating sound events, encompasses both musical craftsmanship and analytic reflection. It may include technologies for sound synthesis and processing, as well as traditional methods for sound generation by musical instruments or the human voice, and also principles of orchestration. Common to many instances of sonic design, is having acoustic components that blend with concurrent real or imagined motion sensations. Thus, sonic design can be understood as a multimodal phenomenon, yet we often lack suitable concepts for differentiating and evaluating these multimodal components. This paper aims to present work on developing a scheme to detect, and actively exploit, generic motion components in sonic design, be that as analytic or creative tools.

Keywords: Sonic design · motion · texture · timbre · orchestration · role analysis

1 Introduction

We may come across the expression ‘sonic design’ in different contexts, but there seems to be a consensus that it designates the activity of generating perceptually salient sound events, be that in music (instrumental, vocal, electronic), multimedia (theatre, movies, videos), human-machine interactions (computers, phones, electronic devices), or even as attributes of consumer products (cars, motorbikes, lighters), or marketing logos and branding in general (e.g., the so-called “James Bond chord”). But whereas sound logos will have significations in the direction of semiotics or narrativity, i.e., conveying some specific meaning beyond the sonic event, the focus of the present chapter will be limited to subjectively perceived sonic features, primarily to sonic design as an instance of musical creativity, be that in performance, improvisation, or composition.

To our knowledge, there have been only a few attempts to apply sonic design perspectives to so-called Western classical music, yet there is arguably a strong affinity between salient perceptual features of Western classical music and several key issues in sonic design. The lack of focus on perceptual sonic features we have seen in mainstream Western music theory is symptomatic of a general focus on the symbolic representations of pitch and duration, i.e., on Western notation-based features, and usually not on features of output sound. However, given recent technological developments, tools for sound generation and analysis are now readily available and applicable to various

A. R. Jensenius (Ed.): SSD 2022, CRSM 12, pp. 3–24, 2024.
https://doi.org/10.1007/978-3-031-57892-2_1

sonic design features, including those of Western classical music, enabling systematic research in this area.

Furthermore, we may see the noun ‘sound’ used interchangeably with the adjective ‘sonic’ and the expression ‘sound design’ used with more or less the same extension as ‘sonic design.’ Given our past work in this area, we shall continue to use ‘sonic design’ here and understand this expression to also include a number of elements in addition to what may be understood as purely acoustic components. This is actually the main point of the present chapter: *Understanding sonic design as a multimodal music-related activity, comprising sensations of sound fused with sensations of motion.* In more detail, I shall present an overview of salient sonic and motion features, and suggest how these can be integrated into useful tools for music creation and analysis. In view of these aims, similar motion components can be found across different instances of music and multimedia, constituting what can be called *generic motion components for sonic design.*

That subjective experiences of music are closely linked with sensations of motion, seems now to be claimed in much music-related discourse, and we have, during the last decades, seen a deluge of publications on such links. But the main focus in these publications seems to be on whole body motion, i.e., on how people move in synchrony with music by so-called *entrainment* (the process of aligning or synchronizing independent rhythmic processes), and less on more small-scale motion of effectors (fingers, hands, arms, tongue, lips, etc.) in sound-producing body motion. Of particular interest in our context are the details of motion we associate with specific sound features (e.g., the rapid back-and-forth shaking motion associated with a tremolo sound), and how sensations of such motion may be integral to our mental images of musical sound. From our own and other work on music-related motion, we have come to believe that what is referred to as ‘sonic design’, is also a matter of detecting and qualifying several motion components. Actually, we believe the links between sound and body motion are so extensive that we may not be able to univocally state what-is-what of sound and motion in our subjective perception of music, and should just accept that we need to explore both sound and motion components in sonic design.

Fortunately, it turns out that based on past and more recent research within music perception and associated cognitive sciences, it is indeed possible to develop some more systematic schemes for detecting and differentiating salient multimodal features that can be useful in sonic design (see, e.g., Godøy and Leman 2010 for an overview of music-related body motion). This is first of all linked with our sensations of the temporal patterns of energy in both sound and body motion, what we could collectively call *energy envelopes* of sound and motion, manifest, e.g., in a protracted sound linked to a sensation of protracted motion, or in a percussive sound linked to a sensation of an impulsive motion. The key issue here is the recognition of basic motion categories based on body motion constraints, such as the biomechanical and motor control differences between a sustained and an impulsive kind of body motion, as well as recognizing several other constraints that serve to shape output musical sound.

The ambition of this chapter is then to contribute to a conceptual, analytic framework for sonic design based on exploiting such sound–motion relationships, including both observable sound-producing motion of performing musicians (blowing, hitting, stroking, bowing, plucking, rubbing, etc.), and more subjectively perceived or imagined motion sensations emerging in listening, such as from various subtle timbral changes in the course of sound events, or from composite textural patterns with several concurrent layers of sound-producing motion, or also from more superordinate dynamic and spectral shapes of composite ensemble sound. And notably, the term ‘motion’ here also includes postures, as posture may be understood as a prerequisite motor control element for motion (Rosenbaum 2017), as well as a component of sound shaping, for instance, in body postures used when performing specific instruments or the posture shapes of the vocal apparatus related to specific vocal sounds.

We shall then, in the next Sect. 2, have an overview of what may be considered salient sonic features in our context, extending from the basic acoustic and signal-based to the more behaviorally and musically significant. Then follows a Sect. 3 on generative features in sonic design, encompassing both traditional instrumental and/or vocal music as well as more technology-based means for synthesis and processing, and a Sect. 4 on the related topic of multimodality, based on the abovementioned belief that sonic design involves more than just ‘pure’ sound. Then follows a Sect. 5 on analytic tools for sonic design, leading on to a Sect. 6 on textures and roles, focused on the distribution of sound events as well as individual musicians’ contributions to the output sound events. These topics lead to a Sect. 7 with some ontological reflections on the perceptual significations of different components of sonic design, as well an encounter with challenges in what we can call *musical translation* in the succeeding Sect. 8, i.e., on the transfer of musical ideas from one instrumental or vocal setting to another, before a concluding Sect. 9 with a brief summary of the main ideas and some thoughts on further work on sonic design.

2 Sonic Features

The expression ‘sonic design’ seems to go back to the pioneering work of Robert Cogan and Pozzi Escot, with the publication of their highly innovative and mainly sound features-based approach to musical analysis (Cogan and Escot 1976). Further work aiming to correlate musical features more directly with concrete sonic features by the use of sonograms and thus breaking out of the confines of the Western notation-based analytic framework was presented in (Cogan 1984). As for exploring significant sonic features of musical sound, current work on sonic design can reap benefits from a very large number of relevant publications, ranging from those focused on physical-acoustic features (e.g., Rossing 2002, Loy 2007), and/or technologies for digital sound synthesis and processing (e.g., Roads 1996, Zölzer 2011), to those focused on auditory perception (e.g., Bregman 1990, Fastl and Zwicker 2007), including publications related to sonic object perception (Griffiths and Warren 2004, Bizley and Cohen 2013), and some highly relevant publications on object formation in human motion (Klapp and Jagacinski 2011, Loram et al. 2014). Concerning motion as integral to sonic design, we should mention publications on sound-motion relationships (e.g., Rocchesso and Fontana 2003, Godøy and Leman 2010, Clayton et al. 2013), as well as on methods for motion capture and subsequent data processing of sound-motion relationships (e.g., Godøy et al. 2016, Godøy et al. 2017, Gonzales-Sanchez et al. 2019), all contributing to a broad background for work on sonic design. We can also benefit from projects of so-called *interactive sonic design* in various artistic and/or entertainment contexts, spurred on by new technical possibilities for multimedia experiences, as well as by the need for enhanced modes of human-machine interaction (Franinovic and Serafin 2013).

But the most significant contribution to a framework for sonic design research here has been the theoretical work of Pierre Schaeffer and co-workers on so-called *sound objects*, theoretical work emerging from compositional activities in the so-called *musique concrète* genre of the late 1940s and following decades. The impetus for this theory development was experiences with musical composition based on sound fragments from a variety of sources, be that human, animal, environmental, mechanical, instrumental, or electronic, and bypassing most features of more traditional Western music, leading to a fundamental revision of the basic principles of music theory (Schaeffer 1952, 1966, Chion 2009, Schaeffer et al. 1998, Godøy 2013, 1997a, 2021a).

The result of this revision was a theory based on the subjective perception of sound objects*,* defined as fragments of sound, typically in the duration range of 0.5 to 5 s. The reason for this focus was initially pragmatic, with the use of looped fragments on discs in the early days of the *musique concrète*, before the advent of the tape recorder. These experiences of listening to innumerable repetitions of such looped sound fragments made Schaeffer and co-workers realize that their perception of the sound fragments changed. Their listening focus shifted away from the anecdotic significations (e.g., a door squeaking signaling that someone is coming) towards the sound features as such (e.g., the glissando feature of the squeaking sound), something that came to be called *reduced listening*. This meant focusing on the overall dynamic, timbral, and pitch-related shapes, as well as various internal details, of the entire sound object, engendering a theory based on exploring perceptually salient features at different timescales within sound objects.

Schaeffer's approach was that of a top–down feature differentiation, following a seemingly naïve Socratic line of questioning as to what we are hearing, and progressively differentiating more and more feature dimensions based on this inquisitive listening with the long-term aim of correlating these subjective sensations with more objective acoustic features. This method ended up with a classification of the overall dynamic and pitch-related shapes of the sound objects, called the *typology of sound objects*, and with a more elaborate classification of the internal features of the sound objects, called the *morphology of sound objects*. The typology had three main categories for the overall dynamics, called *facture*:

- *Impulsive*: a short and percussive kind of sound
- *Sustained*: a prolonged and relatively stable sound
- *Iterative*: a rapidly repeated sound such as in a tremolo

The typology had three main categories for pitch-related content, called *mass*:

- *Tonic*: with a clear and stable sense of pitch
- *Complex*: being strongly inharmonic or noise-dominated
- *Variable*: having a changing sense of pitch

The typology was meant to be a first and coarse classification of sound objects to be further differentiated with the help of morphological features. Any sonic object would be assigned more sub-features, and some also sub-sub-features, in sum, providing a rather progressively more and more elaborate scheme for feature classification.

In Fig. 1, there is a 3x3 typology spectrum illustration of sound objects from tracks 31, 32, and 33 on CD3 in (Schaeffer et al. 1998), sounds made with traditional Western instrumental ensembles. Interestingly, Schaeffer applied these generic motion types to several different sources on CD3, namely the sounds in tracks 31–33, 34–36, 37–39, and 40–42, i.e., in 12 rather different examples, some of which came from synthetic sources, illustrating the universality of this typological categorization scheme.

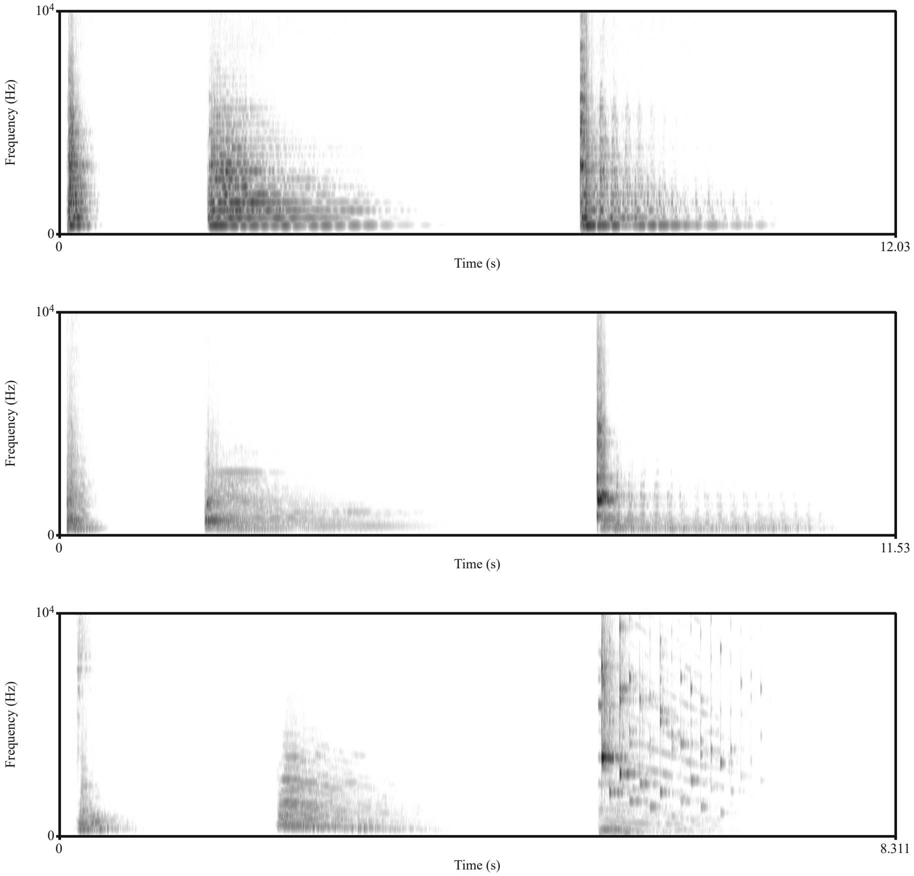


Fig. 1. Three sets of three instrumental ensemble sounds from tracks 31, 32, and 33, on CD3 in (Schaeffer et al. 1998). First line: a *tonic sound* as *impulsive*, *sustained*, and *iterative*. Second line: a *complex sound* as *impulsive*, *sustained*, and *iterative*. Third line: a *variable sound* as *impulsive*, *sustained*, and *iterative*. The sounds may be perceived as different in detail but similar in overall dynamic and spectral shapes.

Schaeffer's concept of *mass* was intended as a scheme for classifying spectral distribution, extending from single clearly pitched sounds to complex spectral sounds, and as harmonic, inharmonic, or noise-dominated, as well as their distribution of components in the spectrum and in time, be that stationary or evolving (as the *profile of mass*). All these elements can be related to a general notion of *shape*, i.e., of pitch, of overall dynamics, of more internal fluctuations, or of spectral content, e.g., to the *wah-wah* sound of an opening-closing of a straight mute on a trumpet, or the shifts between the *sul tasto* and the *sul ponticello* in bowing on a violin.

Going further into the morphology of sonic objects, the initially most salient features are the so-called *gait* and *grain*. Gait denotes the slower motion within the object, as can be seen later in Fig. 3 as the undulating motion in violin 2, viola, and cello. Grain denotes the fast fluctuations within a sound object, also to be seen in Fig. 3 in the subsequent tremolo motion in the violas.

Concerning current and recent research on timbre, the attention devoted by Schaeffer to the evolution of the spectral features within a sound object is remarkable, compared with some notions of timbre associated mostly only with the stationary spectral features, i.e., with what was called *tone color* (Slawson 1985). In the context of sonic design, limiting explorations to stationary spectra is insufficient, as this will miss the rich features of musical sound due to the within-sound motion of different kinds, ranging from various transients to more pronounced textural motion. In sum, the sound object theory of Schaeffer had the ambition of being able to diagnose whatever sound fragment is thrown at us by detecting its most salient perceptual features. That is, by trying to figure out why it subjectively sounds the way it does, and then later on, trying to correlate these subjective features with acoustic data. That said, we presently have readily available tools for going into various details of timbre, providing visualizations (e.g., with *Sonic Visualiser*), enabling extracting various defined features such as spectral flux, spectral centroid, harmonicity, etc. (e.g., *MIRtoolbox* for *Matlab*), and also make *analysis-by-synthesis* simulations of timbral features (e.g. in *Max/MSP*), just to mention some prominent tools here.

Also, within the framework of Western music theory, there are many features, albeit at the level of tones, that arguably could be included in a sonic design framework. This goes for various kinds of voicing and/or distributions of tones in time and spectrum, significantly so when we compare tone distributions, e.g., dense "Beethoven-type" chords with more widely spaced "Chopin-type" chords. Also, chord categories (e.g., triads, fourths chords, polychords, cluster chords, etc.) and modality categories (e.g., Lydian, Phrygian, Messiaen modes, etc.), all have salient effects on the sonic design, something that clearly deserves a more extensive research effort within music theory.

From the mentioned typological and morphological features as well as tone-level features of Western music theory, we see that salient features are manifest at different, and also often concurrent, timescales, making *timescales differentiation* a crucial topic in sonic design (Godøy 2022a, 2022b). The most important timescale for sonic design, following the seminal work of Schaeffer, is that of the sound object (as mentioned above, typically in the 0.5. to 5 s duration range), because this timescale may contain most defining features in terms of style, aesthetics, sense of motion, and affect. The typical sound object duration is also largely sufficient to contain sequentially evolving events, such as an entire tone envelope with its attack, sustain, and decay, thus enabling a cumulative and all-at-once presence of these components in echoic memory, something that is necessary for the holistic perception of a sound object. This insight came to be known through the so-called *cut bell* experience of Schaeffer and co-workers, i.e., that manipulating the attack and sustain segments of a bell sound would radically alter the overall sensory impression of the sound. Understanding how sequentially occurring features fuse into the sensation of a sound object as a holistic entity is a major challenge in sonic design.

3 Generative Features

The holistic nature of sound objects is also manifest when it consists of several tones in succession, as noted by Michel Chion: “[…] a harp arpeggio on the score is a series of notes; but, to the listener, it is a single sound object.” (Chion 2009, p. 33). A series of tone events may certainly be perceived as a coherent entity by way of the proximity of tones, as suggested by gestalt theory (Tenny and Polansky 1980), but even more so if we consider the arpeggio as a single coherent motion unit. In fact, holistically conceived motion units seem also to be optimal for body motion motor control in what may be called *motor gestalts* (Klapp and Jagacinski 2011), as well as implementing the efficacy of so-called *intermittent motor control* (Loram et al. 2014), i.e., a point-by-point scheme for controlling upcoming motion events. These motor control elements seem to converge in suggesting the existence of what we could call *motion objects,* similar to sound objects, and their multimodal combination in what we could call *sound–motion objects.*

The basic idea here is that sound production, albeit variably so, is conditioned by various constraints. These constraints not only determine what is possible, difficult, or even impossible for sound production on various instruments and/or the human voice but, in a more positive sense, contribute to salient features of the resultant sound. Constraints range from various physiological limitations on speed, amplitude, rate, etc., of sound-producing motion, up to affecting emergent sonic features such as spectral flux, harmonicity, spectral centroid, etc. (Godøy 2021b). In particular, motion constraints manifest in the temporal unfolding, i.e., in the envelopes presented above, and in the fusion of otherwise separate events by so-called *coarticulation* (Godøy 2014).

Coarticulation signifies the fusion or contextual smearing of motion components due to the need to prepare for upcoming motion events, e.g., hands need to move ahead of fingers in piano performance to place the fingers in a certain position to hit a key at the right moment in time, and there are also spillover effects from recently made motion, so coarticulation may be affected by both future and past events. Coarticulation is well known in several everyday tasks such as typing and tool use (Rosenbaum 2009), but most of all in speech (Hardcastle and Hewlett 1999), where the vocal apparatus is preparing for the production of upcoming sounds, as well as being moving away from the vocal apparatus shape for the most recently produced sound. Coarticulation means that there is a constraint-based tendency to fuse small-scale motion events into larger-scale motion events, and in music, coarticulation may contribute to the fusion of not only sound-producing motion but also of the output sounds, cf. The Chion example mentioned above. It can be argued that due to coarticulation, there is often a tendency towards object formation both in motion and output sound (Godøy 2022b).

Constraint-based motion may also result in distinct, mutually exclusive categories such as the mentioned impulsive, sustained, and iterative types of sound, e.g., a sustained sound or motion fragment will, by definition, not be impulsive. But there may be transitions between such categories by so-called *phase transition* (Haken, Kelso, and Bunz 1985). For instance, if a sustained sound is incrementally shortened, it may at some point turn into an impulsive sound, or if the rate of separate impulsive sounds is increased, it may at some point turn into an iterative sound. Similarly, there may be phase transitions concerning coarticulation in that decreasing the distance between tone onsets may increase the levels of both anticipatory and spillover motion, and hence also the level of contextual smearing of output sounds (Godøy 2021b).

Including sound-producing motion in research on sonic design is challenging when we try to understand performers' 'tacit knowledge' of shaping sound to their ideals. The tacit sonic design knowledge can, in part, be accessed by recording sound-producing body motion by video or with motion capture systems while also recording physiological data associated with such motion, e.g., muscle tension (EMG), pupillometry, brain activity (EEG) and other brain observation data. Such data can, after extensive preprocessing, be correlated with output sound features, something that should enhance our understanding of sound-motion objects in sonic design.

But as for sound generated by electronic means, what are the production schemas at work in such cases? Schaeffer's (and our) response is that the same typological schemas may apply to all sound events, regardless of origin, and thus also to synthetic sound, as in the examples on tracks 40, 41, and 42 on CD3 in (Schaeffer et al. 1998), because the energy envelopes may be matched to motor schemas, i.e., so that these schemas apply to electronic sounds as if the electronic sounds were made by body motion (see Godøy 2021b for this example).

So-called *physical models* of sound synthesis have a special status in view of sonic design because they are designed to (variably so) simulate energy relationships, differently from more abstract synthesis models. Physical models are diverse, ranging from quite simple to highly complex. Models of the so-called *source-filter* type are particularly interesting in our context in that they are based on the principle of an energy source that produces an output that is passed through various filters, resulting in an output sound with musically interesting features. An instance of this is the so-called *Karplus-Strong* model for simulating a plucked string. A burst of white noise sent into a feedback loop through a delay line with a low-pass filter will sound like a plucked string. It is interesting as a model of a 'real world' instrument in that a quantum of energy (the noise burst) is reverberating and gradually dissipating its energy within a system.

With more abstract sound synthesis models, there is the challenge of navigating to intended output features, given the point of departure that, in theory, any sound, heard or unheard, may be generated by digital synthesis. The extensive work in recent decades on the control of output features in synthesis has documented that some kind of holistic input scheme would be useful, and we have seen physical models that work by simulating the physics of the sound generation as well as having an input that simulates the actual sound-producing body motion (Bounëard et al. 2010).

Given the continuous increase in computational power and widespread activity in developing more responsive interfaces for gestural control, we are probably going to see enhancements in such direct control of physical model synthesis, hence have tools more in line with what we could call an ecological and constraint-based framework for synthesis in sonic design. A number of software tools are now available for experimentation and systematic analysis-by-synthesis explorations of sound-motion features, e.g. such as *Modalys* and *Max/MSP*, as well as useful suggestions for concrete work strategies (Farnell 2010).

4 Multimodality

From the overview of sonic and generative features, it should be clear that sonic design involves more than 'pure' sound. The features we encounter in sound events are related to motion, and sensations of motion are, in turn, composite, involving sensations of effort, energy, quantity of motion, trajectory shapes, posture shapes, and derivatives of motion such as velocity, acceleration, jerk, and the mentioned phenomena of phase transition and coarticulation. Also, secondary and more 'passive' features, such as haptic, proprioceptive, visual, etc. sensations, may all (variably so) be related to sonic design. Given extensive experiences of observing sound-producing motion as well as motion in general, it seems reasonable to suspect that most people may have motion sensation components in sound perception, as has been the claim of the so-called *motor theory* now for more than half a century (Liberman and Mattingly 1985, Galantucci, Fowler and Turvey 2006). I coined the term *motormimetic cognition* to signify this sensing or mental simulation of sound-producing body motion linked to whatever sound we are hearing or imagining in music (Godøy 2001, 2003). The basic tenet here is that sonic design is a multimodal topic, not limited to any idea of 'pure' sound.

Looking closer at what is going on in sound events, we realize that sound onset and sustain are happening because of an energy transfer from the musician to the instrument. This means that, e.g., a rapid drum fill is as much a rapid sequence of mallet-hand-arm-shoulder-etc. motion as a series of sound events (cf. Godøy et al. 2017). How the modalities of sound and motion work together, as well as which is the most important in any listening situation, is still an open question. Could even a silent choreography of sound-producing motion give us some sense of a drum fill? What is crucial in our context is that images of sound-producing motion, in turn with several components, e.g., sensations of muscle contraction, proprioceptive sensations, visual sensations, etc., may all contribute to giving us some salient image of the drum fill. This implies that the motion components can also become a tool for handling the otherwise ephemeral sound sensations.

The following motion features are detectable, measurable, and may be documented in motion data (Godøy 2021b):

- *Quantity of motion* (QoM): the overall sense of energy in sound-producing motion
- *Velocity of motion*: the sensation of displacement speed and direction
- *Acceleration*: the sensation of change in the displacement speed
- *Jerk*: abruptness in the displacement
- *Phase transition*: a qualitative categorical change due to incremental change in amplitude and/or frequency of motion
- *Coarticulation*: the fusion of otherwise separate elements due to spillover and/or anticipatory smearing

A main feature of motormimetic cognition concerns mapping energy envelopes of perceived sound to body motion energy envelopes, be these body energy envelopes based on actually seen body motion (as when attending a performance) or only imagined (headphones listening, eyes closed), so as to become integral sound features**.** Some examples of motormimetic elements relevant to sonic design are evident just by an enumeration of sound features and their possible corresponding sound-producing motion, as in the following:

- *Tremolo*: back-and-forth hand motion
- *Trill*: lower arm rotation motion
- *Gait*: slower undulating paced motion
- *Grain*: rapid back-and-forth or up-and-down motion hand motion
- *Crescendo/decrescendo*: gradual increase/decrease of force and/or amplitude of motion
- *Flam*: double strokes in drumming
- *Glissando*: sweeps by hand, arm, whole torso, on an instrument
- *Sustained sound*: slow, protracted motion
- *Impulsive sound*: rapid ballistic kind of hand/arm motion

Also, when considering articulation elements in music, e.g., *staccato, legato, sforzato, tenuto,* and bowing types, e.g., *martellato*, *spiccato*, etc., see (Halmrast, Guettler, Bader, and Godøy 2010), we may be reminded that they all owe their existence to motion components and that such articulation elements are in fact multimodal phenomena.

Concretely, tracing the typology components of the mentioned *facture* and *mass*, and the morphology components of *gait*, *grain*, and *profiles of mass* as shapes, is a prominent feature of Schaeffer's theoretical work (Schaeffer 1952, 1966), evident in the many conceptual shape images in his publications. Visualizing sonic features as shapes is something we can do in our minds, with pencil and paper, on the computer screen, or just with our fingers and hands in the air, and importantly, regard these tracings as *generic images.* Thinking of shapes as generic means that they may be applied across different modalities and contexts and be useful as practical tools in both analytic and generative contexts, e.g., in musical translations (see below).

In summary, sonic design is not limited to 'tone color,' i.e., not limited to stationary spectra, but includes motion, motion within sound objects, such as various transients, fluctuations (timbral, dynamic, pitch-related), and all sorts of textural patterns, as well as corresponding sound-producing motion. *Shape cognition,* in the sense of depicting all kinds of spectral features, both quasi-stationary and changing spectra, all kinds of within-spectrum motion, all kinds of dynamic envelopes, etc., becomes a prime tool for working with multimodality, with the capacity to translate from one modality to another in the analysis and generative processes of sonic design.

5 Analytic Tools

There is clearly a need to develop better tools for analysis and systematic work strategies in sonic design. Ideally, such tools should 1) help diagnose why/how particular sonic features produce specific aesthetic and affective results, and 2) help realize wished-for aesthetic outcomes.

Existing knowledge of musical acoustics and music technology is useful for grasping many sound features, as are the mentioned tools for research on music-related motion. What we have much less of, is analytic tools for sonic design in more traditional Western composition theory. Western music theory, with its focus on more abstract and symbol-based concepts of pitch and duration, does not tell us much about output sound, and is thus largely inadequate for work in sonic design. Fortunately, we have had important developments of software tools enabling explorations of different kinds of sound features, such as the *MIRtoolbox* for *Matlab* (Lartillot and Toiviainen 2007), and for visualizing sound features, such as the *Sonic Visualiser,* and software that enables more experimental changes to sound features, such as *AudioSculpt,* and the mentioned *Modalys* and *Max/MSP* software for hands-on work with analysis-by-synthesis.

The main challenge in view of analytic tools is that salient aesthetic features are emergent based on a distributed substrate in both the time and frequency domains. Thus, we need, first of all, to map out the different relevant timescales and feature dimensions and then to figure out how to represent the ephemeral emergent features relevant to sonic design. Our response is firstly to make graphic images of unfolding motion and sound, i.e., of both envelopes and spectra, as was suggested by Schaeffer's theory of sound objects, and secondly, to carry out systematic analysis-by-synthesis experiments of holistic, sound object-level features. Also, following Schaeffer, there is converging evidence that the timescale of the sound object is the most important in view of salient features and that other timescales should be seen in relation to this timescale, either as internal features of sound objects or as features of the overall shape of the sound objects (Godøy 2021a). The main arguments in favor of the sound object timescale in our analytic approaches are as follows:

- The object timescale, with its typological categories, is crucial for the overall emergent features of style, sense of motion, and affect, and this may also apply to musical semiotics (Delalande et al. 1996)
- Including entire sound objects in our explorations is crucial for capturing salient features distributed in time, cf. the mentioned *cut bell* experience of salient features that may be non-existent at shorter timescales
- The object timescale contains the morphology features, and various morphology patterns may be further differentiated into sub-sub-features
- Sound fragments longer than the typical sound object duration may contain several competing overall features, making focusing on single object features difficult

As for analytic tools, Schaeffer's approach consists of top-down feature differentiations based on subjective sensations, however, these subjective sensations could later be correlated with acoustic features. Practically, this means:

- Subjective tracing of overall typological shapes of facture and mass
- Subjective tracing of salient morphological shapes, i.e., of various internal features
- Correlating subjective tracings with signal-based representations

This shape-tracing strategy is one of the main ideas of motormimetic cognition and is arguably an extension of Schaeffer's ideas (cf. Godøy 2006), as shape concepts are manifest in the typological facture categories presented above. Such shape tracing may include the assumed sound-producing motion as reflected in the facture of sound objects, i.e., *impulsive*, *sustained*, and *iterative* shapes, and also apply directly to the dynamic and spectral sound object elements.

In addition, there are methods that are only mostly hinted at in Schaeffer's works, as they were not so easily implementable with the available technologies of the 1950s and the following couple of decades (Godøy 2021a), but which are possible now:

1. Analysis-by-synthesis generation of incrementally different variant sound objects by incremental changes in feature dimension values (Risset 1991) to explore categorical limits of salient perceptual features
2. Experimental explorations of phase transition and coarticulation by incremental changes in input and control parameters

The main goal of these analytic schemes is to create feature awareness in the analytic and practical work of sonic design, i.e., to make that which is present in subjective experience more explicit by sketching the subjectively perceived shapes and then naming these shapes, thus analytically differentiating salient features. We see then that images of shape, or what we could call *shape cognition* (Godøy 2019), become a useful part of practical analytic tools, with shape cognition also having a broader foundation in so-called *morphodynamic* thought (Thom 1983, Petitot 1990, Godøy 1997a).

Following the seminal ideas of Schaeffer, we can have a foundation for analytic sonic design tools here, starting with dynamic and spectral shapes applied at the object timescale and continuing downwards to a progressively more detailed differentiation of features as shapes. Also, following Schaeffer's idea of correlating these subjective features with acoustic data, we are working on a bottom-up, signal-based scheme for machine-based typological categorization, with the long-term aim of enabling studies of large collections of sound objects.

6 Textures and Roles

The term ‘texture’ is used in a rather inclusive way in musical contexts for designating the overall appearance of sound, similar to the overall appearance of a fabric of textile, wood, or other materials. Discussions of texture in music are typically few and rather brief, which is odd considering how crucial textural features are in Western musical styles, all the way from the emergence of polyphony to present-day music culture.

That texture has been ignored in much Western music theory is probably due to texture being an emergent property of temporally distributed substrates, i.e., of successive tone events and/or internal tone features unfolding in time. There is thus no simple reduction to be made of texture, as texture rather requires a holistic approach, similar to seeing distributed patterns elsewhere, such as in clouds, in waves, in bouncing objects, or in ornate surfaces, meaning that texture in music, with its distributed basis, only exists on the sound object timescale.

Actually, one of the main points of Schaeffer's typomorphology is about creating what we could call a textural taxonomy, a universally applicable scheme for qualifying perceptually salient textural features, but notably so, both at the tone event and sub-tone event timescale. Theoretically, we can think of a continuum extending from stationary sound (made by additive synthesis with perfectly harmonic spectra) to highly complex sounds with many inharmonic and/or noisy spectral components as well as containing many transients and fluctuations. Along such a continuum, we may find several different kinds of sonic textures, however, they are not always based on tones as the basic ingredients. That said, we do have some useful textural concepts in Western music theory that include various degrees of internal motion, which could initially be classified according to traditional categories:

- Monody: mostly single melodic lines with intermittent accompaniment, however, with variable degrees of embellishments and more spurious sonic events.
- Homophony: mostly as successions of chords, but with various internal fluctuations and sonic events, e.g., as is the case in the example in Fig. 3 below.
- Polyphony: in principle, mostly independent voices, and in some cases, with a fabric of voices so robust that a work can be transferred to different instrumental settings (e.g., as in J. S. Bach's *The Art of Fugue* where the individual voices have such outstanding melodic features and limits in ambit, that the same score may be performed by very different instruments, and with what could be considered musically acceptable results).
- Heterophony: melodic lines in unison or other intervals, and with deviations, found in various non-Western music as well as in jazz and some 20th-century Western music as well as in many non-Western kinds of music (see Godøy 1997a, pp. 219–223 for more on this).

We may also see the coexistence of textural motion features (fast) and sustained harmonic-modal features (slow) both in traditional Western music (baroque to romantic) and in more recent kinds (Messiaen, Lutoslawski, Xenakis, etc.), and in acousmatic music (sustained sounds with superposed transient or fluctuating motion). In principle, these two components, fast and slow, may be separated and explored by an analysis-by-synthesis scheme (cf. Godøy 1997b with Chopin's C major prelude in various variant guises where the texture is kept constant across several variant pitch modes, i.e., C-major, C-minor, C-phrygian, C-lydian, etc.).

As for textural roles, we have the classic Western scheme of a melodic foreground with a homophonic background accompaniment (with varying levels of voicing independence), as we can see in the schematic view of Fig. 2. However, there are, of course, other kinds of role distributions also in Western music, e.g., with more voicing independence, more pronounced polyphonic textures, as well as recently, also heterophonic and complex textures with various statistical distributions of sound events, e.g., as in some music of Xenakis, Lutoslawski, and Ligeti, where sonic textures may be the most prominent design element.

The analytic strategy for textural elements may then consist of a top-down differentiation of roles, going on to designating sub-roles, sub-sub-roles, etc., and then matching these roles with instrumental idioms and evaluating 1) the suitability of chosen instruments in terms of idioms, and 2) the well-formedness of instrument selections and combinations in terms of acoustic results. Figure 3 shows an excerpt from the second movement of Rimsky-Korsakoff's *Capriccio Espagnol,* where some of the mentioned textural features, as well as roles and idiomatic role assignments, are represented, in addition to an acoustical distribution of sound that is remarkable in its sonority and

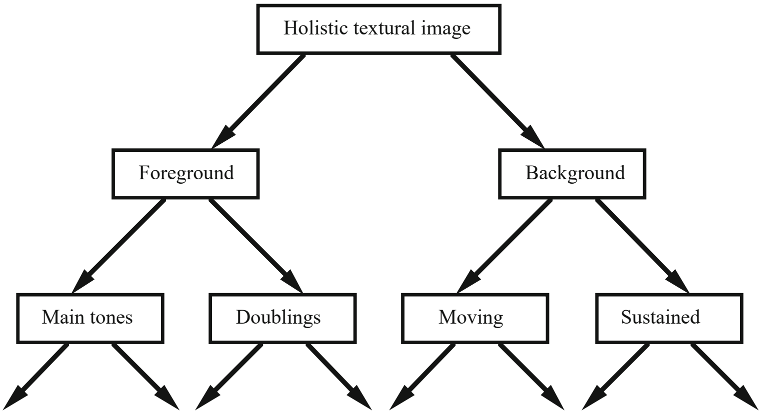


Fig. 2. Textural differentiation. A schematic overview of roles, top-down, from the holistic texture downward to role differentiations. These role differentiations could also continue to more sub-roles and, at the lowest level, be assigned to instruments optimally suited to the roles.

robustness. This robustness is due to Rimsky-Korsakoff's principle of having the mid-range chordal tones close together and having the top-range and the bottom-range tones more spread out, making for maximal sonority and little risks of "bad" sound. It will always sound good even if the musicians are not at their best.

Once we consider the use of instruments and orchestration as related to sonic design, we see that there has indeed been extensive, albeit not so well-articulated, knowledge of sound design issues, sometimes also linked with ergonomic issues, i.e., issues of idioms and issues of what's good and bad registers on instruments, issues of what is easy, difficult, or outright impossible on different instruments. However, Rimsky-Korsakoff (1912), combined ideas on acoustics from Helmholtz with experimentation exploring sound combinations, as well as systematic exploitation of instrumental idioms, in view of both performer well-being and optimal acoustic output. His work also includes ideas on good voicing and good parts, and he admonished orchestrators to delete anything in a score that could not be assigned to a clear role.

Thus, what we may collectively call textural features are fundamentally multimodal in evoking sensations of motion, implying that we really can't isolate a 'purely sonic' component in sonic design contexts. However, we can express something about our intentional feature focus at any time in listening and/or imagining. The different kinds of motion in textures, in particular the categories of sustained background vs. moving gait and faster grain, as well as impulsive attack reinforcements, all contribute to the sonic design, with the sustained giving an impression of a reverberant effects processing mix (in this case, pre-electronic). This may be seen in Fig. 3, where the sustained tones of horns 3, 4, and double bass make up the sustained role as a background to the undulating motion in violin 2, viola, and cello, with the remaining instruments massively doubling the foreground melody with its sustained tones. In general, extensive use of sustained tones can create a sensation similar to heavy reverb use, as in the famous Mantovani string orchestra sound, made by having the string players use lots of divisi with time

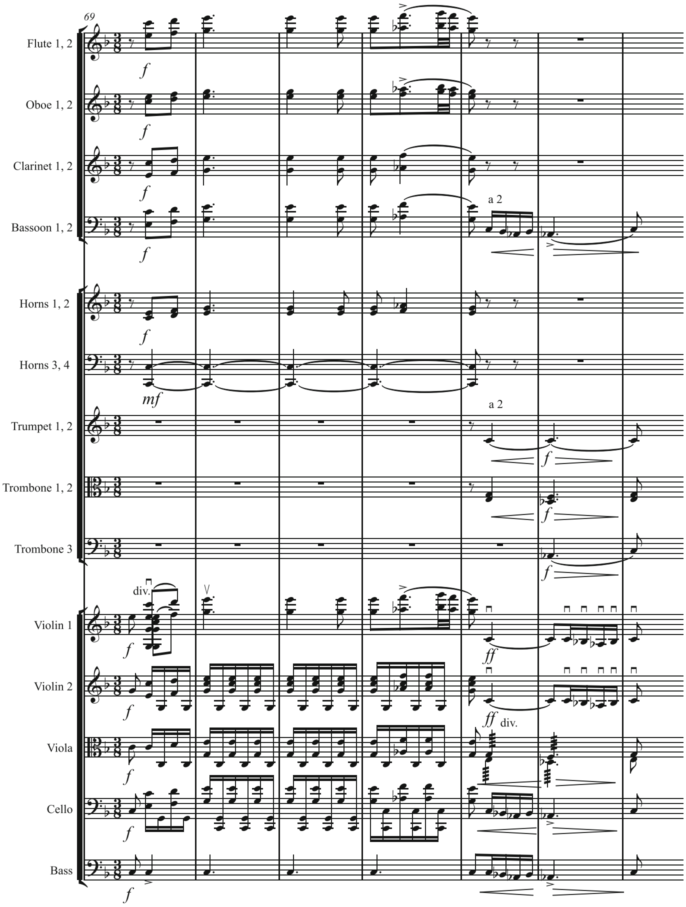



Fig. 3. An excerpt from measures 69–75 of Rimsky-Korsakoff's *Capriccio Espagnol*, second movement. The roles and instrument assignments are as follows: *Sustained background role*, assigned to horns 3 and 4, double bass, and partly trumpets and trombones, *accompanying gait role*, assigned to the undulating motion in violin 2, viola, and cello, with these using open strings and also three-tones broken chords in violin 2 and double stops in viola and cello for maximal sonority, i.e., idioms most practical here in the key of C, and then a *grain role* in the viola in measures 73 and 74, as well as *impulsive onsets role* in violin 1 in measure 69 and partly other string instruments with broken chords in sixteenth notes, and with the prescribed downstrokes in violin 1 and violin 2 in measures 73 and 74. In addition, there are the massively doubled parallel thirds and sixths in the foreground melody in woodwinds and violin 1, with instruments of the same family relatively close in register, and thus highly fused. Such use of close registers within similar instruments is an ideal for acoustic design in Rimsky-Korsakoff's orchestration, i.e., if similar instruments are used in widely different registers, they are less likely to successfully melt together. (The score here is in C.)

lags so as to create an illusion of stuck tones, e.g., in his famous recordings of *Charmine* (this technique was actually invented by Ronald Binge).

7 Ontological Reflections

In Western music theory, we have, with the development of music notation, acquired means for focusing on symbolic features, e.g., typically on pitches and durations, organized by various abstract schemes, schemes not always significant in terms of perceptual salience. A bit simplified, we could list the main features of Western music theory as follows: pitches, durations, intervals, chords, modality, motives, and articulations, and also some composite features such as form, style, and more recently, overall 'sound' and sensations of affect, and then try to have some opinion as to which of these elements are significant for subjective experiences of sonic design. Such evaluations of musical features are cases of *ontological reflections* and could be done in view of revealing the importance attached to sonic design issues.

Within Western music culture, we may often encounter an inherited view of music as having a 'core' of melody, themes, motives, and formal schemes, in short, as what has been referred to as *Formenlehre*, and only a 'periphery' of instrumental and sound design features. In some cases, we have also seen pitch-related features being transformed into abstract relationships, e.g., as in the so-called *pitch class set theory*, with a loss of spatiotemporal modal-harmonic features in favor of numerical relationships that may seem unrelated to subjective perceptual features. We have also seen elaborate schemes for the organization of notation symbols in compositions with little reflection on the perceived output, e.g., often with little or no sound object-level feature awareness, as pointed out in Iannis Xenakis' critique of serial music as lacking macroscopic feature concepts of statistical distributions of sound events (Xenakis 1992). A similar disregard for emergent perceptual issues seems to apply to some more recent and rather naïve cases of sonification, cf. the next section.

The crucial factor is to see composition, music production, and sonic design schemes separate from output perceptual features, i.e., not to confuse production formalisms with perceptually salient features. Jean Petitot (1990) has proposed a general model that is also relevant for sonic design, consisting of a *control sphere* and a *perception sphere*, where changes in the parameters of the control sphere may, or may not, have significant effects on the perception sphere, implying that we need to critically examine the relationships between these two spheres. In sonic design, a typically salient output from control input would be the sense of attack, ranging from 'bowed' to 'percussive' in the perception sphere, based on the incremental shortening of the attack time in the control sphere, whereas a change in pitch class set distributions might not have any significant effect compared with for instance rhythmical-textural elements. With readily available technologies, it is indeed possible to explore such levels of the salience of different structural schemes by an analysis-by-synthesis approach, starting by generating several variant sound objects with incremental changes in control input and then letting listeners evaluate significant, less significant, or insignificant changes. This could, for instance, be applied to sound objects where rhythmical-textural elements are kept unchanged and where pitch and/or modality features are systematically changed to allow evaluations of the relative significance of texture features vs. pitch features (see Godøy 1997b on this).

Taking the emergent features of continuous, fused sound events as primordial for sonic design, the ontological primacy is clearly at the sound object level (chunk level) of holistic shapes, not at the atomistic symbolic notation level. The sound object timescale has a fundamental ontological status, so that in sonic design, we have the primacy of the sound object. As such, there are some affinities with phenomenology in that we make sense of our continuous streams of perceptual input by interrupting these streams into chunks, chunks containing the cumulative images of a certain segment of continuous experience (Husserl 1991, Ricoeur 1981). The important message from phenomenology is that salient features emerge based on the distributed substrate of an entire object (t1 to t2), and that sonic design needs to start considering entire sound objects as coherent entities.

A general point is to think about sonic design in relation to real-world, non-abstract models of sound generation, closer to our common experiences of causality. This also includes motormimetic cognition and generic motion components as fundamental to sonic design, with ‘motion’ here also including various attributes of motion (shapes, effort, velocity, acceleration, jerk, etc.) as well as postures. In summary: ontological reflections are about *sorting out what is what* in sonic design, and we should be careful when mapping features from one domain to another, evaluating the validity of such mappings, something that will be at the core of what we may call *musical translation* in the next section.

8 Musical Translation

We can define the expression *musical translation* as the transfer of a musical idea from one setting, instrumental or vocal, to another, typically as is done in orchestration or arranging. The basic idea is to render an excerpt (or entire work) of music in a new ensemble setting, typically a solo or chamber music work adapted for a full symphony orchestra, with the assumption that the orchestrated music is just an alternative version of the original, hence a case of *musical translation.*

As is the case with language translations, the difficulty with idioms is that they will often become quite awkward, if not outright misleading, in a strictly literal translation, whereas a more reformulated version may actually be truer to the original than a literal translation. In music, this means that typical idioms for any instrument may not work well if transferred note-by-note to another instrument or sets of instruments, but could work well if transformed by either changing to an ergonomically better version or to a version with several instruments cooperating (Godøy 2018).

Musical translation will thus involve 1) an analysis of the original in view of what is the main musical idea(s), 2) a consideration of what (if any) are highly peculiar idioms embedded in the original, 3) considering whether these idioms could be transformed without doing too much harm to the overall aesthetic intention of the original, and 4), rendering this transformed idea in the new setting using optimal idioms of the new ensemble instruments. In other words, musical translation means adapting a generic motion script, with some adjustments of idioms, yet conserving the overall musical intention. Similar to natural language translation, where translating word-by-word is problematic and translating phrase-by-phrase is often much better, translating tone-by-tone in musical contexts may be problematic, whereas translating sound-object-by-sound-object is usually much better.

This is, in particular, the case when translating between instruments with and/without sustained sound, e.g., from piano to strings, wind, or tutti orchestra. For instance, the effect of the piano sostenuto pedal needs to be taken into account in the translation, otherwise, the result will be unduly dry, and in terms of effects processing, the wet-dry balance in different ensembles is really about making transformed, non-literal translations, cf. the mentioned Mantovani example with the reverb imitated by sustained string tones.

Flexibility in translation is possible because although the sound events we are working with in sonic design may have quite salient overall perceptual features, there is also the possibility of variation of the constituent detail features, hence that the categorical boundaries may be flexible, as is one of the hallmarks of categorical perception (Harnad 1987). The limits of tolerance for such variations can be studied empirically through the analysis-by-synthesis method, i.e., by making several incrementally different variants of some sonic object and then having participating subjects judge when there is a transition from one category to another, as in the abovementioned bowed to percussive category boundary exploration.

Similar problems of translation may be found in the domain of *sonification*, mapping elements from one domain to another, typically with data from a different domain than music (e.g., various experimental or observational data) to musical sound, to enable listening to the data rather than having to study large collections of numerical data (Hermann, Hunt, and Neuhoff 2011).

As for the generic, and thus translatable, features of music-related motion, we can see that the same sonic feature may apply to various similar body motion types, e.g., in the famous barber scene in Charlie Chaplin's *Dictator*. In that scene, Chaplin merges the sound-producing motion of Brahms' *Hungarian Dance Number 5* with the everyday motion of shaving: the *sustained* motion as a protracted razor shaving motion, the *impulsive* motion as a rapid flick removing soap, and the *iterative* motion as a rapid back-and-forth motion of rubbing in the soap (Godøy 2010).

For both translation and sonification, the most important questions are: Which sonic feature(s) is (are) the most prominent? And: How can this be somehow tested with an analysis-by-synthesis approach? Musical translation and sonification may then be testbeds for sonic design features, and what survives a transfer and what does not could be a crucial test for perceptual salience and teach us more about generic motion features.

9 Conclusions

In sum, it seems reasonable to claim that perceived and/or imagined motion may be integral to sound perception and may also have the potential to be a useful element in sonic design. Furthermore, such motion sensations may leave traces of both effort (muscle contractions) and vision (postures and trajectories shapes). Common to such sensations is that they may be conceptualized as shapes, shapes unfolding in both the time domain (dynamic envelopes) and the frequency domain (spectral envelopes), shapes that may furthermore be rendered as amodal graphical figures that can enable translation between modalities.

Although many of the issues covered in this chapter remain to be more systematically explored, we have, for the moment, good reason to conclude with the following:

- Focusing on sensations of motion in music perception is an efficient strategy to make us aware of salient features we might otherwise not be aware of
- Exploring generic motion components in sonic design may enhance our capabilities for both systematic diagnosis and enhanced skills for the creation of musical sound

Needless to say, there are also many outstanding issues:

- We need more systematic studies of sound-motion relationships, both because of how motion shapes sound and how listeners perceive sound–motion links
- We should work towards developing machine-based sonic object categorization enabling large-scale studies of music collections
- We need to supplement traditional Western music theory and composition theory with sonic design theory

Yet the current state of knowledge and skills in sound design may be put to use now because:

- Tracing shapes, both of sound-producing motion and postures, as well as of output sound and sound features, can be useful as generative tools in improvisation and composition
- Generic motion components can contribute to revising teaching methods, allowing for more spontaneous and improvisation-like creation of musical sound
- Detecting and qualifying generic motion components in sonic design can advance our understanding of why and how music affects us

References

Bizley, J.K., Cohen, Y.E.: The what, where and how of auditory-object perception. Nat. Rev. Neurosci. **14**, 693–707 (2013)

Bregman, A.: Auditory Scene Analysis. The MIT Press, Cambridge (1990)

Bouënard, A., Wanderley, M.M., Gibet, S.: Gesture control of sound synthesis: analysis and classification of percussion gestures. Acta Acust. Acust. **96**(4), 668–677 (2010). https://doi.org/10.3813/AAA.918321

Chion, M.: Guide to Sound Objects (English translation by John Dack and Christine North). ElectroAcoustic Resource Site (EARS) (2009). http://ears.huma-num.fr

Clayton, M., Dueck, B., Leante, L.: Experience and Meaning in Music Performance. Oxford University Press, New York (2013)

Cogan, R.: New Images of Musical Sound. Harvard University Press, Cambridge (1984)

Cogan, R., Escot, P.: Sonic Design: The Nature of Sound and Music. Prentice-Hall, Englewood (1976)

Delalande, F., et al.: Les unités sémiotiques temporelles: Éléments nouveaux d'analyse musicale. Éditions MIM-DocumentsMusurgia, Marseille (1996)

Farnell, A.: Designing Sound. The MIT Press, Cambridge (2010)

Fastl, H., Zwicker, E.: Psychoacoustics: Facts and Models. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-68888-4

Franinovic, K., Serafin, S.: Sonic Interaction Design. The MIT Press, Cambridge (2013)

Galantucci, B., Fowler, C.A., Turvey, M.T.: The motor theory of speech perception reviewed. Psychon. Bull. Rev. **13**(3), 361–377 (2006)

Godøy, R.I.: Formalization and Epistemology. Scandinavian University Press, Oslo (1993/1997a)

Godøy, R.I.: Knowledge in music theory by shapes of musical objects and sound-producing actions. In: Leman, M. (ed.) Music, Gestalt, and Computing. JIC 1996. Lecture Notes in Computer Science, vol. 1317, pp. 106–110. Springer, Heidelberg (1997b). https://doi.org/10.1007/BFb0034109

Godøy, R.I.: Imagined Action, Excitation, and Resonance. In: Godøy, R.I., Jørgensen, H. (eds.) Musical Imagery, pp. 239–252. Swets & Zeitlinger, Lisse (Holland) (2001)

Godøy, R.I.: Motor-mimetic music cognition. Leonardo **36**(4), 317–319 (2003)

Godøy, R.I.: Gestural-sonorous objects: embodied extensions of Schaeffer's conceptual apparatus. Organised Sound **11**(2), 149–157 (2006)

Godøy, R.I.: Gestural affordances of musical sound. In: Godøy, R.I., Leman, M. (eds.) Musical Gestures. Sound, Movement, and Meaning. Routledge, New York (2010)

Godøy, R.I.: Quantal elements in musical experience. In: Bader, R. (ed.) Sound – Perception – Performance. Current Research in Systematic Musicology, vol. 1, pp. 113–128. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-319-00107-4_4

Godøy, R.I.: Understanding coarticulation in musical experience. In: Aramaki, M., Derrien, M., Kronland-Martinet, R., Ystad, S. (eds.) Sound, Music, and Motion. CMMR 2013. Lecture Notes in Computer Science, vol. 8905, pp. 535–547. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-12976-1_32

Godøy, R.I., Song, M.-H., Nymoen, K., Haugen, M.R, Jensenius, A.R.: Exploring sound-motion similarity in musical experience. J. New Music Res. **45**(3), 210–222 (2016). ISSN 0929-8215. https://doi.org/10.1080/09298215.2016.1184689

Godøy, R.I., Song, M.-H., Dahl, S.: Exploring sound-motion textures in drum set performance. In: Proceedings of the SMC Conferences, pp. 145–152 (2017)

Godøy, R.I.: Motor constraints shaping musical experience. Music Theory Online. **24**(3) (2018). ISSN 1067-3040. https://doi.org/10.30535/mto.24.3.8

Godøy, R.I.: Musical shape cognition. In: Grimshaw, M., Walther-Hansenm M., Knakkergaard, M. (eds.), The Oxford Handbook of Sound and Imagination, vol. 2, pp. 237–258. Oxford University Press, New York (2019)

Godøy, R.I.: Perceiving sound objects in the Musique Concrète. Front. Psychol. (2021a). ISSN 1664-1078. https://doi.org/10.3389/fpsyg.2021.672949

Godøy, R.I.: Constraint-based sound-motion objects in music performance. Front. Psychol. **12**, 1–17 (2021b). ISSN 1664-1078. https://doi.org/10.3389/fpsyg.2021.732729

Godøy, R.I.: Timescales for sound-motion objects. In: Calderon, T., Gorbach, T., Tellioglu, H., Kaltenbrunner, M. (eds.) Embodied Gestures, pp. 39–48 (2022a). TU Wien Academic Press. ISSN 978-3-85448-047-1. https://doi.org/10.34727/2022/isbn.978-3-85448-047-1_5

Godøy, R.I.: Thinking rhythm objects. Front. Psychol. **13** (2022b). ISSN 1664-1078. https://doi.org/10.3389/fpsyg.2022.906479

Godøy, R.I., Leman, M.: Musical Gestures: Sound, Movement, and Meaning. Routledge, New York (2010)

Gonzalez-Sanchez, V., Dahl, S., Hatfield, J.L., Godøy, R.I.: Characterizing movement fluency in musical performance: toward a generic measure for technology enhanced learning. Front. Psychol. **10** (2019). ISSN 1664-1078. https://doi.org/10.3389/fpsyg.2019.00084

Griffiths, T.D., Warren, J.D.: What is an auditory object? Nat. Rev. Neurosci. **5**(11), 887–892 (2004)

Haken, H., Kelso, J., Bunz, H.: A theoretical model of phase transitions in human hand movements. Biol. Cybern. **51**, 347–356 (1985). https://doi.org/10.1007/BF00336922

Halmrast, T., Guettler, K., Bader, R., Godøy, R.I.: Gesture and timbre. In: Godøy, R.I., Leman, M. (eds.) Musical Gestures. Sound, Movement, and Meaning, pp. 195–223. Routledge, New York (2010)

Hardcastle, W.J., Hewlett, N.: Coarticulation: Theory, Data, and Techniques. Cambridge University Press, Cambridge (1999)

Harnad, S.: Categorical Perception. Cambridge University Press, Cambridge (1987)

Hermann, T., Hunt, A., Neuhoff, J.G.: The Sonification Handbook. Logos Publishing House, Berlin (2011)

Husserl, E.: On the Phenomenology of the Consciousness of Internal Time. In: Brough, J.B. (ed.), pp. 1893–1917. Kluwer Academic Publishers, Doredrech (1991)

Klapp, S.T., Jagacinski, R.J.: Gestalt principles in the control of motor action. Psychol. Bull. **137**(3), 443–462 (2011)

Lartillot, O., Toiviainen, P.: A Matlab toolbox for musical feature extraction from audio. In: International Conference on Digital Audio Effects, Bordeaux, 2007 (2007)

Loram, I.D., van De Kamp, C., Lakie, M., Gollee, H., Gawthrop, P.J.: Does the motor system need intermittent control? Exerc. Sport Sci. Rev. **42**(3), 117–125 (2014)

Loy, G.: Musimathics. The MIT Press, Cambridge (2007)

Petitot, J.: 'Forme' in Encyclopædia Universalis. Encyclopædia Universalis, Paris (1990)

Ricoeur, P.,: Hermeneutics and the Human Sciences. Cambridge University Press/Éditions de la Maison des Sciences de l'Homme, Cambridge/Paris (1981)

Risset, J.-C.: Timbre analysis by synthesis: representations, imitations and variants for musical composition. In: De Poli, G., Piccialli, A., Roads, C. (eds.) Representations of Musical Signals, pp. 7–43. The MIT Press, Cambridge (1991)

Roads, C.: The Computer Music Tutorial. The MIT Press, Cambridge (1996)

Rocchesso, D., Fontana, F.: The Sounding Object. Mondo Estremo Publishing (2003)

Rossing, T.D.: The Science of Sound. Addison Wesley, San Francisco (2002)

Rosenbaum, D.A.: Human Motor Control, 2nd edn. Elsevier, Burlington (2009)

Rosenbaum, D.A.: Knowing Hands: The Cognitive Psychology of Manual Control. Cambridge University Press, Cambridge (2017)

Schaeffer, P.: A la recherche d'une musique concrète. Éditions du Seuil, Paris (1952)

Schaeffer, P.: Traité des objets musicaux. Éditions du Seuil, Paris (1966)

Schaeffer, P. (with sound examples by Reibel, G., and Ferreyra, B.) (1998) (first published in 1967). *Solfège de l'objet sonore*. INA/GRM, Paris

Tenny, J., Polansky, L.: Temporal gestalt perception in music. J. Music Theory **24**, 205–241 (1980). https://doi.org/10.2307/843503

Thom, R.: Paraboles et catastrophes. Flammarion, Paris (1983)

Xenakis, I.: Formalized Music, Revised Pendragon Press, Stuyvesant (1992)

Zölzer, U.: Digital Audio Effects, 2nd edn. Wiley, Chichester (2011)

Sound Design and the Theory of Self-augmented Interactions

Marc Leman(✉)

Department of Musicology, IPEM, Ghent University, Ghent, Belgium
Marc.Leman@ugent.be

Abstract. In the past decades, musicology has been evolving at a pace that matches new developments in technology. Underneath this development, a new theory of music emerged, embracing interaction states as a model for understanding how music can be empowering. In the present chapter, sound design is considered from the viewpoint of interaction states, using caregiver–infant communication as a challenging domain of application. Sound design components of interest are identified, as well as human capacities for dealing with them in terms of empowerment. These are related to the concepts of self-augmented interaction and biofeedback-based sound design.

Keywords: Sound design · interaction · music technology · biofeedback

1 Introduction

Imagine the request of a neonatologist, a medical doctor specialized in the care of newborn babies. Imagine that the neonatologist requests us, sound designers, to stimulate premature-born infants with musical sounds as a way to lower stress and stimulate brain development. Stress in premature infants is mainly due to sounds coming from several clinical devices in the NICU (Neonatal Intensive Care Unit), as well as to a lack of body movement and stimulation thereof. Stress is known to cause long-term emotional complications, abnormal brain development, and health alteration (Beltrán et al., 2022). The effects of musical sounds could be measured via physiological outcome indicators of stress, such as heart rate, blood pressure, and oxygen levels, among others. Hence, hard-core evidence could be provided on whether a sound design treatment would be an improvement compared to the current situation. Can we, as sound designers, offer proper musical stimuli?

In this chapter, the neonatologist's request will serve as a common thread in a general reflection on sound design and its underlying theory of musical action and perception, or at least how we interpret this theory. We start with identifying the core components that are typical for music as well as the core human capacities needed for processing these components. Using these components and capacities, we clarify a theory called "self-augmented interaction." With this theory, we aim to explain the power of music in healing, that is, why music could be effective in particular clinical contexts. While the case of preterm infants will leave us with many unanswered questions, the neonatologist's

A. R. Jensenius (Ed.): SSD 2022, CRSM 12, pp. 25–36, 2024.
https://doi.org/10.1007/978-3-031-57892-2_2

request is challenging and fruitful for sharpening our own premature points of view. Towards the end of this chapter, we will expand our viewpoints on biofeedback as a possible way to go for future sound design applications.

2 Strategies: Stimulus-Response or Interaction?

Our challenge is to understand which sound design could work for preterm infants. This promises to be a delicate and difficult task, radically diverging from an earlier but misleading claim that Mozart would make infants smarter (see https://en.wikipedia.org/wiki/Mozart_effect). We expect that a decrease in stress in infants would already be a major achievement. Making them smarter could be understood in terms of brain development through rich stimulations. However, can we be sure that music will decrease rather than increase stress, given an already stressed infant? And what can music offer in terms of brain development? We know that preterm infants, due to limited experience with (prenatal) low-pass filtered speech, have difficulties processing natural speech after birth (François et al., 2021). But would music with a proper sound design work any better?

The sound design approach should probably be inspired by the natural way in which caregiver and infant communicate with each other (Malloch, 1999; Trewarthen and Aitken, 2001; Van Puyvelde et al., 2013). Such communication is interactive, involving speech, gestures, and touch. In observations of speech, the exchange of patterns between the caregiver and infant seems to establish an interaction state that is experienced as meaningful by the caregiver. Such an interaction state opens a window of opportunity for an intense contact with the infant, typically described as an attending state in which synchronized sound exchanges, gesturing, and participation in narratives are important characteristics. The overall picture is that caregiver-infant communication is rooted in musicality.

Given the practice of meaningful interaction, that is, interaction meant to establish contact, attention, and intention, the original question of the neonatologist can perhaps be inverted. Rather than asking for a proper sound design that stimulates preterm infants, the question should be whether infants are capable and willing to interact with a designed musical sound. In the former case, we put the infant in a passive role, adhering to a stimulus-response paradigm in the hope that some measurable effect will be generated. In the latter case, we put the infant in an active role, adhering to an interactive paradigm in the hope that an interaction state can be established, so that opportunities are generated for creating effects. The latter could be far more effective because it simulates the interactive basis of caregiver–infant communication. Obviously, this approach suggests a strategy of sound design based on principles of interaction.

3 Components: Emergent Patterns and Expressive Gestures

Given the fact that caregiver–infant interaction can be described as “musical” (Trehub, 2013; Trevarthen, 2008), it is instructive to look at two rather peculiar components of music and see how they could be incorporated into a proper design.

3.1 Pattern Emergence

Pattern emergence implies that a pattern has structural dispositions conjugate with the human disposition for emergence. For example, the human ear/brain will transform harmonic sounds into auditory patterns experienced as pitch (see Langner, 2015). The structure could be a harmonic pattern at 600, 800, 1000, or 1200 Hz having a disposition for subharmonics (or common periodicity) when seen from the viewpoint of the auditory system. This pattern would elicit a clear pitch percept at 200 Hz due to "Verschmelzung" mechanisms (Schneider, 2018a). These mechanisms may also fuse multiple harmonic patterns into a chord. When such harmonic patterns are played in sequence, their fusion may elicit expectations, leading to tonal tensions and relaxation dynamics. The bottom–up mechanisms of pattern emergence may compete with top–down mechanisms of patterns formed by habits, and it seems that both may influence the perception of tonal tension and relaxation, although the precise contributions of sensory (bottom–up) and cognitive (long-term memory) processing are still debated (Collins et al., 2014; Sears et al., 2019). In speech, however, "Verschmelzung" is less evident at multiple hierarchical levels. It works at the level of single pitches of a voice, but not for pitch complexes like chords, although tonal induction as an emergent outcome of the accumulation of tonal information in speech tone sequences might be considered. Clearly, the ability to form emergent patterns at frequency levels >80 Hz is more prominent in music than in speech. A similar observation can be made for lower frequency structures (<10 Hz) where rhythms, tempi, and meters are formed. Rhythms are made of pulses that subsume a metric, that is, a super-structure that emerges from the lower-level pulse structure. While rhythms can be strongly present in speech as well, rhythmic regularity in music is usually more pronounced than in speech. Similarly, timbres might blend and form emergent texture patterns, a phenomenon that is well-known in orchestration (Schneider, 2018b). Pattern emergence is less obvious for speech because it may just blur the signal, making it less apt for understanding its semantics. Moreover, in music, it often happens that different performers co-regulate their actions to generate pattern emergence at rhythmic, pitch, and timbre levels. Clearly, this phenomenon is far less prominent in speech. Accordingly, for preterm infants, some degree of pattern emergence might be considered as an ingredient of the sound design because we can assume that it will stimulate the infants' brain disposition for pattern emergence and associated predictions.

3.2 Gestural Expression

There is another aspect in which music differs from speech, namely, in gesturing. Gestures are known to accompany speech (Goldin-Meadow, 2005). Yet, in music, gestures form a more explicit part of the sound pattern. Gestures tend to structure musical pitch as a moving sound object (e.g., with portamento and intonation). In a similar way, gestures tend to structure time (e.g., making intervals shorter and longer for adaptations of tempo and meter), articulations (e.g., legato and staccato), sound color, musical narrative, and dynamics (e.g., crescendo and diminuendo). In short, human gestures structure the sound expression, making gestures constituent of music.

Godøy (2003) used the term “motor-mimetic” to specify the gestural interaction with sound endowed with gestural cues. The term is closely related to the term “mirroring.” The basic idea is that music is endowed with expressive cues that composers and performers encode as part of the sound design, and which listeners and dancers decode through gesturing because it offers them an expression-based, and gesture-based, perspective for prediction. Motor mimesis thus means that music has traces of human-encoded gestural patterns that listeners can decode in terms of a corporeal gestural mirroring of these patterns. “Baby talk” is a good example of a kind of communication in which the expressive component is the major vehicle for interacting. Accordingly, for preterm infants, gestures in the sound design may be considered a component of the sound design because we can assume that it stimulates the infant’s disposition to move in response to it.

Both pattern emergence and gestural expression are probably the core components of a proper sound design that would be capable of generating interaction states with the potential to have empowering effects. In our example, we want the sound design to resemble the caregiver–infant interaction, thus stimulating brain and body movement to become more expressive, responsive, and engaging, with a plausible transfer to other brain functions, such as those needed for speech.

4 Capacities: Affordance, Entrainment and Anticipation

What, then, would be the required capacities of an interactive sound design system? Some of the key human capacities for dealing with sound design have already been suggested. We identify them here as affordance, entrainment, and anticipation.

4.1 Affordance Capacity

An affordance is a property of the sound design, and the capacity to act upon an affordance is called: the affordance capacity (Godøy, 2010). Affordances can be understood as invitations to act in a particular way rather than another. The classic example is the design of a door handle, inviting us to open the door by turning left, or right, based on our knowledge of handling doors. In musicology, the notion of affordance has sometimes been linked with the notion of “frozen emotion” and the idea that composers and performers encode “frozen emotion” in music while listeners have the capacity to decode these emotions because they work as affordances. The affordance capacity is a decoding capacity which, in the case of music, and “frozen emotion” in particular, is likely based on mirroring, or (overt or covert) gesturing along with the music. Would preterm infants already have an affordance capacity? That’s an interesting question. If expression is indeed partly innate, then a biological response to sound cues through movement, perhaps somewhat uncontrolled, can plausibly be expected from an infant, although the time of development after birth will obviously be important to build up knowledge for affordance decoding.

4.2 Entrainment Capacity

The entrainment capacity is the capacity for moving along with music, either in a continuous manner when movement flows along with the music or in a discrete manner when movement marks musical events (Clayton et al., 2005). As the word suggests, entrainment implies that there is something in the music that brings the listener in sync with its temporal course through some form of dynamical process of attraction. In recent years, this phenomenon has received much attention as it also implies to (co-)regulated narratives (McGowan and Delafield-Butt, 2022). In the musical domain, entrainment has been studied in the context of (co-regulated) sensorimotor synchronization, where it has been associated with a bias to subliminally reduce prediction errors in the alignment of body movement with sound cues (Phillips-Silver and Keller, 2012). While entrainment is often defined in relation to synchronization, as the dynamic adaptation of sensorimotor behavior due to coupling, entrainment may also be defined in a broader perspective, as the capacity of giving a response to cues (Trost et al., 2017). In view of infant stimulation, the idea is that the stimulus contains cues to entrain the infant's responses. However, capacities for synchronized responding might be limited, subdue to the infant's development.

4.3 Anticipation Capacity

The anticipation capacity is the capacity to predict events. This topic has been widely studied in the context of predictions of sound structures, both in pitch and rhythm (Huron 2008), as well as in the sensorimotor domain related to synchronization (Maes et al., 2014). Obviously, anticipation is also possible in gesturing, where sound cues engage gestures that are intrinsically predictive. When a gesture is initiated, it typically follows a spatiotemporal trajectory based on so-called forward models in the brain. Once initiated, such gestures can become vehicles for anticipating musical events, leading to the phenomenon of reverse causality (Leman, 2016). For example, when listeners are dancing to music, the music and the dance movements are correlated, and typically, the dance movements will anticipate the events that occur in the music. Given the dance–music correlation and the knowledge that dance anticipates the music, the listener may then believe that the dance is, in fact, what causes the music. Obviously, the listener knows that the counterfactual statement "no-dance thus no-music" would fail because the music will continue if the listener doesn't dance. Nevertheless, despite the denial of the counterfactual, the illusion of reverse causality may be very strong, and it is typically associated with strong feelings of control and power.

Furthermore, we can assume that affordance, entrainment, and anticipation are tightly related to each other. For example, a musical pulse at about 2 Hz is a typical affordance, and children and adult listeners and dancers respond to it by moving along with the pulse, thus engaging an entrainment mechanism for the tempo that is based on prediction, making it possible to experience the illusion of reversed causality. However, this affordance assumes that there is a disposition for moving at 2 Hz. Is this disposition available at birth, or do humans acquire it? In relation to preterm infants, it is reasonable to believe that reverse causality (and hence: gesture–sound anticipation) can only occur when a gesture–sound link is already established, and therefore, it is unlikely that a preterm infant has the anticipation capacity to act in this way. Nevertheless, developing such an anticipatory capacity is probably a key goal of caregiver–infant interactions. For sure, after a few months, babies already understand gesture–sound relationships in the sense that they use them in outspoken anticipatory expressions (personal experience of the author).

5 Sound Design and the Theory of Self-augmented Interaction

Having specified musical components and human capacities, our goal is to understand their role in terms of a theory of music interaction. This theory is the backbone of our sound design for preterm infants. This theory, by the way, gradually evolved over the past decades in the slipstream of cognitive science. It aims at understanding why people, through interacting with music, might benefit from it. People's impelling attraction to perform, listen, and dance to music is intriguing, and its empowering effects are documented, albeit poorly understood. The acclaimed beneficial power of music is the reason why our neonatologist wanted a sound design with musical properties. But do we understand where that power comes from?

5.1 What is Self-augmented Interaction?

In what follows, we develop the notion of self-augmentation as a distinctive feature of music interaction. Self-augmented interaction implies that interaction is becoming sustained, richer, and more empowering than other states that do not, or to a lesser degree, have this empowering effect. We may assume that the distinctive feature of a self-augmented interaction state is based on a more optimal functioning of its underlying constituent processes. For example, when a string quartet plays and maintains a particular stable tempo, functional for global musical processing, it is because the members of that quartet have co-regulated their actions such that the required tempo can be created and maintained. The tempo is based on the optimal functioning of the underlying constituent timing and sensorimotor mechanisms. The "self" points to the fact that the musicians want to play together and realize this tempo without any external driver. If a musician plays too fast, the spell of a stable tempo may be rapidly lost, and the enriched state may become dis-integrated. Self-augmented interactions require physical effort, attentional focus, and sharpness of sensorimotor activity.

Such states can be conceived from the viewpoint of complex dynamical systems (Schiavio, Maes, van den Schyff, 2022). In Leman (2016), we introduced the notion of musical homeostasis. While in medicine, homeostasis refers to states of systems that regulate processes in the human body, such as body temperature, or sugar level, musical homeostasis typically requires physical effort and sensorimotor skills. Obviously, some training and learning may be needed before high-end self-augmentation can be achieved, especially when sensorimotor skills are requested. In that sense, the self-augmented interaction state refers to a precarious level (because it can quickly dis-integrate) requiring physical effort in the form of attention, physical activity, and highly skilled sensorimotor gesturing.

At this point, it may appear that we are far from our caregiver–infant narrative. However, based on what we discussed so far, human interaction between caregiver and infant can be understood in terms of a mutual exchange of gestures which, through co-regulation, may lead to a self-augmented interaction state or homeostasis. The hypothesis is that self-augmented interaction states facilitate empowering effects, such as attention spans and specific outcomes, such as bonding, and other psychosocial effects, such as the formation of goal-directed and anticipatory behavior.

5.2 A Model for Self-augmented Interaction

A global model for understanding self-augmented interaction is shown in Fig. 1 (see Leman, 2016, chapter 8). In a nutshell, this model suggests that music engages sensorimotor processes of prediction, embodiment, and expression, which in turn engage emotion-related processes of control and agency, arousal and attention, and pro-social empathy leading to reward-related processes that drive human subjects to engage more with music. As such, a cycle is created that may support the realization of self-augmented interaction states. The connection between prediction and control/agency, the connection between embodiment and arousal/attention, or the connection between expression and pro-social empathic emotions, may be the focus of separate research projects in modern musicology (e.g., Bader, 2018). Yet, the overall picture is that a network of processes, through its mutual influences of streams of information, hormones, and neurotransmitters, can develop into a state that can be characterized as self-augmented because it surpasses the normal state of being. When such a state can be maintained for a while, homeostasis is established, which opens a window of opportunity for effects that are otherwise hard to obtain. Recall that pattern emergence and gestural expression will facilitate the generation of self-augmented interaction states. These components of the sound design fit with the human capacity for affordance, entrainment, and anticipation. In addition, multi-person co-regulation of actions sets a social context that can be motivating, and the formation of a musical narrative can become compelling.

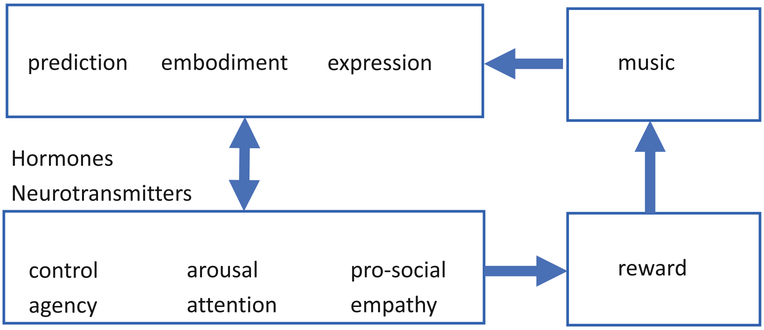


Fig. 1. The Mechanism leading to homeostasis.

Thus, when a caregiver and infant happen to maintain interaction through a co-regulated gestural and sound narrative, it is possible that this interaction develops into a self-augmented interaction state. To maintain that homeostasis, the infant and caregiver rely on their own state of co-regulated sensorimotor-emotive processes, using these with extra effort in view of homeostasis. The assumption is that such an interaction state opens a window of opportunity for training and learning, implying a reinforcement of its underlying processes. The gain, apart from developing better music interaction capabilities, should also be seen in terms of its transfer to other sensorimotor, cognitive, and social functions.

5.3 Foundations in Expression Theory

This interaction theory is a testable theory about human expressive behavior and how it contributes to psychosocial empowerment. However, given its complexity, it may be interesting to decompose the theory into sub-theories. Accordingly, the theory can be approached from the viewpoint of embodiment (e.g., Leman, 2007), predictive processing (e.g., Seth, 2014, Koelsch et al., 2019), and expression (Leman, 2016). While embodiment and predictive processing are well-known viewpoints, expression theory is often less well understood in cognitive science, despite its development by scholars who contributed to the foundations of cognitive science, such as David Hume, Adam Smith, Charles Darwin, and Erving Goffman (see Bonicco-Donato, 2016). Briefly stated, expression theory is based on the idea that expression in person A calls for an expressive response in person B, which in turn serves as a stimulus for expression in A and so on, thus leading to a mutual exchange of expressions that might result in an interaction state, plausible a self-augmented interaction state.

An expression can be defined as a pattern from A transmitted to B. However, a major source of confusion in expression theory concerns the idea that expressions are utterances of some underlying state of being. However, in many contexts, expressions really don't require any inference at all about the expression's underlying state of being (cf., Kahneman, 2011). Instead, the interaction is direct and spontaneous, based on expressive responses to patterns, through alignment, mirroring, including counterpoint gesturing. Obviously, whether inferencing or gesturing is applied largely depends on the context and type of interaction. But the main point here is that expressions do not always have to point to something underlying.

Based on observations about caregiver–infant interactions, it is likely that much interaction is based on intuitive thinking (Kahneman, 2011), plausibly under the umbrella of overall analytic thinking. Therefore, rather than inferring the latent state of being (known as the theory of mind theory), it is more appropriate to speak about gestural responding (Leman, 2016). The real power of expression exchanges is their ability to build up and maintain self-augmented states. Expression theory may thus be understood in terms of an exchange of expressive gestures as patterns that steer up the interaction towards self-augmented states.

6 Biofeedback Systems

The original neonatologist's question was whether we, as sound designers, could develop stimuli to lower preterm infants' stress in the NICU and stimulate their brain development. Based on the above considerations, it is straightforward to consider a sound design that would be able to create and maintain self-augmented interaction states with the preterm infant. Hence, the possibility of having an adaptive sound design, similar to a real human who is adaptive to the infant's responses, can be considered.

6.1 Sound Design in Biofeedback Systems

The development of a sound design with biofeedback for preterm infants may find inspiration in the development of biofeedback systems in other domains. In recent years, several attempts at developing biofeedback systems for sports have been undertaken, such as in running, weightlifting, and biking (Van den Berghe et al., 2021, 2022; Lorenzoni, 2019a, b; Maes et al., 2019), as well as biofeedback systems for physical rehabilitation of patients with Multiple Sclerosis (Moumdjiam et al., 2019). Typically, the sound design would interfere with the human action–perception coupling in view of changing the behavior.

In Van den Berghe et al. (2021, 2022), for example, a biofeedback system is used to relearn the behavior of a recreational runner. A runner's way of running will likely cause knee injuries over time when the impulse levels measured at the tibia exceed a particular threshold. The impulse level is an indicator of unwanted or wanted behavior, and it can be measured with an accelerometer. Based on that information, the interactive sound design can be changed. The current system, called Low-Impact Runner, adds noise to music that is nicely synchronized with the running. Accordingly, a reinforcement learning paradigm applies in which the impulse level drives the amount of noise added to the heard music. If the measured impact level is too high, a high noise level is added; if the measured impact level was high and is later lowered, noise is lowered. As such, it is possible to drive the runner's running style towards a new (self-chosen) running style that generates less impact. Such interactive sound designs are based on principles of intuitive and embodied responses rather than warning signals that would call for analytic thinking and inference. In our example, a balance is regulated between a highly enjoyable and motivating stimulus, that is, preferred music whose tempo is nicely aligned with the regularities of movement during running (using DJogger, see Moens et al., 2014), and a highly annoying disturbance of that same stimulus, based on different noise levels (see Lorenzoni et al., 2019a, b). While music engages the runner in a self-augmented interaction state, noise regulates the degree of annoyance and adaptation. The result is a powerful system having effect sizes in the order of $>25\%$!

6.2 Ethical Considerations of Sound Design

Whether a biofeedback-based interactive sound design is feasible for preterm infants in the NICU is a matter of careful research strategy. It would be possible to detect awake states in infants, and depending on the infant's activity during that state, stimuli could be provided that afford body responses. Such responses can be monitored, and immediate sound feedback can be provided in view of establishing particular interaction states. However, further work will have to tell us whether this type of sound design is feasible and effective in terms of psychosocial empowerment. If we really push forward our theoretical ideas, an interactive design should also consider gestural design together with sound design. Based on the theoretical insights, the neonatologist's request would imply that new types of sound design are not only interactive but also embodied. However, the essence is that they would create interaction states with a window of opportunity for generating beneficial effects.

However, whatever adaptive sound design approach is used, it will raise ethical issues. On the one hand, the failure of a sound design strategy may have severe consequences for the later development of the preterm infant. For example, when the wanted effect of a decrease in stress turns out to be an increase in stress, this may impact the infant's development. On the other hand, if we are almost sure that a sound design would work, can we then set up an experiment where one group serves as a control (not using the sound design)? These and other issues need further consideration and development.

7 Conclusion

The request of a neonatologist to develop a sound design for preterm infants in the NICI was used here as a common thread for considerations about human–sound interactions and their effects. Using the concept of self-augmentation, it was argued that sound design components that match human capacities can lead to interaction outcomes beyond the human's individual reach. It is assumed that such outcomes offer windows of opportunity for possible powerful effects and empowerment. This theory of self-augmented interaction evolved over several decades of research in musicology, and it was developed at pace with trends in cognitive science and technological developments. As a measure of its success, one may count the number of PhDs and big research projects and laboratory facilities created in view of this kind of music research over two decades in Europe. Further development of this theory, or at least our interpretation of the theory, depends on evidence-based applications in domains such as interactive arts, sports, and physiotherapy. As it stands, it seems that multi-media-based biofeedback systems are the key to testing the ultimate power of the theory.

Acknowledgment. The idea that sound design requires a theory of gestural music interaction has been advocated by Rolf Inge Godøy since the early years of the "Arbeitskreiss" (called: ISSM or International Society of Systematic Musicology). Founded in the mid-1990s, the ISSM initiated numerous summer schools and conferences advocating the use of new technologies and novel empirical strategies in musicology. Rolf Inge's contribution to a proactive musicology, in line with the concept of the "Arbeitskreiss," has been and still is much appreciated, and this chapter is a tribute to Rolf Inge's achievements.

References

Bader, R. (ed.): Springer Handbook of Systematic Musicology. SH, Springer, Heidelberg (2018). https://doi.org/10.1007/978-3-662-55004-5

Beltrán, M.I., Dudink, J., de Jong, T.M., et al.: Sensory-based interventions in the NICU: systematic review of effects on preterm brain development. Pediatr. Res. **92**, 47–60 (2022)

Bonicco-Donato, D.: Une archéologie de l'interaction. De David Hume à Erving Goffman. Vrin, Paris (2016)

Clayton, M., Sager, R., Will, U.: In time with the music: the concept of entrainment and its significance for ethnomusicology. In: European Meetings in Ethnomusicology, vol. 11, pp. 1–82. Romanian Society for Ethnomusicology (2005)

Collins, T., Tillmann, B., Barrett, F.S., Delbé, C., Janata, P.: A combined model of sensory and cognitive representations underlying tonal expectations in music: from audio signals to behavior. Psychol. Rev. **121**(1), 33 (2014)

François, C., Rodriguez-Fornells, A., Teixidó, M., Agut, T., Bosch, L.: Attenuated brain responses to speech sounds in moderate preterm infants at term age. Dev. Sci. **24**, e12990 (2021)

Godøy, R.I.: Motor-mimetic music cognition. Leonardo **36**(4), 317–319 (2003)

Godøy, R.I.: Gestural affordances of musical sound. In: Musical Gestures, pp. 115–137. Routledge (2010)

Goldin-Meadow, S.: Hearing Gesture: How Our Hands Help Us Think. Harvard University Press, Cambridge (2005)

Huron, D.: Sweet Anticipation: Music and the Psychology of Expectation. MIT Press, Cambridge (2008)

Koelsch, S., Vuust, P., Friston, K.: Predictive processes and the peculiar case of music. Trends Cogn. Sci. **23**(1), 63–77 (2019)

Kahneman, D.: Thinking, fast and slow. Farrar, Straus Giroux (2011)

Langner, G.D.: The Neural Code of Pitch and Harmony. Cambridge University Press, Cambridge (2015)

Leman, M.: Embodied Music Cognition and Mediation Technology. MIT Press, Cambridge (2007)

Leman, M.: The Expressive Moment: How Interaction (with Music) Shapes Human Empowerment. MIT Press, Cambridge (2016)

Lorenzoni, V., Van den Berghe, P., Maes, P.J., De Bie, T., De Clercq, D., Leman, M.: Design and validation of an auditory biofeedback system for modification of running parameters. J. Multimodal User Interfaces **13**(3), 167–180 (2019a)

Lorenzoni, V., Staley, J., Marchant, T., Onderdijk, K.E., Maes, P.J., Leman, M.: The sound instructor: a music-based biofeedback system for improving weightlifting technique. Plos One **14**(8), e0220915 (2019b)

Maes, P.J., Leman, M., Palmer, C., Wanderley, M.M.: Action-based effects on music perception. Front. Psychol. **4**, 1008 (2014)

Maes, P.J., Lorenzoni, V., Six, J.: The SoundBike: musical sonification strategies to enhance cyclists' spontaneous synchronization to external music. J. Multimodal User Interfaces **13**(3), 155–166 (2019)

Malloch, S.: Mothers and infants and communicative musicality. Music. Sci. **3**, 29–57 (1999)

McGowan, T., Delafield-Butt, J.: Narrative as co-regulation: a review of embodied narrative in infant development. Infant Behav. Dev. **68**, 101747 (2022)

Moens, B., et al.: Encouraging spontaneous synchronisation with D-Jogger, an adaptive music player that aligns movement and music. PLoS ONE **9**(12), e114234 (2014)

Moumdjian, L., et al.: Walking to music and metronome at various tempi in persons with multiple sclerosis: a basis for rehabilitation. Neurorehabil. Neural Repair **33**(6), 464–475 (2019)

Phillips-Silver, J., Keller, P.E.: Searching for roots of entrainment and joint action in early musical interactions. Front. Hum. Neurosci. **6**, 26 (2012)

Schiavio, A., Maes, P.J., van der Schyff, D.: The dynamics of musical participation. Music. Sci. **26**(3), 604–626 (2022)

Sears, D.R., Pearce, M.T., Spitzer, J., Caplin, W.E., McAdams, S.: Expectations for tonal cadences: sensory and cognitive priming effects. Q. J. Exp. Psychol. **72**(6), 1422–1438 (2019). https://doi.org/10.1177/1747021818814472

Seth, A.K.: The cybernetic Bayesian brain. Open mind. Frankfurt am Main: MIND Group (2014)

Schneider, A.: Pitch and pitch perception. In: Bader, R. (ed.) Springer Handbook of Systematic Musicology. SH, pp. 605–685. Springer, Heidelberg (2018a). https://doi.org/10.1007/978-3-662-55004-5_31

Schneider, A.: Perception of timbre and sound color. In: Bader, R. (ed.) Springer Handbook of Systematic Musicology. SH, pp. 687–725. Springer Heidelberg (2018b). https://doi.org/10.1007/978-3-662-55004-5_32

Trehub, S.: Communication, music, and language in infancy. In: Arbib, M.A. (ed.) Language, Music, and the Brain. Strüngmann Forum Reports, vol. 10. MIT Press (2013)

Trevarthen, C., Aitken, K.J.: Infant intersubjectivity: research, theory, and clinical applications. J. Child Psychol. Psychiatry Allied Disciplines **42**(1), 3–48 (2001)

Trevarthen, C.: The musical art of infant conversation: narrating in the time of sympathetic experience, without rational interpretation, before words. Music. Sci. **12**(1_suppl), 15–46 (2008)

Trost, W.J., Labbé, C., Grandjean, D.: Rhythmic entrainment as a musical affect induction mechanism. Neuropsychologia **96**, 96–110 (2017)

Van den Berghe, P., et al.: Music-based biofeedback to reduce tibial shock in over-ground running: a proof-of-concept study. Sci. Rep. **11**(1), 1–12 (2021)

Van den Berghe, P., et al.: Reducing the peak tibial acceleration of running by music-based biofeedback: a quasi-randomized controlled trial. Scand. J. Med. Sci. Sports **32**(4), 698–709 (2022)

Van Puyvelde, M., et al.: The interplay between tonal synchrony and social engagement in mother–infant interaction. Infancy **18**(5), 849–872 (2013)

Musical Meter as Shape: An Embodied Perspective on Metrical Trajectories and Curves

Mari Romarheim Haugen(✉)

Department of Musicology, RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo, Oslo, Norway
m.r.haugen@imv.uio.no

Abstract. The perception of musical rhythm includes not only the *sonic rhythm* but also the *endogenous reference structures,* such as *meter*. Musical meter is often described and understood as points in time or durations between such points. In this chapter, I argue that musical meter also has a *shape*. I propose that we perceive and make sense of musical meter based on our previous musical experiences involving meter-related bodily motion. In other words, the meter-related motion is integral to the perceived meter—they are the same. Meter thus has a *shape* that relates to the embodied sensations of these movements. Also crucial is the notion that musical meter is conditioned by musical culture. This perspective on meter as shape is highly influenced by Godøy's *motor-mimetic perspective on music perception* and *musical shape cognition* and concurs with the multimodal approach to sonic design that acknowledges motion as intrinsic to music performance and perception.

Keywords: Musical meter · rhythm · multimodality · music culture · musical shape cognition

1 Introduction

The perception of musical rhythm involves the interaction between the *sonic rhythm* (also referred to as the *rhythmic surface* (e.g., London, 2012) and *sounding rhythm* (Honing, 2013) and the *endogenous reference structures,* such as *pulse* (also referred to as the *beat* (e.g., Honing, 2013), *regulative beat* (Nketia, 1986), *subjective beat* (Chernoff, 1979), *tactus* (London, 2012), *inner pulsation* (Kubik, 1990), and *internal beat* (Danielsen, 2006)) and *meter*. Such structures are not necessarily represented by sonic events, but instead supply an implicit framework against which one perceives them (e.g., Danielsen 2010; Haugen 2016b; London 2012). Whereas the pulse comprises a single periodicity, the meter groups or organizes that pulse. Meter consists of a minimum of two hierarchically organized periodicities on different time scales: a pulse or tactus (referent) level that is coordinated with one or more levels of organization, for example, the pulse level, an ordering of pulse beats into measures (e.g., double and triple meter), and subdivisions of the pulse (e.g., London 2012).

Meter is often described as successive time points (e.g., Lerdahl & Jackendoff, 1983) or as the durations between the points (e.g., Bengtsson, Gabrielsson, & Thorsén, 1969;

A. R. Jensenius (Ed.): SSD 2022, CRSM 12, pp. 37–48, 2024.
https://doi.org/10.1007/978-3-031-57892-2_3

Kvifte, 2004). In this article, I argue that meter also has a *shape*. This is not an entirely new idea. Some theorists have proposed various conceptions of meter that stress its continuous and dynamic aspects from different perspectives. Meter has, for example, been explored as composer-specific motion curves (e.g., Becking, 1928; Clynes, 1995), an underlying dynamic flow of an "away from–back to" *cycle* (Zuckerkandl, 1956), continuous pulsations of up-and-down motion trajectories (Waadeland, 2000), projection and process (Hasty, 1997), dynamic attending (e.g., Jones, 2019; Large & Jones, 1999), entrainment of attentional periodicities (London, 2012), and beat-bins (Danielsen, 2010, 2019). Fundamental to the present perspective is the conviction that meter is intrinsically related to motion, and that meter perception is influenced by people's previous embodied experiences and music-cultural background. This perspective is highly influenced by embodied music cognition (e.g., Godøy & Leman, 2010) in general and Rolf Inge Godøy's *motor-mimetic perspective on music perception* (e.g., 2003, 2006, 2010) and *musical shape cognition* (2019) in particular. It contributes to the multimodal approach to sonic design that acknowledges that motion is intrinsic to music perception.

2 Music and Motion

Essential to the present view on meter is the notion that human perception is *multimodal* in nature. This refers to how we use multiple senses simultaneously when we explore our environment (e.g., Gibson, 1966). It has also been pointed out that integrating several modalities is the optimal strategy for perception since it achieves a better understanding of the world (Ernst & Bülthoff, 2004). Within this approach, perception is an active process–it is something that we *do*, and which is related to sense-making and based on previous multimodal experiences (e.g., Noë, 2004; Shapiro, 2010; Varela, Thompson, & Rosch, 2016). *Motor theories of perception*, for example, point out that sound perception includes not only auditory input but also an understanding of what we believe caused the sound–that is, the sound's *source* and/or the *action* that produced the sound (e.g., Berthoz, 2000; Laeng, Kuyateh, & Kelkar, 2021; Liberman & Mattingly, 1985). Sound perception, then, includes knowledge of *sound-source* relationships based on previous multimodal experiences. Accordingly, Gaver (1993) proposes an *ecological approach to auditory event perception* and highlights that sound is informative not only about its source but also about the materials involved and their interaction, environment, and location (direction). In the same vein, Bennett Hogg points out that sounds "do not carry meaning in and of themselves, but are the sites of complex and mediated sets of relationships between physical sounds, perceptual systems, personal associations, culturally signifying gestures, bodily and emotional responses, observed actions and reactions, and culturally learned listener expectations" (Hogg, 2011, p. 88).

Clarke (2005) proposes an ecological approach to musical meaning, highlighting perception as sense-making and noting that, when we hear a sound and recognize what produced the sound, we grasp its perceptual meaning. He then criticizes the information-processing approach, which holds that perception starts with stimulus-driven simple features that are subsequently combined into more complex structures (ib. p. 14). He refers to Gibson's (1966) concept of *direct experience* to argue that there is no need for complex processing or the interpretation of stimulus information—instead, the information is directly specified by the structure of the environment. When we perceive a sound, for example, of piano playing, we will immediately recognize the sound as what it is, without any complex processing.

Arnie Cox focuses on the importance of mimetic behavior in music cognition. He hypothesizes that "part of how we comprehend music is by imitating, covertly or overtly, the observed sound-producing actions of performers" (Cox, 2016, p. 12). Rolf Inge Godøy suggests a *motor-mimetic perspective on music perception* (e.g., 2003, 2006, 2010, 2019). Along the lines of Pierre Schaeffer's terminology for describing sonic objects (Schaeffer, North, & Dack, 2017), Godøy suggests that the simulated sound-producing actions that we relate to the perceived musical sounds can be directly related to playing an instrument, but they can also be imitative of *sonic shapes.* In other words, we can perceive a sound as a sonic shape, including a corresponding simulated action with a similar shape. These actions and their corresponding sound shapes, he explains, usually fall into one of three main categories: *impulsive, sustained,* or *iterative.* We perceive these *action–sound shape* relationships as meaningful units due to our multimodal perception. For example, we know from experience that a sustained sound-producing action with continuous energy transfer (e.g., stroking) will produce a sustained sound, whereas an impulsive sound-producing action with a fast attack (e.g., hitting) will produce an impulsive sound (Godøy, 2011). Furthermore, we recognize similarities between sound-producing actions with a particular shape and other kinds of motion with a similar shape, as, for example, in dance and sound tracing (Godøy, Song, Nymoen, Haugen, & Jensenius, 2016). Godøy (2010) exemplifies this relationship with the barbershop scene from Charlie Chaplin's The Great Dictator. In this scene, Chaplin shaves a customer to the accompaniment of Brahms's Hungarian Dance No. 5, and his shaving motions appear to correlate perfectly with the musical sound.

Music-related motion involves not only the actions related to sound production and perception but also the gestural repertoire associated with the specific music culture in question, such as typical movement patterns or dance (e.g., Haugen, 2016b; Naveda, 2011). Here, *music culture* refers to that which arises when multiple people share a repertoire of musical concepts and practices (e.g., Baily, 1985; Blacking, 1955; Clayton, Dueck, & Leante, 2013). It includes everything that allows cultural insiders to recognize a given music genre, such as typical instruments, sonic features, phrasings, timing, ways of singing and/or playing, and signature motion patterns. This understanding of music culture takes into account that our experiences with and general exposure to music are more relevant than our geographical area as such (see also, Jacoby et al., 2020; Trehub, Becker, & Morley, 2015).

3 Meter-Related Motion

The close relationship between meter and motion is often highlighted in the literature. Periodic body motion such as foot tapping, body swaying, head nodding, and dance moves are often labeled "entrained" motion since it follows the perceived meter (e.g., Dahl et al., 2010; Jensenius, 2007; Merchant, Grahn, Trainor, Rohrmeier, & Fitch, 2015). Some researchers have suggested that such repetitive music-related body motions are rooted in *basic gestures.* The concept of the basic gesture can be traced back to Becking

(1928) and defined as a three-dimensional repeating motion pattern of a body part during one period of a repetitive sequence, whereby its shape will be such that the starting point and the ending point will be connected. In an exploratory study by Styns and Van Noorden (2006), people were asked to move a joystick while listening to march music, baroque string music, and a metronome, all played at a constant tempo (120 beats per minute). The analysis showed that most people synchronized their motion to the pulse of the music, but the ways in which they moved varied according to the musical content. Van Noorden (2010) later observed that the participants tended to use a limited set of movement strategies or basic gestures. Visualizations of the participants' motion patterns in space revealed motion patterns shaped like "raindrops", "figure-eights", and "bananas."

Basic gestures have also been investigated in music and dance research. Naveda and Leman highlighted the intimate relationship between music and dance in many music genres and noted that repetitive motion patterns in these dance styles are commonly synchronized with the musical meter (Naveda, 2011; Naveda & Leman, 2009, 2010). They then suggested that such repetitive dance patterns are based on spatiotemporal reference frames or basic gestures. They developed a method through which metrical points derived from the musical sound could be projected onto basic gestures extracted from repetitive motion in the corresponding dance (Leman & Naveda, 2010), then used it to compare basic gestures in performed samba and Charleston dance. The basic gestures were obtained from motion-capture recordings of repetitive motion using the hand, torso, head, and foot in the dances. They observed that certain motion forms (for example, round and arc-like) and periodicities related to different metrical levels.

Several ethnomusicological studies have argued that in music cultures where music and dance have evolved together under mutual influence, the meter must be understood in relation to the musicians' and dancers' bodies. Bengtsson (1974), for example, points out that, in such genres, the underlying meter may be both conditioned by the dance *and* intrinsic to the music, even when the music is detached from the actual dancing. In a study of Brazilian drum patterns, Kubik (1990) explains that the percussionists' "inner pulsation" is often not present in the sound, but is often visible in the performers' and dancers' body motion. Agawu (2003) points out that, in many genres in West and Central Africa, there is an interaction between specific periodic sonic rhythms, often referred to as *time-lines* (*topoi* in Agawu's (2003) terminology), and the meter. In many time-line genres, the music and the dance took shape together, and the pulse in performance is often expressed by the dancers' feet. For cultural insiders, then, the perception of a standard pattern, or time-line, will instinctively and spontaneously incorporate either the actual dancers' feet or an image of their motion. People unfamiliar with the music genre's intrinsic way of moving may perceive and understand its sonic rhythm patterns differently (Agawu, 2003; Naveda, 2011).

Scandinavian folk music is yet another tradition featuring an intimate relationship between music and dance, and scholars often highlight that meter in this music should be understood in relation to the periodic motion in the corresponding dances (e.g., Bakka, 1978; Blom, 1981; Omholt, 2009). Norwegian anthropologist and ethnomusicologist Jan-Petter Blom (1927–2021), for example, was interested in this correspondence and highlighted the influence of music culture on rhythm production and perception. Accordingly, he proposed a *motor theory of rhythm* to capture that "culture-specific movement styles of a social group represent shared kinaesthetic experiences embedded in its musical forms of expression, thus constituting the implicit and shared background knowledge from which socially appropriate rhythmic actions/reactions are generated" (Blom, 2006, p. 79). Blom also emphasized that musical meter should be understood in relation to any corresponding dance and, in the case of Scandinavian folk music and dance, to the vertical motion pattern of the dancers' center of gravity in particular (see, e.g., Blom, 1981, 1993, 2006). Blom observed that the vertical motion of the dancers' center of gravity, caused by bending and stretching the hips, knees, ankles, and joints of the feet, seemed to follow a regular up-and-down pattern that he called the dancers' *libration pattern* or *libration curves*. He noted that this pattern was repeated in each measure regardless of the different steps and tunings in the dance. The execution of the libration patterns in terms of the number of oscillations, position of turning points, and overall shape are considered style-specific and directly linked to the musical meter (Blom, 1981).

4 Toward a Theory of Meter as Shape

Crucial to the perspective on meter as shape proposed in this chapter, is the conviction that meter is intrinsically related to motion, and that meter perception is influenced by personal experience and music culture. Fundamental to it, as well, is an understanding of the experienced rhythm as an interaction between the sonic rhythm and the meter—something I will unpack further below.

As pointed out in the introduction, the experienced rhythm includes not only the perception of sonic events but also endogenous reference structures such as meter. Central to the present perspective is the insight that the *experienced rhythm* emerges via an *interaction* between the sonic rhythm and the meter (see Fig. 1). Note that, in this case, the *experienced rhythm* does not refer to the perceived sound alone but rather to *both* the sonic rhythm and the meter simultaneously. From this perspective, the meter is not derived from the sonic rhythm; instead, the sonic rhythm and the meter are mutually dependent. As a result, we experience the sonic rhythm and the meter not as discrete entities but as aspects of the experienced rhythm. I would argue, then, that musical meter is also learned in this context—that is, during musical *sonic rhythm–meter* interactions. Meter perception is influenced by the sonic rhythm–meter interactions to which we are most often exposed and with which we are most familiar, based on our previous musical experiences and music-cultural backgrounds. This conviction is also in line with Kvifte's (2007) *pattern-recognition* concept, which highlights the importance of the perceiver's knowledge and experience in metrical interpretation, and London's (2012) concept of metric *recognition*, which claims that, in meter perception in familiar music genres, one matches the sonic rhythm against a repertoire of well-known rhythmic/metric templates (London, 2012, p. 67).

Moreover, I argue that meter-related motion, such as foot tapping, body swaying, head nodding, and repetitive dance moves, are not only externalizations of a perceived meter but also the way in which we both learn and shape that meter during musical experiences. In parallel with Godøy's aforementioned *motor-mimetic perspective on* *music perception* (2003, 2006, 2010), which suggests that we make sense of perceived sounds based on our previous experience with how sounds are produced–that is, action–sound relationships, I propose that a *meter–motion shape* relationship is implicated in meter perception. I suggest we do not perceive the meter as one thing and meter-related motion as something else. Instead, we understand meter–motion relationships as meaningful wholes due to previous musical experiences involving meter-related bodily motion. Meter thus has a *shape* that includes sensations of what it feels like to move the body in space in a particular manner—for example, in relation to gravity and/or qualitative motion features such as weight and flow (see, e.g., Laban, 1960, on effort). Once the meter is acquired, one does not need to see or perform its intrinsic periodic body motion shape to perceive it. It is inherent in the perceived meter, either overtly or covertly.

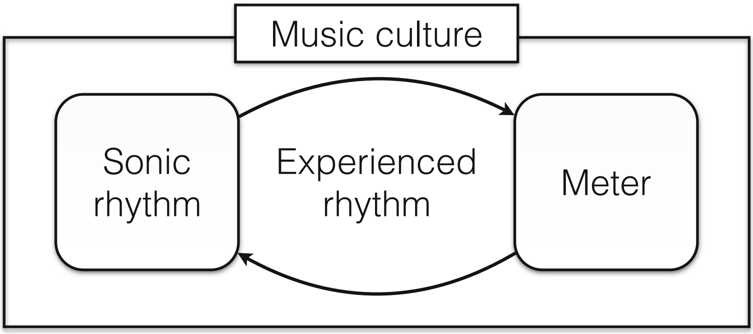


Fig. 1. An illustration of *experienced rhythm* as an interaction between the perceived *sonic rhythm* and *meter*, influenced by music culture.

As stated above, like action–sound shape relationships in sound perception, I propose that meter–motion shape relationships in meter perception are conditioned by previous experiences. However, since musical sound and the motion associated with it differs considerably among genres and music cultures, and musical meter always occurs in musical contexts, metrical shapes are dependent on *music culture*—how one usually moves—and also by each person's embodied experience with the culture's meter-related motion. I will refer to those meters spontaneously perceived by cultural insiders as *culture-specific meters*. This is not to say that this culture-specific meter is the only perceivable meter possible but rather that it, including its shape, is likely to be quite consistent among the people conversant with the music culture in question. A familiar music genre will automatically evoke the culture-specific metrical shape. An unfamiliar music genre with an unfamiliar metrical shape might be experienced within a metrical-gestural framework with which the perceiver is familiar—that is, with a familiar metrical shape. To learn to know a new metrical shape, then, one has to acquire some embodied knowledge of the meter in question. I also suggest that perceived metrical shapes are not necessarily fixed but can vary during a musical performance—for example, due to a perceived stylistic change in the middle of the piece.

5 Metrical Shapes in Norwegian Folk Music and Dance

The aforementioned Scandinavian folk music and dance genres are interesting examples of the importance of culture-specific embodied metrical shapes in music performance and perception. In this final section, I will exemplify the present perspective on meter as shape via the Norwegian folk music genre telespringar. Springar tunes are among the older types of Norwegian folk dances, and telespringar is a springar from the region of Telemark in Norway, performed by couples. It can be sung or played on several traditional instruments but is most often played on a Hardanger fiddle. Telespringar is normally notated in triple meter, but it is commonly understood by cultural insiders that the beats are of uneven duration—what is often referred to as asymmetrical triple meter—and follow a long–medium–short duration pattern (e.g., Blom, 1981; Groven, 1971; Kvifte, 1999). Telespringar derives from oral traditions, and its music and dance developed together under conditions of mutual influence. The intimate relationship between music and dance is often emphasized in rhythm studies of telespringar and, in particular, when it comes to meter. As previously mentioned, it has been suggested that meter in these genres should be understood in relation to performers' periodic motion, and, specifically, the fiddler's foot stamping, which is integrated into this tradition of playing, but also the vertical motion pattern of the dancers' center of gravity in the corresponding dance (e.g., Blom 1981; Kaminsky 2014; Kvifte 2007).

To investigate the meter-related body motion in telespringar, I carried out a motion capture study involving three experienced telespringar performers: a fiddler playing the Hardanger fiddle and a dance couple (Haugen, 2016a, 2017). The presence of an asymmetrical beat-duration pattern was supported by the analysis that revealed that the musician's integrated foot stamping followed a very regular long–medium–short pattern (Fig. 2b). The analysis of the dancers' periodic vertical motion also showed a very regular motion pattern at a beat level that consisted of a small "valley-shaped" down–up motion during the long beat 1, a deeper down–up motion during the medium-long beat 2, and a small up–down motion during the short beat 3 (Fig. 2a).

Interestingly, telespringar dancers do not refer to beat durations. For example, when they teach telespringar dance, they do not talk about a long–medium–short pattern but rather a heavy–heavier–light pattern (Omholt, 2011). This "weight" pattern seems to correspond well to this curve, since the deepest "valley-shaped" beat 2 might feel "heavier" than beat 1, and beat 3, which has a small up–down motion, might feel light. And it is not only the dancers but also the musicians who refer to the beat in terms of felt weight. In a recent study by Mats Johansson (2022), where he interviewed folk musicians about timing-sound interactions in traditional Scandinavian fiddle music, the musicians explained how the integrated foot tapping influences their playing in terms related to force or weight, describing the foot stamping as "heavy" and "light," and even explaining that some beats should be played with an "upwards" feeling.

The notion that insiders experience telespringar meter as patterns of force and/or weight was also supported by the motion capture study (Haugen, 2017), wherein the musician's acceleration curves based on foot stamping revealed a high–higher–low pattern (Fig. 2c). This pattern indicates that more power is put into the first two foot stamps than into the third, resulting in a strong–stronger–weak pattern, which seems to correspond well to the dancers' heavy–heavier–light pattern.

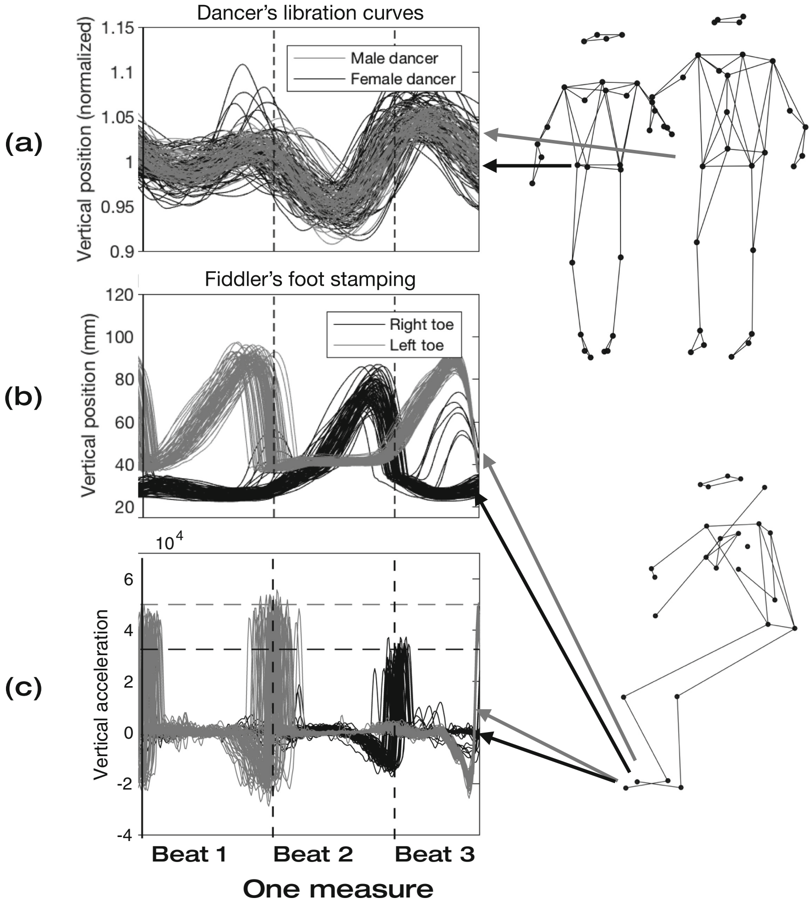


Fig. 2. Plots showing (a) the vertical position of the dancers' hips (libration curves), (b) the vertical position of the fiddler's foot stamping, and (c) the vertical acceleration of the fiddler's foot stamping. All three are chunked into segments of one measure and plotted on the same graph.

The analysis above suggests that all of the performers shared an understanding of the music's metrical shape, which seems to relate to the traditional ways of moving in the particular genre. I suggest that these motion patterns are integral to the culture-specific meter in telespringar. In other words, the meter includes not only points in time but also a shape that relates to the embodied sensations of these motion patterns. In that case, we can assume that people unfamiliar with the motion intrinsic to telespringar music might experience its rhythm differently from cultural insiders.

6 Concluding Remarks

In this chapter, I have presented a perspective on musical meter that highlights the intimate relationship between meter and motion. I suggest that meter is essentially learned and shaped through periodic body motion in musical contexts. Namely, the perceived meter and the corresponding meter-related motion are intrinsically related—they are the same. It follows from this that meter is continuous and has a shape that relates to the embodied sensations of these movements. In other words, I propose that meter perception includes meter–motion shape relationships. This approach is highly inspired by Godøy's (e.g., 2003, 2006, 2010, 2019) motor-mimetic perspective on music perception. It also contributes to the multimodal approach to sonic design, emphasizing the embodied aspects of rhythm production and perception in music, including those intrinsic to meter. I also highlight that meter is conditioned by a person's music-cultural background and embodied experience with the music culture's meter-related motion. It also suggests that individuals with different embodied experiences will perceive the musical meter, and consequently the rhythm, differently. This perspective implies an acknowledgment of the crucial role of embodied knowledge in musical experiences in general. If we have some embodied experience with the gestural repertoire commonly associated with a particular music genre, including its meter-related motion, we might gain a deeper understanding of the music as such.

References

Agawu, V.K.: Representing African Music: Postcolonial Notes, Queries, Positions. Routledge, New York (2003)

Baily, J.: Music structure and human movement. In: Howell, P., Cross, I., West, R. (eds.) Musical Structure and Cognition, pp. 237–258. Academic Press, London (1985)

Bakka, E.: Norske Dansetradisjonar, vol. 15. Samlaget, Oslo (1978)

Becking, G.: Der musikalische Rhythmus als Erkenntnisquelle. Benno Filser, Augsburg (1928)

Bengtsson, I.: On notation of time, signature and rhythm in Swedish polskas. In: Hilleström, G. (ed.) Studia Instrumentorum Musicae Popularis III, pp. 22–31. Nordiska Musikförlaget, Stockholm (1974)

Bengtsson, I., Gabrielsson, A., Thorsén, S.-M.: Empirisk rytmforskning: en orientering om musikvetenskaplig bakgrund, utveckling och några aktuella projekt. Svensk tidskrift musikforskning **51**, 49–118 (1969)

Berthoz, A.: The Brain's Sense of Movement. Harvard University Press, Cambridge (2000)

Blacking, J.: Some notes on a theory of african rhythm advanced by erich von hornbostel. Afr. Music **1**(2), 12–20 (1955). http://www.jstor.org/stable/30249422

Blom, J.-P.: The dancing fiddle. On the expression of rhythm in Hardingfele Slåtter. In: Blom, J.P., Nyhus, S., Sevåg, R. (eds.) Norsk Folkemusikk, vol. VII, pp. 305–312.Universitetsforlaget, Oslo (1981)

Blom, J.-P.: Rytme og frasering - forholdet til dansen. In: Aksdal, B., Nyhus, S. (eds.) Fanitullen. Innføring i norsk og samisk folkemusikk, pp. 161–184. Universitetsforlaget, Oslo (1993)

Blom, J.-P.: Making the music dance: dance connotations in Norwegian fiddling. In: Russell, I., Alburger, M.A. (eds.) Play it like it is. Fiddle and dance studies from around the Noth Atlantic, vol. 5, pp. 75–86. The Elphinstone Institute, University of Aberdeen (2006)

Chernoff, J.M.: African Rhythm and African Sensibility: Aesthetics and Social Action in African Musical Idioms. University of Chicago Press (1979)

Clarke, E.F.: Ways of Listening: An Ecological Approach to the Perception of Musical Meaning. Oxford University Press, Oxford (2005)

Clayton, M., Dueck, B., Leante, L.: Experience and Meaning in Music Performance. Oxford University Press, Oxford (2013)

Clynes, M.: Microstructural musical linguistics: composers' pulses are liked most by the best musicians. Cognition **55**(3), 269–310 (1995). https://doi.org/10.1016/0010-0277(94)00650-A

Cox, A.: Music and Embodied Cognition. Listening, Moving, Feeling, and Thinking. Indiana University Press, Bloomington (2016)

Dahl, S., et al.: Gestures in performance. In: Godøy, R.I., Leman, M. (eds.) Musical Gestures. Sound, Movement, and Meaning, pp. 36–68. Routledge, New York (2010)

Danielsen, A.: Presence and Pleasure: The Funk Grooves of James Brown and Parliament. Wesleyan University Press (2006)

Danielsen, A.: Musical Rhythm in the Age of Digital Reproduction. Ashgate, Farnham (2010)

Danielsen, A.: Pulse as dynamic attending. analysing beat bin metre in neo soul grooves. In: Scotto, C., Smith, K.M., Brackett, J.L. (eds.) The Routledge Companion to Popular Music Analysis: Expanding approaches, pp. 179–189. Routledge/Taylor & Francis Group, New York/London (2019)

Ernst, M.O., Bülthoff, H.H.: Merging the senses into a robust percept. Trends Cogn. Sci. **8**(4), 162–169 (2004). https://doi.org/10.1016/j.tics.2004.02.002

Gaver, W.W.: What in the world do we hear?: an ecological approach to auditory event perception. Ecol. Psychol. **5**(1), 1–29 (1993). https://doi.org/10.1207/s15326969eco0501_1

Gibson, J.J.: The Senses Considered as Perceptual Systems. Houghton Mifflin, Boston (1966)

Godøy, R.I.: Motor-mimetic music cognition. Leonardo **36**(4), 317–319 (2003). http://www.jstor.org/stable/1577332

Godøy, R.I.: Gestural-sonorous objects: embodied extensions of schaeffer's conceptual apparatus. Organ. Sound **11**(02), 149–157 (2006). https://doi.org/10.1017/S1355771806001439

Godøy, R.I.: Gestural affordances of musical sound. In: Godøy, R.I., Leman, M. (eds.) Musical Gestures: Sound, Movement, and Meaning, pp. 103–125. Routledge, New York (2010)

Godøy, R.I.: Sound-action chunks in music. In: Solis, J., Ng, K. (eds.) Musical Robots and Interactive Multimodal Systems. Springer Tracts in Advanced Robotics, vol. 74, pp. 13–26. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22291-7_2

Godøy, R.I.: Musical shape cognition. In: Grimshaw-Aagaard, M., Walther-Hansen, M., Knakkergaard, M. (eds.) The Oxford Handbook of Sound and Imagination, vol. 2. Oxford University Press, New York (2019)

Godøy, R.I., Leman, M.: Musical Gestures: Sound, Movement, and Meaning. Routledge, New York (2010)

Godøy, R.I., Song, M., Nymoen, K., Haugen, M.R., Jensenius, A.R.: Exploring sound-motion similarity in musical experience. J. New Music Res. 1–13 (2016). https://doi.org/10.1080/09298215.2016.1184689

Groven, E.: Musikkstudiar - ikkje utgjevne før. 1. Rytmestudiar. In: Fjalestad, O. (ed.) Eivind Groven. Heiderskrift til 70-årsdagen 8. oktober 1971, pp. 93–102. Noregs boklag, Oslo (1971)

Hasty, C.F.: Meter as Rhythm. Oxford University Press, New York (1997)

Haugen, M.R.: Investigating periodic body motions as a tacit reference structure in Norwegian telespringar performance. Empir. Musicol. Rev. **11**(3–4), 272–294 (2016a). https://doi.org/10.18061/emr.v11i3-4.5029

Haugen, M.R.: Music–dance. Investigating rhythm structures in Brazilian samba and Norwegian telespringar performance. Ph.D. thesis. University of Oslo, 07 Oslo Media, Oslo (2016b). http://urn.nb.no/URN:NBN:no-56252

Haugen, M.R.: Investigating musical meter as shape: two case studies of Brazilian samba and Norwegian telespringar. In: Van Dyck, E. (ed.) Proceedings of the 25th Anniversary Conference of the European Society for the Cognitive Sciences of Music, pp. 67–74. University of Ghent, Ghent (2017)

Hogg, B.: Enactive consciousness, intertextuality, and musical free improvisation: deconstructing mythologies and finding connections. In: Clarke, D., Clark, E. (eds.) Music and Consciousness:

Philosophical, Psychological, and Cultural Perspectives, pp. 79–93. Oxford University Press, Oxford (2011)

Honing, H.: Structure and interpretation of rhythm in music. In: Deutsch, D. (ed.) Psychology of Music, pp. 369–404. Academic Press, San Diego, Calif. (2013)

Jacoby, N., et al.: Cross-cultural work in music cognition. Music Percept.: Interdisc. J. **37**(3), 185–195 (2020). https://doi.org/10.1525/mp.2020.37.3.185

Jensenius, A.R.: Action–sound: developing methods and tools to study music-related body movement. Ph.D. thesis. University of Oslo, Oslo (2007)

Johansson, M.: Timing-sound interactions. Groove-forming elements in traditional scandinavian fiddle music. PULS, 7 (2022)

Jones, M.R.: Time Will Tell: A Theory of Dynamic Attending. Oxford University Press, New York (2019)

Kaminsky, D.: Total rhythm in three dimensions: towards a motional theory of melodic dance rhythm in Swedish Polska music. Dance Res. **32**(1), 43–64 (2014). https://doi.org/10.3366/drs.2014.0086

Kubik, G.: Drum patterns in the “batuque” of benedito caxias. Latin Am. Music Rev./Revista de Música Latinoamericana **11**(2), 115–181 (1990)

Kvifte, T.: Fenomenet “asymmetrisk takt” i norsk og svensk folkemusikk. Stud. Musicol. Nor. **25**, 387–430 (1999)

Kvifte, T.: Description of grooves and syntax/process dialectics. Stud. Musicol. Nor. **30**, 54–77 (2004)

Kvifte, T.: Categories and timing: on the perception of meter. Ethnomusicology **51**(1), 64–84 (2007)

Laban, R.: The Mastery of Movement, 2nd edn. rev. and enl. by Lisa Ullmann. edn. MacDonald & Evans, London (1960)

Laeng, B., Kuyateh, S., Kelkar, T.: Substituting facial movements in singers changes the sounds of musical intervals. Sci. Rep. **11**(1), 22442 (2021). https://doi.org/10.1038/s41598-021-01797-z

Large, E.W., Jones, M.R.: The dynamics of attending: how people track time-varying events. Psychol. Rev. **106**(1), 119–159 (1999)

Leman, M., Naveda, L.: Basic gestures as spatiotemporal reference frames for repetitive dance/music patterns in samba and charleston. Music Percept.: Interdisc. J. **28**(1), 71–91 (2010). https://doi.org/10.1525/mp.2010.28.1.71

Lerdahl, F., Jackendoff, R.: A generative Theory of Tonal Music. MIT Press, Cambridge (1983)

Liberman, A.M., Mattingly, I.G.: The motor theory of speech perception revised. Cognition **21**(1), 1–36 (1985). https://doi.org/10.1016/0010-0277(85)90021-6

London, J.: Hearing in Time: Psychological Aspects of Musical Meter. Oxford University Press, Oxford (2012)

Merchant, H., Grahn, J., Trainor, L., Rohrmeier, M., Fitch, W.T.: Finding the beat: a neural perspective across humans and non-human primates. Philos. Trans. Roy. Soc. London B: Biol. Sci. **370**(1664), 20140093 (2015). https://doi.org/10.1098/rstb.2014.0093

Nketia, J.H.K.: The Music of Africa. Gollancz, London (1986)

Naveda, L.: Gesture in samba: a cross-modal analysis of dance and music from the Afro-Brazilian culture. Ph.D. thesis. Faculty of Arts and Philosophy, Ghent University, Ghent, Belgium (2011)

Naveda, L., Leman, M.: A cross-modal heuristic for periodic pattern analysis of samba music and dance. J. New Music Res. **38**(3), 255–283 (2009)

Naveda, L., Leman, M.: The spatiotemporal representation of dance and music gestures using topological gesture analysis (TGA). Music Percept.: Interdisc. J. **28**(1), 93–111 (2010). https://doi.org/10.1525/mp.2010.28.1.93

Noë, A.: Action in Perception. MIT Press, Cambridge (2004)

Omholt, P.Å.: Regional og typologisk variasjon i norsk slåttemusikk: en kvantitativ tilnærming med et historisk perspektiv. Ph.D. Universitetet i Bergen, Bergen (2009)

Omholt, P.Å.: Rytmen i kryllingspringar. Musikk tradisjon **25**, 47–65 (2011)

Schaeffer, P., North, C., Dack, J.: Treatise on Musical Objects an Essay across Disciplines, 1st edn. University of California Press (2017)

Shapiro, L.: Embodied Cognition, 2nd edn. Routledge, Milton (2010)

Styns, F., Van Noorden, L.: Some basic observations on how people move on music and how they relate music to movement. In: Gritten, A. King, E. (eds.) Proceedings of the Second International Conference on Music and Gesture. Northern Collage of Music, Manchester (2006)

Trehub, S.E., Becker, J., Morley, I.: Cross-cultural perspectives on music and musicality. Philos. Trans. Roy. Soc. B: Biol. Sci. **370**(1664), 20140096 (2015). https://doi.org/10.1098/rstb.2014.0096

Van Noorden, L.: The functional role and bio-kinetics of basic and expressive gestures in activation and sonification. In: Godøy, R.I., Leman, M. (eds.) Musical Gestures: Sound, Movement, and Meaning, pp. 154–179. Routledge, New York (2010)

Varela, F.J., Thompson, E., Rosch, E.: The Embodied Mind: Cognitive Science and Human Experience. MIT Press, Cambridge (2016). Revised edn.

Waadeland, C.H.: Rhythmic movements and moveable rhythms: syntheses of expressive timing by means of rhythmic frequency modulation. Ph.D. Norwegian University of Science and Technology NTNU Trondheim (2000)

Zuckerkandl, V.: Sound and Symbol: @<1@>: Music and the External World, vol. 44:1. Pantheon Books, New York (1956)

Prelude to a Theory of Gestural Time, Proto-Geometry, and Music

Chris Stover(✉)

Queensland Conservatorium, Griffith University, Brisbane, Australia
c.stover@griffith.edu.au

Abstract. This chapter draws on prompts from Rolf Inge Godøy, Edmund Husserl, and a range of Indigenous, queer, and decolonial phenomenological thinkers to frame a theory of gestural time for music that rethinks the relationship between experience and perception. It plays with the distinction between Husserl's "exact" and "descriptive" sciences, putting the latter to work as a productive foil to the drive for empirical exactitude that animates much perception and cognition theory. It does so not to replace exactitude, but to enrich the experiential nexus. Gesture emerges as an at least equally (and perhaps more) plausible first principle for reunderstanding the mechanisms by which perception functions. Focusing on a debate on categorical identity between Rainer Polak and Justin London, it considers the possibility that a turn to affect—understood in Baruch Spinoza's sense of a pre-personal flow of force relations that condition the very possibility of experience and perception in the first place—can work to elide certain kinds of experimental cleavings to a priori category distinctions and to at least provisionally displace perceptual exactitude as the primary location for understanding musical experience.

Keywords: gesture · gestural time · phenomenology · affect · Indigenous knowledge systems

1 Introduction

Among many other things, Rolf Inge Godøy's interventions into how we might understand musical gestures—whether construed as the metaphorical gesture of a musical utterance, the physio-spatio-temporal gesture of a musician's (or listener's) performed action, or the 'gesture' of a perceptual act (or the phenomenological data such an act produces)—open onto manifold possibilities for music analysis, music creation, and artistic research. In the first part of this chapter, I will explore two of those possibilities, deploying Godøy's multivalent usage to think about something we might call gestural time and to pursue the implications of what Edmund Husserl (1983) refers to as a "proto-geometry" that grounds—but, importantly, operates outside the bounds of—what he calls the "exact sciences." I have recently begun to explore the capacity of this concept for thinking about temporal processes in music from the Black radical tradition (Stover 2021b). The second half of the chapter will inquire into how such a turn can help us better understand certain kinds of gestural qualities in music and their affective implications, setting

A. R. Jensenius (Ed.): SSD 2022, CRSM 12, pp. 49–66, 2024.
https://doi.org/10.1007/978-3-031-57892-2_4

these ideas into a dialogue with queer, postcolonial, and Indigenous-epistemological phenomenological practices. In doing so, I have three aims. First, to insist on addressing the cultural and political implications of any phenomenological apparatus, and to bring these concerns productively into the discussion on how to 'do' phenomenology (Ihde 1986; Spiegelberg 1975; van Manen and van Manen 2021). Second, to consider the gestural texture of any act of phenomenological engagement as a Husserlian first principle. Natalie Depraz, Francisco Varela, and Pierre Vermersch (2003) make this notion explicit when they describe virtually all of Husserl's key concepts in gestural terms: the epoché as a "gesture of suspension" (p. 26), the "gesture of reduction" (p. 45), the "gesture of placing the habitus in suspense" (p. 216; in Husserlian language, bracketing the natural attitude), and so on. And third, to use this work to recuperate and put to work a controversial claim by Senegalese poet and philosopher Léopold Sédar Senghor, in which 'Hellenic reason' is counterposed with 'African emotion' (Senghor 2003, p. 288), by suggesting, borrowing a concept from Martin Heidegger (1962), that the gestural or affective qualities of temporal events are covered over by rationalist epistemological frameworks and that we would do well to strive to 'clear' or 'unconceal' the gestures that precede and ground quantitative analysis.

2 From Gestural Objects to Gestural Time

> To 'co-incide' suggests how different things happen at the same moment, a happening that brings things near to other things, whereby the nearness shapes the shape of each thing. (Ahmed 2006, p. 39)

To open the concept onto a somewhat broader range of inquiry applications, Godøy transforms Pierre Schaeffer's (1966) well-known "sonorous object" (*objet sonore*) into a more generalized "gestural object" (Godøy 2006, p. 149) more precisely located in phenomenological experience than in any kind of material-factical 'object-in-itself'. Gesture in this way becomes a mode of engagement with the "meso-level" of musical experience (Godøy 2017; see just below), which includes perceiving a received acoustical signal *as* gesture (the way the latter term is most often described in music theory and analysis; see Hatten 2004; Gritten and King 2006), the gestures that afford different kinds of musical performance (e.g., a conductor's movements or the way a player moves their body to achieve a certain performed task; see Stone 2007), and the ways in which we use physical or metaphorical gestures to describe, entrain to, or otherwise respond to musical data (e.g., dancing or toe-tapping or hand gestures to illustrate the 'shape' of a musical phrase).

Musical gesture generally, and the specific notion of framing a musical utterance as a gestural object, underscores music's temporality in an important way. This is probably too obvious to even need to say. But the turn to gesture was and remains a crucial intervention into a broader music-theoretical discourse on musical shape (see, for example, Tenney 1992), which relied on a more or less static metaphor that could only adequately describe the temporality of a musical utterance with some labor. A gesture is an action in time. Robert Hatten focuses this probably too-simple definition, defining gesture

> rather inclusively as *any energetic shaping through time that may be interpreted as significant*. By significant, I mean that for some interpreter, a gesture will convey information with respect to affect, modality, and/or communicative meaning. (Hatten 2006, p. 1, italics in original)

Hatten's provisional definition offers three important points for consideration. First is the "energetic" nature of a gesture, which we should interpret in differential terms, as a transfer of energy from one temporal location to another that is enacted precisely through and because of that gesture. As a gesture rather than a categorically precise shape, the end point of the energetic transfer is only provisionally known. Second is the subtle way he describes what is going on as a shap*ing*, which transforms the spatialized 'shape' into an active gerund. Third is what a gesture does for some interpreter: what is communicated, or (more important) what kind of a change in affective valence is made manifest. Gesture, in this sense, resides on the temporal-object side of the phenomenological experiencer-experienced nexus, and interpretation is what happens when one encounters the gesture. (This, we'll soon see, is close to the way I'll be focusing on the term/concept below.)

Crucial to this formulation, of course, is a gesture's temporal nature. Godøy's invocation of the gestural object and, soon—synthesizing Schaeffer's word that started it all—the *gestural-sonorous object* draw upon Edmund Husserl's (1991) well-known consideration of the temporal extendedness of what constitutes the 'now' of any experience. Husserl famously invokes a simple musical melody to illustrate this point, which has been taken up in phenomenologically-oriented musicology and music theory in manifold ways (for example, in Schutz 1976; Lochhead, 1982; Clifton 1983). Without laboring over all the intricate details, what is important here is the horizon of temporal experience, which, as Husserl describes, operates via two asymmetrical processes of retention (the re-presentation of a past experience in a lived present) and protension (the opening of experience onto an imminent range of possible futures). For Godøy, what he describes as the meso-level of musical experience (0.5 to 5 seconds; Godøy 2017) is the most logical timescale for conceiving, perceiving, and investigating music's gestural-sonorous objects, since that is the scale at which we can relatively unproblematically hold even a complexly composite event together in consciousness as a whole. The meso-level refers, then, to what Eric Clarke characterizes as a present that "can be dilated for as long as it is possible to hold a temporal object in a single 'nexus of apprehension'" (2011, p. 8).

Godøy alludes to the possibility that the musical macro-level can operate as a kind of gestural object too, but does not pursue the implications very far, except in the very important sense that a (meso-level) gesture's larger context matters for perception and meaning-making. Missing in a lot of the literature about musical gesture is something like the way Roger Sessions defines musical phrase, which William Rothstein also takes up: "What … is a so-called 'musical phrase' if not the portion of music that must be performed, so to speak, without letting go, or, figuratively, in a single breath?" (Sessions 1950, 13; see also Rothstein 1989, 3–4), meaning the phrase-like qualities of longer musical spans (the "so to speak" and "figuratively" of Session's provisional definition). It remains an open question as to how might we analogously consider longer 'gestures', perhaps cognitively afforded by repetition, developmental trajectories, culturally-marked syntactic behaviors, and the like.

But perhaps more important than all these temporal-categorical considerations is what a given gesture's temporal profile is *doing* in any given event. To turn back to Hatten's definition, a gesture is an "energetic shap*ing*" (emphasis added), which means a transfer of energy from one spatial and/or temporal location to another is taking place. A gesture, therefore, is defined by the fact that its specific kind of temporality is, again, one of energetic displacement: from *a* to *b* via the directed motion *i*, to put it in David Lewin's (1987) terms.

Godøy's gestural object resonates with an important concept from Husserl: the temporal object. For Husserl, temporal objects "are not only unities in time but … also contain temporal extension in themselves" (1991, p. 24). Importantly, this is an essential feature of *all* objects, as Alfred North Whitehead famously makes clear (e.g., Whitehead 1964, p. 165–167), even if it is not always immediately apparent. Husserl takes great care to clarify what is temporal about temporal objects and why it matters to think of them so, in doing so, playing with the multiple, perhaps seemingly contradictory ways in which we must strive to understand what time is in the first place. Time is, in one perspective, the medium in which events take place: "temporal objects … spread their matter over an extent of time, and such objects can become constituted only in acts that constitute the very differences belonging to time" (Husserl 1991, p. 41). From another perspective, though, the movement of events are what create time, hence Aristotle's dictum in Book IV, §12 of *Physics* that "time is a measure of motion and of being moved"; this is evident in many of Husserl's formulations, such as the notion of an "act-continuum" that engenders any temporal unfolding. (Some sources translate Aristotle's κινήσεως as "change" rather than "motion"; see Aristotle 2008 (109) and Bostock 2006.) Paradoxically, both of these perspectives are at once true for Husserl, and their co-constitutive nature is part of what makes the whole enterprise of trying to understand how "time-consciousness" operates—and indeed time's very ontological status—so endlessly complex.

But this is not a chapter on the ontology of time, it is about certain kinds of temporal phenomena that engender what we can now start calling *gestural time*. Gestural time is a particular way of being-in (or being-of) time. It is a form of time that, in Husserlian terms, is *anexact*; "essentially, rather than accidentally, inexact" (Husserl 1983, p. 166), meaning that, as gesture, it possesses a kind of qualitative precision not necessarily capturable using quantitative tools. (Or, better, quantitative tools fail to capture what matters about a temporal gesture.) This precision is temporal—some measure of time is either traversed or produced, depending on one's ontological commitment—but also, importantly, affective, in the sense of producing changes in an interlocutor's capacity to act. Ideas are multiplying here, so let me clarify what I mean by these two interrelated modalities.

On one hand, a temporal anexactitude—"roundness" as opposed to a circle or sphere is a spatial example given by Husserl, which Jacques Derrida (1978), Michel Serres (2018), and Gilles Deleuze and Félix Guattari (1987) put to work in varying ways—is a gesture in a context that produces a particular range of effects, which is irreducible to an abstract type. Another way to put this is that it presents a range of entrainment-affordances. An example is the "fork" (*garfo*) gesture in Brazilian samba. The fork is a repeated short–long–short figure, often notated as sixteenth–eighth–sixteenth notes, but in practice, each element of which is stretched slightly, such that each of the two shorts is slightly longer than half a long and the long conversely is slightly shorter than notation would suggest. From another taxonomical perspective, the fork may be conceptualized as a triplet figure the middle term of which is slightly elongated. See Gerischer (2006) and Haugen and Danielsen (2020) for more on microtemporally fluid figures in samba. It is important that the stretching is neither quantitatively precise nor consistent from one iteration to the next, but rather the figure's gestural quality—its *forkness*—is continually being produced. This leads to the second hand: *how* the fork's gestural nature is produced in any given instantiation has to do with an ongoing flux of affective relations at play between performed gestures by the samba ensemble. As I have described elsewhere in the context of Cuban rumba (Stover 2018), the specific ways in which a given iteration of a repeated figure like the fork are stretched have to do with how the ongoing microtemporal flux of the music is being taken up, largely *precognitively*, by the player. (I'll turn below to a recent way in which microtiming behaviors have been presented by Rainer Polak and Justin London to think further through the stakes of these two considerations.) This is a tenet of theories of enactivist cognition, even if not always framed in precisely these terms, and serves as an important counterpoint to representational theories through which cognition drives embodied responses. So the affective affordances of an earlier or ongoing gesture have an effect on how one plays the fork, which in turn functions as an expression of the affective genealogy that partially conditions how its particular identity is staged. That expression is, again, anexact: it is inexact in that it cannot be known just how a given player will respond to a received stimulus, but essentially so in that an effect—a change of valence—is ever in the process of transpiring.

I've gone through this far too quickly, but have developed these ideas elsewhere (Stover 2018, 2021a). Some key points are worth delineating, however. My account of how affect operates stems from the long Spinozist tradition through which (1) the word affect is a shorthand term for the double movement of relational flows between interacting bodies; (2) affects are produced by those bodies and (3) also continually reconstitute them; (4) interacting bodies in this sense may be said to be acting on one another; (5) therefore what changes when a body is reconstituted within a nexus of affective flows is its "capacity to act" (Spinoza 2002) or its *valence* (its capacity to enter into new affective connections; see Varela and Depraz (2005) and Stover (2021a)). In other words, a body's affective valence is precisely what engenders both its actual actions and the ways in which responds to proximal actions, in an ongoing flow. In the fork example, it is precisely the ongoing gestural flux of microtiming pullings to and fro that conditions bodies (of performers, of the musical gestures themselves) to unfold in a particular shape in any given instantiation.

Gestural time is an important intervention into conceptions of musical temporality. As a concept, it is neither radical nor rare. For example, any time a classical performer transforms the more or less fixed notation of a musical score into a multiply-directed microtemporal expression, gestural time is being enacted. Most music, indeed, is gestural in this sense, but there are more overtly clear examples that help us understand why a turn to gesture matters: the shifting metric flux of Hardanger fiddle music, the nuanced prosody of Mississippi delta blues singing, the gestural rhetoric of Gagaku court music, each of which very effectively resists notational representation. Nancy Murphy's (2023) recent work on flexible meter illustrates this vividly, if not precisely in these terms. What is phenomenologically important in this account takes us back to Godøy's work, and just what it is we are experiencing when we turn our attention to the gestural quality of any music, beyond or alongside any kind of purported isochronous representation that we might try to use to model our experience. Godøy wishes to clarify empirically "our capacity to capture and handle the ephemeral and temporally distributed features of music" (2017, p. 10). While we should raise our eyebrows at the notion of *capturing* anything, which has complex and fraught ethical implications, the underlying premise is promising: how deeply and to what degree of detail can we come to understand our *experience* of gesture in its very ephemeral and temporally-distributed nature? Further, how can we come to understand our experience of an experience of gesture—the stuff of what I call second-order phenomenological methodology—seeking to understand the lived experience of an interlocutor, in this case, the musical interlocution of an observed performer?

3 Husserl's Proto-Geometry

Early in the 1905 lecture that opens *On the Phenomenology of the Consciousness of Internal Time*, Edmund Husserl makes an astonishing point, the cognitive implications of which remain to be fully unpacked. In this passage, he explains that

> sensed 'synchrony' is not simply equivalent to objective simultaneity; sense equality of temporal intervals, given phenomenologically, is not straightaway objective equality of temporal intervals; and the sensed absolute datum is, again, not immediately the being-experienced of objective time (this is true even of the absolute datum of the now). (Husserl 1991, p. 8)

In other words, how one comes to experience synchrony or equality is separable from whatever we might call the objective data of that which is experienced. This is a crucial point that underlies my theory of beat span (Stover 2009) and Anne Danielsen's (2010) theory of beat bins (see Danielsen, Johansson, and Stover (2023) for a comparison between these analytic orientations). Both of these theories orbit around and seek to explain what we might call the *near-simultaneities* of two or more discrete acoustic events, which, regardless of the quantitatively precise locations of their onsets or perceptual centers, are held by a perceiver to be constituent parts of the same temporal gesture, for example the same beat, as a temporally-extended phenomenon. In other words, we can perceive them as synchronous even in their *objective* non-synchrony. We can choose to do this, and we can also find ourselves doing it without actively thinking about it.

In his book-length meditation on Husserl's "The Origin of Geometry," Jacques Derrida leans into Husserl's (admittedly brief) development of the concept of anexactitude. First of all, he considers Husserl's history-of-science account of how geometry, as an ideal science, coalesced from a more generalized "pregeometrical world," a "world of things disposed … according to an anexact space and time" (1978, 122). But this world "is a cultural world already informed by predictions, values, empirical techniques and the practice of measurement and inductiveness which themselves have their own style of certainty" (120). It's easy to read this as a naïve prehistory that was eventually overcome by a more precise scientific episteme, but I would argue that is an incorrect and colonialist reading. As Derrida insists, “the protogeometer always already ha[s] at [their] disposal anexact spatiotemporal shapes and essentially ‘vague morphological types” (123), and it is equally naïve to consider “this anexactitude of the object or concept to be… a ‘defect’.” To this end, we should always be on guard when terms like “deviation” or “discrepancy” are evoked, which (intentionally or not) pathologize anexact musical gestures as aberrations.

Derrida goes on to quote Husserl, from the passage in *Ideas*, volume I, where the concept of anexactitude is first spelled out:

> The most perfect geometry and its most perfect practical mastery of it cannot enable the descriptive natural scientist to express (in exact geometrical concepts) what he expresses in such a simple, understandable, and completely appropriate manner by the words ‘notches’, ‘scalloped’, ‘lens-shaped’, ‘umbelliform’, and the like—all to them concepts which are *essentially, rather than accidentally, inexact*, and *consequently* also non-mathematical” (Husserl 1983, p. 166; also in Derrida 1978, p. 122; italics in original)

The section where this crucial quote occurs marks something of a material-ontological shift through which descriptive “morphological concepts” are shown to coalesce in ways that always remain vague and fluid. Husserl insists that their vagueness is, again, not a defect but rather is an essential (and, importantly, “legitimate”) quality. Husserl takes care to clarify two types of morphological essences: one that stems from “exactness of ideal concepts” (Husserl 1983, 167), which is the proper purview of what he calls the exact sciences, and one that flows from what he calls a “firmness and … pure distinguishability of generic concepts … which have their extension in the realm of fluidity,” which is the purview of a more originary descriptive science. Exact and descriptive sciences can and do overlap in key ways—as, for example, Godøy’s work has long demonstrated—but according to Husserl have very different aims, procedures, and animating questions. Isabelle Stengers meditates on a similar idea in her work on Alfred North Whitehead, also foregrounding the aesthetic (yet highly technical) nature of what I would call Husserl’s descriptive orientation. Stengers writes:

> Between the most concrete experience and the various abstractions, there is no hierarchy for Whitehead. The artist’s perception is not more authentic, it is different; and, what is more, it testifies to a trained eye. Nor is there anything painfully paradoxical about the the very fact that, when testifying that ‘it’ is never the same [referring to Whitehead’s examination of Cleopatra’s needle’s relatively fixed or unfixed location on the Charing Cross embankment], she must say ‘it’, implying the stability that she nevertheless denies. The artist’s testimony concerns the experience of a contrast but does not provide weapons to a contradiction. (Stengers 2011, p. 76)

The artist’s experience will become relevant below as well.

Godøy’s usage, in fact, is crucial for understanding why the tension between these two perspectives matters. In his earlier work, Godøy (1997) develops a “morphodynamical theory” of musical shape, drawing upon the work of René Thom and Jean Petitot.

According to the theory, "human perception is a matter of consolidating ephemeral sensory streams (of sound, vision, touch, and so on) into somehow more solid entities in the mind, so that one may recall and virtually re-enact such ephemeral sensations as various kinds of shape images" (Godøy et al. 2016, p. 2). That is, we perceive temporal events as examples of categorical types, and our ability to do so is an important way in which we make sense of the world. What Godøy and others (including Pierre Schaeffer) want to do is clarify the boundaries of a perceptual shape-category, which has led to many valuable studies that seek empirically to test those boundaries, asking what can change and how much before a gestural-sonorous object can no longer rightly be classified within a particular category. Elsewhere, for example, Godøy (2017, 10–11) draws upon Petitot's methodology to develop what he calls a "control space" and a "morphology space" in order to be very meticulous about the "what changes and how much" question. So far, so good: this exemplifies the exact-science trajectory in Husserl's account and resonates with gestalt theories of spatial-temporal recognition.

Husserl, however, admonishes us to resist this particular categorical imperative: "we experience 'bodies'—not geometrical-ideal bodies but precisely those bodies that we actually experience, with the content which is the actual content of experience" (1970, 25). To frame an experience in terms of its cleavage to a predetermined categorical model is precisely what the epoché is intended to, at least provisionally, elide. My contention extends from this: a turn, within a larger project of phenomenological variation, to the anexact gestural qualities of a perceived temporal event can put Husserl's notion of descriptive science to work as a productive foil to empirical exactitude. Not to replace the latter, but to enrich the experiential nexus. There are two reasons such a turn is important. First, it can allow us to call into question the particular ideal shape that we may be claiming underlies all the 'distorted' performed/perceived instantiations. This is an extraordinarily important political claim that I will turn to in the last section of this chapter. Second, as the following analysis will make clear, it can give us tools to resist certain kinds of assumed categorical a priori, especially those grounded in received ideas about cognitive limits, which, as I've suggested above, affect theory, with its focus on pre-cognitive processes, elides. In the next section, I'll stage my engagement with these two notions around a simple experiential question: are there two or three durational categories operating at the beat subdivision level in a performance of West African drum-dance music?

Both of these rationales are grounded on the fundamental phenomenological principles of reduction and imaginative variation. In terms of reduction, the first important step (as in all phenomenological inquiry) is to bracket the natural attitude—in this case, the epistemological presuppositions of a certain constellation of empirical practices and methods—and return to the experience to ask what else? Under what alternative experiential rubric is the temporal object knowable? Sara Ahmed (2006, p. 27) richly illustrates how our "bodies are directed in some ways and not others" (and Frantz Fanon (2008) clarifies just some of the hegemonic forces at work in orienting our bodies in particular directions), so the stakes of working, even provisionally, to bracket constraining or oppressive forces are very high. In terms of variation, then, the task is to deliberately and creatively shift one's orientation toward the object of experience, in order to produce novel experiential relations with it. Those relations, ultimately, change us, as Ahmed poignantly puts it: “The ‘new’ is what is possible when what is behind us, our background, does not simply ground us or keep us in place, but allows us to move and allows us to follow something other than the lines that we have already taken” (2006, p. 62–63).

4 Micro-gesture and Phenomenological Variation

I’d like to turn now to Rainer Polak’s (2010) empirical study of jembe music from Bamako, Mali. In this study, Polak suggests, among many other things, that a relatively consistent expressive-timing pattern occurs across trios of played events in manjanin, a 12-cycle drum-dance piece. Polak shows how, in a number of performances, this onset sequence maps very well onto a short–medium–long ratio. In some of his examples, this taxonomy seems quite clear-cut, for example, ratios of 26:32:42, 27:32:41, and 23:32:45 (see Polak’s Table 4); whereas in others—for example, 25:36:39—he hedges, suggesting that perhaps a S–L–L taxonomy might be more appropriate. Polak compares four players’ renderings of a repeated échauffement figure, which is important to the dramatic intensification of a jembe–dance dialogue. Also important here is categorical (non)overlap between *ranges* of the second and third pulses (the pulses which call into question the need for a medium–long versus long–long distinction). In his first three examples (for which Polak suggests a medium–long ratio) the ranges are nonoverlapping: 26–38 and 39–47 in the first case, 26–36 and 38–44 in the second, and 28–38 and 41–49 in the third. In the fourth example that problematizes this framework, the ranges overlap—32–40 and 36–41—calling further into question their categorical distinctness. Polak’s concern about categorical slippage ultimately materializes as what he calls a short–flexible–long ratio, where the expressive lengths of the two outer onsets are relatively determinable, whereas the length of each middle onset is more fluid. I’ll return to what I see as a productive liminality already built into Polak’s taxonomical hesitation.

In his commentary on Polak’s study in the same special issue of *Music Theory Online*, Justin London (2010) insists that S–M–L might not work as a practicable beat subdivision taxonomy since the timing distinctions are too small to be perceived according to these categories. London is probably correct according to the perceptual frameworks he enlists to stage his arguments. But at the same time, he acknowledges the persistent empirical there-ness of the timing ratios. How do we work through this interesting perceptual–empirical paradox?

In order to understand what is at stake here, both methodologically and ontologically, I’ll quote London at length. London writes:

> Polak’s approach challenges my arguments on [two] grounds, (a) that one can have three distinct subpulse-classes … and (b) [that] these distinct subpulse-classes may be defined qualitatively rather than quantitatively. I think he is correct on the latter, but not on the former. I am convinced from both Polak’s empirical data and from his ethnographic reports that jembe players and listeners recognize categorical differences amongst subpulses….
>
> Where we disagree is whether or not one may have three distinct classes of beat subdivision. I believe Polak’s data [show] that there are two, and that his [medium] category represents expressive timing variants of underlying short … or long …

> subdivision units. To be clear, I think Polak's data clearly indicate that jembe performers consistently play different subpulses with different durations depending on their position in the metric cycle. But Polak is making a stronger claim: not that these are simply expressively timed versions of one or two subpulse-classes, but that they manifest three categorically distinct subpulse classes.

The tension hinges on the word 'categorical', which is at best an unfortunate word choice and at worst a colonial insistence that things be put into categories in the first place: one of music theory's original sins. What is at stake in cleaving to a two- or three-category beat-subclass taxonomy? Very little, I'd say: except to the extent that one argument draws upon an epistemological apparatus built around what we understand a priori to be perceivable, as I have described above. I suggest this places too big an epistemological burden on perception, the way we currently understand it to function. This is where affect comes in. If affect does indeed function as a pre-personal—and therefore pre-cognitive—flow of force relations that changes one's capacity to act within an ongoing interactive context, and if affect's effects are observable through the empirically measurable events that unfold in that context, then we might be able to ascertain at least some of the ways any given musicking participant is being affected by attending to the very particular ways in which what they do changes over the time of the performance. This, then, involves both doubling down on attention to empirical details like timing ratios between trios of played events and taking stock of ongoing music-environmental stimuli that might have effected a subtle change in performance orientation: a 'call' that invites some kind of 'response'.

This is also where tempo comes in. The music Polak examines is very fast—faster than the speed of thought, Gilles Deleuze would say. As London makes clear, it's fast enough that we cannot categorically distinguish between discrete event-duration categories, even while we can—especially upon close, repeated listening—vividly and accurately describe how a particular part 'feels' using qualitative terminology. Polak demonstrates this beautifully with his examples where he extracts individual cycles and even individual instrumental parts and loops them in order to draw the listener's attention to specific timing details—a parallel can be made to Godøy's "control" and "morphology" shapes. But tempo might actually be crucial here. If affect likewise moves faster than the speed of thought, how does it function? Henri Bergson (1999) provides a possible framework, which has been instrumental in how Léopold Sédar Senghor and others have theorized communal interplay in African performance practice. Bergson theorizes an affective 'zone of indeterminacy' between reaction and action, an infinitesimal timespace within which we are affected, and before cognition and perception take place. Patricia Clough similarly describes the timespace of affect's operation as "the indeterminacy of autonomic responses" (2010, p. 209) within which consciousness can only be a "substractive" iteration that necessarily reduces away from affective complexity; there will always be an affective "remainder" to conscious perception. We act in this timespace before we realize it. In the dense, rapidly repeating context of Bamako jembe music (for one example of many), we might say we never quite have time to do the cognizing that follows and makes sense of (or categorizes) action. In short, again, we 'feel' it: according to the affect theory orientation I subscribe to, feeling always precedes and conditions perception. What we feel is precisely the improvisational interaction between participants—the little or big extemporaneous gestures that continually redirect the music's trajectory. We feel the 'call' of those changes in affective valence, and 'respond' in some way. We feel, to bring another stream of affect theory into the conversation, a certain kind of emotional valence that might simply result in us continuing to do what we're doing because everything is feeling 'right'.

What follows from all this is that both Polak and London are correct according to the terms of their epistemological vantage points. Having studied extensively in Mali and being himself a high-level jembe practitioner, Polak is considerably closer to the ground than his research collaborator, which I'll suggest shortly is important. But indeed, the very way he hedges about that "flexible" beat subclass suggests a productive opening of what we might call the taxonomical imperative onto other, phenomenologically valent experiential modes. If cognition theory reveals a perceptual limit to how we can identify the categories that performed gestures fall into, then it seems imperative to consider those gestures from different experiential perspectives, perhaps not as discrete events that work together to parse a given beat in a particular way, but as a composite gesture that moves through that beat, enacting a transference of energy from one beat onset to the next. This requires shifting attention away from discrete events (measured as ratios or IOIs) toward the relations that emerge and are engendered between them. Phenomenological philosopher Françoise Dastur describes this deliberate shift in perspective "let[ting] the constitutive operation appear" and, even more germanely, "let[ting] appear the temporal character of what is given to us" (2000, p. 180).

The jembe music Polak analyzes exhibits a fascinating productive aporia. On one hand, like so much cyclic drum–dance music from Africa and the African diaspora, a continuous sense of intense forward motion takes place throughout any given performance, one ramification of which is that the music very often speeds up, sometimes considerably, as it builds to a climax. On the other hand, in this particular case, that forward motion seems in every beat iteration to be slightly arrested as each of the three played beat subdivisions slows down slightly. The energetic displacement that results is a kind of halting gestural time at the micro-level that belies the longer-scale intensive trajectory of the music. The relationship between these two temporal trajectories matters; precisely how so remains the task of future research.

5 Experience Matters

> "Phenomenological explanation deals not only with given data, but with potentialities." (Dastur 2000, p. 184)

From the perspective of many Indigenous epistemologies, knowledge is active and dynamic, and objects and concepts are identified, in Indigenous North American scholar Shay Welch's terms, "according to their relationship to other things in an active process" (Welch 2019, p. 41). Further, "the things we know emerge from the ways in which we participate as embodied beings" (p. 43), which, using the phenomenological language I've been orbiting around above, means (potentially) bracketing one's epistemological preconceptions, immersing oneself in the affective flow of an ongoing context, conceiving of cognition and perception (or intentionality) as imaginative processes (see just below), and remaining open to what might result.

Welch describes a tendency among what she refers to as "those mired in Western post-positivistic scientific and philosophical ideology" (p. 75) to "perceive basic level conceptual categories as objective, self-evident descriptions of mental phenomena" (p. 74). She suggests instead a storytelling methodology that moves away from, for example, the raw empirical data of an object of experience toward an ever-richer engagement with what it is the object of experience is doing, both "in-itself" (to borrow Husserl's language) and for the experiencer. What Welsh hopes to shed is any effort to force data into pre-formed categorical boxes, which is precisely to foreclose the possibility that one might be able to experience differently. "[W]ithout the recognition of the possibility of multiple ways of being, there can be recognition of multiple ways of knowing" (p. 83). Storytelling, according to Welsh, is a method that potentially cuts across discursive and conceptual boundaries and, in doing so, makes possible the discovery of different ways of being through which new knowledge forms can begin to take shape. Beyond this, though—and far more important for what I have been staging in this chapter—is the possibility that different expressive media, namely the gestural medium of dance (or music!), can function as deeply communicative, even if non-narrative, storytelling modes. In short, Welsh aims to clarify "how dancing creates meaning" (p. 105). Here Welsh's conception aligns with the gestural orientation I have been staking out thus far. Welsh suggests that.

> gestures are embodied symbolic communication—a sort of 'oral motility', as [Shaun] Gallagher puts it—that are esssential to narrative praxis. Gestures are naturally and innately communicative quite independently of verbal language…. (p. 105; internal quote from Gallagher 2006, p. 107)

From a more general engagement with the kinds of gestures that can be found co-occurring with spoken communication, Welsh soon pivots to how gestural language, in itself, can function as "a form of embodied and implicit knowing within and as storytelling" (p. 113). She is most interested in understanding how dance, as a gestural language that operates outside of verbal discourse, functions as a primary mode of meaning-sharing in Indigenous knowledge systems and beyond. The 'beyond' is important here, as Welsh is careful not to draw too fine a distinction between Indigenous knowledge systems and whatever we might characterize as their oppositional twin (see pp. 118–119). Dance, according to Welsh, has an immediacy that verbal language cannot reach, which operates before or below the level at which language is able to engage:

> The kinetic bodily logos of thinking in movement are another way of conceiving of the preverbal or nonverbal nature of movement as procedural meaning-making and communicative action. In fact … while verbal prose may frequently be ambiguous … embodied dynamics are precise. This is because verbal language is not and does not constitute experience. Therefore, attempts to verbalize experiences obscure the fine qualitative and affective constituents of experiences that make them so rich and unique to the individual. (p. 120)

I would add that “precise” here should be read in exactly the proto-geometric sense that so attracted Husserl, Derrida, Serres, and Deleuze and Guattari, as described in Sect. 2 above.

Beyond its active, dynamic, embodied, gestural nature, Welch suggests that knowledge and its production are communal (p. 32) and relational, and therefore ethical (p. 33). This perspective flows through a great deal of global Indigenous epistemology, and is a hallmark of what has recently become known as Africana phenomenology (Henry 2005). An early exemplar of this latter philosophical perspective is the theoretical, artistic, and practical work of Léopold Sédar Senghor, especially his notion of knowledge as communally produced, which he frames as a particularly African modality but which we can think of in more generalized affective terms as well. Senghor’s “law of participation” is as well-known as it is controversial, and Senghor spent a great deal of effort in his later writings correcting what he saw as egregious misinterpretations of perhaps his most oft-cited statement, “Classical European [or sometimes ‘Hellenic’] reason is analytical and makes use of the object. African reason is intuitive and participates in the object” (Senghor 1965, p. 34). The notion of participating in an object of experience is, of course, a profoundly phenomenological claim. But beyond participating “in the object,” which among other things is, crucially, an assumption that the object possesses a kind of relational agency (Bennett 2010), knowledge for Senghor is produced within a community of practice; that is, it is distributed, liminal, and creative. For Senghor, this is a *rhythmic* and *cyclical* process, and additionally a *lyrical* one. “The call is not the simple reproduction of the cry of the Other; it is a call of complementarity, a *song*: a call of harmony to the harmony of union that enriches by increasing *being*” (Senghor 1965, p. 63). Welch similarly foregrounds the relational nexus: “[o]ur contextualized positions are a field of possibilities and opportunities, and as we think and act, we create and structure meaning by creating connections” (Welch 2019, p. 57).

Time, then, is for Senghor both rhythmic and lyrical: two musical metaphors that undergird his entire relational metaphysics. It is an iterative ordering force (hence the significance of cycles in so much African music) that produces existence. But the kind of force it is sensible rather than material:

> This ordering force … is rhythm. It is the most sensible and least material thing. It is the vital element par excellence. It is the primary condition for, and sign of, art, as respiration is of life—respiration that rushes or slows, becomes regular or spasmodic, depending on the being’s tension, the degree or quality of the emotion…. It is not a symmetry that engenders monotony; rhythm is alive, it is free. (Senghor 2003, p. 296)

Here the proto-geometric, gestural nature of musical rhythm becomes especially apparent and profoundly meaningful. Likewise, the significance of attending apodictically to rhythm’s protogeometricity, to do the work to learn to experience a gestural-sonorous object on its own productive terms, as sensible (gestural, expressive) rather than material (durational, taxonomical). Again the jembe example from above illustrates this point vividly: the argument about beat-subclass types and concomitant appeals to lowest perceptual limits misses the point of what it is the repeating (or “respiring”) musical gestures are producing through their ongoing iteration.

What all this amounts to is an appeal to expand our conception of what it means to experience and how to do so, and to turn to the concept of gestural design (and, in music, to gestural-sonorous-objects operating in proto-geometric space-times) as a productive exploratory timespace. Experience is an iterative process, the operations of which are perspectival, positioned, and relational. These three themes are grounded on the fact, crucial to phenomenological philosophy, that we have (or are) bodies from which we experience and that our experiencing bodies get extended through different kinds of affordance-relations, including tools, the development of layers of awarenesses through repetitions of familiar actions, cultural emplacements, and more. Sara Ahmed is interested in "how bodies are directed in some ways and not others" (2006, p. 27); that is, how through those foldings of experiences and emplacements our subjectivities are constructed such that some next actions are viable and others less so. There are important connections in Ahmed's account to Frantz Fanon's (2008) theorization of how bodily orientations and capacities are foreclosed by ideological, historical, hegemonic, and other oppressive forces. "Orientations involve directions toward objects that affect what we do" (Ahmed 2006, p. 28), which is not necessarily a conscious process:

> We move toward and away from objects depending on how we are moved by them.... Turning toward an object turns 'me' in this way or that, even if that 'turn' does not involve a conscious act of interpretation or judgment. (p. 28)

In other words, the experience-in-motion is what does the "turning," and our orientations from this perspective are pre-cognitive in the sense I have described above. Rather than actions performed *by* subjects that pre-exist them, we are always already "in" those actions.

In ordinary modes of moving about and experiencing, this iterative process "de-distances" (Heidegger 1996, p. 104) certain aspects of the world, making them familiar and "available." Certain aspects of the world become, in some way, known to us: they become part of the foreground figuration of the world, the natural way we come to expect the world to be. But, as Ahmed makes clear, "[t]he figure 'figures' insofar as the background both is and is not in view. We single out this object only by pushing other objects to the edges or 'fringes' of vision" (2006, p. 37). In other words, the very practice of de-distancing inevitably engenders new distances by pushing other objects or perspectives or attitudes out of one's understanding. This is a crucial point to keep in mind when appealing to any empirical account of perceptual experience: what is left out when a particular framework is set in place?

Ahmed's overt project is to *queer* phenomenology (although she makes a convincing argument that phenomenology has always been queer in the way it disrupts "straight" modes of relational perception). What makes Ahmed's phenomenology queer is a concerted effort to "dis-identify" (Muñoz 1999) with received perceptual frameworks, to orient and re-orient ourselves such that we are able to resist the kinds of nearnesses that foreclose possibilities and potentials. Ahmed suggests that

> [w]hat is reachable is determined precisely by orientations that we have already taken.... The surfaces of bodies are shaped by what is reachable. Indeed, the history of bodies can be rewritten as the history of the reachable. Orientations are about

> the direction we take that puts some things and not others in our reach. (Ahmed 2006, p. 55–56)

To dis-identify with this form of orientational foreclosure is, first of all, a radical political stance. It is to identify and resist the force of ideology and to insist that there are other modes of being and doing that can be animated by new orientational enactments.

What, then, has all this to do with gestural time?

First of all, gestural time can be counterposed with measured time in the particular sense that it grounds the latter: first there was gesture, as a gloss on what Senghor refers to as a "humid and vibratory" *logos* that has been covered over by "the analytic turn of thought" or the "*ratio*" (Diagne 2019, p. 25; the temptation to read Polak's durational ratios as a contrafact to Senghor's fecund ontology is strong). As Souleymane Bachir Diagne writes of Senghor's signal theoretical contribution, this amounts to an "illumination, beneath the analytic intelligence—the faculty that understands by analyzing and separating parts external to each other (*partes extra partes*)—of the faculty of vital knowledge, which in a single immediate and instantaneous cognitive gesture can comprehend a composition that is living, not mechanical, and therefore cannot be decomposed" (p. 23–24). Not short–medium–long (nor any other categorical determination), but a composite decelerating gesture that does affective work within its contextual trajectory. Here, to turn back to a discursive strategy we encountered in Depraz, Varela, and Vermersch's (2003) work at the beginning of this chapter, the act of phenomenological engagement is described in gestural terms. If we follow Husserl's project through which we strive to make experience increasingly apodictic with that which is experienced, then the more gestural we can make cognition, as well as (or via) phenomenology's primary methodological tools (the *epoché* as a "gesture of suspension," the "gesture of reduction," and so on as described above), the more closely we may be able to map the essential nature of the gestural-sonorous object. From a musical perspective, this means practicing hearing gesturally in order to apodictically map the gestural design of the music we are experiencing.

Second is the way in which a turn to gesture necessitates a rethinking of what empirical measurement can reveal that is meaningful about music. The argument against a gestural-phenomenological method is that we should be attending to music-temporal phenomena that are given to perception, hence the argument by London and many others for various kinds of perceptual thresholds that limit what we ought to be able to say about minute microtiming measurements. But the fact that gestural qualities may not be immediately given to perception is precisely the point: as Martin Heidegger insists, "just because the phenomena are proximally and for the most part *not* given, there is need for phenomenology" (Heidegger 1962, p. 60). Phenomenology is the study of the structure of experience, but it is also, equally, a practice of expanding or otherwise transforming the nature and scope of what is experienceable. Phenomenology is, essentially, necessarily creative. So, again: what new listening modalities are afforded by adopting a gesture-orienting listening posture, and what new details might be hearable by doing so?

Third, lastly, my turn to queer, Indigenous, and decolonial phenomenology in the last part of the chapter, and my reading of them as extensions (rather than rejections) of Husserl's foundational phenomenological project, amounts to an appeal for phenomenological researchers of all stripes to deeply engage what I'll hesitatingly gloss as the existential stream of phenomenological theory and practice. What does this mean? It means to take seriously the ways in which social, cultural, and historical contexts affect how we are able to perceive the world. How the world has been and continues to be given shape by forces outside us. Beyond this, it means to activate and vigorously practice ways of contesting what decolonial feminist theorist Françoise Vergès calls “epistemicide”: to join an ongoing struggle against “a system that has dismissed scientific knowledge, aesthetics, and entire categories of human beings as non-existent” (Vergès 2021, p. 13). Decolonial phenomenology has much to offer both of these imperatives, e.g., the drive expressed by Frantz Fanon in the opening pages of *Black Skin, White Mask*: “What I want to do is help the Black man to free himself of the arsenal of complexes that has been developed by the colonial environment” (Fanon 2008, p. 19). Understanding the nature of music’s gestural processes is, to be fair, many orders of magnitude less urgent than liberating human beings from oppression. But to give the final word to Senghor, art is one of the most potent expressions of the vital force that he understands to flow through all relational human connections. As Diagne phrases it, “[i]t is in art that we can find a premonition of what it is we must become” (2019, p. 49); this requires “the capacity to ‘produce only in freedom’” (Senghor 1964a, b).

References

Ahmed, S.: Queer Phenomenology: Orientations, Objects, Others. Duke University Press, Durham (2006)

Aristotle: Physics, Book IV. Trans. by R.P. Hardie and R.K. Gaye. Internet Classics Archive (n.d.). http://classics.mit.edu/Aristotle/physics.4.iv.html

Aristotle: Physics. Trans. by R. Waterfield. Oxford University Press, Oxford (2008)

Bennett, J.: Vibrant Matter: A Political Ecology of Things. Duke University Press (2010)

Bergson, H.: Matter and Memory. Trans. by N.M. Paul and W. Scott Palmer. Zone Books (1999)

Bostock, D.: Space, Time, Matter, and Form: Essays on Aristotle’s Physics. Oxford University Press, Oxford (2006)

Clarke, E.: Music, phenomenology, time consciousness: meditations after Husserl. In: Clarke, D., Clarke, E. (eds.) Music and Consciousness: Philosophical, Psychological, and Cultural Perspectives, pp. 1–28. Oxford University Press (2011)

Clifton, T.: Music as Heard: A Study in Applied Phenomenology. Yale University Press (1983)

Clough, P.T.: The affective turn: political economy, biomedia, and bodies. In: Gregg, M., Seigworth, G.J. (eds.) The Affect Theory Reader, pp. 206–225. Duke University Press (2010)

Danielsen, A.: Here, there and everywhere: three accounts of pulse in D’Angelo’s ‘left and right’. In: Danielsen, A. (ed.) Musical Rhythm in the Age of Digital Reproduction, pp. 19–36. Ashgate (2010)

Danielsen, A., Johansson, M., Stover, C.: Beat bins, beat spans, rhythmic tolerance: three theories of musical microtiming. Music Theory Spectrum **45**(2), (2023)

Dastur, F.: Phenomenology of the event: waiting and surprise. Hypatia **15**(4), 178–189 (2000)

Deleuze, G., Guattari, F.: A Thousand Plateaus: Capitalism and Schizophrenia. Trans. by B. Massumi. University of Minnesota Press (1987)

Depraz, N., Varela, F., Vermersch, P.: On Becoming Aware. John Benjamins Publishing Company (2003)

Derrida, J.: Edmund Husserl’s Origin of Geometry: An Introduction. Trans. by J.P. Leavey, Jr. University of Nebraska Press (1978)

Diagne, S.B.: Postcolonial Bergson. Trans. by L. Turner. Fordham University Press (2019)

Fanon, F.: Black Skin White Masks. Trans. by R. Philcox. Grove Press (2008)

Gallagher, S.: How the Body Shapes the Mind. Clarendon Press, Oxford (2006)

Gerischer, C.: *O Suingue Baiano*: rhythmic feel and microrhythmic phenomena in Brazilian percussion. Ethnomusicology **50**(1), 91–119 (2006)

Godøy, R.I.: Formalization and Epistemology. Scandinavian University Press (1997)

Godøy, R.I.: Gestural-sonorous objects: embodied extensions of Schaeffer's conceptual apparatus. Organ. Sound **11**(2), 149–157 (2006)

Godøy, R.I.: Key-postures, trajectories and sonic shapes. In: Leech-Wilkinson, D., Prior, H.M. (eds.) Music and Shape, pp. 4–29. Oxford University Press (2017)

Godøy, R.I., Song, M., Nymoen, K., Haugen, M.R., Jensenius, A.R.: Exploring sound-motion similarity in musical experience. J. New Music Res. **45**(3), 1–13 (2016)

Gritten, A., King, E. (eds.) Music and Gesture. Routledge (2006)

Hatten, R.: Interpreting Musical Gestures, Topics, and Tropes: Mozart, Beethoven, Schubert. Indiana University Press, Bloomington (2004)

Hatten, R.: A theory of musical gesture and its application to Beethoven and Schubert. In: Gritten, A., King, E. (eds.) Music and Gesture, pp. 1–23. Routledge (2006)

Haugen, M.R., Danielsen, A.: Effect of tempo on relative note durations in a performed samba groove. J. New Music Res. **49**(4), 349–361 (2020)

Heidegger, M.: Being and Time. Trans. by J. Macquarrie and E. Robinson. SCM Press (1962)

Heidegger, M.: Being and Time. Trans. by J. Stambaugh. State University of New York Press (1996)

Henry, P.: Africana phenomenology: its philosophical implications. C.L.R. James J. **11**(1), 71–101 (2005)

Husserl, E.: The Crisis of European Sciences and Transcendental Phenomenology: An Introduction to Phenomenological Philosophy. Trans. by D. Carr. Northwestern University Press (1970)

Husserl, E.: Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy, First Book: General Introduction to Phenomenology. Trans. by F. Kersten. Kluwer Academic Publishing (1983)

Husserl, E.: On the Phenomenology of the Consciousness of Internal Time. Trans. by J.B. Brough. Kluwer Academic Publishers (1991)

Ihde, D.: Experimental Phenomenology: An Introduction. State University of New York Press (1986)

Lewin, D.: Generalized Musical Intervals and Transformations. Yale University Press (1987)

Lochhead, J.: The temporal structure of recent music: a phenomenological investigation. Ph.D. dissertation, State University of New York at Stony Brook (1982)

London, J.: Commentary. Music Theory Online **16**(4), (2010)

Massumi, B.: Parables for the Virtual: Movement, Affect, Sensation. Duke University Press, Durham (2002)

Moustakas, C.: Phenomenological Research Methods. Sage Publications (1994)

Muñoz, J.E.: Disidentifications: Queers of Color and the Performance of Politics. University of Minnesota Press (1999)

Murphy, N.: Times A-Changin': Flexible Meter as Self-Expression in Singer-Songwriter Music. Oxford University Press (2023)

Polak, R.: Rhythmic feel as meter: non-isochronous beat subdivision in Jembe music from Mali. Music Theory Online **16**(4), (2010)

Rothstein, W.: Phrase Rhythm in Tonal Music. Schirmer Books (1989)

Schaeffer, P.: Traité des objets musicaux. Éditions du Seuil (1966)

Schutz, A.: Fragments on the phenomenology of music. In: Joseph Smith, F. (ed.) Search of Musical Method. Gordon & Breach (1976)

Senghor, L.S.: Liberté, vol. 1: Négritude et humanisme. Seuil (1964a)

Senghor, L.S.: On African Socialism. Praeger, Westport (1964b)

Senghor, L.S.: Prose and Poetry. Trans. by J. Reed and C. Wake. Heinemann African Writers Series (1965)

Senghor, L.S.: What the Black Man Contributes. Trans. by M.B. Mader. In: Bernasconi, R., Cook, S. (eds.) Race and Racism in Continental Philosophy, pp. 287–301. Indiana University Press (2003)

Serres, M.: The Birth of Physics. Trans. by D. Webb and W. Ross. Rowman & Littlefield (2018)

Sessions, R.: The Musical Experience of Composer, Performer, Listener. Princeton University Press, Princeton (1950)

Spiegelberg, H.: Doing Phenomenology: Essays on and in Phenomenology. Martinus Nijhoff (1975)

Spinoza, B.: Ethics. In: Spinoza: Complete Works, Trans. by S. Shirley. Hackett (2002)

Stengers, I.: Thinking with Whitehead: A Free and Wild Creation of Concepts. Trans. by M. Chase. Harvard University Press (2011)

Stone, R.: Shaping time and rhythm in african music: continuing concerns and emergent issues in motion and motor action. Transcult. Music Rev. **11** (2007)

Stover, C.: A theory of flexible rhythmic spaces for diasporic African music. Ph.D. Dissertation, University of Washington (2009)

Stover, C.: Affect and improvising bodies. Perspect. New Music **55**(2), 5–66 (2018)

Stover, C.: Musical interaction in the tetravalence. In: de Assis, P., Giudici, P. (eds.) Machinic Assemblage of Desire: Deleuze and Artistic Research, pp. 55–75. University of Leuven Press (2021a)

Stover, C.: The queer rhythm of Cecil Taylor's 'enter evening.' Deleuze Guattari Stud. **15**(3), 363–385 (2021b)

Tenney, J.: Meta/Hodos and META Meta/Hodos. Frog Peak Music (1992)

van Manen, M., van Manen, M.: Doing phenomenological research and writing. Qual. Health Res. **31**(6), 1069–1082 (2021)

Varela, F., Depraz, N.: At the source of time: valence and the constitutional dynamics of affect. J. Conscious. Stud. **12**(8–10), 61–81 (2005)

Vergès, F.: A Decolonial Feminism. Trans. by A.J. Bohrer. Pluto Press (2021)

Welch, S.: The Phenomenology of a Performative Knowledge System: Dancing with Native American Epistemology. Palgrave MacMillan (2019)

Whitehead, A.N.: The Concept of Nature. Cambridge University Press, Cambridge (1964)

A Grammar of Expressive Conducting Gestures

Sylvie Gibet(✉)

IRISA, Campus of Tohannic, Université Bretagne Sud, 56000 Vannes, France
sylvie.gibet@irisa.fr

Abstract. In recent years, research on sonification has paid more attention to sound variations induced by expressive gestures. This chapter focuses on conducting gestures, emphasizing expressive gestures performed by the non-dominant hand. It is assumed that these gestures implicitly correspond to musical nuances partially encoded in the scores and convey a meaning based on a grammatical structure specific to gestural languages. We, therefore, propose to analyze these gestures in light of linguistic mechanisms that govern signed languages. In particular, we are interested in the processes of sign formation from the combination of elementary components and the inflection processes that apply to these components to efficiently generate rich and expressive sentences. Based on this grammatical theory underlying sign language and a sound-tracking methodology, we create and evaluate a new dataset of expressive conducting gestures.

Keywords: Expressive gestures · Conducting · Grammar · Sign language

1 Introduction

This chapter presents a study focused on expressive musical gestures that represent the interpretation of musical works. Our search for a gestural language to drive digital sound systems naturally led us to consider conducting, the gestural art of directing musical performances by orchestras or choirs in rehearsal or concert situations. Conducting relies on proven techniques that have evolved over the centuries, from the ancient art of *chironomy*—a conducting technique that uses hand gestures to direct musical performances, typically Gregorian chants in choirs—to recent orchestral techniques developed for classical music or other music ensembles. In this context, the orchestra can be considered a "meta-instrument," where the performers master their instrument and are guided by the conductor's gestures to perform according to the musical intention.

Conducting gestures are fascinating because they embody a deep understanding of the musical piece. The conductor, often a skilled musician, can indeed internalize the musical intent of the work as it was imagined and conceived by the composer. They can encode sound images to apprehend the organization and the streaming of the musical discourse, the parallel melodic lines, the rhythms,

A. R. Jensenius (Ed.): SSD 2022, CRSM 12, pp. 67–91, 2024.
https://doi.org/10.1007/978-3-031-57892-2_5

the variations and breaths, the dynamics of the sounds, the quality of the timbre, etc. Then, through their body language and gestures, the conductor directs and motivates the musicians, efficiently conveys the organizational and temporal elements of the work, ensures its metrical development, and transfers its expressive and dynamic strengths.

Finally, conducting gestures are designed for effective communication. Although idiosyncratic, a set of them can be shared by a large community of musicians [24]. On the other hand, when semantically close gestures are characterized by different realizations, they share similar features in form or kinematics. This sharing, as well as transmission over time, requires an encoding of gestures [34] governed by rules of economy specific to gestural languages. In this chapter, we hypothesize that these rules constitute a grammar of conducting gestures.

Whereas most studies of conducting gestures focus on the gestures made by the dominant hand, i.e., the beating gestures that indicate the structural and temporal organization of the musical piece (tempo, rhythm), this chapter is on expressive conducting gestures performed by the non-dominant hand. These gestures show other aspects of music performance and interpretation, including variations in dynamics and intensity, musical phrasing or articulation, accentuation, entrances and endings, sound quality and color, and more generally, they reflect musical intent and expressiveness. Following the hypothesis that there exists a set of meaningful gestures or features shared by conductors, we propose a grammar of expressive gestures that draws directly from the grammatical foundations of sign languages for the Deaf. These gestures have some common properties with conducting gestures, as they are both visual and gestural languages; that is to say, they use the sensorimotor system to produce the gestures and the visual receptors to receive the information. Moreover, similar mechanisms can be observed, both in conducting and sign languages, including iconic dynamics and spatial referencing mechanisms to describe and manipulate metaphorical or metonymic entities. Both use space, whether the body space or the space in which the gesture unfolds, thus promoting expression within the narrative or along the musical discourse. We, therefore, propose to analyze conducting gestures in the light of sign language gestures.

After positioning our approach with sound-related gestures and conducting gestures, we propose in this chapter to analyze the linguistic similarities between the conductor's gestures and those of sign languages [10]. This approach leads us to define a repertoire of expressive gestures classified into four main categories (Articulation, Dynamics, Attack, and Cut-off) corresponding to classically used sound modulations. Within each category, we define several expressive variations. Our methodological approach can be linked to the theory of sonorous objects [32], and by extension, to that of gestural-sonorous objects [12]. Following this methodology, sound objects are first defined and grouped into functional categories, and then gestures and their variations are identified. We then present our data collection and propose a qualitative evaluation of our gestural dataset using machine learning before briefly reviewing the research challenges for gesture recognition and motion-to-sound transformation systems.

2 Related Work

2.1 Sound-Related Gestures

As mentioned in [13], many works concern the study of musical gestures in fields such as musicology, cognitive science, gesture linguistics, computer music, and human-computer interaction.

Research in gestural control of musical instruments, both physical and simulated, has highlighted the many possibilities to control sound features with various mapping schemes. Beyond the analog or digital relationship between gestural features (including geometrical, kinematic, or physical features), and sound features (including temporal, spectral, or psycho-acoustical features), there are cognitive relationships based on abstract representations of mental images of sounds or movements. This can be connected to the theory of musical sounds presented in [32], where the acoustic substrate of sounds is potentially associated with perceptual images. This theory has been extended to the concept of embodied gestural-sonorous objects [12].

There have been proposals for the classification of sound-related gestures [7]. Four functional aspects of musical gestures are usually considered [15]:

1. Sound-producing gestures, including excitatory gestures such as hitting, bowing, plucking, blowing, and modifying gestures such as continuous modulations of pitch or timbre
2. Sound-facilitating gestures that support the sound-producing gestures and include support, phrasing, and entrained gestures
3. Sound-accompanying gestures that follow the music
4. Communicative gestures, intended for communication

For sound-producing gestures, the relationship between sound and body motion is well understood by musicians. Godøy et al. [12] argue that different theories can explain the gesture–sound link. According to the ecological perspective, auditory perception exploits cues from previous experiences to produce patterns that give meaning to sound. Other researchers share the idea that motor production is involved in the perception of sound. More specifically, the *motor theory of speech perception* [22] holds that the listener recognizes speech by activating the motor programs that would produce sounds like those that are being heard. This theory can be transposed to sign language gestures with the *motor theory of sign language perception* [9]. In this case, the linguistic knowledge is embodied into sensory-motor processes, where sensory data may be visual clues (iconic gestures) or perception of action, and motor commands put into action the multiple degrees of freedom of the articulated system. Our approach builds on these theories from a linguistic point of view.

2.2 Conducting Gestures

Since Mathews' research on conductor programs [3], much work has focused on conducting gestures, from the analysis and recognition of gestures to their use in gesture-controlled sound systems [16,28]. Many different sensors have been used to capture conductor's gestures, from commercial sensors (e.g., accelerometers, gyroscopes, infrared cameras, and electromyographic (EMG systems), to sensors designed specifically for conducting, such as the MIDI Baton [17]. Gesture follower systems have been developed, for example, the *Conductor Follower* [6], or interactive systems using sensor gloves for capturing expressive gestures [27]. Other approaches have led to the recognition of gestures, notably using Hidden Markov Models. This is the case of the system that follows both the rhythm and the amplitude of the right hand, as well as the expressive gestures of the left hand [18], or the system that follows and recognizes conducting gestures by real-time warping of the observed sequence to the learned sequence [1]. These capturing devices and gesture tracking and recognition models of conducting gestures have led to multiple systems that map gestures to sound synthesis [33]. These include systems for live performances, home entertainment, interactive public installations, or even conductor training systems [16]. More recently, a sound system that follows conducting gestures has been proposed [19], and machine learning approaches have been developed in music conducting [26,31].

Our approach goes beyond existing studies of conducting gestures, which aim to map gestures to sound systems. Instead, we return to a structural analysis of gestures with sound objects, focusing on the characteristics of gestural languages that encode information at different levels of abstraction and at different time scales.

2.3 Motivation

In our study, we focus on conducting orchestral or choral gestures and, more specifically, on expressive conducting gestures performed by the non-dominant hand. These gestures are not predefined but are hand signs that have been created over the centuries to direct a group of musicians. Thus, constrained by the structure of the musical work, the message conveyed by each gesture corresponds to a desired sound function, and the quality of the movement responds to a clear and understandable musical intention.

The expressive conducting gestures differ significantly from other musical gestures. Unlike sound-producing gestures, they do not involve any interaction with a physical instrument. If we exclude the beating gestures performed with a baton, conducting gestures involve all the degrees of freedom of the conductor's arm and torso and possibly their gaze and facial expression. On the other hand, these are anticipatory gestures based on a predictive reading of the musical score. They are anchored at key moments of the musical discourse, indicating variations of dynamics, attacks, temporal phrase variations (slowing down, acceleration, cuts), and qualitative sound variations (e.g., timbre). These are concise and efficient gestures that anticipate the sound flow in real-time while remaining synchronized with the rhythm of the music. These qualities are also those sought in gesture-controlled digital sound systems.

In this chapter, we are interested in the linguistic dimension of the conducting gestures, in the sense that they are structured in several layers of abstraction, following linguistic principles. These layers and rules define the basis of a language whose linguistic structure is similar to signed languages. The economy of representation proper to any language can be expressed by grammatical processes at different levels. First, we will see that the conducting gestures are structured according to a limited number of basic components. Modifying one of these components modifies the gesture's meaning, which can lead to expressive nuances of the musical interpretation. This linguistic specificity is also characterized by grammatical rules based on the iconic and spatial dimension of gestures, which brings the conducting gestures closer to those of signed languages.

The linguistic extension of the gestural-sonorous objects [12] is a first step towards understanding the underlying grammatical structure of expressive conducting gestures, from hand sign formation to musical phrasing. The objective is not to find a unique repertoire of expressive gestures that would be shared by all conductors; these gestures differ according to the style of the conductor and the type of music. Instead, the aim is to identify structural elements and invariant features that constitute the foundations of these gestural languages and to formalize their rules of production. By extension, this structuring might facilitate the spontaneous understanding between conductors. The examples chosen are partially inspired by those presented in [4]. Our contribution concerns the comparative linguistic study between conducting and sign language gestures, based on a formal grammar of French sign language (LSF) [25].

3 Similarities Between Sign Language and Conducting Gestures

Interestingly, there are strong similarities between conducting and sign language gestures. This similarity can be explained by the fact that these gestural languages both rely on visual and gestural modes of communication and on processes of **spatiality** and **iconicity** to build the meaning of the sequence of gestures (utterances in sign language or phrases in conducting). Spatiality is one of the fundamental elements of gestural expression, as the gestures are executed in the 3D geometric space surrounding the body. Iconicity is characterized by the more or less close resemblance between the imagined concept and the performed gesture.

Although sign languages differ from country to country, we find iconicity processes at all levels (phonological, lexical, syntactic–semantic). For example, rain can be represented in sign language with a claw handshape and a hand movement from top to bottom; variations of this sign make possible the creation of the signs river, torrent, or waterfall. Such movements can also control the sound synthesis of natural phenomena such as rain with various strengths in different environments [5]. In *Play of the Waves* (*La Mer, Debussy*), the conductor can move back and forth as if a wave was moving through the orchestra. These wave movements are very similar in sign language. It should be noted that iconic signs in sign language are not mimicry. Although they metaphorically imitate particular objects, situations, or actions, they follow specific conventions and rules.

Conducting gestures use similar conventions. For example, the conductor, like the signer, uses their body and frontal space efficiently so that the musicians can distinguish the gestures and understand their meaning. Furthermore, the signer or conductor remains in place and refers spatially to static or dynamic entities in this abstract space.

Several aspects explain the richness of expression that both gestural languages offer. First, their multimodality allows the parallel use of information conveyed by different articulatory channels (including handshapes, hand movements, torso orientation, head movements, facial expressions, and eye gaze). Second, the gestures can be broken down into meaningful components that are then recombined to form signs or phrases. Moreover, similar grammatical mechanisms can be observed, both in conducting and sign language gestures.

In this chapter, we will consider three main grammatical processes:

- The structuring into elementary components that we will call phonological components
- The spatiality
- The iconicity

We will review these three mechanisms by showing examples of the similarity between conducting and sign language gestures. In what follows, we will use Millet's grammar of French Sign Language to describe the structural aspects of both sign language and conducting gestures [25]. This grammar, very flexible and generic, can be extended to different sign languages. We will show how it can apply to both gestural languages at the lexical, syntactic, or semantic levels using inflected processes.

3.1 Phonological Components

In sign language, we can identify minimum units, called phonological components, that are structured to form the signs and that take a limited number of values. One of the basic assumptions is that two distinct signs can be differentiated when only one of the components is changed (the so-called minimal pairs). These phonological components are expressed simultaneously in multiple channels, including manual and non-manual. Manual components contain Placement (PL), Hand Configuration (HC), and Hand Movement (HM), and non-manual ones include facial expression and eye gaze. The components of conducting gestures are similar to those of sign language; we will also call them phonological components.

Both sign language and conducting gestures use a limited set of hand configurations; Millet identifies about 41 in French sign language. In conducting, the number of handshapes regularly used for expressive gestures is about 10 (included in the sign language set), but this depends on the conducting technique and style. Similarly, although continuous, the hand movements in sign language are characterized by typical trajectories that belong to a finite number of shapes. Traditionally, we consider simple elementary movements (pointing, line, arc, ellipse) or complex ones (spiral, waves, etc.). These movements can be achieved at various locations (Locus) in the signing space (starting and target points), according to the three biomechanical planes (see Sect. 4). They can be unitary or repeated movements. Conducting gestures use similar hand movements but are limited in number (mainly pointing, line, arc, and ellipse). Later, we will also see that the location of gestures, in both sign language and conducting, can take values in a finite discrete set of areas surrounding the signer or the conductor. These components, combined in parallel with the other components, form signs that convey meanings that may vary if at least one of the phonological components is modified.

For example, in sign language, the Fist or Pursed hand configuration may be used to pick up a purse or a sheet of paper. Pursed, associated with placement near the mouth and an alternating hand movement of opening and closing the fingers, becomes the sign [DUCK]. In conducting, the attack gesture with the same Fist handshape, associated with a straight downward movement, means to hit hard. The same attack gesture with a Pursed handshape, associated with repeated and precise movements of small amplitude, means beating the bar in *staccato* mode.

4 Spatiality

In sign languages, signs and sentences are organized in space. We differentiate the signer's space from the signing space. The signer's space can be divided into discrete areas along the three dimensions axis: height, distance, radial, as shown in Fig. 1 (Left), or it can be described relative to the three biomechanical planes: sagittal, frontal, and transverse (Fig. 1, Right).

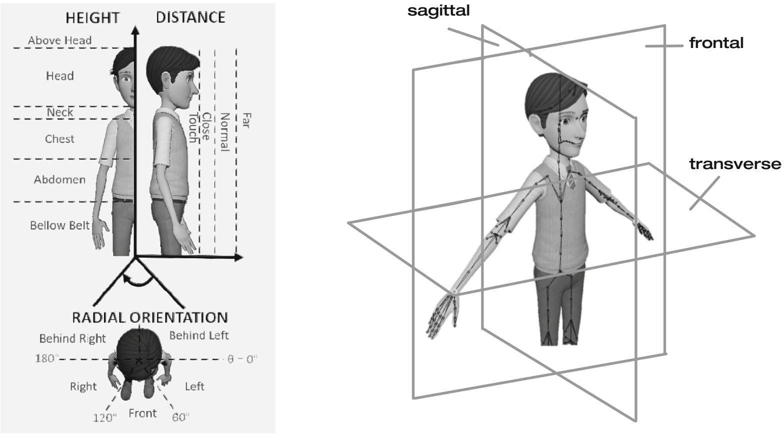


Fig. 1. The signer's space. Left: discretization along the height, distance, and radial axis (extracted from [29]). Right: the three anatomical planes: sagittal, frontal, and transverse

The signing space goes beyond the physical or geometrical signer's space—it is an abstract and delimited space, which makes spatial thinking possible. Through *spatialization*, signs can be signed at spatial references created and organized in the signing space. Locations, called Locus, become the referent locations of the entity. Deictic gestures may designate this entity by pointing with the index finger, the hand, or even with an eye gaze. Moreover, the entities can be placed relative to each other, with simplified and meaningful hand configurations, called *proforms*. This is also the space in which the discourse is deployed, which allows the syntactic consistency of sentences. For example, verbs can use trajectories in the signing space that link entities or express syntactic variations by changing personal pronouns. In French sign language, the signing space is divided into discrete pre-semantic areas (Fig. 2, Left).

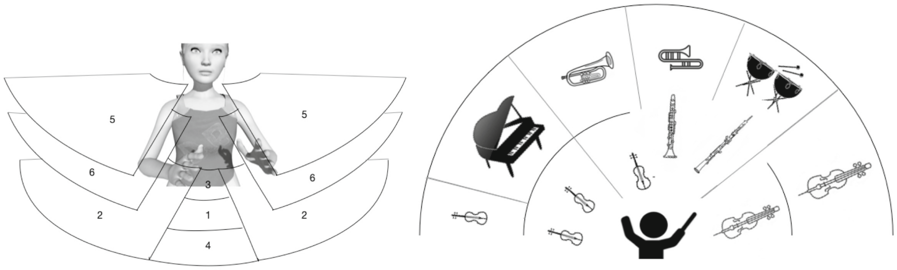


Fig. 2. Left: The pre-semanticized signing space in French sign language. 1: Neutral space; 2: Pro-3 (pronoun he/she); 3: Pro-1 (pronoun I); 4: Inanimate (goal); 5: Indefinite agent; 6: Locative linked to the verb. Right: The conducting space in a symphony orchestra

We can define the *conducting space* as a delimited, abstract space that represents the stage, with musicians and groups of instruments (Fig. 2, Right). The conductor's stage can be compared to a metaphorical surface (for the plan of the orchestral scene) or volume (for the sound), in which some entities can be designated or manipulated. There are spatial metaphors associated with this space. They can be found in the orchestra (e.g., a soloist, the timpanists, the string players), in showing, pointing, occupying space, following lines or curves, etc. They can also be found in the sound, in manipulating the instruments (pulling, pushing, gathering, etc.), or in the sound qualities (evoking a specific timbre, augmenting the brightness, etc.). During the performance, the metaphorical gestures used by the conductor are understood and translated into sounds. The musical discourse is thus elaborated in this space through spatial referencing (Locus), use of deictic gestures, following lines, paths, etc.

5 Iconicity

Iconicity is at the heart of sign languages and, more generally, gestural languages. In this section, we will explore the different types of iconicity involved at three levels: lexical, syntactic, and semantic, both for sign language and for conducting gestures.

5.1 Iconicity at the Lexical Level

At the lexical level, iconicity has an illustrative purpose in sign language, what Cuxac calls "signing by showing" [8]. The signs are thus represented by concrete objects, symbols, or metaphorical concepts. Two kinds of mechanisms can be used to modify the meaning of the signs by changing very few components.

- *A derivative-based* mechanism designates a family of signs with a similar component attached to the same meaning. For example, signs located on the side of the forehead have a meaning related to psychic activity, such as [CONCEPT], [TO THINK], or [TO INVENT] (see Fig. 3). The placement is identical, while the hand configuration and hand movement are different.
- *Inflected mechanisms* allow to modify a sign by changing one specific component in this sign:
  - *Size-and-Shape Specifiers* use hand configuration, wrist orientation, and hand movement to describe the shape and size of an object. For example, the sign [BOWL] (Fig. 4, Left) becomes a [BIG-BOWL] (Fig. 4, Right) if the shape or size of the hand trajectory is modified.
  - Although listed in the previous category, *spatialization* is implicitly included in iconic processes. An entity signed at a specific place will designate it at this location: "This bowl at this place."
  - *Proforms* represent animated entities (e.g., a person or object) characterized by a limited number of hand configurations. They function as pronouns, thus avoiding naming an entity multiple times. For example, the [PERSON] proform can be positioned in the narrative scene. In addition, one person can be represented in different postures associated with different hand configurations (e.g., a raised finger for a standing person or a curved one for a sitting person). Also, several people can be represented in space (around a table, for example) or a conference room.

Fig. 3. Derivative-based signs in sign language with the same placement on the head. Left: [CONCEPT]. Middle: [TO THINK]. Right: [TO INVENT]

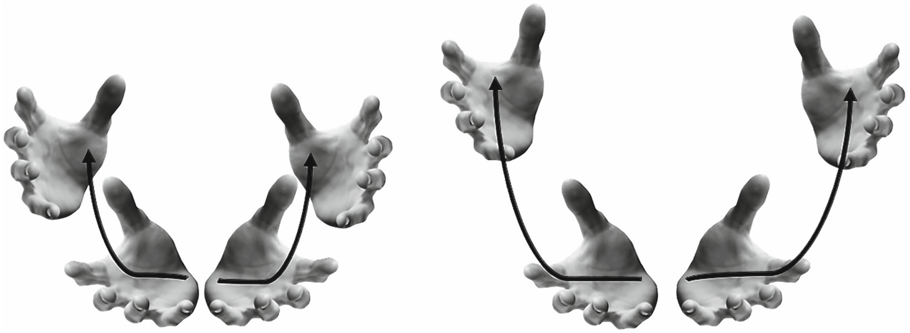


Fig. 4. Left: The standard sign [BOWL]. Right: the sign [BIG-BOWL], with size and shape specifier (extracted from [29])

In conducting gestures, we find similar iconic mechanisms. One of the significant components is the placement. Deictic gestures show locations on the stage, indicating, for example, a group of musicians. The handshape can be a traditional deictic index finger, a V handshape or a flat handshape, or even a slightly curved handshape. These deictic gestures can also be performed with different body parts, such as the head or the eye gaze. For example, the sign [LOOK-AT-ME] shown in Fig. 5 (Left), which can be used in both sign language and conducting gestures, involves a V handshape coupled with a pointing hand movement. During the execution of this gesture, the torso and the head move synchronously with the hand. In the same way, the phrase “I am looking at you” implies the same V handshape with a reversed hand movement, while the gaze is directed towards the target representing the entity to be seen, for example, a solo musician. This V-hand configuration can be considered derivative-based for a series of signs involving vision. We also find the various inflection mechanisms mentioned earlier. In the previous example, changing the gaze target or the hand trajectory changes the meaning of the gesture “I look at you.” A conducting gesture can also be performed at a specific location in the conducting stage, thus specifying an instruction to a specific group of musicians (spatialization). Furthermore, when it comes to expressing the radiating quality of an orchestral sound or a bright timbre, the movement can be more or less ample (*Size-and-Shape Specifier*) (Fig. 5, Right).

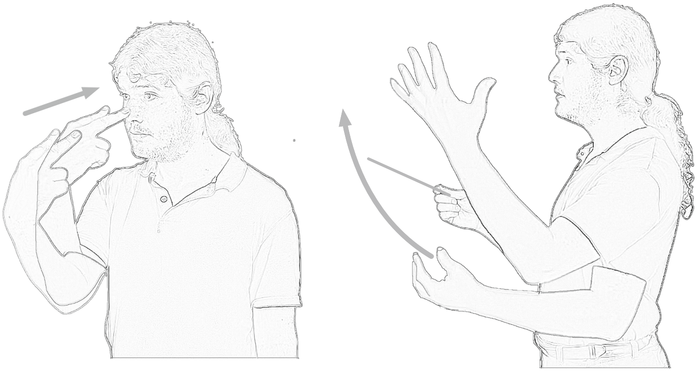


Fig. 5. Left: the sign [LOOK-AT-ME], with the V handshape, used both in conducting and sign languages. Right: the conducting gesture for increasing the brightness of the timbre

Among the functional conducting gestures, some concern dynamic gestures associated with the intensity with which the instruments play. These dynamic gestures are generally executed along vertical paths in the frontal plane: louder for an upward gesture and softer for a downward one. An inflected mechanism can be applied to the handshape, with a flat hand stretched upwards or slightly bent downwards, released at the end of the movement. Another inflection can be expressed by the kinematics of the movement: a strong acceleration will accompany a fast *crescendo* of large amplitude (from p to f). At the same time, a smooth decreasing speed will be observed for a soft *decrescendo*. Thereby, the expression of dynamics in expressive conducting gestures uses a combination of phonological components and inflectional processes similar to the *size-and-shape specifiers* of sign language.

Attack gestures can be represented by arc or line paths. Here also, the inflection can be applied to the handshape and the movement quality. A Fist handshape can express “Hitting hard”, indicating a powerful sound strike. The quality of the movement can also modulate the type of attack, with more or less weight given to the arm movement. To simulate a softer attack, the handshape can be modified, such as an open, flat hand with the palm facing down. In addition, by changing the movement and orientation of the hand, one can more closely imitate actions on specific materials (metal, wood, etc.) and use this metaphor to indicate different qualities of attack (representing, for example, various *staccato*). Again, these examples show the inflectional mechanisms used in conducting gestures, similar to sign language.

5.2 Iconicity at the Syntactical Level

In sign language, at the syntactic level, the relations between the entities of the scene are embedded in the signing space. For example, this iconicity can be represented by i) a relative Placement of the objects: e.g., “The ball is under the table,” where that proform [BALL] is shown under the proform [TABLE], the signs ball and table having been signed before, or ii) verbs described by trajectories in the signing space, also called *Indicating verbs*. Different inflected mechanisms exist for such verbs. The first one is linked to the hand configuration, which represents, for transitive indicating verbs, the direct object. For example, in the two sentences, “I give you a glass” or “I give you a coin,” the [GLASS] or the [COIN] are represented by different hand configurations: a cylindrical one for a glass or a pursed one for a coin (Fig. 6).

Fig. 6. Indicating verb with direct objects. “I give you a glass”, performed with a cylindrical HC meaning [GLASS]; with the Pursed HC, it becomes“I give you a coin”

The second inflected process for indicating verbs is achieved by changing the trajectories of the hands in the signing space, according to the agent and the recipient of the verb, respectively. Thus, in the sentence “You give me a glass,” the hand movement follows a line from a point in front of the signer to a point on their chest, whereas in the sentence “I give him a glass,” the line goes from the chest to the right side of the signer, symbolizing the 3rd person pronoun [PRO-3]. The hand configuration representing the direct object [GLASS] (cylindrical hand configuration) is identical.

In conducting gestures, conductors also use indicating verbs, as illustrated in Fig. 7 (Right) with the phrase “I propose you prepare to start” corresponding to the sign [PROPOSE-PRO2] ([PRO2] being the 2nd person). Here, the conductor uses this sign to tell the flutist: “I propose you prepare your breath to start playing.” The hand movement goes from the chest towards the flutist, and the hand spreads from closed to open. This expressive gesture is very similar to the indicating verb [OFFER-PRO2], meaning “I offer you” used in different sign languages (Fig. 7, Left).

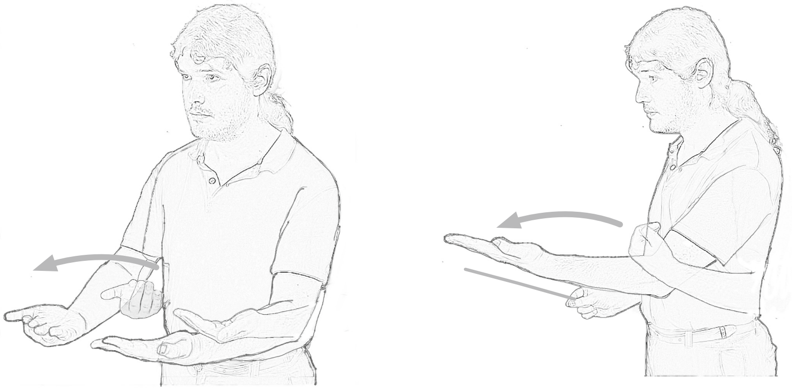


Fig. 7. Left: the French sign language indicating verb [OFFER-YOU]: “I offer you”. Right: the conducting gesture [PROPOSE-PRO2]: “I propose you prepare to start”

Many other indicating verbs borrowed from sign language are frequently used by conductors, with different meanings according to the context, for example, the sign language signs [TO INVITE], [TO BRING], [TO CARRY], etc. The expressive conducting gestures are not identical, but the inflected mechanisms follow the same rules. They involve primarily changes of the handshape, movement trajectory (direction, start and end locations that change the agent/beneficiary), and kinematics (dynamical quality). Thus, in the gesture “Pulling out an object” (Fig. 8, Left), the hand moves along a straight line from a musician on the stage toward the conductor. This means metaphorically “Pulling a sound.” It may be performed differently according to the direct object represented by the handshape. For “Pulling a full sound,” the Spread-bent handshape represents a specific brass instrument. Note that the French sign language sign [TO-ATTRACT] is very close to this expressive gesture (Fig. 8, Right).

The substitution of the Spread-bent handshape by the Pinched handshape in Fig. 9 (Right) can be used to indicate the entrance of flute sounds or vocalists (“thinner” sounds). The handshape may represent the envelope of the instruments’ spectrum. In this gesture, the other components remain the same (movement and orientation of the hand), except the placement that might express a higher pitch. This gesture is similar to the French sign language sign [TO CHOOSE] executed with the dominant hand (Fig. 9, Left).

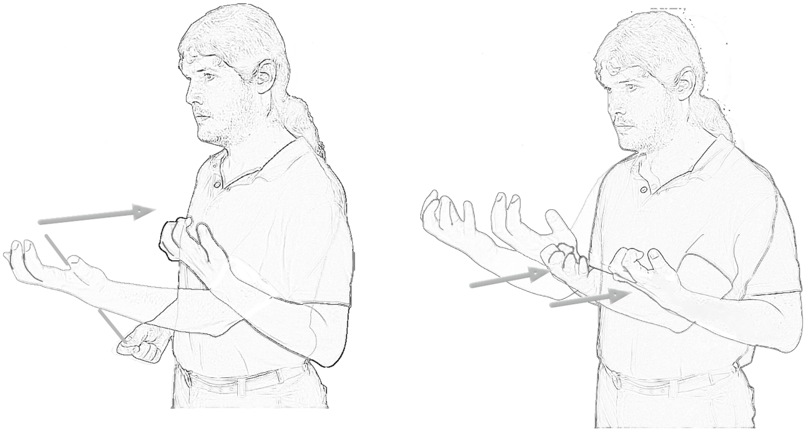


Fig. 8. Left: The conducting gesture "Pulling a brass sound". Right: The indicated verb [TO-ATTRACT] in French sign language

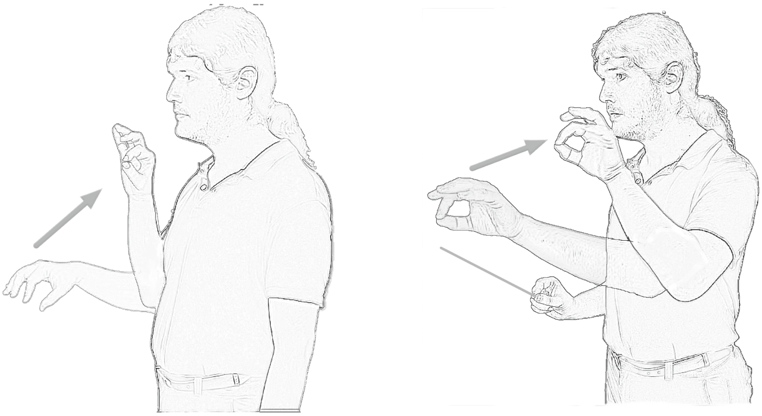


Fig. 9. Left: The indicating verb [TO CHOOSE] in French sign language. Right: The conducting gesture "Pulling a flute sound"

5.3 Iconicity at the Semantic Level

When preparing the orchestration, the conductor must understand the structure of the musical work, both in space (instruments) and time (musical development). In this preparation phase, the score is broken down into essential phases, using points of articulation or other markers (signs, text) located at the level of the instrumental ensemble or a specific group of instruments. This results in a constantly changing combination of instruments that come in and out at different times. Hence, dynamic changes, radiating quality of sound, and timbre are often achieved by the addition or the removal of instruments. During the performance, the conductor can then convey the most important cues of musical development through his gestures. These structural aspects are linked to the semantics of gestures. We distinguish spatial and temporal aspects, as well as aspects specific to the sound texture and quality.

From the point of view of spatial semantics, many gestures indicate musical paths. In particular, they show where a musical phrase begins and ends and in which direction it develops. These paths can be inscribed in the conducting space, showing, for example, the movement from one group of musicians to another. They may also represent melodic lines executed by the movement of the hand, such as a direct line, an arc, or a wave curve. The inflected elements considered here are mainly the placements or trajectories of the hand. The quality of the movement, especially the way the hand moves from one group to another, can also change to inform the musical evolution: slow, abrupt change, etc. Similarly, these trajectories can be found in the sign language narrative. For example, the French sign language sign [REGULAR] may indicate the steady flow of a crowd of people or a herd of gazelles. More generally, the movement of a vehicle or an animated entity can be represented in sign language by a trajectory between several target points in the signing space.

Specific aspects of the temporal structure of the musical work give rise to conducting gestures that indicate essential points of articulation in the development of the music. For example, the conductor, using a circular movement, can tell the musicians to keep moving at a specific tempo. In the same category of temporal semantics, similar circular and repeated trajectories can be found in the signs [TO CONTINUE] or [TO START AGAIN] in French sign language, which can also be used by conductors.

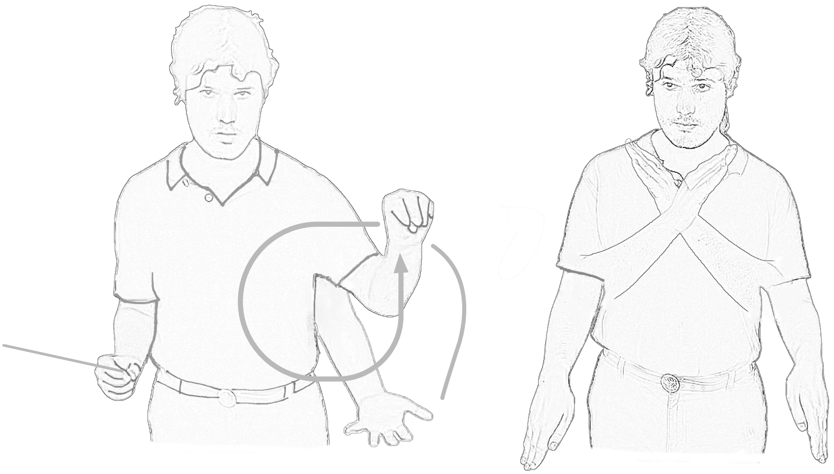


Fig. 10. Temporal semantics. Left: the conductor's gesture indicates to cut off. Right: the signs [TO STOP] in French sign language

Also temporally, the end of a musical phrase can be indicated by the conductor with a cut-off gesture (Fig. 10, Left), which can be modulated by modifying the amplitude of the trajectory, using the whole arm or only the hand, or by closing the hand more or less rapidly. The French sign language sign [TO STOP] (Fig. 10, Right) can also be used by the conductor. Numerous other gestures warn the musicians of places in the score where they should pay attention, which can result, for example, in a deictic gesture with the index finger pointing upwards or a gesture mimicking a pivot zone in a musical passage by drawing it. These gestures are similar to those used in sign language.

Fig. 11. Sound quality. Left: the conducting gesture "support an object" for a sustained sound. The French sign language signs [HEAVY] (Middle) and [LIGHT] (Right)

Semantic conducting gestures can also express aspects of the sound content or quality (timbre, brightness, spectral envelope, etc.). For example, a conducting gesture mimicking the touch of a flat surface can be used to obtain a homogeneous sound quality. This gesture has similarities with the sign [FLATENED] in French sign language, which can be used with the spread–flat hand configuration and a movement in a horizontal plane to qualify the flat structure of a surface. In the same way, a squeaking sound can be represented with a slow movement and a claw handshape to evoke a thick substance corresponding to a rough spectral texture. Such a material could be represented by the same sign in sign language, for example, to knead a more or less thick and viscous dough. In contrast, a soft material would be associated with the French sign language sign [SOFT]. Finally, the sustained (Tenuto), heavy or light quality of the sound can be expressed by the conducting gesture meaning "Supporting an object" (Fig. 11, Left), or by the signs [HEAVY] or [LIGHT] in French sign language (Fig. 11, Middle and Right).

This presentation of conducting gestures closely related to sign language is far from being exhaustive. It would be interesting to extend this study by analyzing several conducting systems and systematizing the link between expressive conducting gestures and the grammatical mechanisms presented in this chapter.

In the following, we use some examples mentioned above to build our gesture–sound database.

6 Repertoire of Four Classes of Conducting Gestures

In this section, we were interested in the definition of a restricted subset of meaningful and expressive gestures borrowed from the vocabulary of conductors. These correspond to effective sound variations, particularly those transcribed on musical scores. We, therefore, proposed a case study to analyze conducting gestures performed by the non-dominant hand. For this purpose, based on the previous study, we created a dataset of expressive gestures to control the interpretation of musical excerpts, and we evaluated this dataset following Laban's Theory of Effort [20,23]. Our motivation was twofold. First, we wanted to transfer the nuances written on orchestra scores to expressive gestures. We thereby oriented our choice towards gestures inspired by orchestral conducting for their ability to represent meaningful and expressive sound variations. Second, we relied on the grammar of French sign language [25] to take into account the elements of gestural structuring presented in Sects. 4 and 5. To create this dataset, we followed the sound-tracking methodology [30]. We defined a limited set of sound objects belonging to traditional functional categories and derived gestures that reflect these categories with appropriate expressive variations.

6.1 Sound Categories and Variations

The challenge of the conductor is to have a global idea of the composer's musical intention, to imagine sounds and colors, and to read and understand all the scores of all the instruments. Besides the information contained in the temporal organization (tempo, rhythm) of the musical excerpt, we focused on four main categories: Articulation, Dynamics, Attack, and Cut-off.

- The Articulation category is related to the phrasing of the musical discourse, which is strongly dependent on the style of the piece. It expresses how specific parts of a piece are played from the point of view of musical phrasing and how they are linked and co-articulated, taking into account the synchronization and quality of the musical sequencing. Among the techniques of articulation, we have retained in our case study three of them: *Legato* (linked notes), *Staccato* (short and detached notes), and *Tenuto* (held and sustained notes). In our examples, we know these terms and their meaning might differ according to the instrument and the musical context.
- The Dynamic category, also called Dynamics or Intensity in musicology, characterizes the music's loudness. In our study, we were interested in variations of dynamics. These variations can be progressive (smooth) or abrupt, with an increase or decrease in intensity. Four dynamic variations have been retained: *Long Crescendo*, *Long/Medium Decrescendo*, *Short Crescendo*, *Short Decrescendo*.

- The Attack category gathers different types of accents, which are indicated in the score by different symbols, but also by terms such as *sforzato* (*sfz*). In our study, we identified two primary distinctive attacks: *Hard hit*, *Soft Hit.*
- The Cut-off category expresses the way a musical phrase ends. We have retained two main variations within this last category: *Hard Cut-off*, *Soft Cut-off.*

Table 1. Repertoire of gestures: four categories (Articulation, Dynamics, Attack, Cut-off), described by their hand movements (HM) and hand configurations (HC). In each category, there are several classes. Attributes and possible values are given for each class. To simplify the table, we use Bent instead of Spread-Bent and Flat instead of Spread-Flat

Gesture categories	Classes	HM			HC
		HM-value	Plane	Quality	HC-value
Articulation	Legato	Lemniscate	Frontal	Smooth, Light	Flat/Bent
	Staccato	Line	Horizontal	Jerky, Abrupt	Pursed/O
	Tenuto	Line	Frontal	Slow, Heavy	Flat/Bent
Dynamics	Long crescendo	Arc	Frontal Up	Large, Slow,	Flat
	Medium Decrescendo	Arc	Frontal Down	Large, Slow	Bent
	Short crescendo	Arc/Line	Frontal Up	Short, Rapid	Flat
	Short Decrescendo	Arc	Frontal Down	Short, Smooth	Bent
Attack	Hard Attack	Arc/Line	Frontal Down	Rapid, Heavy	O/Pursed/Flat
	Soft Attack	Line/Arc	Frontal Down	Rapid, Light	Fist/Bent
Cut-off	Hard Cut-off	Ellipse	Frontal	Rapid, Abrupt	Flat to O
	Soft Cut-off	Ellipse	Frontal	Smooth, Slow	Flat to Pursed

6.2 Grammar of Gestures and Their Modulation

We defined a lexicon of gestures and their discrete variations according to the four categories mentioned above: Articulation, Dynamics, Attack, and Cut-off. The gestures in the Dynamic category are generally isolated actions performed in the frontal plane, upward or downward (*crescendo* or *decrescendo*), with varied duration, depending on the variation of the sound intensity (short, medium or long). The gestures in the Attack category correspond to short actions, so they can be used isolated or repeated a limited number of times, depending on the nature of the sound accents. The gestures in the Cut-off category are isolated actions performed at the end of musical phrases. They follow an elliptical trajectory that closes at the end, with a handshape that changes from open to closed. The amplitude, duration, and kinematic quality of these gestures change according to the end of the musical phrase. Unlike the other gestures, those of the Articulation category are continuous gestures repeated over one or more cycles. For this category, we considered three gestures involving various hand movements and handshapes performed in different planes with various kinematics.

The structure of these gestures is that of sign language gestures, defined by the parallelized composition of phonological components, which gather manual components (Hand Placement, Hand Movement, Hand Configuration) and non-manual components (facial mimicry, eye gaze, mouthing). In this case study, our gestural corpus is composed only of hand–arm gestures. The number and nature of hand configurations and hand movements change according to the context (nature of the musical passage, style of the conductor, etc.). In our expressive gesture dataset, we selected five basic hand configurations that can be seen in Fig. 12 (Spread-Bent, Fist, Pursed, Spread-Flat, O). We retained four hand movements: Line, Arc, Ellipse, and Lemniscate (In 2D geometry, a *lemniscate* is any of several eight-shaped curves).

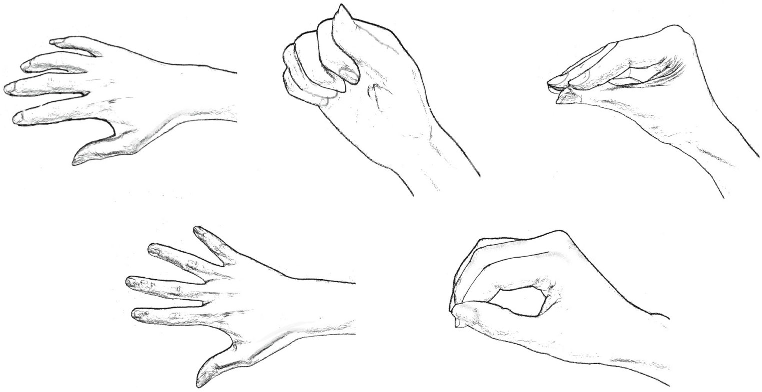


Fig. 12. List of the five selected hand configurations. Top, from left to right: Spread-Bent, Fist, Pursed. Down, from left to right: Spread-Flat, O

Combining these parallel components results in gestures with specific meaning and expressiveness. The modification of one or more components can lead to the alteration of the gesture's expressiveness. For example, an Attack can be represented by a generic gesture meaning "Hitting an object." It is mainly characterized by a vertical Arc component in the frontal plane, the type of the hand configuration indicating the nature of the object being hit, and the quality of the motion indicating the strength of the hitting.

Table 1 illustrates our gesture repertoire according to the four categories and the two dimensions. For each category, several discrete classes have been identified, associated with a set of discrete attributes and values. Moreover, the modification of the quality of the movement (above all, the kinematic quality, such as speed and acceleration, or the dynamic quality, such as the variation of the effort impelled in the gesture) modulates the gesture and, consequently, the sound nuance.

6.3 Data Acquisition Protocol

How the datasets of gestures or sounds corresponding to the categories described above are constructed is essential insofar as it determines the richness of the resulting sound and gesture variations, in particular, the quality, precision, and subtlety of expressive nuances. In the following, we describe the data acquisition methodology we adopted in our case study.

Our approach is directly inspired by *sound tracing* experiments on digital tablets whose goal was to produce 2D kinematic tracings related to sounds categorized by Pierre Schaeffer's typology of sound objects [11]. Other experiments extended this principle in 3D by exploiting motion capture technologies based on markers detected by infrared optical cameras, leading to very accurate recordings [30]. In these experiments, the gestures were performed freely while listening to sound examples with a limited number of sound features (pitch, spectral centroid, dynamic envelope). In our experiments, we also adopted this sound tracing methodology, but we focused instead on higher-level cognitive sound features, using musical excerpts related to the interpretation categories presented above. Several aspects of the sound characteristics intervene simultaneously (dynamics, timbre, etc.). Still, we have selected musical excerpts so that each of them highlights, more specifically, one of the categories identified above. In addition, to limit the variability of gestures, these were determined based on a lexicon of sign language gestures approved by conductors. These gestures, especially those involving iconic dynamics, are similar from one sign language to another and may be shared by different conductors.

Our data collection comprises two kinds of musical excerpts, for a total of 50 musical excerpts:

- 30 orchestral classical music, mostly taken from conducting scores [24]
- Two musical phrases with different variations played on a piano (one variation at a time, keeping the same tempo of 80 bps), extracted from the work of J. S. Bach: Prelude No. 1 in C Major and Cantate Bwv 147

These excerpts cover the four sound categories (Dynamics, Attack, Cutoff, Articulation), and each sound variation is represented in different musical excerpts (at least three excerpts per variation). Moreover, within the same musical excerpt, different nuances of the same variation can be present at different times of the excerpt (for example, several attacks or several cut-offs). An expert conductor validated these musical excerpts and the corresponding chosen gestures.

Our motion data was recorded thanks to a motion capture system based on passive markers and infrared cameras, which measures very precisely the position of the markers located on the body (20 markers) and the hands (8 markers per hand) with a frame rate of 200 Hz. Three subjects participated in the recording session: one expert musician with a good level of conducting, one expert musician in classical music, and one non-expert subject.

For each musical excerpt, there was a preliminary training phase in which the excerpt was played several times, and the participant was instructed to perform a given gesture while listening with their non-dominant hand. Then, the executed movement was recorded along with the corresponding sound excerpt. During each recorded sequence, the user repeated the gesture at least five times. This process was repeated for each musical excerpt. After pre-processing and manual segmentation, the dataset comprises 1265 gesture samples for each subject. Even though we got synchronized gesture and sound data, there are several drawbacks with this experimental protocol. In particular, the data were recorded in a studio and not in a real orchestral performance situation. It does not allow for analyzing the anticipation specific to the conducting gestures. In the following, we will only analyze the data of the expert subject.

6.4 Evaluation and Research Methodology

We used questionnaires to evaluate the gesture and sound databases. Questions concerning the expressive quality of gestures were related to the Effort parameters from the Laban Movement Analysis theory [20,23]. This theory identifies semantic components that describe the structural, geometric, and dynamic properties of human motion. The Effort components focus more specifically on qualitative movement aspects regarding dynamics, energy, and intent [21]. It comprises four sub-categories (Weight, Time, Space, and Flow), which vary continuously in intensity between opposing poles. The Weight Effort parameter refers to physical movement properties, the two opposing weights being *Strong* (powerful, forceful) or *Light* (gentle, delicate, sensitive). The Time Effort parameter represents the sense of urgency and has been defined by two opposing dimensions: *Sudden* (urgent, quick) and *Sustained* (stretching the time, steady). The Space Effort parameter defines the directness of the movement, which is related to the attention to the surroundings: *Direct* (focused and toward a particular spot) and *Indirect* (multi-focused and flexible). Finally, the Flow Effort parameter defines the continuity of the movement: *Free* (fluid, released) and *Bound* (controlled, careful, and restrained).

Within the preliminary study, we were interested in classifying the expressiveness of the performed Articulation gestures. These gestures were evaluated through Laban's Effort parameters (Weight, Time, Space, Flow). We used two types of questions: i) questions based on Laban Effort parameters, expressed quantitatively on a Likert scale from 1 to 7; ii) questions based on semantic terms (at least three terms per opposite pole). A total of 21 subjects answered the questionnaires. The qualitative variables were coded as numeric variables.

This allowed us to propose a classification method of expressive gestures according to the three expressive classes (Legato, Tenuto, Staccato). To classify the expressive qualities, two machine learning methods were used: Logistic Regression and Random Forest. We found an accuracy of 86% and 84%, respectively, for both sets of questions, which encourages us in our approach.

This preliminary evaluation appears relevant, as it allows us to discriminate the different expressive classes of the Articulation category. It constitutes a methodological approach that can be used in gesture recognition for sonification systems to validate the choice of gestures and their variations. If it applies to Articulation gestures, it can also be adopted for other categories.

7 Conclusion

Expressive conducting gestures are essential for guiding musicians and can be the starting point for gesture recognition systems used for sonic interaction. Such an interactive system involves recognizing the gestures being performed, adapting to their variations in real-time, and finding the most effective and meaningful mapping algorithms to match gesture and sound parameters. The specificity of such an approach and related research challenges can be summarized as:

- Multichannel structure: meaningful gestures can be defined as spatial and temporal structural patterns. These patterns contain multiple channels running in parallel, such as hand configurations and movements, eye gaze, and facial expressions. Within these patterns, we can identify stable and static areas (hand configurations and facial expressions), dynamic areas (hand movements), and transient areas (co-articulation within patterns). For example, the Cut-off gesture can be represented by an elliptical movement (hand movement channel) or a handshape (hand configuration channel) evolving between the Spread-Bent and the Pursed shape.
- Segmentation: motion capture data is represented as a multidimensional time series, which needs to be synchronized with sound to identify meaningful phases and those that constitute transitions between gestures.
- Annotation: structured and segmented gestures can be labeled and annotated (like syllables or words in a language), thus allowing the identification of meaningful motion chunks. This annotation process can be done manually or automatically. A sequence of postures can be represented by a symbolic sequence similar to a written phrase in natural language and following the indications written on musical scores.
- Expressive qualities: variational aspects of expressive gestures are inscribed into patterns that can be temporally adjusted according to the musical context and the expressive intention of the conductor. For example, the Tenuto articulation can be realized by a gesture that follows an elliptic trajectory similar to a Legato gesture; it is the variation of speed and acceleration that determines the expressive modulation.

Many unsolved questions remain. One of the central issues in recognition systems concerns gesture adaptation and anticipation, which is necessary to control time-constrained sound processes. Several approaches which exploit different types of models (Hidden Markov models, dynamic time warping, particle filtering, etc.) have been developed [1,2,14]. Another issue is related to the gesture–sound mapping process. The very different nature of sound and motion signals makes it difficult to identify the best characteristics of each and propose mappings between them.

A large amount of data is needed to learn the high variability of expressive gestures of conductors. The advent of neural architectures using deep learning opens up new possibilities for gesture recognition and mapping. Due to the time series nature of gestures, sequence-to-sequence approaches should be successful for recognizing both gestures and their variations, as far as enough data is available to train their deep architectures. The structuring of gestures into patterns might improve the performance of these neural networks. However, the data available for training such models is still limited. Moreover, there is a lack of aligned resources between motion and audio feedback that would be required to provide parallel resources for training models.

Beyond the analysis of conducting gestures, this chapter provides insights for building gesture–sound datasets for studying expressive gestures with strong semantics. The proposed methodology opens up the possibility of creating new systems of gestural interaction for sonification; it facilitates the learning of gestures and their sharing by many musicians and non-musicians and contributes to the effectiveness of semiotic communication by exploiting grammatical mechanisms specific to gestural languages.

References

1. Bevilacqua, F., Zamborlin, B., Sypniewski, A., Schnell, N., Guédy, F., Rasamimanana, N.: Continuous realtime gesture following and recognition. In: Kopp, S., Wachsmuth, I. (eds.) GW 2009. LNCS (LNAI), vol. 5934, pp. 73–84. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-12553-9_7
2. Caramiaux, B., Montecchio, N., Tanaka, A., Bevilacqua, F.: Adaptive gesture recognition with variation estimation for interactive systems. ACM Trans. Interact. Intell. Syst. **4**(4), 18:1-18:34 (2015)
3. Mathews, M.V.: The radio baton and the conductor program, or: pitch-the most important and least expressive part of music. Comput. Music. J. **15**(4), 37–46 (1991)
4. Braem, P., Bräm, T.: A pilot study of the expressive gestures used by classical orchestra conductors. J. Conduct. Guild **22**(1–2), 14–29 (2001)
5. Brizolara, T., Gibet, S., Larboulette, C.: Elemental: a gesturally controlled system to perform meteorological sounds. In: Proceedings of the International Conference on New Interfaces for Musical Expression, pp. 470–476 (2020)
6. Buxton, W., Reeves, W., Fedorkov, G., Smith, K.C., Baecker, R.: A microprocessor-based conducting system. Comput. Music. J. **4**(1), 8–21 (1980)
7. Wanderley, M.M., Battier, M. (eds.) Trends in Gestural Control of Music. Ircam-Centre Pompidou (2000)

8. Cuxac, C.: La Langue des Signes Française: les voies de l'iconicité. Faits de langues, Ophrys (2000)
9. Gibet, S., Marteau, P.-F., Duarte, K.: Toward a motor theory of sign language perception. In: Efthimiou, E., Kouroupetroglou, G., Fotinea, S.-E. (eds.) GW 2011. LNCS (LNAI), vol. 7206, pp. 161–172. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-34182-3_15
10. Gibet, S., Marteau, P.-F.: Signing avatars - multimodal challenges for text-to-sign generation. In: Face and Gesture Recognition (FG 2023), Workshop on Socially Interactive human-like Virtual Agents (2023)
11. Godøy, R.I., Haga, E., Jensenius, A.R.: Playing "air instruments": mimicry of sound-producing gestures by novices and experts. In: Gibet, S., Courty, N., Kamp, J.-F. (eds.) GW 2005. LNCS (LNAI), vol. 3881, pp. 256–267. Springer, Heidelberg (2006). https://doi.org/10.1007/11678816_29
12. Godøy, R.I.: Gestural-sonorous objects: embodied extensions of Schaeffer's conceptual apparatus. Organ. Sound **11**(2), 149–157 (2006)
13. Godøy, R.I., Leman, M.: Musical Gestures: Sound, Movement, and Meaning. Routledge, London (2010)
14. Héloir, A., Courty, N., Gibet, S., Multon, F.: Temporal alignment of communicative gesture sequences. Comput. Animat. Virtual Worlds **17**, 347–357 (2006)
15. Jensenius, A.R., Wanderley, M.M., Godøy, R.I., Leman, M.: Musical gestures: concepts and methods in research. In: Godøy, R.I., Leman, M. (eds.) Musical Gestures: Sound, Movement, and Meaning, pp. 12–35. Routledge, New York (2010)
16. Johannsen, G., Marrin-Nakra, T.: Conductor's gestures and their mapping to sound synthesis. In: Godøy, R.I., Leman, M. (eds.) Musical Gestures: Sound, Movement, and Meaning, pp. 12–35. Routledge, New York (2010)
17. Keane, D., Smecca, G., Wood, K., The MIDI Baton II. In: Proceedings of the International Computer Music Conference (ICMC 1990), Glasgow, UK, pp. 151–154 (1990)
18. Kolesnik, P., Wanderley, M.M.: Recognition, analysis and performance with expressive conducting gestures. In: ICMC Proceedings, pp. 572–575 (2004)
19. Lemouton, S., Borghesi, R., Haapamäki, S., Bevilacqua, F., Fléty, E.: Following orchestra conductors: the IDEA open movement dataset. In: Proceedings of the 6th International Conference on Movement and Computing (MOCO 2019), New York, USA, Article 25, pp. 1–6 (2019)
20. Laban, R.: The Mastery of Movement. Plays, Inc. (1971)
21. Larboulette, C., Gibet, S.: A review of computable expressive descriptors of human motion. In: Proceedings of the International Conference on Movement and Computing (MOCO 2015), pp. 21–28 (2015)
22. Liberman, A.M., Mattingly, I.G.: The motor theory of speech perception revised. Cognition **21**, 1–36 (1985)
23. Maletik, V.: Body, Space, Expression: The Development of Rudolf Laban's Movement and Dance Concepts. Mouton de Gruyte (1987)
24. Meier, G.: The Score, the Orchestra, and the Conductor. Oxford University Press, Oxford (2009)
25. Millet, A.: Grammaire descriptive de la langue des signes française: dynamiques iconiques et linguistique générale. UGA Éditions, Grenoble (2019)
26. Lee, M.: Deep neural network based music source conducting system. In: ICMC Proceedings, ICMC 2018 (2018)
27. Morita, H., Hashimoto, S., Ohteru, S.: A computer music system that follows a human conductor. Computer **24**(7), 44–53 (1991)

28. Marrin-Nakra, T.: Inside the conductor's jacket: analysis, interpretation and musical synthesis of expressive gesture. Ph.D. Dissertation. Massachusetts Institute of Technology, Program in Media Arts and Sciences (2000)
29. Naert, L.: Capture, annotation and synthesis of motions for the data-driven animation of sign language avatars. Ph.D. Dissertation. University Bretagne Sud (2020)
30. Nymoen, K., Godøy, R.I., Jensenius, A.R., Tørresen, J.; Analyzing correspondence between sound objects and body motion. ACM Trans. Appl. Percept. **10**(2), 9:1–9:22 (2013)
31. Sarasua, A., Caramiaux, B., Tanaka, A.: Machine learning of personal gesture variation in music conducting. In: ACM CHI Proceedings, pp. 3428–3432 (2016)
32. Schaeffer, P.: Traité des objets musicaux. Editions du Seuil, Paris (1966)
33. Van Nort, D., Wanderley, M.M., Depalle, P.: Mapping control structures for sound synthesis: functional and topological perspectives. Comput. Music. J. **38**(3), 6–22 (2014)
34. Rudolf, M.: The Grammar of Conducting: A Comprehensive Guide to Baton Technique and Interpretation. Schimer-Wadsworth, a part of Cengage Learning, Inc., Belmont (1995)

Performance Perspectives

Different Attitudes of Expressive Movement Awareness in Professional Musicians

Annamaria Minafra(✉)

Conservatorio "N. Piccinni", Bari, Italy
a.minafra.14@alumni.ucl.ac.uk

Abstract. This chapter explores professional musicians' awareness of expressive bodily movements, referring to Godøy's concept of *sound-action awareness* in music. Three professional musicians (a pianist, a violinist, and a guitarist) performed three tasks, each corresponding to a phenomenological reduction. Data were collected using a phenomenological approach through semi-structured interviews, observations, and audiovisual recordings. The analysis revealed three different attitudes to expressive movement awareness. The pianist showed theatrically unsynchronised expressive movements, with her musical intentions remaining at a level of pre-reflective self-awareness, perhaps due to her lack of introspective competence. The violinist became aware of his body parts involved in playing but was unaware of his performed expressive movements. The guitarist gradually reduced the expressive movements to achieve optimal performance. This study may encourage expert musicians to explore new practising procedures by developing body self-awareness. Self-reflecting on movement and its kinaesthetic feedback may contribute to achieving sound-action awareness in music, positively affecting musicians' performance and enabling them to self-correct inappropriate postures.

Keywords: Professional musicians · sound-action awareness · expressive movements · phenomenological approach

1 Introduction

Musicians become experts after years of practice (Ericsson et al. 2018; Hallam 2008). They achieve a high degree of fluency and automaticity by embedding expressive bodily movements within technical movements (Davidson 2005, 2011). When performing, musicians are mentally free and able to manage aspects 'in the moment' related to expressiveness or other problems that could emerge (Davidson and Malloch 2009), often through unconscious or pre-reflective self-awareness (Petitmengin et al. 2017). They can execute their movements effortlessly and intuitively as if unaware of their body parts while intentionally performing without any introspective process (Montero 2016). They perform at a non-conscious physiological level during which "movement and postural control are governed by a more automatic process" (Gallagher 2005, p. 73). However, the lack of reflection could undermine the development of movement and body awareness. This may cause musicians to execute unnecessary movements, restricting their chance

A. R. Jensenius (Ed.): SSD 2022, CRSM 12, pp. 95–114, 2024.
https://doi.org/10.1007/978-3-031-57892-2_6

to improve their performance (Montero 2016). When mainly directing their attention to the produced sound, musicians move their bodies while receiving various sensations integral to the musical experience (Godøy 2011, p. 231). This chapter refers to Godøy's concept of *sound-action awareness in music*, which claims that music cognition is a fusion of auditory and motor sensations. From the embodied cognition perspective, this study explores professional musicians' awareness of expressive bodily movements and whether this may assist them in phrasing.

More specifically, this chapter aims to answer the following research questions:

- How can awareness of expressive movements in professional musicians be explored?
- How can awareness of expressive movements be developed?
- How can awareness of expressive movements assist in phrasing?

To address these questions, three case studies from a larger number of cases from the author's Ph.D. dissertation are presented (Minafra 2019). The study is based on a phenomenological approach inspired by (Vermersch 2002) and qualitative thematic analysis by (van Manen 1990; 2014). The chapter starts with an overview of the theoretical foundations of embodied cognition before the case studies are presented and discussed.

2 Kinesthesia, Habits, and Sound-Action Awareness in Music

When performing, musicians reveal and shape "all mental states, both conscious and unconscious" (Davidson and Malloch 2009, p. 565). This includes conveying both technical and expressive information and their musical intentions. Two main types of performance-related body movements have been identified: *instrumental actions* and *expressive movements* (Nusseck and Wanderley 2009). Instrumental actions refer to technical aspects of musical gestures that musicians must learn to reach their expertise, such as fingering, pressure, and energy (Cadoz and Wanderley 2000, p. 73). The instrumental actions include *excitatory actions* that transmit "energy from our bodies to resonating objects such as strings, plates, tubes, and membranes" (Godøy 2011, p. 233) and *modulatory actions* employed to change the sound, such as vibrato, or modify the resonance, such as changing the bow position (Godøy et al. 2006a).

Practising hours every day for years, musicians acquire technical and expressive skills that are expressed through *expressive* or *ancillary movements* (Nusseck and Wanderley 2009). These appear to facilitate performance and are related to motor control and expressiveness (Godøy et al. 2006a). These movements also reveal the body's involvement in performing and communicating expressive musical ideas (Nusseck and Wanderley 2009).

Movement also generates sensations, so-called kinesthesia, which refers "specifically to a sense of movement through muscular effort" (Sheets-Johnstone 2011, p. 73). Kinesthesia occurs spontaneously and is generated by tactile-kinesthetic consciousness. It unfolds in a "spatiotemporal-energic flow of movement each time the person 'moves,' 'does,' and 'accomplishes' something" (Sheets-Johnstone 2020, p. 6). When playing, musicians receive continuous kinesthetic feedback, which stimulates new sensory-motor reactions and generates musical intentions through sound. Changing the quality of these movements affects expressive intentions and the interpretation of music performance (Santos 2019).

The embodied cognition research program maintains that our minds interact in perceiving external stimuli actively and continuously (cf., Varela et al. 1993; Leman 2008; Leman et al. 2018; Lesaffre et al. 2017; Newen et al. 2018; Shapiro 2019; Tomás et al. 2022). Furthermore, knowledge is embodied and cannot be separated from the sensory-motor system (Gallese and Lakoff 2005). During this process, auditory perception appears fundamental to understanding various gestures, actions, and visual information, and it seems that "we can make sense out of what we hear because we guess how the sounds are produced" (Godøy et al. 2006b, p. 258). This spontaneous phenomenon is activated by previously acquired and memorised experiences (Tomás 2022). Musicians' daily practice promotes an independent kind of body memory "consolidated into motor programs [or] muscle memories" (James 2018, p. 4). This is how musicians develop habits, embodied through "implicit memory" (Fuchs 2012). Godøy argues that musicians access sensorimotor information and internal representations by mentally simulating the movements they believe generate that sound:

> to understand musical sound as inseparable from body movement and, more precisely, to understand any sound and/or feature as actually included in some sound-producing action trajectory (Godøy 2011, p. 235).

By directing attention toward an object—in this case, sound—musicians are led to the state of consciousness. This state may relate to what Gallagher (2005) calls "performative awareness," as performers "forget" their bodies. When musicians play without any introspective process, they act through unconscious or pre-reflective self-awareness. This is "an immediate, implicit and irrelational, non-objectifying, non-conceptual and non-propositional self-acquaintance" (Zahavi 1998, p. 23) preceding any reflective act. Furthermore, musicians operate through a body schema system that includes a set of motor programs entailing complex movements and consists of.

> certain motor capacities, abilities, and habits that both enable and constrain movement and the maintenance of posture. It continues to operate, and in many cases operates best, when the intentional object of perception is something other than one's own body (Gallagher 2005, p. 24).

After years of practising, musicians build musical memory related to "procedural memory." Nijs et al. (2013, p. 471) argue that instrument-specific movements "become constituents of the dynamic structure of the body (body schema) and thereby part of the somatic know-how of the musician." However, although "awareness in music [is] an active mental process" (Godøy, 2011, p. 241) in which movement is an essential component, musicians may not be aware of these movements. This has inspired the present study, investigating professional musicians' bodily awareness.

Methods

This chapter focuses on three case studies (Yin 2018)—a pianist, a violinist, and a guitarist—who were part of a larger research project (Minafra 2019). Their subjective experience—"that which appears" (Aspers 2009, p. 1)—was examined by adopting an empirical phenomenological approach to understand the musicians' experience from a first-person perspective (Martiny et al. 2021, p. 3). Data were collected from multiple sources, semi-structured interviews, observations, and audio-visual recordings, and triangulated in the analysis process (Creswell and Miller 2000). First-person and third-person data were combined, referring to each musician's verbal responses and nonverbal information. The first-person method was carried out through verbalisation and offered easy access to subjective data. Through this procedure, "preverbal and pre-reflective aspects of subjective experience (…) are available for intersubjective and objective (biobehavioural) characterization" (Lutz and Thompson 2003, p. 37). Observation of nonverbal responses provided information for identifying the musicians' intersubjective experiences (Thompson and Zahavi 2007). This lent validation and reliability to the study (Høffding et al. 2022). The methods adopted to answer the research question applied the same procedures and tools and followed the same steps to collect and analyse data; thus, they may lead to conceptual and theoretical generalisations (Petitmengin et al. 2013).

2.1 Interviews

A phenomenological approach was adopted for the semi-structured interviews through which first-person data were collected (inspired by Vermersch 2002; Depraz et al. 2003). Phenomenology facilitates the analysis and understanding of complex aspects of consciousness and investigates how individuals experience reality (Zahavi 2010), where a specific kind of reflection or "attitude" is required to be conscious. This may occur from shifting the focus of attention from the *know-that*—the content of the action—to the *know-how*—the way of performing (Varela 1999). The first-person method offers easy access to empirical subjective data and lets participants become aware of their lived experiences (Vermersch 2002). The focus of the musicians' attention was not on the "what"—content of their experience—but on the "how"—the appearance of this content, which "usually remains unrecognized, unnoticed, or pre-reflective" (Petitmengin 2017, p. 142).

Across the interviews, musicians, re-evoking their experiences by suspending judgments, viewed their own lived experience as an observed object external to them. They moved away from their 'natural attitude' (Finlay 2014) of seeing the world with their "familiar acceptance of it" (Merleau-Ponty 2002, p. xv). In this process, the interviewer—the second person—guided the musicians to reflect on their bodies and movements, along with "slowing down" their mental activity (Petitmengin-Peugeot 2002, p. 47). The interviewer shared the first person's experience intersubjectively, including sensorimotor patterns, sensitivity, emotions, body language, language, and cultural elements (Varela and Shear 2002). While sharing experiences through a structured interview protocol, the subjectivity of the interviewer and interviewee met, generating a reciprocal relationship fundamental to understanding each other's perspectives (Høffding and Martiny 2016). Moreover, to achieve validation, the interviewer monitored the truthfulness of this re-evoking act through nonverbal signals, such as using "the present tense and unfocusing of the eyes" (Petitmengin 2017, p. 142).

2.2 Observation

The second method applied was observation. During the interview, the researcher "can empathetically grasp" and share the participants' bodily experiences by gathering direct information about nonverbal behaviour (Finlay 2006, p. 23). Having a professional musical background, the researcher based the interview questions and the musicians' observations on her embodied experience. She focused on body postures, movements, gestures, and other nonverbal indications that participants expressed when referring to playing or those parts of the body involved in playing that were intrinsically part of their lived experience. This observation was fundamental to exploring how musicians made sense of the experiences they were living during the interview and allowed the researcher to "validate the messages" conveyed through their words (Robson 1993, p. 192). Carried out from a third-person perspective, observation was undertaken in two stages. The first was conducted narratively immediately after each interview through descriptive and reflective field notes related to the musicians' non-verbal behaviours, such as gaze direction, unconscious movements, smiling, gestures, and the main ideas they expressed. The second one was based on the audio-visual recordings considering the existing literature on body language and expressive musical gestures (see Davidson 2005; 2012; Keltner 2005; McNeill 2005).

2.3 Audio-Visual Material

The audio-visual recordings of each interview allowed the researcher to analyse nonverbal behaviour, facilitating a comparison of verbal and nonverbal behaviours during the social interaction (Erickson 2011). Movements and gestures often communicate meanings that words cannot express and contribute to the shaping of utterance (Goldin-Meadow 2003). It was possible to triangulate verbal introspective information and nonverbal data, enabling the researcher to better understand the musicians' behaviour. Each interview was video-recorded in a studio with a clean background, consistent lighting, and a fixed camera near the interviewee (Jensenius 2018).

2.4 Participants

The three musicians (a pianist, violinist, and guitarist) presented in this study are expert musicians, with formal classical music training, who perform regularly. All of them also work as music teachers. When teaching, musicians mainly communicate information verbally to their students. This suggests that music teachers, being used to formalising their thoughts, might be aware of their movements while playing and be able to provide an accurate description of such. The musicians presented in this chapter were chosen because their performance displayed emblematic attitudes in showing expressive bodily movements. The pianist is a woman from Greece, the violinist a man from Spain, and the guitarist a woman from Italy, all between the ages of 36 and 42. At the time of data collection, they all taught in state music schools. To secure their anonymity, the musicians are referred to according to their instrument and a number indicating the order in which they were interviewed in the main study (Minafra 2019). The duration of each interview was between 30 and 40 min.

2.5 Procedure

The musicians were asked to perform three tasks, which involved slight modifications to playing the same piece of music. These tasks aimed to explore.

- whether musicians were aware of their movement (instrumental and expressive movements)
- whether movement awareness could be developed during the tasks
- whether developing awareness of expressive movement could affect performance

Each task represented a phenomenological reduction process in which the interviewer, through a non-judgmental conversation, asked the musicians to describe their experiences with breathing, physical tensions, relaxation, touch, mood, mental images, and anything else they felt important during their performance. Before the interview, the musicians were asked to choose the beginning (an eight-bar phrase) of an easy, slow piece to focus on the produced sound and technical movements. They were asked to perform the piece three times from memory and play it by heart to reduce the cognitive performance load (Watson 2006, p. 536). The musicians chose the following pieces of music:

- Piano-2: Chopin, Phantasie Impromptu No. 4 in C sharp Minor
- Violin-3: Mozart, Adagio Concerto No. 5
- Guitar-3: Smith-Brindle, ‘Country dance.’ In Guitarcosmos.

The first task consisted of simply performing the piece and immediately describing their feelings. Before performing the second time, the musicians were asked to mentally simulate the performance actions and what they perceived. This mental rehearsal consists of imagining an action without physical execution through “an active process during which the representation of an action is internally reproduced within any overt output” (Malouin and Richards 2010, p. 241). During this practice, complex abilities such as generating mental locomotor activities linked to the memory of sound are activated. They are based on the internal representation of movement previously acquired from a performer’s experience (Tomás 2022). Through mental simulation of the movement, the neuronal correlates of action are activated in the brain similarly to when the real action is performed (Gallese 2006).

In the third task, before performing the piece again, musicians were invited to execute it through “air instrument playing,” that is, mimicking sound-producing actions in the air as if they were playing the instrument (Godøy et al. 2006b, p. 256). This practice may assist musicians in developing kinesthetic imagination and muscle memory (Liao and Davidson 2016, p. 5), essential aspects of music performance. Immediately after the simulation, without verbalising, they were asked to play the piece again and observe their movements, breathing, sound quality, tensions, kinds of touch, possible images, possible differences with the previous performances, and whatever else they wished to communicate.

2.6 Data Analysis

Data were analysed using phenomenologically oriented qualitative thematic analysis (van Manen 1990; 2014), in which themes emerged from multiple readings of the transcriptions of the musicians' verbal and nonverbal responses. The analysis began by identifying and assembling answers that reflected the groupings of the questions related to each task before transcribing video data. The criteria for transcribing words and gestures were set after watching and re-watching the video many times. It was decided to transcribe only those gestures that referred to the body or parts of the body involved in playing (Minafra 2019). After listening to the interviewer's questions, behavioural components, emotions, and feelings expressed through words and gestures were simultaneously read in connection with each other.

3 Findings

Across the three performances, each musician showed different levels of expressive bodily movement awareness. This led to identifying three main attitudes:

- Piano-2: *Theatralization*
- Violin-3: *Automatic repertoire of expressive movements*
- Guitar-3: *Exploring movements*

These classifications were generated by considering the frequency of three main kinds of expressive bodily movements, such as head nods, trunk sway, and specific instrumental expressive movements exhibited by the musicians while performing the three tasks. To indicate the absence or presence of each specific expressive bodily movement, a quantitative measurement scale was developed: 1 = not at all; 2 = very little; 3 = little; 4 = much; 5 = very much. In the next sections, each of these attitudes will be considered.

3.1 Theatralization

The attitude shown by Piano-2 was classified as *Theatralization.* This definition came from observing her apparent "theatrical" way of swaying in all three performances of the first eight bars of Chopin's Phantasie Impromptu Op. 66. In the first execution (see Figs. 1, 2, 3), she slightly swayed side to side every two sextuplets, accompanied by a little head nod while bending forward. Although these movements are unnecessary for sound production, they seem to facilitate her motor control (Godøy et al. 2006a). They may assist her in keeping time and feeling the pulse better while trying to communicate her involvement in the music to the audience.

Piano-2 scarcely exhibited any forearm lifting and rotation, typical movements for pianists, as her forearms followed the swaying of her trunk. After the first performance in the first verbalisation process, Piano-2 did not mention any movements as if she had not paid attention to them. When asked to describe her feelings, she avoided answering and manifested discomfort, smiling for no apparent reason (Keltner 2005). This response was perhaps provoked because she was unprepared for the questions posed. She reacted as if she misunderstood the question and instead reported positive feelings about the piece.

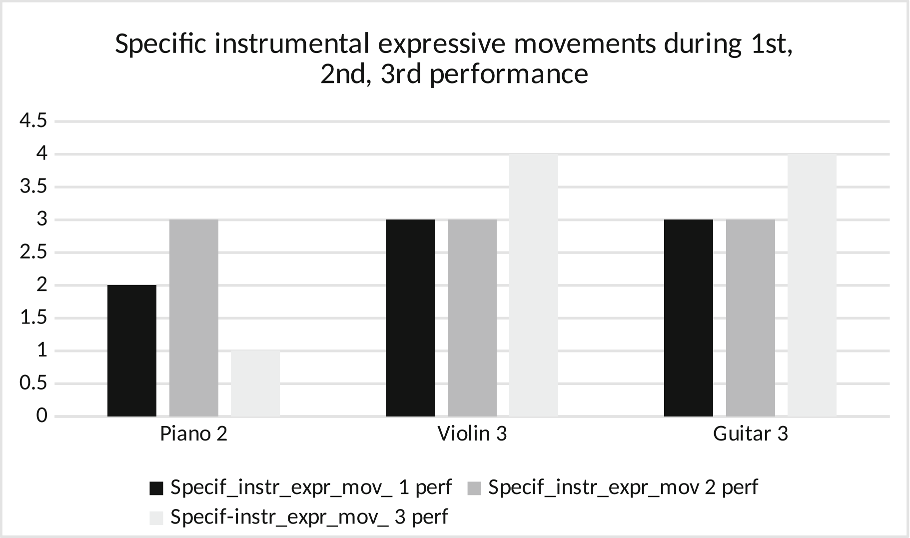


Fig. 1. Specific instrumental expressive movements during the 1st, 2nd, and 3rd performances.

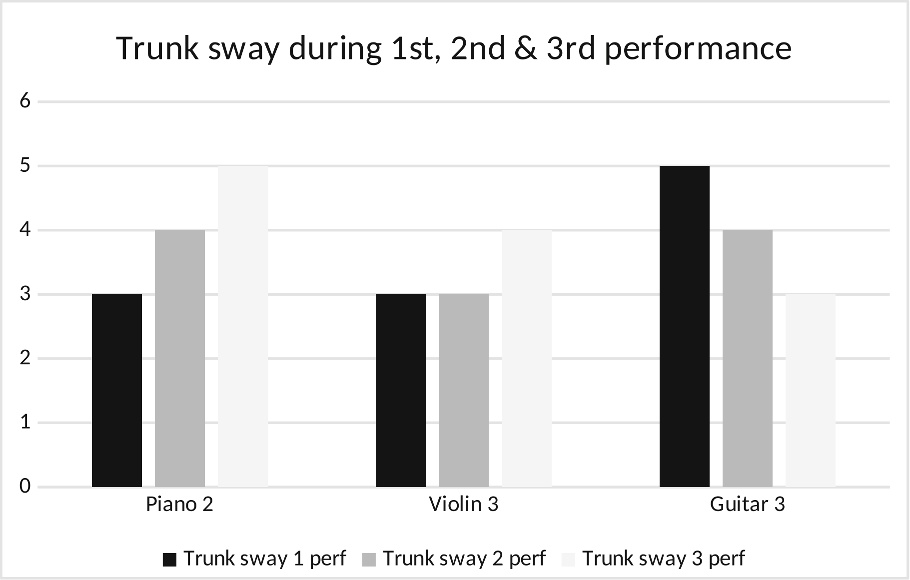


Fig. 2. Trunk sway during the 1st, 2nd, and 3rd performances.

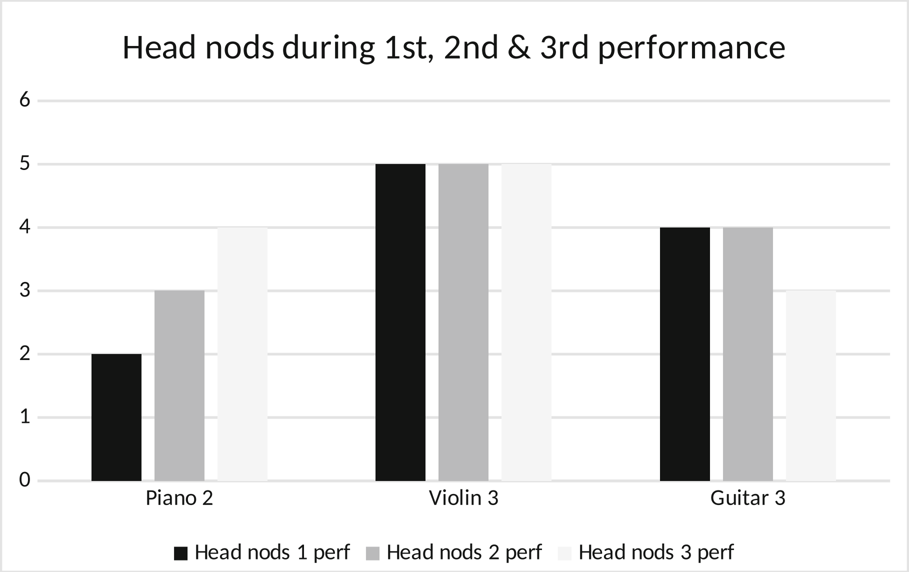


Fig. 3. Head nods during the 1st, 2nd, and 3rd performances.

> **Piano-2**: I love this piece [*smiling, closing her eyes*]. So every time [*looking in space, smiling*] I listen to … this piece or I play this I [*bringing both hands to her heart*] feel… incredibly nice.
>
> **Interviewer**: Yes, but what about your breath for example [*while she slowly laterally rocked, smiled and looked into space*].
>
> **Piano-2**: [*closing her eyes, simulating the heart beating on the chest*] oh-oh I have usually when…even-even it's not you know hum …if I play… Ca-lm you know, but usually when you finish the piece [*still smiling, simulating the heart beating in the chest*] I feel you know my heart [*continues simulating the heart beating, smiling*] here beating.

This musician showed reluctance when asked to perform the second task through mental rehearsal, misunderstanding the instructions. She thought she had to concentrate on and remember her movements during the previous performance. This misunderstanding may be caused by her unfamiliarity with the concept of mental rehearsal and her lack of confidence to ask for further explanation. These factors seemed to contribute to her discomfort.

> **Piano-2**: [*whispering*] Ah, ok … [*looking at the interviewer, scratching one hand, then touching her head, shaking her head*] and I can-not imagine my bo-ody, it's difficult, just my hands and see my…
>
> **Interviewer**: What I am asking you to do is to think about playing the piece mentally, to imagine yourself playing it… then we will speak about that.
>
> **Piano-2**: ok… yea…yea [*after 5 seconds*] I have done it.

Her nonverbal signals of frequently looking into space and smiling indicated her embarrassment (Keltner 2005). After listening again to the explanation of the task from the interviewer, she executed mental rehearsal and played the piece the second time. Compared to the first performance, the effect of theatrical movements was more evident (see Figs. 1, 2, 3). Piano-2 slightly increased instrumental expressive movement, such as lifting and rotating her forearms and emphasised both swaying and head nodding, making them wider at the same points in the music than the first performance. When asked to verbalise after the second execution, she continued to smile very frequently and manifest discomfort and uncertainty (Keltner 2005), saying that she liked mental rehearsal:

> **Piano 2**: yea,… it's like another performance… it was very nice.

She reported feeling much more relaxed than when playing the first time and, claiming not to remember it, could not make a comparison. Although she declared to be more aware of her body in the second performance, she did not describe any of the movements she performed, as the following extract shows:

> **Piano-2**: now I felt my body engaged, mo-ore than the the first time… I-I realized that the my bodyyy was… There!…. Now I was aware of my body, bu-t… the first … time I don't know what I did.

When asked to execute the third task of playing "in the air" and paying attention to movements, suspending her 'natural attitude', Piano 2 was embarrassed and hesitant:

> **Piano-2**: Oh [*whispering*] It's difficult [*looking at the interviewer, frowning, looking at the piano…shaking her head, smiling, whispering, and looking at the interviewer*] I don't know if I can do that… [*laughing*] I've never done this …. ok [*trying to start, immediately stopping and laughing*] It's strange [*laughing, and positioning her hand again looking at it*].

She started playing "in the air" but stopped when she encountered difficulties with the right hand entering at the fifth bar (see Fig. 1). Then, she continued to simulate playing but stopped again, saying sincerely that she found the task difficult. When she started playing "in the air," her head nodded, she slowly inclined her trunk forward on the first two bars, and then, from the third bar, she started swaying her pelvis side to side. Before starting the third execution, for the first time, Piano-2 displayed a "pre-gesture" moving the forearms widely, seeming to transmit her body's energy to the audience and prepare the initial sound (Lizarazu 2022). Then, like her previous performances, she started swaying at the third bar. Although her swaying was more redundant in this performance since the pelvis was involved, no expressive variations, such as sound dynamics, occurred. However, the discomfort she felt during the interview may have provoked that reaction of increased swaying. During the third verbalisation process, Piano-2 did not describe any movements and confirmed her feelings of embarrassment as revealed by her nonverbal cues. She said she was more tense and very embarrassed in the last performance and that she did not enjoy the task at all.

> **Piano-2**: It's ve-e-ery [*smiling*] dii-ifficult for me to to pla-a-y without the clavier… Aaaa but I I'm trying to think if this is because I didn't have the so-o-und or because I didn't have the clavier… I mean that I tried to ima-agine …how… this could be… ahem …in [*playing in the air*] clavier… but without sound… But [*lightly touching the clavier looking at that*] having notes here I think this could be much easier for me… Ye-ea, but [*playing in the air*] like thi-is I don't have sound I don't have the clavier so…and I am trying now to think how it cou-ld be… If you know [playing in the air] I …had sound without the clavier […] I didn't enjoy that performance… as much as the se-econd one… n-no I didn't like it.

Moreover, although she claimed to feel strange when focusing on movement related to sound and was disconcerted about the lack of physical contact with the instrument and sound during "air playing," her fluency in fingerings improved. This could result from performing the piece the third time and to the effect of "air playing."

3.2 Automatic Repertoire of Expressive Movements

Violin-3 played the first six bars of Adagio- Mozart's Concerto No. 5. Across the three performances and "air playing," he very frequently showed expressive movements typical for his instrument, such as swaying the torso back and forth, head nods, and/or moving the instrument up and down (Davidson 2012; Glowinski et al. 2014). His way of performing was classified as an *automatic repertoire of expressive movements* because his movements appeared unconscious, embedded in his instrumental actions, and seemed part of movement schemes that he unconsciously chose.

Violin-3 preferred to play the piece while sitting down. In each task, he executed the same expressive movements at specific points while showing bowing fluency, often accompanying bow changes with two kinds of head nods. Small head nods were manifested when he started the piece while breathing, and on each upbeat, sometimes also raising his eyebrows, as if these movements assisted him in preparing the beat for a new bar. In doing so, he narrowed his lips. Wider head nods were shown at each new bar, and when the music would have been more intense and forte, such as D sharp at the third and fourth bar. Here, he indicated the climax of the piece by slightly bending upwards, swaying his trunk forward, raising his eyebrows on the C sharp—the third bar—and then narrowing his lips on the E—the fourth bar. In the first verbalisation process, he initially avoided answering by directing attention to how he played rather than focusing on his bodily sensations.

> **Violin-3**: Calm, yea and elegant. I tried to… feel elegant […] there, there is ok after a long day and the first… I realized I was thinking in this moment…I'm tired

When verbalising after the second performance, he reported feeling more implicated in the music while showing the simulated gesture of glissando. The term "simulated gestures" has been chosen to indicate simulating playing from which, similarly to kinesthetic gestures, the musicians received multisensorial feedback while verbalising. These kinds of gestures seemed to assist him in self-reflecting and expressing his feelings to re-live the experienced sound quality through sensory-motor perception in his "procedural memory".

> **Violin-3**: Yea I feel more implicated in the performance. I-I-I …now I was in-side the music… I think…. I was more immersed because it wasn't musical the first time… I'm not sure how it sounded but I feel better, the sound was warmer… Warmer more romantic even high glissando [*simulating the glissando*] than the first performance.

He also said he was more conscious of breathing, which was an effect of mental rehearsal. This contributed to developing a sort of kinesthetic thinking in which kinesthesia was fundamental for formalising his thoughts:

> **Violin-3**: for the first note …I-I-I imagined breathing … and I breathed better than the first time… I imagined hm …being conscious [*simulating bowing*] …the resistance of the string with the bow to the harmony of the music … [*touching his temple*] helped me when I played to prepare to fee-eel this kind of hm … this kind of… density …maybe more conscious of playing here [*while simulating holding the bow moving his right elbow up and down*] and here.

When asked to play "in the air," his face did not express tense cues such as narrowing lips or lowering eyebrows. This was the case even though he focused on each technical movement showing more movement fluency, bowing correctly, fingering all the notes, and executing vibrato. In the third verbalisation process, Violin-3 reported perceiving parts of his body better during playing, often expressing his thoughts through "air playing" rather than words. This confirms what was observed during his "playing in the air" and suggests that he became aware of the technical movements. The kinaesthetic and sensory-motor feedback generated while verbalising and "air playing" seemed to assist him in shaping his thoughts and developing body self-awareness.

> ***Violin-3***: I didn't feel better in my whole arms, but I felt better in *[simulating the playing position with his left hand and indicating his left wrist*] in my joint … hmhm [l*ooking and moving his right hand on his left elbow*] it was [*simulating vibrato*] here [*looking at the vibrato and touching his left wrist*] for this hand and for this [*still simulating vibrato*] wrist…movement.

The "air playing" seemed to help him re-live the experienced sound quality through sensory-motor perception.

3.3 Exploring Movements

Guitar-3 played the whole piece *Country Dance* from Guitarcosmos 1 by Reginald Smith Brindle. During the interview process, she carefully observed and explored the quality of movements she executed to improve her sound quality, hence the labelling "Exploring movements." It appeared like she explored the kinaesthetic experience produced from the sensory-motor feedback related to the sound. She seemed to economise her movements, removing the automatic elements shown during Performance 1 and 2 while increasing instrumental expressive movement, such as rotating her left elbow (see Figs. 1, 2, 3).

In the first performance, she frequently nodded her head and swayed her right knee and trunk on each melodic half note. On the quarter notes, she slightly and theatrically swayed her head side to side as if to stress the ascending or descending melody line and communicate her musical involvement. She scarcely performed the instrumental expressive movement of rotating her left elbow. This movement avoids building tension in the arm and assists in executing "flexible and fluid movements" (Bosi 2017, p. 5). After playing, she took a few seconds to reflect on her feelings. When she started speaking, she accompanied her words with iconic gestures (McNeill 2005) to better describe her feeling of stiffness and the solution she found to eliminate it.

> **Guitar-3**: I felt and I am feeling some stiffness [*while moving her right fingers on her palm*] some of them instinctively kick off so I tried [*moving her right wrist and hand completely relaxed*]… to relax them … I tried to control my hand movement [*moving right hand completely relaxed*], when I play I imagine where my hand is going… both hands… I see them… I see the fingerboard, I see the gestures…I think of the sound, I remember its color and I sing… while I am playing, yes all these things.

In the second task, with her eyes closed during mental rehearsal, she slightly swayed her trunk from side to side. She knew this movement since she asked if she could do it before executing mental rehearsal. When she started the second performance, Guitar-3 showed some changes, such as positioning her right hand closer to the guitar hole, producing a different sound than in the first performance, and reducing her knee and trunk swaying until the sixth bar. However, she again started swaying from the seventh bar with the same frequency as the first task. She continued to rotate her elbow and nod with the same frequency. After playing, she said she felt more relaxed when playing the second time and perceived her body better. At this interview stage, she was more aware than the first verbalisation, mainly about her breathing.

> **Guitar-3**: Without the guitar I perceived other things more, my breath… hmmm my breathing while I sang played the piece in my mind, I perceived the sounds of this piece… I felt much more my breathing [*touching her diaphragm area*] when I have the instrument, I perceive I perceive maybe less and… I tried to be very [touching both shoulders] relaxed…
>
> When playing I perceived my body more… more than before, hmm… I combined what I experienced without the guitar this allowed me to feel things that the Fi-irst time I didn't feel….I combined things hmmm some things were so strong…it was so different… it was a completely different sensation… much stronger.

Guitar-3 seemed to explore the movements and their kinesthetic quality related to sound also when verbalising as she simulated the playing:

> **Guitar-3**: I remember the feeling when I embed [*simulating playing with her right thumb*] my finger, the pleasure, the nail [*simulating and singing some notes*]… then I remember the pleasure in embedding my finger [*still simulating*] in the string, then… I felt my breathing much more.

When performing the third task of playing "in the air," looking up at some fixed point, she focused on exploring all the movements she was executing, particularly on the right hand, pinching the strings and softly lifting her wrist. Then, four seconds later, she closed her eyes, including in the simulation her left-hand fingers, as if she realised she had forgotten them. Her main expressive movement was slightly swaying her trunk side-to-side and back-and-forth on upbeats. She also integrated small head nods into a slight swaying, which seemed to assist in phrasing. In the third performance, she was very focused on the playing and seemed to replicate the expressive movements with the same frequency executed while playing "in the air." She seemed to economise her movements, removing the automatic elements shown during Performances 1 and 2 while increasing instrumental expressive movements such as rotating her left elbow (see Figs. 1, 2, 3). She reduced swaying as if she had realised it was an unnecessary movement while increasing the rotation of her left elbow. This appeared to assist in executing flexible movements and relaxing her left arm and shoulder. She maintained the right hand's position on the guitar hole as in the second performance. In the third verbalisation process, Guitar-3 reported that she had attempted to overlap all her experiences in the three tasks.

> **Guitar-3**: These different sequences of working stages allowed me to bring with me some sensations that, compared to the first time, assisted me in playing…. Hmmm I'm not used to… When I started playing I got distracted because I had to overlap all these experiences, the memory of these experiences… because each of them left me a different memory of myself… the third time I tried to put them all together… hmmm I understood that there are some communicating channels…. But… but …. But sometimes when I play I closed them…I don't perceive everything…These channels should be opened… because they help…

4 Discussion

The findings showed three different attitudes, one for each musician. *Theatralization* was the attitude identified in Piano-2 due to her "theatrical" way of swaying. Davidson (2005) refers to the *centre of moment theory* to explain the role of swaying in pianists claiming that "the pianist's waist region functions as the central physical core for the musical expression" (Davidson 2005, p. 219). This movement stimulates the vestibular activity, arousing pleasure and constituting the top of a hierarchic process in which all the other expressive movements are integrated. In the first performance, swaying might have assisted Piano-2 in keeping time and feeling the pulse better. It also appeared to consolidate into her motor programs as an "implicit memory" in playing that piece. Her swaying increased across the other two performances, perhaps due to the self-reflection on movements. The introspection process, being new to Piano-2, who perhaps lacked introspective competence (Vermersch 2009), may have disturbed her. She had difficulties monitoring her movements because she chose a fast piece unsuitable for the task. Playing "in the air" was new for her; therefore, she had difficulties linking her inner playing with the technical movements needed to execute and monitor. This provoked embarrassment that increased in the third task. She tried to hide through the *theatricalisation* of swaying when playing, which assisted her in removing her attention from concerns about the task. Her embarrassed smiling also increased when verbalising. Piano-2's attitude suggests that her playing was based on the "just-do-it principle" (Montero 2016), with a lack of movement reflection. For this musician, sound-action awareness in music remained at the status of pre-reflective self-awareness in which her body stayed in a sort of marginal awareness (Toner et al. 2016).

Violin-3's attitude was identified as an *automatic repertoire of expressive movements*. He showed the same expressive movements at specific points in the piece across the three performances. This included head nods and moving up and down the instrument, typical for violinists (Davidson 2012; Glowinski et al. 2014). When guided to self-reflect, he experienced a sort of introspective "journey." For Violin-3, combining these three tasks with the introspection process effectively achieved awareness of instrumental movements and breathing. Particularly, the "air playing" seemed to help him re-live the experienced sound quality through sensory-motor perception without expressing any facial tension. This suggests that the lack of physical contact with the instrument while playing "in the air" made him move more smoothly, appearing to release tension. However, when he played the piece the third time, he again showed the tense cues of narrowing lips and lowering eyebrows, as they are seemingly embedded in his movement repertoire. These movements appeared to be performed unconsciously and were inconsistent with the piece's character. In the verbalisation process, he said he became more aware of the body parts involved in playing. However, he did not mention or change any expressive movements when playing. Lowering and/or raising his eyebrows and narrowing his lips, embedded in his movement repertoire, could cause tension and negatively affect performance. His attitude in executing gestures suggests that this is how he understands and communicates the musical structure. If trained to self-reflect on movement, Violin-3 could develop sound-action awareness related to instrumental movements and become aware of unnecessary and tense cues.

The attitude of Guitar-3 was described as *exploring movements*. While experiencing an "introspective journey" that began in the first task, Guitar-3 gradually shifted from pre-reflective to reflective self-awareness (Petitmengin et al. 2017). This was manifested when she described her motor imagery related to the trajectory of movement that she needed to produce sound in the first verbalisation process. In the second performance, her behaviour suggests she realised the knee and trunk swaying was unnecessary and attempted to eliminate them. However, although she did not have the power to completely remove them, she seemed to explore the movements and their kinesthetic quality related to sound. The exploration of this feeling continued when verbalising. Guitar-3 simulated the playing that generated tactile-kinesthetic feedback (Sheets-Johnstone 2011), merging auditory and motor sensations. This assisted her in going on to develop the process of sound-action awareness. In the third verbalisation process, Guitar-3 reported attempting to overlap all the lived experiences in the three tasks. However, although she tried to organise and provide continuity from the sequential succession of what Godøy (2011, p. 237) calls "sound-action chunks" in her sensory experience, this process was difficult. This could be explained by the fact that there is a basic discontinuity in motor control in the generation and control of action (Godøy 2011, p. 239). Across the three performances, she progressively economised her movements removing the automatic elements, such as swaying her trunk or right knee, while increasing instrumental expressive movement by rotating her left elbow. This avoided creating tension in her arm, helped her movement fluency, and improved her performance. The attitude displayed by Guitar-3 across the three tasks suggests that she explored her "procedural memory" related to gestures, auditory, and memory of sound, bringing her to develop sound-action awareness in music.

5 Conclusions

Due to the small number of participants, the findings from this research cannot be generalised. However, the method of inquiry allowed exploring how musicians experienced their movement awareness from their “inside” and “outside” (Høffding et al. 2022). The procedures adopted are reliable and may lead to conceptual and theoretical generalisations that may be developed with further research, adopting a “phenomenological mixed method” in which qualitative phenomenological data are combined with quantitative data (Martiny et al. 2021). This study may encourage expert musicians to explore new practice procedures by training them to develop movement and body self-awareness. Mental rehearsal and playing “in the air” while self-reflecting on movement and its kinaesthetic feedback may contribute to achieving what Godøy (2011) calls sound-action awareness in music to positively affect musicians’ performance. This process may assist them in becoming aware of their tensions, enabling them to self-correct inappropriate postures.

Acknowledgements. The author wishes to thank all the musicians who participated in this study and Professor Susan Hallam for her valuable advice.

References

Aspers, P.: Empirical phenomenology: a qualitative research approach. (The cologne seminars). Indo-Pac. J. Phenomenol. **9**(2), 1–12 (2009). https://doi.org/10.1080/20797222.2009.11433992

Bosi, B.: Left-hand injuries in guitarists: literature review and some solutions. Per Musi. 37 (2017). https://doi.org/10.35699/2317-6377.2017.5238. Accessed 10.09.2022

Cadoz, C. Wanderley, M.M.: Gesture -Music. In M. M. Wanderley, & M. Battier (eds.), Trends in Gestural Control of Music (pp. 71-94). Paris: Ircam - Centre Pompidou (2000)

Creswell, J.W., Miller, D.L.: Determining validity in qualitative inquiry. Theory into Practice. **39**(3), 124–130 (2000). https://doi.org/10.1207/s15430421tip3903_2

Davidson, J.W.: Bodily communication in musical performance. In: Miell, D., MacDonald, R., Hargreaves, D. (eds.) Musical Communication, pp. 215–237. Oxford University Press, Oxford (2005)

Davidson, J.W.: Movement and collaboration in musical performance. In: Hallam, S., Cross, I., Thaut, M. (eds.) The Oxford handbook of music psychology, pp. 364–376. Oxford University Press, Oxford (2011)

Davidson, J.W.: Bodily movement and facial actions in expressive musical performance by solo and duo instrumentalists: two distinctive case studies. Psychol. Music **40**(5), 595–633 (2012). https://doi.org/10.1177/0305735612449896

Davidson, J.W. Malloch, S.: Musical communication: the body movements of performance. In: Malloch, S., Trevarthen, C. (eds.) Communicative musicality. Exploring the Basis of Human Companionship, pp. 565- 583. Oxford University Press, Oxford (2009)

Depraz, N., Varela, F.J., Vermersch, P. (eds.): On Becoming Aware: A Pragmatics of Experiencing. John Benjamins Publishing, Amsterdam (2003)

Erickson, F.: Uses of video in social research: a brief history. Int. J. Soc. Res. Methodol. **14**(3), 179–189 (2011)

Ericsson, K.A.: An Introduction to the Second Edition of The Cambridge Handbook of Expertise and Expert Performance: Its Development, Organization, and Content. In: Ericsson, K.A., Hoffman, R.R., Kozbelt, A., Williams, A.M. (eds.) The Cambridge Handbook Of Expertise and Expert Performance, 2nd edn., pp. 3–20. Cambridge University Press, Cambridge, UK (2018)

Finlay, L.: The body's disclosure in phenomenological research. Qual. Res. Psychol. **3**(1), 19–30 (2006). https://doi.org/10.1191/1478088706qp051oa

Finlay, L.: Engaging phenomenological analysis. Qual. Res. Psychol. **11**(2), 121–141 (2014). https://doi.org/10.1080/14780887.2013.807899

Fuchs, T.: The Phenomenology of Body Memory. In: Koch, S.C., Fuchs, T., Summa, M., Müller, C. (eds.) Body Memory, Metaphor and Movement, pp. 9–22. John Benjamins Publishing Company, Amsterdam (2012)

Gallagher, S.: How the body shapes the mind. Oxford University Press, Oxford (2005)

Gallese, V.: Corpo vivo, simulazione incarnata e intersoggettività. Una prospettiva neuro-fenomenologica. In: Cappuccio, M. (a cura di), Neurofenomenologia. Le scienze della mente e la sfida dell'esperienza cosciente, pp. 293–326. Milano: Mondadori (2006)

Gallese, V., Lakoff, G.: The brain's concepts: the role of the sensory-motor system in conceptual knowledge. Cogn. Neuropsychol. **22**(3–4), 455–479 (2005). https://doi.org/10.1080/02643290442000310

Glowinski, D., et al.: Analyzing expressive styles and functions of bodily movement in violinist performance. In: MOCO 2014, Jun 16–17, 2014. Paris, France: ACM 978–1–4503–2814–2/14/06 (2014).https://doi.org/10.1145/2617995.2618023, Accessed 23 Sept 2017

Godøy, R.I.: Sound-Action Awareness in Music. In: Clarke, D., Clarke, E. (eds.) Music and consciousness, pp. 231–243. Oxford University Press, Oxford (2011)

Godøy, R.I., Haga, E., Jensenius, A.R.: Exploring music-related gestures by sound-tracing. - a preliminary study. In: 2nd ConGAS International Symposium on Gesture Interfaces for Multimedia Systems, Leeds, May 9–10 2006 [Paper] (2006a)

Godøy, R.I., Haga, E., Jensenius, A.R.: Playing "air instruments": mimicry of sound-producing gestures by novices and experts. In: Gibet, S., Courty, N., Kamp, J.-F. (eds.) GW 2005. LNCS (LNAI), vol. 3881, pp. 256–267. Springer, Heidelberg (2006b). https://doi.org/10.1007/11678816_29

Goldin-Meadow, S.: Hearing Gesture: How our Hands Help Us Think. Belknap Press of Harvard University Press, Cambridge (MA) (2003)

Hallam, S.: Music Psychology in Education. Institute of Education, London (2008)

Høffding, S., Martiny, K.: Framing a phenomenological interview: what, why and how. Phenomenol. Cogn. Sci. **15**, 539–564 (2016). https://doi.org/10.1007/s11097-015-9433-zAccessed27.01.2017

Høffding, S., Martiny, K., Roepstorff, A.: Can we trust the phenomenological interview? Metaphysical, epistemological, and methodological objections. Phenomenol. Cogn. Sci. **21**, 33–51 (2022). https://doi.org/10.1007/s11097-021-09744-zAccessed23.08.2022

James, B.: Pianism: performance communication and the playing technique. Front. Psychol. **9**, 2125 (2018). https://doi.org/10.3389/fpsyg.2018.02125Accessed28.08.2022

Jensenius, A.R.: Methods for studying music-related body motion. In: Bader, R. (ed.) Springer Handbook of Systematic Musicology. SH, pp. 805–818. Springer, Heidelberg (2018). https://doi.org/10.1007/978-3-662-55004-5_38

Keltner, D.: Signs of appeasement: evidence for the distinct displays of embarrassment, amusement, and shame. In: Ekman, P., Rosenberg, E.L., (eds.) What the face reveals. Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS), 2nd edn., pp.133–157. Oxford University Press, Oxford (2005)

Leman, M.: Embodied music Cognition and Mediation Technology. The MIT Press, London (2008)

Leman, M., Maes, P.-J., Nijs, L., Van Dyck, E.: What is embodied music cognition?. In: R. Bader (ed.) Springer Handbook of Systematic Musicology. Springer Handbook, pp. 747–760. Springer, Heidelberg (2018). https://doi.org/10.1007/978-3-662-55004-5_34, Accessed 10.08.2022

Lesaffre, M., Maes, P.-J., Leman, M. (eds.): The Routledge Companion to Embodied Music Interaction Routledge. Routledge, New York (2017)

Liao, M.-Y., Davidson, J.W.: The effects of gesture and movement training on the intonation of children's singing in vocal warm-up sessions. Int. J. Music. Educ. **34**(1), 4–18 (2016). https://doi.org/10.1177/0255761415614798

Lizarazu, H.: Pre-gesture, gesture and sound on [No] Piano: music from somewhere. In: Tomás, E., Gorbach, T., Tellioğlu, H., Kaltenbrunner, M., (eds.), Embodied Gestures, pp. 21–28. Wien: TU Academic Press (2022). https://doi.org/10.34727/2022/isbn.978-3-85448-047-1 3. Accessed 08.08.2022

Lutz, A., Thompson, E.: Neurophenomenology integrating subjective experience and brain dynamics in the neuroscience of consciousness. J. Conscious. Stud. **10**(9–10), 31–52 (2003)

Malouin, F., Richards, C.L.: Mental practice for relearning locomotor skills. Phys. Ther. **90**(2), 240–251 (2010). https://doi.org/10.2522/ptj.20090029

Martiny, K.M., Toro, J., Høffding, S.: Framing a phenomenological mixed method: from inspiration to guidance. Front. Psychol. **12**, 602081 (2021). https://doi.org/10.3389/fpsyg.2021.602081Accessed26.07.2022

McNeill, D.: Gesture and Thought. The University of Chicago Press, London (2005)

Merleau-Ponty, M.: Phenomenology of Perception. Routledge, New York (2002)

Minafra, A.: Exploring Kinaesthetic and Body Self-Awareness in Professional Musicians [Ph.D. Dissertation]. University College of London-Institute of Education, London (UK) (2019). https://discovery.ucl.ac.uk/id/eprint/10068662/

Montero, B.G.: Thought in action. Expertise and the Conscious Mind. Oxford University Press, Oxford (2016)

Newen, A., De Bruin, L., Gallagher, S. (eds.): The Oxford Handbook of 4E Cognition. Oxford University Press, Oxford (2018)

Nijs, L., Lesaffre, M. Leman, M.: The musical instrument as a natural extension of the musician. In: Castellengo, M., Genevois, H., Bardez, J.-M., (eds.) Music and its Instruments, pp. 467–484. Sampzon, Editions Delatour France (2013)

Nusseck, M., Wanderley, M.M.: Music and motion—how musicrelated ancillary body movements contribute to the experience of music. Music Percept. Interdisc. J. **26**(4), 335–353 (2009). https://doi.org/10.1525/mp.2009.26.4.335

Petitmengin-Peugeot, C.: The Intuitive Experience. In: Varela, F.J., Shear, J., (eds.), The View from Within: First-Person Approaches to the Study of Consciousness, pp. 43–77. Thorverton, UK: Imprint Academic (2002)

Petitmengin, C.: Enaction as a lived experience towards a radical neurophenomenology. Constructivist Found. **11**(2), 138–147 (2017). Accessed 27 Dec 2018

Petitmengin, C., Remillieux, A., Cahour, B., Carter-Thomas, S.: A gap in Nisbett and Wilson's findings? A first-person access to our cognitive processes. Conscious. Cogn. Int. J. **22**(2), 654–669 (2013). https://doi.org/10.1016/j.concog.2013.02.004

Petitmengin, C., van Beek, M., Bitbol, M., Nissou, J.-M., Roepstorff, A.: What is it like to meditate? Methods and issues for a micro-phenomenological description of meditative experience. J. Conscious. Stud. **24**(5–6), 170–198 (2017)

Robson, C.: Real World Research. A Resource for Social Scientists and Pratictionner-Researchers. Blackwell, Oxford, UK (1993)

Santos, T.F.: The reciprocity between ancillary gesture and music structure performed by expert musicians. In: Proceedings of the International Conference on New Interfaces for Musical Expression, UFRGS, pp. 292–297 (2019).https://doi.org/10.5281/zenodo.3672966. Accessed 09.08.2022

Shapiro, L.: Embodied Cognition, 2nd edn. Routledge, New York, NY (2019)

Sheets-Johnstone, M.: The Primacy of Movement. John Benjamins Publishing Company, Amsterdam (2011)

Sheets-Johnstone, M.: The body subject: being true to the truths of experience. J. Specul. Philos. **34**(1), 1–29 (2020). https://doi.org/10.5325/jspecphil.34.1.0001Accessed12.05.2020

Thompson, E., Zahavi, D.: Philosophical issues: Phenomenology. In: Zelazo, P.D., Moscovitch, M., Thompson, E. (eds.) The Cambridge handbook of consciousness, pp. 67–87. Cambridge University Press, Cambridge (2007)

Tomás, E.: Preface. In: Tomás, E., Gorbach, T., Tellioğlu, H., Kaltenbrunner, M., (eds.), Embodied Gestures, pp. xi–xvi. Wien: TU Academic Press (2022). https://doi.org/10.34727/2022/isbn.978-3-85448-047-1. Accessed 08 Aug 2022

Tomás, E., Gorbach, T., Tellioğlu, H., Kaltenbrunner, M.: (eds.). Embodied Gestures. Wien: TU Academic Press (2022). https://doi.org/10.34727/2022/isbn.978-3-85448-047-1. Accessed 08 Aug 2022

Toner, J., Montero, B.G., Moran, A.: Reflective and pre-reflective bodily awareness in skilled action. Psychol. Conscious. Theory, Res. Pract. **3**(4), 303–315 (2016)https://doi.org/10.1037/cns0000090. Accessed 4 Jan 2018

Van Manen, M.: Researching lived Experience. Human Science for an Action Sensitive Pedagogy. State University of New York Press, Albany, NY (1990)

Van Manen, M.: Phenomenology of Practice. Meaning-Giving Methods in Phenomenological Research and Writing. Left Coast Press, Walnut Creek (CA) (2014)

Varela, J.F.: Ethical Know-How: Action, Wisdom, and Cognition. Standford University Press, Standford (1999)

Varela, F.J., Thompson, E., Rosch, E.: The Embodied Mind: Cognitive Science and Human Experience. MIT Press, Cambridge (1993)

Varela, F.J. Shear, J.: First-person accounts: why, what and how. In: Varela, F.J., Shear (eds.) The View from Within. First-Person Approaches to the Study of Consciousness, pp. 1–14. Thorverton: Imprint Academy (2002)

Vermersch, P.: Introspection as practice. In: Varela, F.J., Shear, J., (eds.) The View from Within: First-Person Approaches to the Study of Consciousness, pp. 17–42. Thorverton (UK): Imprint Academy (2002)

Vermersch, P.: Describing the practice of introspection. J. Conscious. Stud. **16**(10–12), 20–57 (2009)

Watson, A.H.D.: What can studying musicians tell us about motor control of the hand? J. Anat. **208**(4), 527–542 (2006). https://doi.org/10.1111/j.1469-7580.2006.00545.x

Yin, R.: Case Study Research and Applications. Design and Methods, 6th edn. Sage, London (UK) (2018)

Zahavi, D.: The Fracture in Self-Awareness. In: Zahavi, D. (ed.) Self-awareness, temporality and alterity, pp. 21–40. Kluwer Academic Publishers, Dordrecht (1998)

Zahavi, D.: Naturalized Phenomenology. In: Gallagher, S., Schmicking, D. (eds.) Handbook of Phenomenology and Cognitive Science, pp. 3–19. Springer, London (2010)

Empirical Analysis of Gestural Sonic Objects Combining Qualitative and Quantitative Methods

Federico Visi[1,2](✉), Rodrigo Schramm[1], Kerstin Frödin[1], Åsa Unander-Scharin[1],
and Stefan Östersjö[1]

[1] Luleå University of Technology, School of Music in Piteå, GEMM (Gesture Embodiment and Machines in Music), Piteå, Sweden
{federico.visi,rodrigo.schramm,
kerstin.frodin}@associated.ltu.se, {asa.scharin,
stefan.ostersjo}@ltu.se
[2] Universität der Künste Berlin, Berlin Open Lab, Berlin, Germany

Abstract. In this chapter, we describe a series of studies related to our research on using gestural sonic objects in music analysis. These include developing a method for annotating the qualities of gestural sonic objects on multimodal recordings; ranking which features in a multimodal dataset are good predictors of basic qualities of gestural sonic objects using the Random Forests algorithm; and a supervised learning method for automated spotting designed to assist human annotators. The subject of our analyses is a performance of *Fragmente^2*, a choreomusical composition based on the Japanese composer Makoto Shinohara's solo piece for tenor recorder *Fragmente* (1968). To obtain the dataset, we carried out a multimodal recording of a full performance of the piece and obtained synchronised audio, video, motion, and electromyogram (EMG) data describing the body movements of the performers. We then added annotations on gestural sonic objects through dedicated qualitative analysis sessions. The task of annotating gestural sonic objects on the recordings of this performance has led to a meticulous examination of related theoretical concepts to establish a method applicable beyond this case study. This process of gestural sonic object annotation—like other qualitative approaches involving manual labelling of data—has proven to be very time-consuming. This motivated the exploration of data-driven, automated approaches to assist expert annotators.

Keywords: Gestural sonic object · multimodal analysis · machine learning · music performance · choreomusical composition

1 Introduction

The chapter begins with an introduction to central topics: the gestural sonic object, multimodal analysis of music performance, and machine learning in music practice and analysis. Then we describe the analysed piece and the methods adopted for data

A. R. Jensenius (Ed.): SSD 2022, CRSM 12, pp. 115–137, 2024.
https://doi.org/10.1007/978-3-031-57892-2_7

collection and analysis before reporting on the results of feature ranking for sound and gestural modalities and automated spotting of gestural sonic objects qualities. We discuss the implications of using the notion of gestural sonic objects in artistic practice and present some practical and conceptual considerations arising from our experience with annotating gestural sonic objects. Finally, we propose some interpretation of the feature ranking results and overall implications of these studies.

1.1 Gestural Sonic Objects

The sonic object is generally associated with the electroacoustic composition practice known as musique concrète, particularly with the work of Pierre Schaeffer and his collaborators (Schaeffer, 1966). Essentially, sonic objects are defined as fragments of musical sound approximately in the 0.5–5 s duration range that can be perceived holistically as a coherent and meaningful unit (Godøy, 2018). The concept was extended from an embodied perspective informed by motor theory by Rolf Inge Godøy (2006). From this viewpoint, sonic objects are extended with the gestural affordances of musical sound into *gestural sonic objects*. We consider the concept of gestural sonic object as a useful tool for research and artistic practice, as it allows for an analysis that uses perception as the starting point for explorations of sound and body movement in music. This resonates with the attitude of Schaeffer and collaborators, as described by Godøy, who notes that subjective perception of sound is the most important tool for research, while correlations between subjective perception and acoustic signals are mapped only at a later stage (Godøy, 2018, p. 762).

Godøy (2018, p. 768) also notes that the "motor theory suggests that production schemas are projected onto what we hear", indicating that characteristics of the gesture involved in sound production may affect how the resulting sound is experienced. The idea of such resonances between gesture and produced sound is investigated further by Godøy et al. (2016). This is done in relation to the three basic dynamic envelopes of sonic objects suggested by Schaeffer: sustained (continuous transfer of energy from the body to the instrument, resulting in a more or less continuous sound), impulsive (sudden peak of effort resulting in a sudden attack in the sound followed by a decay), and iterative (rapid back and forth motion, resulting in fast ripple-like features in the sound). These categories are effectively illustrated by Godøy et al. (2016) using the graphical representation we report in Fig. 1. Similarities between sound and motion related to these typological categories are central to the analysis we propose in this chapter.

In a project titled Music in Movement, Östersjö (2016) initiated a series of multimedia productions that sought to combine the practices of musical composition and choreography, building on a multimodal understanding of music perception and on an analytical approach to performance built on the concept of gestural sonic objects. This entailed researching how qualitative and quantitative data could be combined in the composition process. The outcome was a series of works comprising choreographies (performed by musicians, with and without their instruments), new music (for Vietnamese and Western instruments), installations, and video art, all drawn from analysis of gesture as seen in "Go To Hell", a multimedia production based on Östersjö's performance of the guitar composition Toccata Orpheus by Rolf Riehm (1990). In a PhD project carried out as a

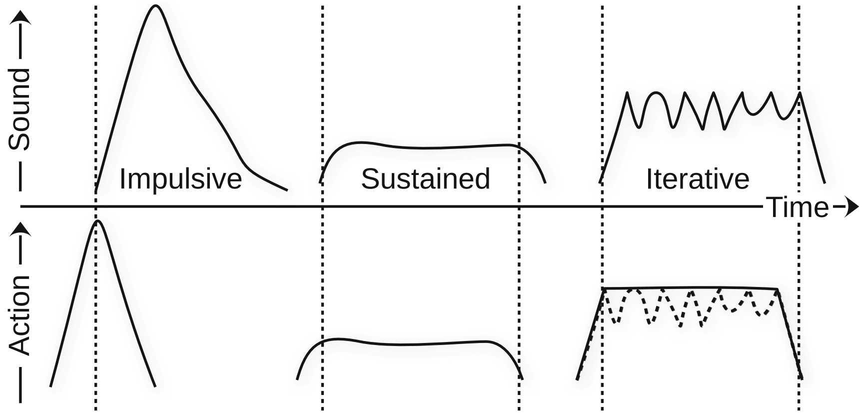


Fig. 1. Schematic illustration of the three basic dynamic typological categories of sound (top) and the corresponding motion effort types (bottom) from Godøy et al. (2016).

part of Music in Movement, Nguyễn (2019) further observes how "gesture in musical performance can be reflective of societal constructions of gender, but also holds the potential to create a platform for critique and the proposition of social change" (p. 42). Her artistic PhD project explored how the analysis of gestural sonic objects can provide the material for a compositional practice driven by the aim of producing artworks that also enact a performative critique of embodied practices of composition and performance. Such artistic application of a multimodal analysis of gestural sonic objects also informs the work discussed in the present chapter.

1.2 Multimodal Analysis of Music Performance

Embodied perspectives of human cognition have shifted scholarly understandings of the experience of music (Clayton & Leante, 2013; Leman, 2012) and have established the notion of music as a multimodal phenomenon, i.e., engaging multiple perceptual channels. Several other studies have employed multimodal data to study music performance with the premise that music is a multimodal phenomenon. To mention a few instances, the quantity of motion has been related to expressiveness (Thompson, 2012) and has been used to study the dynamic effects of the bass drum on a dancing audience (Van Dyck et al. 2013), while contraction/expansion of the body has been used to estimate expressivity and emotional states (Camurri et al. 2003).

In a previous study (Visi et al. 2020), we started developing a method for analysing music performance by combining qualitative and quantitative data. We used the stimulated recall technique, affording phenomenological variation through repeated listening. This allowed the listener to approach the listening situation, for instance, from a first- or third-person perspective (Ihde, 2012; Stefánsdóttir & Östersjö, 2022). The study argued that it is necessary to develop methods for combining qualitative and quantitative to fully understand expressive musical performance. The work presented in this chapter develops the observations by Visi et al. (2020) by proposing a method for qualitative annotations based on gestural sonic objects and techniques for quantitative data analysis aimed at supporting their empirical analysis of music performance. For the study presented in this chapter, we have recorded multimodal data from a full performance of Fragmente[2], focusing on the data obtained from the flute player. Gestural sonic objects were annotated in direct collaboration with Frödin and Unander-Scharin, who were also able to provide insight into their experience as composers and performers of the piece.

1.3 Machine Learning in Music Practice and Analysis

Machine learning has been extensively used in the context of music information retrieval, music performance analysis and generative music (Miranda, 2021). Recent machine-learning approaches require large amounts of data to train robust models. This requirement, while commonly addressed in some music-related tasks such as automated music segmentation (McCallum, 2019), deep learning-based generative music (Engel et al. 2020), and automatic chord recognition (Bortolozzo et al. 2021), is often a challenge with multimodal analysis tasks that rely on small datasets that are only partly labelled. To circumvent the limitations caused by the need for large datasets, some interactive machine-learning techniques allow the user to interact with the machine-learning model and the feature selection algorithm to guide the system towards the expected output (McCallum, 2019). Alternatively, or in combination with interactive machine learning, automated feature learning can drastically reduce the need for manual feature engineering (Yosinski et al. 2014). In this study, we have investigated several methods for automated feature selection (or ranking) and compared prediction results to better understand the relationships between features and gestural sonic object qualities.

Currently, there are several machine-learning approaches to building models that use multimodal data as input for classification tasks (Bishop, 2006). However, they usually suffer from overfitting when high data dimensionality is present and only a very low number of samples is available for training. When overfitted, a model can predict samples that are identical or very similar to the ones present in the training dataset, but it fails to generalise the unseen data distribution. In other words, the model memorises the training data instead of learning to classify new data.

There are well-known strategies for avoiding overfitting by means of regularisation and pruning (Duda et al. 2001), and the use of an external dataset is a common approach to evaluate the overfitting of a model. When overfitting, the model accuracy over the training/test dataset will usually still increase, while accuracy decreases on the evaluation dataset (unseen data). In this study, we do not have an external dataset for validation, which imposes extra difficulty when selecting the machine learning models and respective feature sets. To mitigate overfitting issues caused by small training datasets, we have considered alternative solutions already applied in the machine learning field, such as domain adaptation (Redko et al. 2019), zero/few-shot learning (Fu et al. 2020), weak supervision (Paul et al. 2018), and robust feature selection (Xie et al. 2019).

Unfortunately, domain adaptation and techniques designed to handle weakly labelled datasets still require a considerable amount of training samples to achieve robust models. One could argue that feature engineering and machine learning models could be trained on generic gesture recognition datasets (Estévez-García et al. 2015; Ruffieux et al. 2014; Tits et al. 2018) and then be transferred to the gestural sonic object context.

However, the nature of the *Fragmente* ^{2} multimodal data recording, which contains a particular configuration of sensors, combining synchronised audio, video, motion, and electromyogram, imposes restrictions and incompatibilities to a direct application of the aforementioned machine learning approaches.

In real-world applications, automated feature extraction methods usually generate redundant and noisy features. Moreover, further analysis of high-dimensional features is problematic as we cannot easily retain the physical meanings of these features. Dimensionality reduction and feature selection-based techniques have the power to discard redundant and noisy features, as well as highlight understandable data properties that can be easily connected to the studied phenomenon.

Given the reasons mentioned above, and to combine dimensionality reduction and feature selection, we have employed a wrapper method (Li et al. 2018) as our feature engineering strategy. The wrapper method uses a predefined learning algorithm (Random Forest in our case) to evaluate the quality of selected features based on the predictive performance. The strategy iterates over two steps: a) searching for a subset of features and b) evaluating the selected features. These two steps iterate until a stop criterion is satisfied. This approach worked well in this case of study, however, it is worth mentioning that wrapper methods can have an impractical search space (for d features, it is $2^{d}!$) when the number of features is very large. The rationale for a methodology that combines predictive machine learning models and feature selection is that the optimisation of these models is intrinsically connected to a good feature selection.

2 Gestural Sonic Object Multimodal Analysis

2.1 The Piece: *Fragmente*^{2}

Fragmente^{2} is a composition by Kerstin Frödin and Åsa Unander-Scharin for a solo musician and a dancer, based on the Japanese composer Makoto Shinohara's solo for tenor recorder *Fragmente* (1968). An initial artistic aim for the two artists was to explore how the musical and choreographic components could be combined in a compositional process in which neither is given less prominence than the other.

Makoto Shinohara (b. 1931) belongs to the first generation of Japanese composers who engaged with the European avant-garde movement, with a particular interest in electronic music and musique concrète. His studio work is also clearly reflected in his compositions for acoustic instruments, and this may explain how analysing musical objects in the score became a central vehicle for creating the new composition. Shinohara's score to *Fragmente* is an open-form composition consisting of 14 short fragments, in which extended techniques on the recorder are a central component.

In addition to Shinohara's 14 fragments, *Fragmente*^{2} contains three additional movement-based fragments, carried out in (relative) silence. The title, *Fragmente*^{2} (2021), suggests that the new composition widens the perspective from the sonic objects in the original score to choreomusical and gestural sonic perspectives. The notion of gestural sonic object was central in the artistic process, which also included analyses of gestural objects in the choreography of the two performers. In *Fragmente*^{2}, the joint compositional work was largely carried out on an object level, counterpointing gestural and sounding materials, while seeking independence for each part. The musical score and the dancer's choreography have a similar density of activity. Obviously, the musician's choreography does not hold the same level of refinement as the dancer's, it is instead worked out from other principles: firstly, what movements were possible to execute while playing, and secondly, how the compositional content could be further enhanced by adding choreographed movement to the musical performance. The creative process made the two artists more aware of the sounds that were produced by their moving bodies, and these were eventually integrated into the compositional structure.

An example of how the compositional methods were directly related to the analysis of different types of objects can be seen in Fig. 2, which provides a display of how the artists developed what they called "object maps," which indicate how composed objects are gesturally and temporally related in a particular fragment.

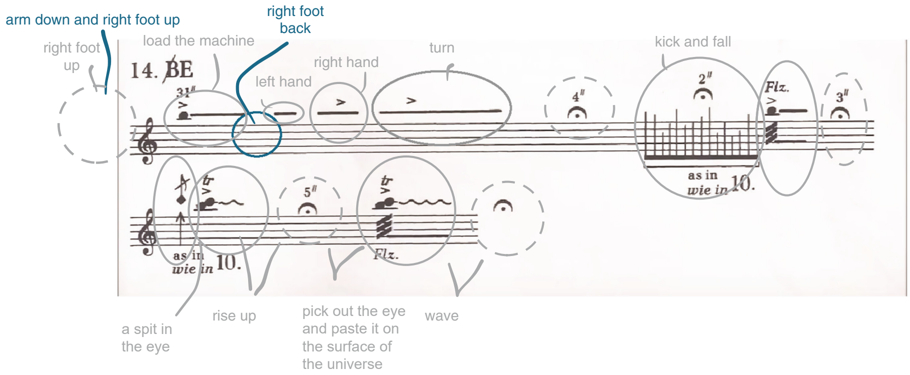


Fig. 2. Object map of Fragment 14. The musician's gestural sonic objects are marked with grey circles (dashed circles represent silence), the gestural objects carried out by the musician are further marked in blue, whereas the gestural objects of the dancer are indicated only with grey text. The score is © 1974 Schott Music, London. With kind permission of Schott Music, Mainz, Germany.

As seen in object map F14 in Fig. 2, the fragment begins with gestural objects in both parts, preceding the first sound. These first gestural objects are carried out as synchronised movement; both performers lift their right foot and put it on the left lower leg. As can be seen at the beginning of the recorder player's part, when the first gestural sonic object is played, a gestural object follows, wherein the musician's right foot returns to the floor. This leads straight into the next three gestural sonic objects (a repetition of the first note), each synchronised with gestural objects in the dancer's part. In the second line of Fragment 14, the interaction is different and starts out as cause-and-effect-like relations, leading to a more contrapuntal structure in the final objects. In this particular fragment, the form is derived from an interpretation of the original score, and the choreography both reflects and enhances these structures. While the second line activates a contrapuntal relation, the choreography still follows the original phrasing of the music. It should be noted that the relation between the original score and the new composition is different across fragments and, therefore, not always as closely related to the original piece, but sometimes seeking novel possibilities in how the different objects can be related and combined.

After establishing a working method based on object analysis, the two artists observed that the compositional process could be understood as a set of phenomenological variations of first and third-person perspectives (Ihde, 2012) when exploring and performing the relationships between movements and sounds. Hence, methods similar to those applied in the qualitative analysis of the process also form part of the artistic methodology.

Regarding the use of object analysis, the possibility of activating objects in different spatial configurations indicated the structural impact of a particular space in the compositional process. Bodily action in a particular space often decided how sonic and gestural objects were connected, and constitutes one example of phenomenological variation in the artistic process.

2.2 Quantitative Data Collection

We recorded multimodal data throughout a full performance of *Fragmente*[2]. This included multichannel audio (three channels: separate clip-on condenser microphone for the flute and a stereo recording of the hall ambience) and video (two cameras placed on the left and on the right of the performance space). Full-body motion capture, EMG (finger flexors, oblique muscles, trapezius, and deltoids), and two insole pressure sensors were captured in a configuration similar to the one adopted in a previous study by some of the authors (Visi et al. 2020).

We focused the first data collection session on the flute player, obtaining measurements of kinematics, kinetics, and muscle activity using a mobile movement analysis system comprising wireless inertial sensors and EMG electrodes (Noraxon, United States, see Fig. 3). Full body kinematics were measured with a wireless MyoMotion (Noraxon, United States) system comprising 16 inertial sensors. Sensors were mounted on the head, upper arms, forearms, hands, upper thoracic (spinal process below C7), lower thoracic (spinal process above L1), sacrum, upper leg, and lower leg and feet. The sampling rate was set to 100 Hz. The ground reaction force from the feet was measured bilaterally with wireless pressure sensor insoles (Medilogic, Germany), with a sampling rate of 100 Hz. Muscle activity was measured with EMG using a Noraxon MiniDTS (Noraxon, United States) wireless eight-sensor system. Skin preparation was done according to the Surface ElectroMyoGraphy for the Non-Invasive Assessment of Muscles (SENIAM) protocol, including shaving and rubbing with chlorhexidine disinfection. Bipolar, self-adhesive Ag/AgCl dual surface electrodes with an inter-electrode distance of 20 mm (Noraxon, United States) were placed on flexor digitorum (Blackwell et al. 1999) anterior deltoids, oblique muscles, and upper trapezius bilaterally. The EMG sampling rate was 1,500 Hz. EMG data of the finger flexors allowed us to capture finger movements, which would be difficult to capture by means of optical or inertial sensing. This way, we obtained movement-related data describing key interactions between the musician and the instrument. All the data was synchronised and imported into ELAN (Version 6.4, 2022).

Fig. 3. Wireless EMG and motion sensor setup.

2.3 Gestural Sonic Object Qualitative Annotation Method

Qualitative annotations related to gestural sonic object timing and basic typological categories (see Sect. 1.2) were added to the ELAN timeline alongside quantitative data during collaborative annotation sessions. We devised a method to annotate gestural sonic objects in an audiovisual recording of a music performance. Firstly, the performance is segmented by identifying salient events occurring in the meso timescale (approx. 0.5 – 5 s), as it is in this range that sequences of tones and movements can form a coherent object with a shape (Godøy, 2018). In this first step, segments in the meso timescale are selected and played back to determine where a gestural sonic object begins and ends. This is not a trivial task, as oftentimes, the boundaries of a gestural sonic object are not obvious. In approaching this empirical analysis task, we often referred to the fundamental characteristics of a gestural sonic object, asking ourselves:

- Is the segment long enough to perceive salient basic features such as pitch and timbre as well as elements of rhythm, texture, melody, and harmony?
- Can we perceive the segment as a whole, or is it too long?
- Does it *feel* like a single object or a sequence of objects?
- Can we describe a clear shape in the movement of the performer?
- Can we describe the performed movement as a single action?

The gestural sonic objects identified through this procedure were then analysed for the purpose of spotting basic typological categories of the dynamic envelopes (impulsive, sustained, iterative) for two modalities. This resulted in seven tiers containing time-based annotations: one indicating the gestural sonic objects and six containing the timings of the dynamic envelopes for each category and modality, as shown in Fig. 4.

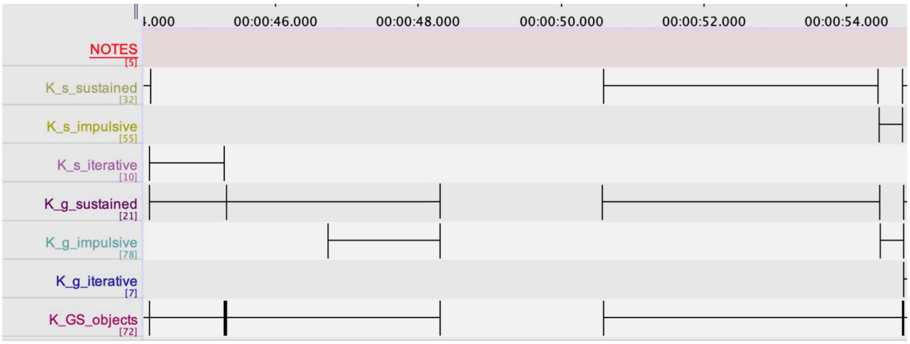


Fig. 4. Detail of the ELAN timeline showing tiers identifying gestural sonic objects (bottom) and the basic categories of dynamic envelopes for sound and gesture modalities. The tier labelled “K_GS_objects” contains the beginning and end of gestural sonic objects. The tiers with names starting with “K_g_” contain the start and end points of the respective dynamic envelopes in the gestural domain, while the tiers with names starting with “K_s_” contain the start and end points of the respective dynamic envelope for the sound domain.

For each modality, iterative, sustained, and impulsive components are annotated, thus describing how each gestural sonic object is structured. For the flautist, the analysis focused on movements related to instrumental sound production. In case of doubt or disagreement among the annotators, we referred to the questions above to reach a consensus.

This method for manual annotation of gestural sonic objects developed and tested in the present project was built on earlier experience of stimulated recall analysis (Östersjö, 2020; Visi et al. 2020). There are important similarities between this method and phenomenological approaches to music research, such as Christensen’s (2012) method of “experimental listening,” designed as “repeated listenings, guided by deliberately varied music-focusing strategies and hermeneutical strategies, and clarified by intersubjective inquiry” (p. 46). We see our annotation method as an intersubjective inquiry through what could be conceived of as a series of phenomenological variations, making deliberate use of the specific intentionality of the audio and video technologies used in the playback situations (Ihde, 2009; Verbeek, 2008).

2.4 Feature Ranking Using the Random Forest Algorithm

With the data on the gestural sonic object categories obtained using the qualitative labelling method described above, we explored the quantitative data for the purpose of understanding relationships between the typological categories of gestural sonic objects and data describing sound and body movement. We extracted features from the quantitative data and used the Random Forest algorithm to rank the best predictors for each gestural sonic object category. Random Forest is a popular ensemble learning algorithm that combines multiple decision trees to improve the accuracy and robustness of the model by reducing overfitting and increasing generalisation. It randomly selects feature subsets and data samples to train each decision tree independently before aggregating their predictions to make a final prediction.

From the motion capture data, we extracted low-level descriptors based on kinematic features, including position and its derivatives (velocity, acceleration and jerk) and contraction index. From the pressure sensors, we measured the performer's balance between the feet. From the EMG data, we calculated the root-mean-square (RMS) in order to measure the intensity of muscular activation related to the performer's finger movement while playing the instrument. From the audio recorded using the microphone mounted on the flute, we used RMS as a measurement of sound energy. We additionally extracted pitch, which is applied to capture the melodic envelope of gesture sound objects. These features contribute to a total of 134 continuous signals (audio and motion) sampled (or resampled to) 1000 times per second. The final dataset contains a single multimodal recording with a duration of 560 s, with 305 gestural sonic object annotations, including their respective gestural and sonic qualities. With the aim to capture different time resolutions of gestural sonic object events in the time sequence, we built the dataset by scanning the signal with sliding windows of multiple durations (10 ms, 100 ms, 500 ms, and 1000 ms), and fixed hop size (20 ms) for all windows.

We extracted statistical descriptors from each analysis window, independently of the signal source. The statistical descriptors reduced the dimensionality of raw data at the cost of losing time localisation. The statistical descriptors are: mean, variance, minimum and maximum values, skewness and kurtosis. The system uses a total of $K \times N \times M = 3216$ features, where $K = 6$ is the number of statistics, N is the 4 window sizes, and $M = 134$ is the number of input signals. Still, a high dimensional feature set and manual feature selection would not be a reasonable procedure. For this reason, we applied the wrapper method (Li et al. 2018), in which we randomly evaluated subset feature combinations by measuring their prediction capacity on a machine-learning model. We first reduced the original feature set dimensionality from 3216 to 50, which is computationally more manageable. To do so, we did not use Principal Component Analysis (PCA) in order to avoid losing the direct interpretation of the original data. Instead, we ranked features through the Random Forest method.

Our initial feature ranking is based on the correlation coefficient among all the variables and their respective individual variance. Thus, features that have a high correlation with several other variables are removed. The resulting feature set is further pruned such that features with very low variance are excluded from the dataset. These methods select the feature subset without any transformation that could distort the original feature interpretation. The reduced feature set is then screened by a wrapper method based on Random Forest, allowing the model to embed nonlinear relationships into a lower dimensional space, giving us a direct view of the most important features. The Random Forest prediction model is implemented with 500 trees, trained to detect gestural sonic objects and their respective qualities. To minimise overfitting, we applied cross-validation (40% for training, 40% for testing, and 20% for validation) and pruning procedure (maximum depth = 8 and maximum number of features at each split = 3). The random forest model is configured with Gini impurity for the splitting node procedure. The final feature ranking is based on the average feature score over 1000 random experiments.

2.5 Multimodal Spotting

Based on the feature selection procedure described in the previous section, we also analysed the spotting capabilities of the produced feature selection. Spotting is a technique used to identify specific patterns or events within a larger data set by applying algorithms or filters to the data. In the context of time series data, spotting techniques are often used to identify onsets and offsets of specific events or behaviours, which can then be used to segment the data and extract meaningful insights. In this work, we define the spotting procedure as detecting each starting (onset) and ending (offset) time point of a gestural sonic object.

Since the dataset is based on a single recording and, therefore, is quite small for generalisation, we do not expect to have high accuracy on the onset and offset detections. With this in mind, we trained a Random Forest-based classifier designed to maximise the onset/offset detection accuracy, that strongly penalises false positives. The result, even with a low detection rate of gestural sonic objects, can be used to semi-automatically aid the annotation process of new multimodal recordings. In this case, onset and offset detections can be used as cue points, and these first estimates can be manually confirmed or refined by experts.

3 Results

We have performed experiments to evaluate the capabilities of minimal feature sets for gestural sonic object classification. The goal was to find a considerably small set of representative features while keeping as high as possible the gestural sonic object classification accuracy. There were two reasons for a small feature set: a) fewer features can help avoid overfitting; b) low data dimensionally is more feasible to interpret.

As mentioned in Sect. 2.4, we guided the feature selection through an iterative process that ranks the best features while creating a Random Forest-based machine learning model. This process is known as the wrapper method (Li et al. 2018). Given our initial high dimensional feature set, the classifier is trained to recognise the gestural sonic object qualities as annotated in the dataset by experts. These qualities consist of the three basic dynamic envelopes for the two gestural modalities, resulting in a total of six classes. We will refer to the gestural sonic object qualities with codes shown in Table 1.

Table 1. The six main gestural sonic object qualities used in the model.

gesture sustain = 'g_sus'	gesture iterative = 'g_ite'	gesture impulsive = 'g_imp'
sound_sustain = 's_sus'	sound iterative = 's_ite'	sound impulsive = 's_imp'

Annotators might have overlapped some gestural sonic object quality labels during the annotation process. In order to accommodate these cases, the coding scheme also includes possible permutations generated from the initial six gestural sonic object qualities (e.g., ['s_sus' AND 'g_sus'] and ['s_sus' AND 'g_imp']). Figure 5 shows the sample distribution regarding each class in our annotated dataset. We have a total of 17 classes, plus the null class (NC). The NC is related to all data samples that were not labelled by the annotators. This means that part of these samples might not have been correctly assigned to a specific gestural sonic object quality and were unequivocally put in the NC fold. Since the NC is predominant in the dataset, and to avoid the excessive influence of unreliable samples and unbalanced class partitions, we randomly selected and kept only 10% of the original NC data in the final dataset.

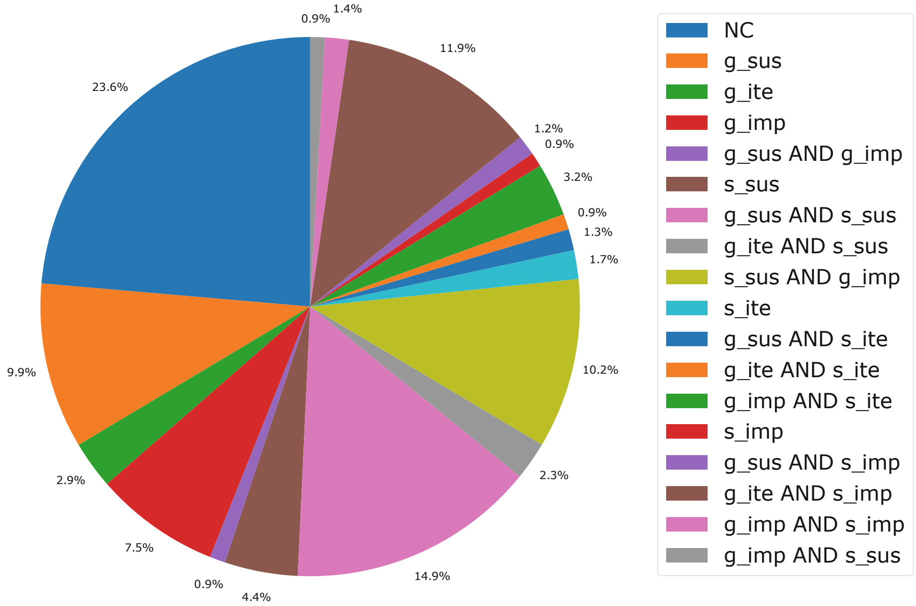


Fig. 5. An overview of the gestural sonic object quality class distribution.

3.1 Feature Ranking Per Modality

Throughout the course of the training process, the Random Forest algorithm ranks the features based on their capability to better separate the data distributions. To minimise the inevitable influence of overfitting, we applied pruning to the classification trees. A grid search experiment was used to find the minimal tree depth while keeping accuracy above 90%. Figure 6 shows the classification accuracy versus the tree depth. We found a max depth of 8 as a good compromise, reaching approximately 90% of accuracy.

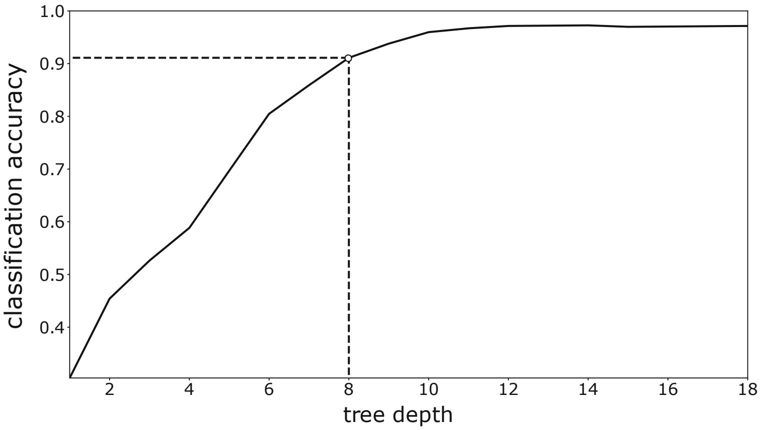


Fig. 6. Random Forest pruning: Classifier accuracy is kept over 90% to avoid overfitting.

Once we defined the maximum tree depth, we performed feature selection experiments on the following targets: a) only gesture qualities, b) only sound qualities, and c) gesture and sound qualities. Given the extremely high data dimensionality, no brute force approach would be feasible to find the best small subset of input features. Instead of an exhaustive search, the Random Forest algorithm randomly selects features from the dataset. It is worth mentioning that we cannot guarantee that we will have the optimal final feature subset. In order to increase the chances of a good feature selection, the Random Forest is configured with 500 trees, each doing random feature selection with the configuration described in Sect. 2.4. We ran each experiment 1000 times with distinct random seeds. This procedure helped to increase the feature variability and gave us a better cover of the feature search space. Tables 2 and 3 summarise the set of features that were mostly chosen across the 1000 experiments and were scored among the top 10 features while predicting qualities in the gesture, sound, and gesture-sound domains, respectively. In other words, we selected the top 10 features based on the top score occurrence frequency of the 3216 features from the initial feature set over all experiments.

In the second experiment, we used the top 50 features from the first experiment. A new Random Forest model was trained on this new subset, and we ranked the resultant top 10 features again. Tables 4 and 5 show the selected top 10 features based on their highest score for gesture, sound, and gesture-sound domains, respectively.

Table 2. The top 10 most frequently selected features for the gesture and sound domains, separately.

Rank	Gesture domain			Sound domain		
	Signal	Statistic	Window Size	Signal	Statistic	Window Size
1	RT_finger_flex	max	100	audio pitch	min	500
2	RT_finger_flex	min	500	audio pitch	mean	1000
3	RT_finger_flex	max	1000	audio RMS	mean	500
4	RT_finger_flex	min	10	audio pitch	var	500
5	RT_finger_flex	min	1000	audio pitch	min	1000
6	RT_finger_flex	max	10	audio pitch	mean	500
7	RT_finger_flex	mean	100	audio pitch	mean	100
8	RT_finger_flex	min	100	audio RMS	min	100
9	RT_finger_flex	max	500	audio RMS	min	500
10	RT_finger_flex	mean	10	audio RMS	mean	100

Table 3. The top 10 most frequently selected features for the joint gesture and sound domains, concomitantly.

Rank	Gesture-Sound domain		
	Signal	Statistic	Window Size
1	audio pitch	min	500
2	audio pitch	mean	1000
3	m1_RT_ext_oblique_rms	max	1000
4	audio pitch	min	1000
5	audio RMS	min	500
6	m1_RT_ext_oblique_rms	max	500
7	m1_RT_ext_oblique_rms	mean	1000
8	audio RMS	mean	500
9	audio pitch	mean	500
10	m1_RT_ext_oblique_rms	mean	500

Selecting features using decision trees can be challenging due to the potential for high variance and overfitting, which can lead to suboptimal performance and reduced generalisation ability of the model. A small change in the data can have a big influence on the feature selection. However, based on the k-fold cross-validation and multiple random experiments, we found consistent features that were selected repeatedly most of the time. An additional relevant observation was the importance of the multi-scale/resolution of each feature window analysis. The feature selection process picked not only a specific characteristic of the input signal but also its distinct time resolutions.

Table 4. The overall top 10 most frequently selected features for the gesture and sound domains, separately.

Rank	Gesture domain			Sound domain		
	Signal	Statistic	Window Size	Signal	Statistic	Window Size
1	m1_insoles_sum	max	1000	audio pitch	min	500
2	m1_RT_finger_flex_rms	max	1000	audio pitch	mean	100
3	m1_RT_finger_flex_rms	min	1000	audio pitch	min	1000
4	pitch_mean	mean	1000	audio pitch	mean	100
5	pitch	min	1000	audio pitch	mean	1000
6	m1_RT_ant_deltoid_rms	min	1000	audio pitch	var	500
7	Hand_tip_LT_vel	max	1000	audio pitch	mean	500
8	m1_LT_ext_oblique_rms	max	1000	CoM_3D_Z	min	1000
9	CI_movmean	mean	10	audio pitch	min	10
10	CI_movmean	max	100	CoM_3D_Z	min	500

Table 5. The overall top 10 most frequently selected features for the joint gesture and sound domains.

Rank	Gesture-Sound domain		
	Signal	Statistic	Window Size
1	audio pitch	min	1000
2	audio pitch	min	500
3	audio pitch	mean	1000
4	CI	max	1000
5	m1_RT_ext_oblique_rms	max	1000
6	audio pitch	mean	100
7	CI_movmean	min	1000
8	audio pitch	mean	500
9	m1_RT_finger_flex_rms	max	1000
10	CI_movmean	min	1000

3.2 Online Learning Investigation of Gestural Sonic Objects

A challenge has been that, due to the limited amount of data available, we faced a general sensitivity to overfitting. In order to minimise this, incremental and iterative supervision of the annotation process can be integrated with online learning models. A direct application of this kind of strategy could be extracting cue marks that indicate where gesture sonic objects are present in the timeline. Annotators could then validate and correct these cue points to improve the transcription of the recording session. Thus, in addition to the gestural sonic object quality classification task, we also investigated the spotting capabilities of our proposed multimodal feature selection method. The prediction of onsets and offsets for individual gestural sonic object qualities can potentially be used as cue points to assist an iterative and semi-supervised annotation process.

We evaluated the model's capability to increase its accuracy in the case of using our proposed multimodal feature subset and a Random Forest classifier. Figure 7 shows the accuracy while we increase the ratio of training data versus testing data. This procedure emulates an iterative annotation approach, where annotators iteratively add more valid labels to the dataset. In our experiments, the proposed model has over 70% of classification accuracy with only 3% of the training data. When using 20% of training data, the model improvement increases accuracy to over 85%. It is worth noting that because of tree pruning, the accuracy of our model has asymptotic behaviour at approximately 90%. The asymptotic behaviour of the model's accuracy at approximately 90% suggests that even as more training data is added, the model's performance is unlikely to improve beyond this level. This could be due to the limitations of the features used to train the model or the inherent complexity of the underlying patterns in the data.

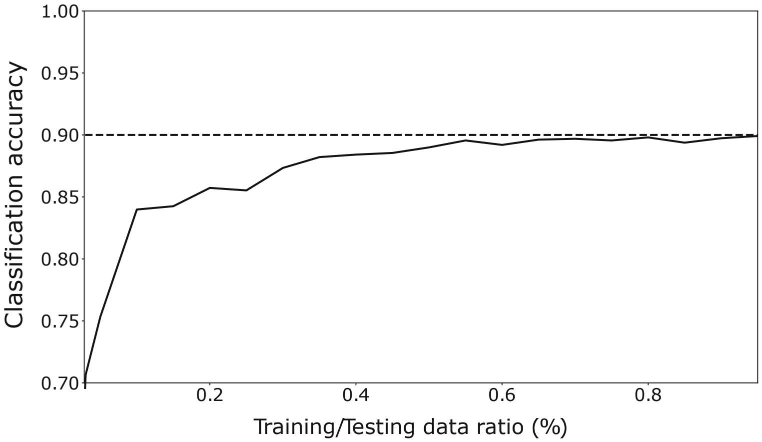


Fig. 7. A measurement of accuracy while varying the ratio of training/testing data.

Figures 8 and 9 show the classification result of an online learning approach on an excerpt of the *Fragmente*^{2} piece. This excerpt is 90 s long, and there are 17 gestural sonic objects in the performance section. Gestural sonic object qualities are annotated on the top tiers, and automatically spotted (predicted) onsets/offsets are indicated on the bottom track of each plot. In Fig. 8, the classifier was trained with a dataset split of 20% for training and 80% for testing, while in Fig. 9, the data split was 90% for training and 10% for testing. The amount of NC was randomly reduced to 10% of its original distribution. A clear improvement in classification when adding more training samples can be observed. This result paves the way for future work since it supports the assumption that adding new recordings with respective annotations would improve the performance of the gestural sonic object spotting and quality classification.

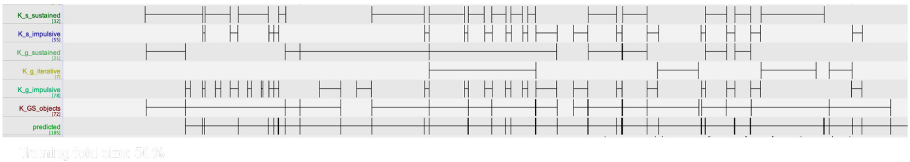


Fig. 8. An illustration of the classification results with a training fold size of 20%. The “predicted” tier (bottom) shows the onsets and offsets automatically predicted for gestural sonic object qualities (impulsive, sustained, iterative) from top tiers K_s (sound) and K_g (gesture).

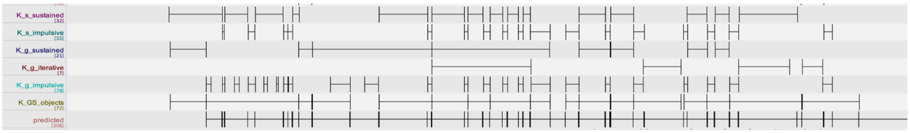


Fig. 9. An illustration of the classification results with a training fold size of 90%. The “predicted” tier (bottom) shows the onsets and offsets automatically predicted for gestural sonic object qualities (impulsive, sustained, iterative) from top tracks K_s (sound) and K_g (gesture).

4 Discussion

The work we presented so far looked at the notion of a gestural sonic object within three different contexts: its use as a conceptual tool for choreomusical composition; the exploration of the concept, and the boundaries of its definition for the development of a method for empirical analysis of embodied music performance; and the use of the data obtained through such analyses for training classifiers capable of predicting the quality of recorded gestural sonic objects through multimodal quantitative data. In this section, we propose some considerations that arose through this interdisciplinary research trajectory that, we believe, might inform further theoretical work.

4.1 Using Gestural Sonic Objects in Artistic Practice

The fact that Shinohara’s score is composed in an object-oriented manner facilitated the artistic process. Therefore, “thinking” in objects became central to the development of the piece: first, in the interpretation of the score; second, in the continued creation of the choreography; and third, in the analysis of gestural sonic objects and gestural objects as they emerged. As noted earlier, it has been possible to enhance the musical structure but also to intentionally explore possible new relations between different object types through an interpretation of the original score built on multimodal object analysis. This object-oriented method, which entailed a close examination of each object, increased the awareness of the choreomusical relation between dancer and musician in performance. It was also instrumental in the rehearsal process since a deepened understanding of the other performer's part emerged from the analysis. This, in turn, enabled a certain plasticity in the rendering of the individual parts in relation to the whole, as in a closely rehearsed chamber music performance. We find the relation between analysis and creation in the compositional process to be a factor that allowed a deepened interaction between the performers, the original score, and its rework. This relies on an embodied understanding of the original score enabled by a multimodal perspective facilitated by the concept of the gestural sonic object. Such an approach significantly helped to enhance the rhythmical relation between all the parts in performance.

4.2 Annotation Method and Theory: Practical and Conceptual Considerations

Implementing the gestural sonic object annotation method on the recordings of *Fragmente*^{2} has led to some reflections regarding the method itself and the concepts it is based on. Firstly, it provided an occasion for examining the definition of gestural sonic object empirically on multimodal music performance recordings. The definition of a gestural sonic object is relatively broad. Godøy (2018, p. 761) posits that "[a] sonic object may encompass a single tone or chord, a short phrase of several tones and/or chords in succession, a single sound event [. . .], or a more composite but still holistically perceived sound event". In other words, a sonic object can be many different things; what is crucial is that it is perceived as a coherent entity. The broad definition and the focus on perception entail that, in practice, determining what a sonic object is and what it is not involves a fair amount of subjectivity. The open coding sessions involving multiple observers we ran to annotate *Fragmente*^{2} made this aspect even more evident. Discussing where sonic objects begin and end in the recordings often led to going back to the literature in order to attempt to adhere to the definition as consistently as possible. We paid particular attention to the 0.5–5 s meso timescale when looking at the duration of the annotated objects, and appreciated that segments shorter than this time range effectively lose discernible timbral qualities that would allow us to identify the source of the sound or the overall musical style of the recording. Segments longer than 5 s are experienced as composite sound segments, thus losing the holisticity that characterises sound objects. We were initially doubtful about how to handle long, sustained drone sounds that were several seconds long. Can they be individual objects on their own despite their duration? We do not have a definitive answer, particularly given that we focused on a single composition. However, in the case of the long sustained notes played by the flute in *Fragmente*^{2}, we regularly found small pitch and timbral articulations that, in a way, worked as a "seam": the point where two long, sustained sound objects fuse. Such aspects were also sometimes found in the corresponding movements of the performer, possibly confirming segmentation and a point of coarticulation at the point where the object fuses. This leads to reflections regarding the "gestural" in gestural sonic objects. The way Godøy extends the notion of the sonic object to comprise gestural and kinematic qualities is underpinned by assumptions of the body being central in the experience of music and the existence of gestural affordances in musical sound (Godøy, 2006, 2010). As exemplified above with the segmentation of long sustained sounds, observing the performer's movements had a crucial role in forming our understanding of gestural sonic objects in *Fragmente*^{2}. This should come as no surprise, given that gestural sonic objects are multimodal by definition. Yet it was challenging that the sound expressed certain qualities that we could not find in the movement and vice-versa. That led us to label the typological categories of the dynamic envelopes separately, which gave us sufficient flexibility to maintain the labels coherent and consistent even when the envelopes of movement and sound appeared different. The two modalities having different envelopes do not contradict the definition of the gestural sonic object and resonate with other empirical studies (Godøy et al. 2016). Another aspect that emerged while developing the method and annotating *Fragmente²* is that, quite frequently, one can find more than one dynamic envelope within the same object without affecting the fact that the object is perceived as a whole. This points to the possibility of gestural sonic objects having an internal dynamic structure that may affect higher-level phenomena such as phase transition, chunking, and phrasing.

In practical terms, the annotation procedure was, as expected, very time-consuming. The annotations of the recordings required more than 10 h of work involving two to four people. We expect the amount of labour required to label a similar recording to decrease significantly as the labelling method is consolidated, given that many of the open coding sessions we carried out were actually focused on developing the method itself. Yet, it is not realistic to think that many researchers and practitioners would be able to invest a similar amount of time to obtain high-quality quantitative data. This calls for tools to support the work of human annotators in ways that help accomplish repetitive tasks whilst not removing human subjectivity from the picture. The way we approached the use of quantitative multimodal data and machine learning algorithms is an attempt to work towards the development of such tools.

4.3 Interpretation of the Feature Ranking Results

The search for multimodal features that can best describe input signals is challenging. Deep learning models have proven to be robust in finding good feature representations as well as producing accurate machine learning models. However, this robustness is tied to the assumption of having access to very large datasets. Unfortunately, this is not the case in the present study. In this work, we have used a series of methods to perform feature extraction as well as feature selection. Initially, we obtained 3216 features from 134 input signals. Such dimensionality is huge compared to the small amount of data samples. Nevertheless, it could also be easily extended to hundreds of thousands of features by applying transformations and additional feature extraction on the input signals. Yet, what is the smallest interpretable feature set that could support a machine learning task to target the classification of gestural sonic objects?

Our approach utilising Random Forest appears to be effective based on the evaluation results. Although we can not ensure the optimal subset feature selection, the proposed iterative process of randomly ranking features by their scores and selection frequency has presented coherent results. The top 10 features, selected among several thousands of experiments, were able to achieve approximately 90% accuracy in the classification of 17 distinct gestural sonic object quality classes. We also kept a very shallow Random Forest model by pruning the trees to a maximum of 8 levels. This helped to diminish overfitting while keeping high accuracy.

The feature rankings in Tables 2, 3, 4 and 5 suggest some interpretations. In the first experiment, audio pitch and audio RMS were top-ranked in the sound domain. This supports the expectation that basic sound features such as pitch and loudness have a predominant effect on how annotators label gestural sonic object qualities in the sound domain. There are many other audio features in the time and frequency domains that could also be used (Lerch, 2012) and that could be explored in future studies.

In the gesture domain, in the first experiment, features of the EMG of the right finger flexor ranked at the top. In the second experiment, other EMG features, as well as insole sensors and motion features, ranked at the top. This indicates that data related to body motion are better predictors than audio features for the classification in the gesture domain. Among the top-ranked features, RT_finger_flex, M1_insoles_sum, m1_RT_ext_oblique_rms, CI_movmean and audio_pitch were the most frequent. CI_movmean (Contraction Index) also appeared many times on the top-50 rank for both sound and gesture domains, being more frequent in the gesture domain. Notably, in the gesture domain, the classifier also ranked the audio_pitch feature within the top 10. This suggests a cross-modal correlation between audio pitch and gestures that contributes to shaping gestural qualities.

Using multiple time resolutions was an important factor in the feature selection process, as we can see in Tables 4 and 5. Most selected features were extracted through a sliding window with a hop size of 20 ms and a time duration that covers 1000 ms of the respective input signal. Larger windows were the majority in the gesture domain, while in the sound domain, distinct resolutions were selected for the audio pitch feature. Obviously, large windows can capture longer gesture and sound envelopes, while shorter windows better capture quick performance articulations and details.

5 Conclusions and Implications

We find that the empirical gesture analysis has implications in several contexts. Firstly, this study was a way for us to engage with the gestural sonic object concept in both artistic practice and music analysis, thereby showing its usefulness and, possibly, its limitations.

More broadly, we seek to explore how the combination of qualitative and quantitative analysis and phenomenological variation may enable more dynamic working methods for cross-disciplinary collaboration. We believe that developing multimodal methods for artistic research may be particularly useful in choremusical practices. We are especially interested in the potential of methods that also engage in how the intentionality of audio and video technologies can be addressed through phenomenological variation. This entails an engagement with different modalities of listening and a design that allows for embodied, multimodal and performative approaches to the experience of sound. More empirical analysis beyond the scope of this study would help refine the methods we have proposed and inform further theoretical developments in the study of gestural sonic objects.

Finally, we believe the findings on feature ranking can inform future work on feature selection and gestural sonic object analysis by supporting decisions on the type and placement of sensors for multimodal data acquisition. Using supervised learning to automate the annotation of gestural sonic objects could lead to a system to assist annotators and save them hours of labour when annotating gestural sonic objects manually. While we are aware of the implications that the use of machine-learning approaches may have—particularly with regard to the introduction of bias and other costs that data-driven practices may involve (Crawford, 2021)—we advocate for approaches that assist rather than replace human expert annotators, thereby keeping humans in the loop while enabling new agencies and approaches.

Acknowledgements. We would like to thank Ulrik Röijezon for making the multimodal recordings of the *Fragmente*^{2} performances possible.

References

Bishop, C.M.: Pattern Recognition and Machine Learning. Springer (2006)

Blackwell, J.R., Kornatz, K.W., Heath, E.M.: Effect of grip span on maximal grip force and fatigue of flexor digitorum superficialis. Appl. Ergon. **30**(5), 401–405 (1999). https://doi.org/10.1016/S0003-6870(98)00055-6

Bortolozzo, M., Schramm, R., Jung, C.R.: Improving the classification of rare chords with unlabeled data. In: ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3390–3394 (2021). https://doi.org/10.1109/ICASSP39728.2021.9413701

Camurri, A., Lagerlöf, I., Volpe, G.: Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques. Int. J. Hum. Comput. Stud. **59**(1–2), 213–225 (2003). https://doi.org/10.1016/S1071-5819(03)00050-8

Christensen, E.: Music Listening, Music Therapy, Phenomenology and Neuroscience [PhD]. Aalborg University (2012)

Clayton, M., Leante, L.: Embodiment in music performance. In Clayton, M., Dueck, B., Leante, L. (eds.) Experience and Meaning in Music Performance, pp. 188–207. Oxford University Press (2013). https://doi.org/10.1093/acprof:oso/9780199811328.003.0009

Crawford, K.: Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press, New Haven (2021)

Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. Wiley, Hoboken (2001)

ELAN (Version 6.4). Max Planck Institute for Psycholinguistics; The Language Archive (2022). https://archive.mpi.nl/tla/elan

Engel, J., Hantrakul, L., Gu, C., Roberts, A.: DDSP: differentiable digital signal processing (2020). https://doi.org/10.48550/ARXIV.2001.04643

Estévez-García, R., et al.: Open data motion capture: MOCAP-ULL database. Procedia Comput. Sci. **75**, 316–326 (2015). https://doi.org/10.1016/j.procs.2015.12.253

Fu, Y., et al.: Vocabulary-informed zero-shot and open-set learning. IEEE Trans. Pattern Anal. Mach. Intell. **42**(12), 3136–3152 (2020). https://doi.org/10.1109/TPAMI.2019.2922175

Godøy, R.I.: Gestural-sonorous objects: embodied extensions of Schaeffer's conceptual apparatus. Organised Sound **11**(2), 149–157 (2006). https://doi.org/10.1017/S1355771806001439

Godøy, R.I.: Gestural affordances of musical sound. In Godøy, R.I., Leman, M. (eds.) Musical Gestures: Sound, Movement, and Meaning. Routledge (2010)

Godøy, R.I.: Sonic object cognition. In: Bader, R. (ed.) Springer Handbook of Systematic Musicology. SH, pp. 761–777. Springer, Heidelberg (2018). https://doi.org/10.1007/978-3-662-55004-5_35

Godøy, R.I., Song, M., Nymoen, K., Haugen, M.R., Jensenius, A.R.: Exploring sound-motion similarity in musical experience. J. New Music Res. **45**, 1–13 (2016) https://doi.org/10.1080/09298215.2016.1184689

Ihde, D.: Postphenomenology and Technoscience: The Peking University Lectures. SUNY Press, Albany (2009)

Ihde, D.: Experimental Phenomenology: Multistabilities, 2nd edn. State University of New York Press, Albany (2012)

Leman, M.: Musical gestures and embodied cognition. In: Dutoit, T., Todoroff, T., D'Alessandro, N. (eds.) Actes des Journées d'Informatique Musicale (JIM 2012) (Issue Jim, pp. 5–7). UMONS/numediart (2012). http://www.jim2012.be

Lerch, A.: Audio Content Analysis: An Introduction. Wiley, Hoboken (2012)

Li, J., et al.: Feature selection: a data perspective. ACM Comput. Surv. **50**(6), 1–45 (2018). https://doi.org/10.1145/3136625s

McCallum, M.C.: Unsupervised learning of deep features for music segmentation. In: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 346–350 (2019). https://doi.org/10.1109/ICASSP.2019.8683407

Miranda, E.R. (ed.): Handbook of Artificial Intelligence for Music. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-72116-9

Nguyễn, T.T.: The Choreography of Gender in Traditional Vietnamese Music [PhD]. Lund University (2019)

Östersjö, S.: Go to hell: towards a gesture-based compositional practice. Contemp. Music. Rev. **35**(4–5), 475–499 (2016). https://doi.org/10.1080/07494467.2016.1257625

Östersjö, S.: Listening To The Other. Leuven University Press, Leuven (2020)

Paul, S., Roy, S., Roy-Chowdhury, A.K.: W-TALC: weakly-supervised temporal activity localization and classification. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11208, pp. 588–607. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01225-0_35

Redko, I., Morvant, E., Habrard, A., Sebban, M., Bennani, Y.: Advances in Domain Adaptation Theory. ISTE Press Ltd, Elsevier Ltd (2019)

Riehm, R.: Toccata Orpheus. Ricordi (1990)

Ruffieux, S., Lalanne, D., Mugellini, E., Abou Khaled, O.: A survey of datasets for human gesture recognition. In: Kurosu, M. (ed.) HCI 2014. LNCS, vol. 8511, pp. 337–348. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07230-2_33

Schaeffer, P.: Traité des objets musicaux: Essai interdisciplines. Edition du Seuil (1966). http://www.seuil.com/livre-9782020026086.htm

Stefánsdóttir, H.S., Östersjö, S.: Listening and mediation: of agency and performantive responsivity in ecological sound art practices. Phenomenol. Pract. **17**(1), 115–136 (2022). https://doi.org/10.29173/pandpr29464

Thompson, M.: The Application of Motion Capture to Embodied Music Cognition Research Marc Thompson [PhD Thesis, University of Jyväskylä] (2012). http://urn.fi/URN:ISBN:978-951-39-4690-6

Tits, M., Laraba, S., Caulier, E., Tilmanne, J., Dutoit, T.: UMONS-TAICHI: a multimodal motion capture dataset of expertise in Taijiquan gestures. Data Brief **19**, 1214–1221 (2018). https://doi.org/10.1016/j.dib.2018.05.088

Van Dyck, E., Moelants, D., Demey, M., Deweppe, A., Coussement, P., Leman, M.: The impact of the bass drum on human dance movement. Music. Percept. **30**(4), 349–359 (2013). https://doi.org/10.1525/mp.2013.30.4.34

Verbeek, P.-P.: Cyborg intentionality: rethinking the phenomenology of human–technology relations. Phenomenol. Cogn. Sci. **7**(3), 387–395 (2008). https://doi.org/10.1007/s11097-008-9099-x

Visi, F.G., Östersjö, S., Ek, R., Röijezon, U.: Method development for multimodal data corpus analysis of expressive instrumental music performance. Front. Psychol. **11**, 576751 (2020). https://doi.org/10.3389/fpsyg.2020.576751

Xie, Q., Luong, M.-T., Hovy, E., Le, Q.V.: Self-training with Noisy Student improves ImageNet classification (2019). https://doi.org/10.48550/ARXIV.1911.04252

Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? In: Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, K. Q. Weinberger (eds.), Advances in Neural Information Processing Systems (vol. 27). Curran Associates, Inc (2014). https://proceedings.neurips.cc/paper/2014/file/375c71349b295fbe2dcdca9206f20a06-Paper.pdf

Studying Performances with Digital Musical Instruments: A Case Study of *Ritual*, a Piece for Solo Karlax

Benjamin Lavastre[1](✉) and Marcelo M. Wanderley[2](✉)

[1] CIRMMT, IDMIL, DCS, McGill University, Montreal, Canada
`benjamin.lavastre@hotmail.fr`
[2] CIRMMT, IDMIL, McGill University, Montreal, Canada
`marcelo.wanderley@mcgill.ca`

Abstract. This chapter discusses ways to study sonic design from the perspective of musical performances with Digital Musical Instruments (DMIs). We first review the specificities of DMIs in terms of their unique affordances and limitations and comment on instrument availability, longevity, and stability issues, which impact the use of DMIs in musical practice. We then focus on the Karlax, a commercial device used in several musical performances for over a decade. We present an analysis of excerpts from three performances of D. Andrew Stewart's piece *Ritual* for solo Karlax, discussing the variability of performers' gestures and the musical choices made. We conclude by suggesting practice exercises to develop performance techniques with the Karlax and discussing musical composition and performance issues with DMIs.

Keywords: Gestures · Digital Music Instruments (DMI) · Music Performance · Computer Music

1 Introduction

Gestures and sounds are tightly coupled in musical performances (Cadoz & Wanderley 2000, Wanderley 2002, Leman & Godøy 2010, Dahl et al. 2010). The sounds produced by a well-known acoustic musical instrument, such as the piano, immediately suggest the general characteristics of the gestures used to play it (Godøy, 2009). Indeed, research suggests that "even listeners with little or no formal musical training can have images of sound-producing movements that reproduce both the effort and the kinematics of the imagined sound-production actions" (Godøy & Jensenius 2009 p. 46). This is possible as there are unequivocal links between gestures and sounds in acoustic musical instruments, i.e., performing the same gestures will likely produce the same sounds.

Godøy and colleagues' observations on performer movements shed light on the richness and complexity of musical performances (Jensenius et al. 2010). Similarly, their work on sound-tracing gestures and the analysis of performances with air instruments (Godøy et al. 2006) show that gestures are embedded in musical ideas, even in the absence of the instrument itself. By tightly combining gesture and sound analysis, they

A. R. Jensenius (Ed.): SSD 2022, CRSM 12, pp. 138–154, 2024.
https://doi.org/10.1007/978-3-031-57892-2_8

propose novel ways to understand better the relationship between musicians and their (acoustic) instruments.

In acoustic musical instruments, gesture–sound relationships are given by the physical behaviors of vibrating structures (e.g., strings, membranes, bars, reeds, columns of air). These structures vibrate in specific ways described by their mechanical properties. In other words, though involving complex vibration patterns, strings, membranes, reeds, etc., can only vibrate in a finite number of ways. Performer gestures and resulting sounds are inextricably coupled by physical laws.

Digital musical instruments (DMIs) are typically composed of an input device connected to a sound-generating device, both linked by mapping strategies defining the relationship between performer actions and resulting sounds (Miranda & Wanderley 2006). In this text, when talking about a gestural controller, we mean the input device itself, with its physical properties, affordances, embedded sensing techniques, and (sensor) data generated. A DMI implies a complete instrument with defined sound characteristics, which might or not be generated in a separate device or embedded in the controller. Therefore, when mentioning the Karlax as a device, we will refer to it as a gestural controller, whereas the same device (Karlax), for instance, in the context of *Ritual*, will be considered a digital musical instrument, since mapping strategies and sound generation have been defined by D. Andrew Stewart. In DMIs, the sound generation algorithm determines the “vibrations” that the instrument produces. What the algorithm will do and how it will relate to the performer’s actions is arbitrarily defined by the instrument designer, the composer, or the performer (or eventually by all of them in one person) (Wanderley 2017). In DMIs, there is no inherent or natural connection between the actions of a performer (performer gestures) and the sound resulting from them. Indeed, there is an infinite number of possibilities allowing for the relationship between gestures and resulting sounds. Performer gestures and resulting sounds need to be coupled by the instrument designer.

2 Digital Musical Instruments in Context

There is a large set of possibilities for a musical performance with a DMI, as DMIs do not need to be played similarly to acoustic instruments. In other words, they do not necessarily produce musical notes as a result of performer gestures. They can instead be used to manipulate pre-recorded note sequences as input devices in live-coding contexts or many other contexts (Malloch & Wanderley 2017).

As DMIs do not necessarily produce a unique sound with given intensity, frequency, and timbral characteristics when excited by a performer’s gesture, hearing a sound produced by a DMI does not univocally bring up the image of a particular gesture.

Though several hundred gestural controllers and DMIs have been proposed in the literature, with more than a hundred controllers already known before the New Interfaces for Music Expression (NIME) conference in 2002 (Piringer 2001, Wanderley & Battier 2000), access to DMIs might be severely limited. There are few examples of gestural controllers and DMIs made in large quantities or readily available, for instance, in a musical instrument shop (with the exception of keyboard controllers and matrix-based controllers such as Ableton’s Push). Obtaining DMIs, when possible, might imply substantial expenses compared to entry-level acoustic and electric musical instruments.

Furthermore, many new controllers and DMIs proposed commercially aim at beginners and are marketed as "enabling anyone to make music, regardless of experience" (McPherson et al. 2019, p.8), raising questions about how much expertise development they allow.

Finally, performances with DMIs are often geared toward novelty, where the performance of a new piece sometimes takes precedence over the choice of existing works in the repertoire. If multiple performers do not repeatedly perform pieces, analyzing invariants and the variability of gesture–sound relationships is impossible.

3 Implications for DMI Design and Performance

3.1 Improved Design

New instruments might not be ready (stable enough) for intensive performance use, with many of the interfaces and DMIs proposed in the literature remaining laboratory prototypes. The move from an initial prototype developed to test a concept to a full-fledged instrument, which is inherently responsive, stable, and robust, is far from evident (Miranda & Wanderley 2006). Though it has been claimed that "Musical interface construction proceeds as more art than science, and possibly this is the only way that it can be done" (Cook 2017, p. 4), in practice, a balance between design and engineering is essential, as DMIs "are meant for real-time performance, instrumentation techniques providing stable, robust, accurate, reproducible and fast response are essential" (Medeiros & Wanderley 2014, p. 14).

Another reason preventing widespread, long-term use of DMIs might be the lack of subtle control (Morreale & McPherson 2017) or fine details of instrument craft in many instruments (Armitage et al. 2017), as "most (new interfaces for musical expression) NIMEs are viewed as exploratory tools created by and for performers that they are constantly in development and almost in no occasions in a finite state" (Morreale et al. 2018, p. 168). The trade-off between craft and engineering is essential, with unhealthy results when one side is overly considered at the expense of the other.

This context calls for DMI designs that aim to produce instruments beyond laboratory prototypes and can become tools for long-term musical expression. Medium- and long-term research projects such as the McGill Digital Orchestra aimed to make such instruments: "In the Digital Orchestra, we hoped to develop a methodology for the process of creating DMIs that would increase the likelihood of their being adopted by performers other than the instrument's designer" (Ferguson & Wanderley, 2010, p. 19). But design alone is not enough: "sophisticated musical expression requires not only a good control interface but also virtuosic mastery of the instrument it controls." (Dobrian & Koppelman 2006, p. 277). A balance between improved design and musical performance is, therefore, essential.

3.2 Accessibility

The path can be rough for musicians interested in acquiring a DMI, learning to perform with it, and eventually developing "virtuosic mastery." First, one needs to decide on an instrument. Typically, this would be done after watching a concert or a video of a performance. Then one needs to get a hold of the DMI (the physical controller, the sound synthesis software, and the mappings used). Alas, getting a copy of a DMI might be a significant limiting issue as controllers are not necessarily available in (physical or virtual) music shops. Only once this step is done can musical practice start. But where should it start?

3.3 Musical Practice

Contrary to acoustic instrument performance in classical music settings, the path to learning a DMI is not well charted, though a few works have tackled this issue, see, for instance, (Butler, 2008; Ferguson & Wanderley, 2010; Hochenbaum & Kapur, 2013; Marquez-Borbon, 2020; Tomás, 2020).

Learning to play DMIs typically relies on musicians watching live performances or through videos, similar to the context of popular, folk, or rock music. Yet, contrary to those, there are few in-person opportunities to make music with DMIs in groups, perhaps except for the control of live loops in club settings. Building communities of practice is crucial to creating the conditions for widespread DMI performance (de Laubier & Goudard, 2006; Fukuda et al., 2021).

3.4 Longevity

A critical issue in the NIME community is the number of interfaces and DMIs that attain some longevity from the total number of instruments proposed each year (Marquez-Borbon & Martinez-Avila, 2018). In NIME, several instruments are proposed that do not establish themselves as performance devices (Morreale & McPherson, 2017) or that might have a "performer base of one" (Ferguson & Wanderley, 2010), i.e., being only played by their inventors. Researchers have pointed out many reasons for this situation, including "the lack of a proper instrumental technique, the inadequacy of the traditional musical notation, and the non-existence of a repertoire dedicated to the instrument" (Mamedes et al. 2014).

3.5 Musical Novelty

Establishing a repertoire of pieces performed multiple times is essential to allow comparisons of expert performers' musical outcomes. As discussed above, this is far from the case with DMIs, somehow implied in the title of the main event on these instruments (*NEW* Interfaces for Musical Expression). Does playing an interface that was proposed several years ago count as NIME? How "new" should an interface be? How long can a performer keep the same instrument? Does one necessarily need to abandon "old" DMIs? (Masu et al., 2023) How can one foster the performance of existing pieces in the repertoire? In which contexts could this happen?

In the rest of this chapter, we will focus on one successful commercial interface that fulfills several of the above requirements, the Karlax.

4 The Karlax

The Karlax (www.dafact.com) is a gestural controller created by Rémi Dury, a well-known composer and performer active in the new music scene in France since the 1980s. At the time of the Karlax development in the early 2010s, Dury already had substantial experience performing with electronic instruments as part of Puce Muse/Espace Musical, an association created together with Roland Caen, Serge de Laubier, and Philippe Leroux (Couprie 2018), most notably performing the Méta-Instrument (de Laubier & Goudard, 2006) in a duo with Serge de Laubier.

The Karlax concept is a device both hands hold, like a clarinet or soprano saxophone. It includes various sensors: 10 continuous keys and 8 pistons, an inertial measurement unit, and several switches. It also includes a rotary axis with bends at each end, allowing the performer to rotate the controller's axis, an action earlier explored in Cook's Hirn Controller (Cook 2017). In its original form, the Karlax is a gestural controller that generates control messages from the various sensors' outputs, not sounds. To become an instrument, i.e., to play sounds with the Karlax, such control messages must be mapped to sound synthesis parameters, and combining a Karlax and its mappings to a synthesizer becomes a DMI.

The Karlax received substantial funding from the industry. This funding allowed for the development of a series of prototypes by professional designers and engineers, an exceptional situation in the context of new interfaces for musical expression. Around seventy Karlax units have been produced, costing several thousand euros each, putting it at the expensive end of the electronic musical instrument's cost range.

Given the confluence of the above, the Karlax has a special place in music technology history. It was developed by an experienced musician who had a clear goal in mind, with substantial financial and technical support over several years, yielding a high-quality commercial product manufactured in multiple (several dozen) copies and performed by dozens of musicians over more than a decade (Lavastre & Wanderley, 2021). These numbers are very far from the situation with traditional acoustic musical instruments played by thousands or millions of people over hundreds of years. Yet, the Karlax is pretty unique in digital musical instruments. Musical performances with the Karlax include solo and mixed pieces, including acoustic instruments, in composition and improvisation settings. The confluence of these unique facts makes it an ideal candidate for evaluating DMI performances.

5 A Comparative Analysis of Interpretations of *Ritual* for Solo Karlax

Comparative music performance studies have developed considerably with the rise of audio and video recording. They have allowed the renewal of the musicological approach towards a multi-disciplinary field, including psychology, music history, analysis, and music theory (Donin, 2005, Lerch et al., 2021). However, comparative studies of interpretations with digital instruments are still marginal. Though musical performance with digital musical instruments can take different forms, from improvisations to imitations of performances with acoustic musical instruments, only a few devices have aroused genuine interest among performers and composers and allowed the development of original approaches to composition, notation, and performance as did the Karlax (Mays & Faber, 2014, Stewart, 2016).

Composer D. Andrew Stewart's piece *Ritual* for Karlax solo from 2015 features detailed notation and developed playing techniques based on a gestural repertoire. The composer has made a significant effort to ensure that the piece can be performed again (notation, explanation, software versions, video recordings). On the other hand, it is one of the only pieces for this instrument in which different filmed versions exist.

This section examines the musical and gestural expressive variations in three interpretations of the piece. We have identified three excerpts at the beginning of the piece, each requiring different instrumental techniques and containing different levels of control. By comparing the different interpretations of the same piece, we aim to highlight the expressive strategies chosen by performers, better understand the aspects of the piece that performers focus on, and how they decide to interpret specific musical gestures in the score.

5.1 *Ritual* for Solo Karlax

Ritual uses physical model synthesis (*Sculpture*, in *Logic Pro*), a type of sound synthesis that emulates the physical properties of acoustic instruments to create sound waves. The piece is based on a specific gestural vocabulary and original mapping strategies developed by the composer. MIDI data from the Karlax are processed and used in algorithms to identify particular gestures (e.g., *shake* or *thrust* as named by the composer). The mappings are created in Cycling '74's *Max,* thanks partly to the *Digital Orchestra Toolbox* library (Malloch et al. 2018). The mapping that associates the raw and conditioned data to the sound synthesis parameters is realized thanks to *libmapper/Webmapper* (Wang et al., 2019). Thus, to perform the piece, the interpreter must combine the appropriate versions of three programs: *Max*, *Logic,* and *Webmapper*.

The score is presented in detail in (Stewart, 2016, p. 3) and describes the required physical gestures, notational symbols, information related to traditional forms of music notation, audible output, and any necessary technical details.

6 Analysis

We analyzed video recordings of three performances (available here: https://youtube.com/playlist?list=PLyCL8KtgnNS-eEdFAhBhg9gylbGKj1YTP):

- V1, performed by the composer in 2015 at the University of Lethbridge
- V2, performed by the composer at the 2018 Crossing Boundaries Symposium / Interactive Art, Science, and Technology (IAST) at the University of Lethbridge
- V3, played by Vlad Baran in 2021 at McGill University

The piece lasts 10 to 15 min and contains six parts. We focused on the introduction, the first page of the score, annotated as "ceremonious awakening." In this relatively free part, we have identified three excerpts corresponding to typical sound morphologies considering musical phrasing, dynamic envelope, and spectral content.

1. Attack/resonance with resonance control
2. Melodic play with control of resonance and timbre
3. Crescendo followed by a terminal accent with control of amplitude and timbre

We used the piece's score notation as a reference. The composer comments on creating the score in great detail in the description, adopting a prescriptive approach (Kanno, 2007). Although, the score contains essential descriptive elements such as durations, rhythms, tempi, nuances, or density. The score is conceived as a succession of specific gestures represented by original symbols associated with sounds. Sometimes traditional symbols are used in different ways: the notes in the staves indicate fingerings, or the numbers at the beginning of the staff indicate octaviation. Furthermore, the sounds are described literarily in the description.

In the case of *Ritual*, the performer has the score and video recordings available on the internet to recreate the piece. The composer's website also provides information on sound synthesis, mapping, and gesture programming stages.

We chose three excerpts from *Ritual* with varying "levels of control." By level of control, we imply the number and complexity of the gesture–sound associations related to the mapping strategies chosen by the composer.

- In A, a low level of control, with a sound activation followed by the control of the sound resonance.
- In B, a more complex control, with the activation of the pistons and the control of timbre and amplitude by the coordinated action of several gestures.
- In C, a moderate control, with a sequence of gestures that modifies the timbre (distortion, modulation). In this last case, the response to the *shake* gesture (notated by *stir* in the score) seems less direct. This is an example of a convergent or many-to-one mapping, where the amplitude is controlled by both the tilt of the Karlax and the rotation of its axis.

For each excerpt, we investigate "expressive variations" made by the performers, the diversity of the interpretation of musical qualities such as dynamics, timbral variations, phrasing, note ~~accuracy,~~ accuracy and gestures (Cadoz & Wanderley, 2000) or the performance of the gesture-sound link (i.e., *transparency*) (Fels, 2002).

In the following, first, we describe the different sound morphologies defined by the composer and compare the three interpretations. Then, we discuss the results by looking at the expressive variations according to the levels of control.

6.1 Attack/Resonance

The first musical gesture of the piece is a kind of gong strike (a *thrust* gesture), with control of the resonance by shaking the instrument (a *shake* gesture). We will use the gestural terminology determined by the composer hereafter (Fig. 1).

The *thrust* gesture is described as follows in the score:

> "This technique requires a coordination of gestures: (1) holding down a single piston and (2) thrusting the Karlax in the direction of the right hand (...) A thrust onset generates a realistic bell tone in this composition."

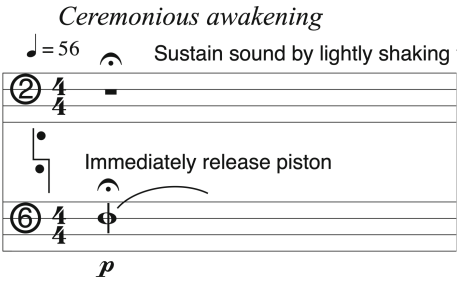
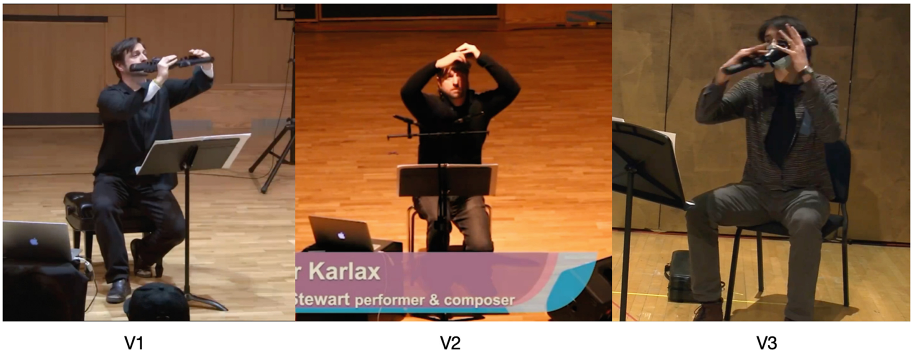


Fig. 1. The top four-line staff notates actions made with the left hand (e.g., pressing a key or a piston), while the bottom staff notates those with the right hand (Measure 1). Further details on the notation are presented in (Stewart 2015).

The *thrust* gesture triggers a complex gong-type sound with a dominant pitch. The shaking of the instrument controls the sustain of the resonance. The more the device is agitated, the more sustain there is.

This introductory gesture, quasi-theatrical, presents no specific difficulty and involves a basic level of control with the initial triggering of the sound and continuous management of the resonance (Fig. 2).



V1 V2 V3

Fig. 2. Introductory gesture for the three versions

By comparing the three versions, we note significant differences in duration, which correspond to the interpretation of the fermata. V3 differs from the two other versions: the pitch of the gong sound is a half-tone higher (D#2 instead of D2), and ancillary bell sounds accompany the resonance. Though these bell sounds appear later in the two other versions, they are absent at this moment in V1 and V2. We also note that the performer in V3 performs rotational movements, which may have triggered the ancillary bell sounds as they appear later, whereas, in the first two versions, this gesture does not occur.

6.2 Melodic Play

The second excerpt is a descending melody in the high register composed of three notes, interpreted by fingering combinations on the continuous keys "The Karlax keys are used similarly to the keys of a piano keyboard in this composition. They are treated as discrete on and off signals." (Stewart 2016). The notes in the staves do not indicate pitches but keys to be pressed (Fig. 3).

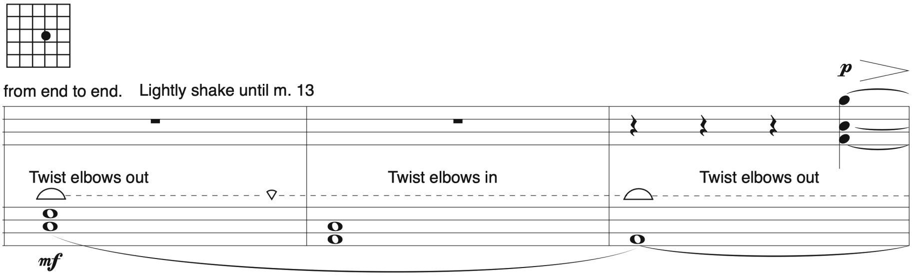


Fig. 3. Melodic play. Whole notes represent keys to be pressed. Timbre is controlled by rotating the axis and by tilting and rolling the Karlax. The grid with the dot indicates the Karlax inclination. Measures 2–4.

Bar 2 corresponds to simultaneously pressing keys 2 and 3 of the right hand until the end. Like in a standard MIDI keyboard, the keys are processed discretely, i.e., pressing the key until the end is necessary to get a signal.

Regarding the actual notes generated, in V1 and V2, we have the sequence Bb6, Ab6, and G6 (descending major second, then a minor second). In contrast, in V3, we have sequences A6, Ab6, and Gb6 (descending minor second, then a major second). This excerpt demands a higher level of control. A first gesture consists of playing pitches using specific fingerings, with control of timbre achieved by the rotation of the Karlax axis in combination with tilting and *rolling* (named by the author) (the roll angle also affects note sustain) and the shaking of the device (which produces a tremolo). Sound intensity is controlled by tilting and rolling the Karlax and rotating the instrument axis (Fig. 4).

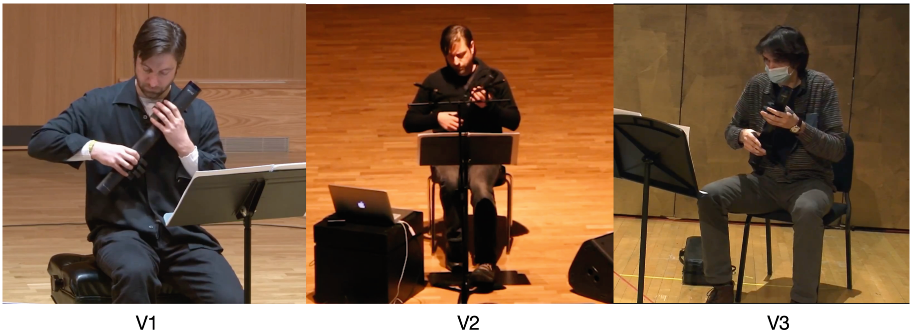

V1 V2 V3

Fig. 4. Gestural posture for the three versions during melodic play (Measures 2–4)

This melody is played differently in the three versions: V1 is clearly articulated, with regular tremolo. V2 contains volume accents; in V3, sound intensity is lower, without tremolo. The indications “Lightly shake” and “Twist elbows in (or out),” the latter corresponding to the rotation of the device’s axis, are freely interpreted. As before, there is a variation in pitch for V3 and the presence of extraneous bells.

6.3 Crescendo Followed by a Terminal Accent

The third excerpt consists of a progressive crescendo followed by an accent. This gesture is made up of several phases: first, the agitation of the instrument (named in the score *stir*) causes distortion, then the rotation of the axis initiates a crescendo amplified and modulated by the activation of the bend (maximum torque on the spring-like sensors at the end of each stroke of the Karlax axis), allowing to saturate the instrument’s timbre (Fig. 5).

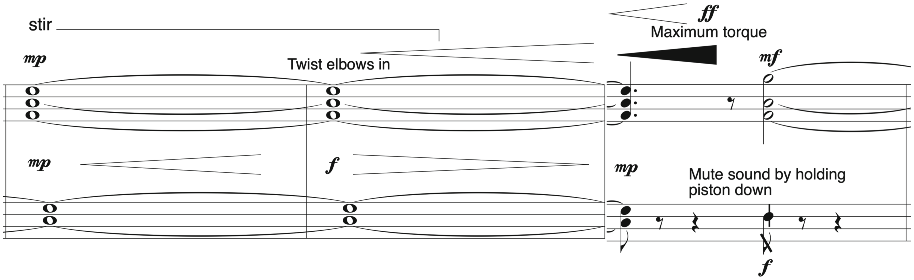


Fig. 5. Crescendo with a terminal accent (Measures 17–19). Whole notes correspond to the keys to be pressed. Amplitude and timbre are controlled by stirring the Karlax and turning the axis. The terminal accent is achieved by rotating the bend at the end of the axis rotation (Fig. 6).

In the score, the composer details these techniques as follows:

> “Stirring produces a dramatic sound color distortion of any sustained bell tones.”
>
> “(...) the resistive twist space is referred to as “maximum torque” and is notated in the score as an opaque triangle, resembling a traditional crescendo symbol that has been filled in. (...) The result is increased loudness and a timbre modulation of the sound.”

The three performance versions present a very different gestural expression with a large amplitude for V2, resulting in an important distortion effect. At the same time, V3 is much more contained, with little change in timbre.

7 Discussion

The analysis of the three excerpts shows a variety of musical and gestural expressions. Though these excerpts do not require virtuoso techniques, they contain subtleties in the control of timbre and the sequence of instrumental techniques that contribute to the richness of the sonic result. For the first part of extract A (the gong thrust), one can hear similar sound results even though the gestures are very different.

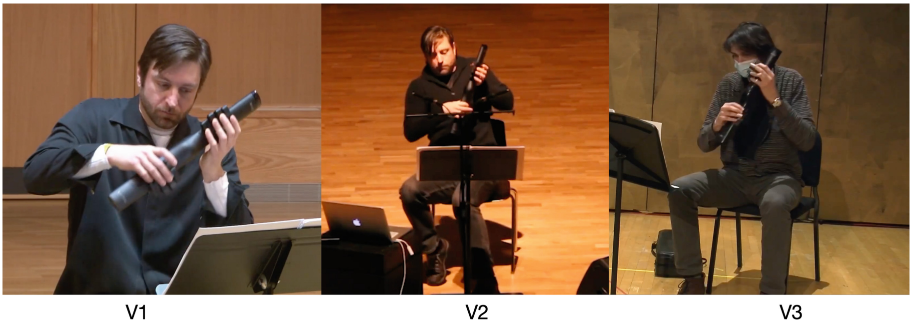


Fig. 6. Gestural posture for the terminal accent for all three versions (Measure 19).

Regarding body expression, the first two versions are more concentrated, and the gestures are slower, especially when moving from the attack to the control of the resonance. In contrast, V3 is more abrupt and shorter. There is a difference in the grip of the instrument for the attack in the two versions of the composer. In the 2015 version, the composer holds the Karlax by the center at the axis, facilitating the gong stroke. The performer in the V3 version rotates the instrument to control the resonance (likely to have triggered the ancillary bell sounds), whereas the composer moves the device longitudinally in the V1 version.

Furthermore, it would be interesting to investigate, in this specific case, to what extent the visual component of this gesture influences the perception of sound duration (Schutz & Lipscomb 2006). In other words, would one perceive the sound rendering in V3 as louder because the performer's gesture is more abrupt?

In the second part of extract A (the resonance), one notices that the link between gesture and sound is not evident because of the important latency between the gesture (agitation of the instrument) and the sound rendering (tremolo or sustain). In terms of control, the signal generated by the agitation of the device based on the algorithm for effort recognition is somewhat approximate because it seems difficult to obtain and maintain intermediate values of the signal accurately. These distortions between the visual and auditory parts complicate the identification of a specific character for each version. It is then challenging to determine expressive variations on a standard basis.

However, there are several solutions to this latency problem. The first and most obvious is to practice. By repeating the same gesture, the performer will get closer to the desired musicality, resulting in more confidence. Also, a calibration stage of the sensors may be necessary. Finally, a solution could be to adjust the effort recognition algorithm (scale, leak speed, and smoothing parameters) to be more responsive, so the gesture more closely matches the sound results. Moreover, it is interesting to note that the same phenomenon of latency can be observed in the acoustic instrumental world, especially with percussion instruments such as the tam-tam or the spring drum.

In excerpt B, one can perceive a more substantial similarity between the performances and more explicit interpretation choices, such as the calmer and more peaceful character of V1 and V3, compared to V2. However, the gestures are relatively different between the composer's version and V3. In V1 and V2, there is a longitudinal movement of agitation of the instrument, whereas, in V3, the instrumentalist makes almost no agitation.

In excerpt C, there are important gestural and musical amplitude variations across the three versions. The interpretation of V3 is more economical than the composer's intentions described in the score, especially in terms of nuances and distortion of the sound. If we focus on gestures, there are significant differences. The first two versions offer similar gestures even though the musical phrasing is different, with the instrument waving longitudinally and the device being raised and then leaning back during the accents. In the V3 version, the performer realizes rotations/gyrations, a sort of paddle stroke. Also, the link between gesture and music is unclear for the same reasons as in excerpt A.

Some gestures allow for limited control, such as the gong strike. Other gestures require the performer to listen to sound results to adjust the gesture, for example, during the sound saturation at the end of the third excerpt.

It turns out that the version that seems to possess the maximum expressive variation is the one that we identified as containing the highest level of control (V2). But one can easily imagine versions with a high level of control but low gesture/music legibility or transparency. So, what if we only focused our comparisons on the musical content? Would one have found the same differences?

The differences in the three performances highlight the rehearsal work needed to incorporate and control the different types of sound morphologies, but also possible technical issues with the particular interface used. Some of the differences in V3 are likely due to technical problems in the device used, the performer having reported issues with that Karlax, including piston malfunction. The many subtleties of the piece allow for a great deal of progression, developing expertise and providing both a sense of control and freedom that are the foundation of the pleasure of instrumental playing.

7.1 Suggestions for Practice Exercises Based on Gestures in *Ritual*

Let us return to the first gesture of the piece, the "gong strike" in excerpt A. The gesture *following* the activation of the sound ("sound-producing" (Jensenius et al. 2010)) could also be considered to have a communicative side. As already observed, this gesture contains an important theatrical aspect.

Can we imagine other types of control? For example, the indications of the Karlax's inclination (on the x and y axes, according to the dot on the tablature above the staves) could differentiate the note attack by favoring specific components (transients). Then it would be necessary to map the tilt data to the characteristics of the sound attack (e.g., distortion, sustain, resonators, etc.).

One could imagine that the stroke acceleration controls another parameter or that its combination with another gesture further differentiates the sound rendering. For example, the speed of a piston pressing could control the sound amplitude. In this case, if the performer activates a piston with more or less speed, this gesture could condition the amplitude of the resulting sound when the impulse is triggered. Some such features have been implemented in *Ritual*. For instance, the rotation speed is associated with the decay structure: a slow rotation produces a longer sound, while a fast/abrupt rotation produces a very short sound (Stewart, 2023).

Generally speaking, with more control complexity, the performer would approach this introductory gesture with a more precise "idea" of the resulting sound. Presumably, this could result in more significant expressive variation between each version. But is increased control complexity the composer's goal? In this musical context, the gong stroke seems to represent a kind of entry into the ritual like Claude Vivier's *Et je reverrai cette ville étrange*. It delimits the time of the ritual with a signal that does not require any particular change in the sound result. In any case, such suggestions could constitute the basis of a series of extended performance exercises with increased complexity, providing performers with material to develop dexterity and musicianship with the Karlax.

7.2 A Few Thoughts on the Use of Digital Instruments and Contemporary Music Creation

The use of digital musical instruments constitutes an opportunity for composers and performers to develop original approaches and to invest in a new field of musical creation (Ferguson and Wanderley 2010). The challenges related to DMI concerning sound synthesis, mapping, the link between gesture and music, and interaction strategies (Lavastre & Wanderley 2021) lead to rethinking the writing and interpretation practices.

Furthermore, instrumental identity has constantly evolved and adapted to musicians' and composers' needs and requirements. Moreover, listeners/viewers/audiences/experiencers also have an essential role in this evolution because composers and performers *play* with the audience's expectations. The cultures of playing, composing, and listening are interconnected and generative. In composing, it is with the development of *musique concrète instrumentale*, notably with the composer Helmut Lachenmann (2009), that the expansion of the notion of instrumental identity is one of the most spectacular. *Pression* for solo cello, *Guero* for solo piano, or *Salut für Caudwell* for two guitars are examples of this composer's extensive exploration of unusual playing techniques that question instrument identity. Therefore, it may be interesting to ask in what sense these playing techniques have or will have consequences for future instrument making. On the other hand, mixed pieces with an electronic part—or augmented instruments equipped with sensors—redefine and blur the instrumental identity.

In this context, digital musical instruments like the Karlax offer new perspectives. For this instrument, the cultures of composing, performing, and listening are still in their infancy, limited by the number of available instruments and a restricted repertoire. Consequently, this leads composers and performers to develop a trajectory—which seems to be the reverse of Lachenmann's—reinforcing the instrument's identity, notably by providing gesture–sound legibility during a phase of "acclimation" (Stewart, 2023). This phase is to convince the audience of the gesture-to-sound relationships within the context of a composition. Also, it seems that composers and performers need to develop a set of rules by developing original techniques and strategies, mapping, signal processing, sound synthesis, notation (Faber & Mays, 2014), (Stewart, 2015) and interaction with other instruments if applicable (Lavastre & Wanderley, 2021).

If we look at the compositional level, one of the challenges of writing with DMI will be to explore how the instrument "responds" to the composer's musical ideas and how it inspires them. Take the example of polyphonic writing with two voices with the Karlax. Two continuous keys control the amplitudes of these two voices, and the axis rotation and the sensors of the inertial unit control the timbre features. The independence of the timbre of each voice will not be perceptible because the same sensors are used for both voices. Therefore, the performer must find other ways to clearly render this musical idea.

The simplest way to achieve a polyphonic type of composition with the Karlax is to assign different sensors to each voice. But that supposes limited levels of control because the Karlax is made for holistic control of simple gestures. However, the many keys and pistons allow the performer to play a variety of key-activation techniques for polyphonic playing. The composer can also consider gate-type many-to-one strategies. In this case, some sensors are activated only when combined with other sensors. This allows differentiation of the control level for the same gesture but requires alternating the control for each voice. By adjusting each of the triggers and varying the control levels of each voice, the instrumentalist can create the illusion of polyphonic writing with independent control.

With this example, we show how limitations linked to the interface are circumvented to achieve musical intentions. The composition process is deeply dynamic, and the choices made by the composer result from balancing between the domains of performance, programming, musical composition, notation, etc., that transcend the mere idea of sound–gesture associations.

On the other hand, reinterpreting a piece with a digital musical instrument or gestural controller such as the Karlax can be much more demanding technologically than a piece for an established acoustic instrument. Among other things, the interface, the computer running the sound synthesis, the conditioning of Karlax's data, and the mapping must be brought together. Similarly, fitting typically in the context of contemporary/experimental music, performances with DMIs might also suffer from the newness aesthetics in performances where the reproducibility of existing repertoire might not necessarily be the first objective of composers, performers, or ensembles.

Paradoxically, an ensemble composed of acoustic instruments and a Karlax, the Fabrique Nomade, has performed and recorded many times the pieces in their repertoire, some of the pieces more than 30 times since 2013 (Faber, 2022). This is a unique situation, many pieces composed for contemporary music ensembles are rarely reperformed after their creation. In that case, the questions of longevity and reproducibility are intrinsic to their performance by Fabrique Nomade and are addressed very early on by the ensemble. Furthermore, collaborations with composers are often longer than average in similar situations; an average collaboration lasts two years with the ensemble Fabrique Nomade.

8 Conclusion and Future Work

Acquiring, learning to play, and keeping a medium- to long-term performance practice with DMIs can be challenging, though examples show they are possible. In this study, we performed a comparative performance analysis focusing on a piece written for a particular gestural controller. The analysis was based on Rolf Inge Godøy's notion of sound–gesture relationships, focusing on the importance and diversity of gestural activity in instrumental performance. In this context, the piece *Ritual* by D. Andrew Stewart for solo Karlax offers an exciting study object as gestures constitute the source of the writing. We compared different interpretations of some excerpts of the piece from sound morphologies and identifiable musical phrasings. Though these excerpts do not present specific technical difficulties, they contain subtleties in the timbral control that allow for variations in interpretation.

In future work, a closer examination of existing analytical tools for a comparative study of interpretations with video seems particularly important in our approach. Furthermore, it would be interesting to compare different interpretations of pieces where the Karlax interacts with acoustic and digital instruments. By allowing a differentiated level of control for each part of the piece, the composer also gives the performer and the listener something to conceive, organize, and perceive musical ideas. Rather than a "virtualization" seeking the precision of the control of an acoustic musical instrument, it seems that it is in the interplay of the relations between the composer, the performer, and the listener that the challenges and the richness of the interpretation with a gestural controller lie.

References

Armitage, J., Morreale, F., McPherson, A.: The finer the musician, the smaller the details: NIMEcraft under the microscope. In: Proceedings of the 2017 International Conference on New Interfaces for Musical Expression, pp. 393–398 (2017)

Butler, J.: Creating pedagogical etudes for interactive instruments. In: Proceedings of the 2008 International Conference on New Interfaces for Musical Expression, pp. 77–80 (2008)

Cadoz, C., Wanderley, M.M.: Gesture - Music. In: Wanderley, M.M., Battier, M. (eds.) Trends in Gestural Control of Music. IRCAM, Centre Pompidou (2000)

Cook, P.: 2001: principles for designing computer music controllers. In: Jensenius, A.R., Lyons, M.J. (eds.): A NIME Reader. Fifteen Years of New Interfaces for Musical Expression, vol. 3, pp. 1–13. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-47214-0_1

Couprie, P.: Le Méta-Instrument: Genèse et évolution d'un nouvel instrument. In: *Musique, Images, Instruments*, vol. 17, pp. 230–245. Ed. Klincksieck, Paris (2018)

Dahl, S., et al.: Gestures in performance. In: Godøy, R.I., Leman, M. (eds.) Musical Gestures, Sound, Movement, and Meaning, pp. 36–68. Routledge, New York (2010)

Dobrian, C., Koppelman, D.: The E in NIME: musical expression with new computer interfaces. In: Proceedings of the 2006 International Conference on New Interfaces for Musical Expression, pp. 277–282 (2006)

Donin, N.: Problèmes d'analyse de l'interprétation: Un Essai de Comparaison Assistée Par Ordinateur d'enregistrements Du Premier Prélude Du Clavier Bien Tempéré. Musurgia **12**(4), 19–43 (2005). http://www.jstor.org/stable/40591409. Accessed 28 Jan 2023

Faber, F.: Personal Communication with Benjamin Lavastre (2022)

Fels, S., Gadd, A., Mulder, A.: Mapping transparency through metaphor: towards more expressive musical instruments. Organised sound **7**(2), 109–126 (2002)

Ferguson, S., Wanderley, M.M.: The McGill digital orchestra: an interdisciplinary project on digital musical instruments. J. Interdiscip. Music Stud. **4**(2), 17–35 (2010)

Fukuda, T., Meneses, E., West, T., Wanderley, M.M.: The T-stick music creation project: an approach to building a creative community around a DMI. In: Proceedings of the 2021 International Conference on New Interfaces for Musical Expression (2021)

Godøy, R.I., Haga, E., Jensenius, A.R.: Playing "air instruments": mimicry of sound-producing gestures by novices and experts. In: Gibet, S., Courty, N., Kamp, J.-F. (eds.) GW 2005. LNCS (LNAI), vol. 3881, pp. 256–267. Springer, Heidelberg (2006). https://doi.org/10.1007/11678816_29

Godøy, R.I.: Geometry and effort in gestural renderings of musical sound. In: Sales Dias, M., Gibet, S., Wanderley, M.M., Bastos, R. (eds.) GW 2007. LNCS (LNAI), vol. 5085, pp. 205–215. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-540-92865-2_23

Godøy, R.I., Jensenius, A.R.: Body movement in music information retrieval. In: Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009), pp. 45–50 (2009)

Hochenbaum, J., Kapur, A.: Toward the future practice room: empowering musical pedagogy through hyperinstruments. In: Proceedings of the 2013 International Conference on New Interfaces for Musical Expression, pp. 307–312 (2013)

Jensenius, R.A., Wanderley, M.M., Godøy, R.I., Leman, M.: Musical Gestures, Concept and Methods in Research. In: Gogøy, R.I., Leman, M. (eds.) Musical Gestures, Sound, Movement, and Meaning, pp. 12–35. Routledge, New York (2010)

Kanno, M.: Prescriptive notation: limits and challenges. Contemp. Music. Rev. **26**(2), 231–254 (2007)

Lachenmann, H.: Écrits et entretiens. Editions Contrechamps. Genève (2009)

de Laubier, S., Goudard, V.: Meta-instrument 3: a look over 17 years of practice. In: Proceedings of the International Conference on New Interfaces for Musical Expression, pp. 288–291 (2006)

Lavastre, B., Wanderley, M. M.: Mixed writing with Karlax and acoustic instruments: interaction strategies from computer music. In: Proceedings of the 15th International Conference on Computer Music Multidisciplinary Research (CMMR). Tokyo, Japan (2021)

Leman, M., Godøy, R.I.: Why study musical gestures? In: Gogøy, R.I., Leman, M. (eds.) Musical Gestures, Sound, Movement, and Meaning, pp. 3–11. Routledge, New York (2010)

Malloch, J., Wanderley, M.M.: Embodied cognition and digital musical instruments: design and performance. In: Lesaffre, M., Maes, P.-J., Leman, M. (eds.) The Routledge Companion to Embodied Music Interaction, pp. 440–449. Routledge, Milton Park (2017)

Malloch, J.W., Schumacher, M., Sinclair, S., Wanderley, M.M.: The Digital orchestra toolbox for max. In: Proceedings of the 2018 International Conference on New Interfaces for Musical Expression (NIME), pp. 255–258 (2018)

Mamedes, C.R., Rodrigues, A., Wanderley, M.M., Manzolli, J., Garcia, D.H., Ferreira-Lopes, P.: Composing for DMIs - Entoa, music for Intonaspacio. In: Proceedings of the 2014 International Conference on New Interfaces for Musical Expression (NIME), pp. 509–512 (2014)

Marquez-Borbon, A., Martinez-Avila, J.P.: The problem of DMI adoption and longevity: envisioning a NIME performance pedagogy. In: Proceedings of the 2018 International Conference on New Interfaces for Musical Expression, pp. 190–195 (2018)

Marquez-Borbon, A.: Collaborative learning with interactive music systems. In: Proceedings of the 2020 International Conference on New Interfaces for Musical Expression, pp. 581–586 (2020)

Masu, R., Morreale, F., Jensenius, A.R.: The O in NIME: reflecting on the importance of reusing and repurposing old musical instruments. In: Proceedings of the International Conference on New Interfaces for Musical Expression. Mexico City (2023)

Mays, T., Faber, F.: A notation system for the Karlax controller. In: Proceedings of the 2014 International Conference on New Interfaces for Musical Expression (NIME), pp. 553–556 (2014)

McPherson, A., Morreale, F., Harrison, J.: Musical instruments for novices: comparing NIME, HCI and crowdfunding approaches. In: Holland, Simon, Mudd, Tom, Wilkie-McKenna, Katie, McPherson, Andrew, Wanderley, Marcelo M. (eds.) *New Directions in Music and Human-Computer Interactio*n. SSCC, pp. 179–212. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-92069-6_12

Medeiros, C.B., Wanderley, M.M.: A comprehensive review of sensors and instrumentation methods in devices for musical expressions. Sensors **14**, 13556–13591 (2014)

Miranda, E.R., Wanderley, M.M.: New Digital Musical Instruments: Control and Interaction beyond the Keyboard. A-R Editions, Middleton (2006)

Morreale, F., McPherson, A.P.: Design for longevity: ongoing use of instruments from NIME 2010–14. In: Proceedings of the 2017 International Conference on New Interfaces for Musical Expression, pp. 192–197 (2017)

Morreale, F. McPherson, A.P., Wanderley, M.M.: NIME Identity from the performer's perspective. In: Proceedings of the 2018 International Conference on New Interfaces for Musical Expression, pp. 168–173 (2018)

Piringer, J.: Elektronische Musik und Interaktivität: Prinzipien, Konzepte, Anwendungen. *Diplomarbeit*. Institut für Gestaltungs und Wirkungsforschung der Technischen Universität Wien (2001)

Schutz, M., Lipscomb, S.: Hearing gestures, seeing music: vision influences perceived tone duration. Perception **36**, 888–897 (2006)

Stewart, D.A.: Ritual for Karlax. Personal Score Publication (2015)

Stewart, A.: Personal Communication with Benjamin Lavastre (2023)

Tomás, E.: A playful approach to teaching NIME: pedagogical methods from a practice-based perspective. In: Proceedings of the 2020 International Conference on New Interfaces for Musical Expression, Birmingham City University, pp. 143–148 (2020)

Wanderley, M.M.: Expert commentary: Perry Cook's principles still going strong. In: Jensenius, A.R., Lyons, M.J. (eds.) A NIME Reader. Fifteen Years of New Interfaces for Musical Expression, pp. 11–13. Springer, Cham (2017)

Wanderley, M.M., Battier, M.: Trends in Gestural Control of Music. Ircam – Centre Georges Pompidou, Paris (2000)

Wang, J., Malloch, J., Sinclair, S., Wilansky, J., Krajeski, A., Wanderley, M.M.: Webmapper: a tool for visualizing and manipulating mappings in digital musical instruments. In: Proceedings of the 14th International Conference on Computer Music Multidisciplinary Research (CMMR). Paris, France (2019)

Artistic Exploration

Embodied Sonic Design: Sound and the Sensory Apprehension of Movement

Gemma L. Crowe(✉)

Emily Carr University of Art + Design, Vancouver, BC, Canada
`gcrowe44729@ecuad.ca`

Abstract. Sound acts as an extension of the body, created by movement and received as vibration. I am focused on the removal of a visual representation of the body as a template; to instead facilitate an embodied experience. As an embodied practitioner, I create immersive sound and media installations derived from recordings of my own moving body. The movement of sound depicts the presence of a body in motion through sensory illusion. Through embodied sonic design, my sound recordings decontextualize, abstract, and reframe the auditory experience. I physically manipulate the recording of sound to perceptually rematerialize the moving physical form during playback with two techniques: *sound shadows* and *embodied binaural spatialization*. These techniques encourage the listener to perceive sound and space with the same awareness that situates their body, such as sensation and proprioception. The perceived physical interaction within the reception of this sound is akin to a kinesthetic projection and is an engagement in spatial thinking, activating mirror neurons and kinesthetic empathy. Creating awareness through physical attunement can regulate systems out of balance by offering the embodiment of alternative states: shifting how one thinks and feels in a particular setting. My research seeks to recognize the listener's unique perspective through their individual body.

Keywords: Sound–Motion Relationship · Perception and Embodiment · Kinesthetic Empathy · Auditory Perception · Embodied Process

1 Introduction

Working with embodied sonic design, I consider how the method defines the product. My work aims to create a novel sensory experience where sound, image, and movement converge. I create immersive sound installations using multi-sensory stimulation and by generating an awareness of the body in space through proprioception. My sound works are created using body movement in unconventional ways. This chapter is dedicated to describing the way I work with sound using an embodied process—outside of instrumentation, vocalization, or electronic sound practices—and prioritizing whole-body integration. I find sound to be an effective artistic medium to profoundly engage the body. In addition, I integrate conventions of film and dance. My artistic research, along with an investigation of embodiment and perception, have paved the way for the experience of my work to be that of listening to see, seeing to feel, and feeling to hear. This chapter is a record of how I work to depict physical presence through sound.

A. R. Jensenius (Ed.): SSD 2022, CRSM 12, pp. 157–174, 2024.
https://doi.org/10.1007/978-3-031-57892-2_9

Rolf Inge Godøy's work has encouraged my investigation of the relationship between sound and motion as an area of serious study. Examining sound–motion similarity in musical experience, defining sound tracing, conceptualizing sound–motion objects (Godøy et al. 2016), and developing the Sonomotiongram technique (Jensenius and Godøy 2013), have inspired foundational concepts essential to my practice-based research. What follows is an artistic reflection on sound, movement, and embodied sonic design as an exploratory practice for generating and sensing the details of sound more deeply. Through embodied sonic design, I consider the effect movement has on the recording and playback of sound as well as the experience of the listening body and how we understand differently the implications and the circumstances of sound as presented.

This chapter will detail the trajectory of my artistic research, which incorporates movement improvisation, theories of perception and embodiment, as well as research on the relationship between sound and movement. Through my artistic research, two embodied sonic design techniques have emerged: *sound shadows*, a source-blocking technique that features the presence of a moving body through the absence of sound, and *embodied binaural spatialization*, a recording method that captures the perspective of a moving body in a choreographed sonic experience for the listening body. Kinesthetic empathy clarifies the potential for embodiment and makes way for a different sensory process, which I explain through the sound installation *the Presence in my Absence*, a felt experience in three-dimensional space. I explain how I work with sound perceived as a moving physical form and reflect on the creation of *the third movement,* a work that explores the kinesphere of individual listening bodies. To close, I position my continued development of these techniques for embodied sonic design as a hopeful practice, inviting different perspectives and lived experiences to become the subject of my work.

2 From Theory to Practice-Based Research

My intention is to engage the audience as a whole, that is, to integrate sensations from the body. I use movement to produce a sonic experience for the listening body. Along with analogue-adjacent processes, I use the movement of my body to physically manipulate and record sound. In sensing movement, I am inviting the audience to situate their individual bodily perspective within the work, allowing the lived experience to temper the work in the absence of visual representation. In order to approach the medium of sound differently, I employ a filmic perspective, and I find it helpful to visualize my sound experiments through descriptions using the conventions of filmmaking. Furthermore, pitch, dynamic, attack, and timbre work to determine movement in a way akin to highlighting, shading, and other visual effects used in animation (Clarke 2005, p. 73).

My first experiment compelled me to create the experience of movement by the body, *self-motion.* I considered what kind of sound–movement would most clearly illustrate self-motion, as opposed to movement *around* the listening body, and endeavored to create the sensation of spinning. I placed a recording device, standing in for the listening body, in the seat of a spinning office chair with a set of stereo speakers playing an ethereal-sounding piece of music nearby. The speed the chair was spinning, and the rate at which the sound panned from the left to right channels depicted a continuous motion. The same decay rate I would have sensed through inertia also helped to establish the effect.

Besides speed and realistic motion, I considered other factors influencing the spinning effect. First, the *Doppler effect* alters a sound's perceived frequencies based on its motion relative to the listener. I understand this as the "bending" of sound waves as they move through space, resulting in a change in pitch. The Doppler effect is responsible for the perception of speed and the location of sound. I charted these variances by moving a small speaker towards and away from my ear, then from a recording device, and found the Doppler effect illustrated the quality of my movement. Listening live through headphones connected to the recording device, I recognized that I could vary the pitch with a subtle increase or decrease in speed and at certain points in space. Refining the desired pitch modulation was an act of precision that I reproduced through movement and muscle memory rather than auditory feedback. Using different sounds and musical samples throughout these trials also demonstrated that certain frequencies compounded the effect.

Convolution resonance, the unique resonance of sound in a specific environment, works with resonating sound waves to create realistic sound environments based on features of that space. I was struck by how the architecture I moved through while recording could be sensed while listening back. I could describe the feeling of the space even if I could not completely determine it. The environments I moved my body in were further situating the listening body. The sound of the music playing in the room aided the spinning effect, situating the sound source as stationary in a specific environment rather than spinning around the head of the listening body.

I have found that sound panning left and right alone is enough to engage my body. This technique is used in Autonomous Sensory Meridian Response (ASMR), a sensory practice where audio recordings, sometimes paired with video, can create the sensation of live action. Some listeners experience physical sensations or "chills" when listening to ASMR (Lochte et al. 2018). In music studies, this effect is referred to as "musical frisson" and can affect skin conductance and the release of dopamine. ASMR practices led me to use in-ear binaural microphones to create a spatial relationship between sound and the listening body and to think through a three-dimensional sound experience.

Our systems of perception and the sites they occupy in the body are important points of inquiry in the study of sound and embodiment (Clarke 2005, p. 75). The presence of both the vestibular organs, responsible for balance and orienting the body in space, and sound reception in the ears is of great interest to my work, although I do not investigate the biological relationship within this research. I treat the listening body as a whole unit.

Two techniques for embodied sonic design emerged through my practice-based research: sound shadows and embodied binaural spatialization. Sound is an indicator of movement and can depict the conditions of its creation (Clarke 2005). The effect of music can be drawn from its performance: the force, speed, and intention of the musician, among many other factors (p.75). I use these sonic design techniques to create sound installations that encourage motion perception in the listening body.

3 The Role of Movement and Improvisation

I approach sound as it relates to the body: as a vibrational experience and as spatial composition. My work is musical but should not be considered as music in a traditional sense. Further, I use the term *movement* rather than *dance* to highlight its transference onto different media. With a background in contemporary dance, I have trained in movement improvisation practices for years. My movement practice now informs how I approach the listening subject through embodied sonic design. Dance is a discipline that explores codified movement, interpretation, and improvisation as expression and performative communication. I use movement to generate ideas, facilitate sensation, and bring attention to the present moment. In my artistic practice, I create movement with sound and light and incite a sensation of dancing in the listening body; the work represents the effects of dance but not the act of dancing itself.

To improvise is to act without preparation, to proceed intuitively. I engage in movement improvisation in conjunction with the generation and manipulation of sound. It is integral to my research and is both a practice and a process. Susan Kozel, a contemporary phenomenologist, describes Maurice Merleau-Ponty's *pre-reflection* and *hyper-reflection* as bodily experiences (Susan Kozel 2008, p. 23, as cited in Merleau-Ponty 1989, Part Two, Chapter 3). The pre-reflective state is an individual's response that precedes judgement, like a reflex. Kozel deems pre-reflection to be a "challenge to language and logic" within the "domain of corporeality" (Kozel 2008, p. 21) because of how we experience it, viscerally. She uses the term hyper-reflection to describe a looping process where the observer considers their own role while in-action (Kozel 2008, p. 22).

Improvised movement and sound emerge non-virtuosic in a pre-reflective state. I suspend judgement by thinking through my body, becoming aware of my actions after they occur, and continuously acting *before* I have a chance to reflect. It is important to me that the pre-reflective experience is embodied by the listening body. Movement improvisation brings me to a state of action where I think only through my body. By recording improvised movement—as audio or video—I am layering it onto the present moment, making the process for both of my embodied sonic design techniques hyper-reflective.

David Borgo describes music improvisation as an engagement in complexity; ideas are disordered and conflicting but are of little consequence to the musician's life beyond playing (2006, p. 22). Borgo suggests that engaging in this kind of play can increase a person's capacity for discordance and, perhaps, discomfort. If sensing my movement improvisation transmits a pre-reflective state, perhaps it can empower the listener with the ability to move through uncertainty.

4 Perception is Art

Barbara Tversky writes about how action and perception are intrinsically linked, crediting much of our understanding of the world to how we move (2019). The reason I aim to engage the senses through movement is because it makes us think differently, offering the listening body a new state of mind. Through embodied sonic design, my sound recordings decontextualize, abstract, and reframe the auditory experience. This is how I break down the content as sensory information, removing it from its circumstance so that it may be taken up differently through the senses.

There is potential for greater sensitivity, and therefore understanding, in how we absorb information from the world around us. I examine affect by initiating effects on the body through listening to recorded sound (as opposed to the performance of live sound) to sustain the act of perceiving and to repeat the experience of listening. The sound works I create are ambiguous, recognizable by the senses, but resisting definition. In abstraction, we create our own meaning—much like we note pathways and patterns in space by simply tracking movement. The movement of sound delineates spatial information, which we consider in relation to ourselves.

Audio recordings act as context in my work, framing the experience. Once established, these recordings are an experience to respond to rather than a product of the moment. The concept of "putting thought into the world," which Barbara Tversky (2019) suggests we do through language, gesture, and graphics, allows us to "transcend the here and now" (p. 99). An audio recording is a permanent and repeatable experience that we can build on. This is also the path of knowledge production.

Recordings offer a state of removal from an immediate event, which welcomes interpretation, reflection, and a moment to shift from an external to a more internal sense of awareness. The difference in how I work with sound is that the listening body must track perceived action: listening for the changing location of an invisible moving sound source requires a heightened sensitivity. Making sense of something through a collaboration of the senses focuses attention on our relationship to sound. This type of gestalt perception is something different, something unique to the experience. I call this *kinesthetic projection.*

Recognizing the process of perception is important here. I have likened my work to optical illusions, where—by design—something is perceived differently than it appears in reality. In realizing the discrepancy, we learn something new about how our eyes work.

5 Embodiment

Embodiment theories deem our corporeal experiences as rich sites of information to aid in the understanding of ourselves and others. The term *embodied cognition* acknowledges the sensory, bodily information of the cognitive process. Treating something outside of ourselves as if it were our own motivates my work. For me, *feeling* is essential. I am preoccupied with embodiment being the experience of sonic design, and I employ a phenomenological perspective when it comes to evaluating the perception of sound installations. Julie Herndon (2022) writes of *embodied composition* as a musical practice in which the creation of sound relates deeply to the body (p. 1). Herndon notes the relationship between movement and sound to increase a sense of connectedness through "internal and external awareness," derived from this practice (p. 6).

Hearing accommodates simultaneous activity. We can consume sound with different levels of engagement; our bodies can be in any orientation to the source, and we can be in almost any physical or mental state and continue to listen. This allows us to connect our immediate personal experience with what we are hearing. The nature of recorded sound encourages an imaginative engagement, and my embodied sonic design techniques invite the listening body's unique perspective. Democratizing experience in this way leads to a more sensitive account, furthermore, embodiment is part of the meaning-making process. With my work, I encourage meaning-making on an individual level by engaging the body and its multiple senses.

Embodiment through sound is not a novel concept. This is a large part of what makes us move to music, for example. Low-frequency bass sounds help to exemplify embodied listening here; they are usually felt more in the body than heard through the ears. Beyond music and entertainment, religious traditions and healing rituals that participate in chanting, singing, music, or sound use the affective qualities of vibration as a means of transcendence and worship. People also use sound to change their state of mind outside of faith practices. Listening to music or recordings of rainfall or birds singing can change our mood or promote relaxation. Binaural music is an emerging genre that uses separate audio tracks in each ear to achieve *entrainment,* synchronizing brainwave frequencies to achieve different mental states (Lochte et al. 2018). These diverse sound applications demonstrate different embodied relationships to sound.

6 Audible Kinesthetic Empathy

Watching movement incites a visceral, kinesthetic response (Wood 2015). When we watch others move, our bodies respond empathetically. We recognize movement by feeling as if we are executing the same movement. Barbara Tversky's (2019) investigation of mirror neurons, and entrainment, explains simply that "body-to-body communication is more direct than word-to-word" (p. 62). Here, *direct* refers to explicit and instantaneous by way of a different mode of cognition. Tversky builds on the theory of *motor resonance*, which is how movement is recognized through our own movement patterns. As motor resonance suggests, the body recognizes sound by the action that produced it (Godøy et al. 2016, p. 4). Godøy's research on *sound–motion similarity in musical experience* measured how different people moved to music and found culturally entrained responses. This inspired the *sound-tracing* paradigm, studying how the dancing body moves in unison with musical attributes.

Kinesthetic empathy is key to the embodied perception of movement. Tversky asserts that abstract thought—essential to activities such as problem-solving, creation, knowledge production, and the conceiving of possibilities—depends on an ability to think in spatial terms (Tversky 2019, p. 41). As Barbara Tversky uses the term, *spatial thinking* indicates mental reasoning through movement visualization. The example she uses is of mental rotation, where determining whether a distorted letter *F*, placed next to a correctly oriented one, is a rotation or mirror image of the correct position. Mental rotation theorizes that mental visualization moves the figure to reorient it properly and then recognizes the discrepancy (pp. 48–52).

I propose that tracking the movement of sound through space might also create a kinesthetic reaction. This is similar to how watching movement creates reciprocal activity in mirror neurons. As an example, a binaural audio track that mimics a swaying pattern from left to right might affect the listening body as if something is actually swaying in front of them. At this point, it might make sense to deem my practice as motion sonification. However, it is not the movement that is being translated into sound, movement is instead made apparent through its effect on sound. I use kinesthetic empathy here to create more dimensions in listening. The audibility of kinesthetic empathy enables other interpretations of sound, as different semiotic representations are embedded in performance.

The way certain music makes us want to move our bodies and the various translations of affective information produced through kinesthetic empathy, ASMR, and musical frisson demonstrate an ability to embody what we hear. The sensation of physical engagement from the perception of movement is what I aim to achieve through embodied sonic design.

7 Sound Shadows

I developed sound shadows as a technique for recording and presenting movement using sound, illustrating the effect that movement has on the environment. The sound shadows are created with a source-blocking technique using the audible interference produced by my moving body, partially obstructing the reception of sound by a recording device. Figure 1 represents the sound shadow recording configuration with a sound source positioned opposite a recording device with enough space between the speaker and microphone to capture the full moving body. The moving body between the sound source and the recording device blocks some of the sound, indicated by the arrows moving to the right of the image. This is akin to the way a body produces a shadow as it blocks light from a surface. The sound shadows are distinguished as the sound waves are eclipsed by the body. The conjecture is that during playback, the sound shadows could perceptually rematerialize the moving physical form blocking the sound during recording.

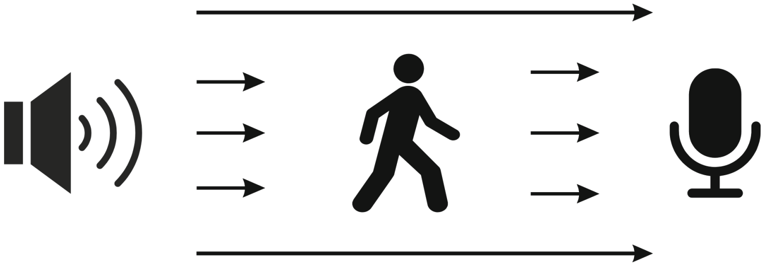


Fig. 1. Diagram representing the recording configuration used in the sound shadow embodied sonic design technique

My sound shadow technique is similar to the sound installation *remnant* by Miller Puckette and Hagan (2019). Puckette and Hagan experimented with the effect physical bodies had on soundwaves and aimed to represent the absent bodies by recreating the effect their presence had on the sound and the environment. The shadow produced by bodily interference was measured, as well as the reflection of sound coming off the body, which was described as "scattering." The two describe the process as "making the absent bodies themselves audible as acoustic reflections and shadows" (p. 1).

remnant used four musicians who played alto flute, trombone, piano, and percussion. My sound shadow technique uses white noise, sound with an equal amplitude distribution in all frequency bands. White noise fills the environment from veritable sonic events and punctuating sounds, which makes it effective as “noise cancellation.” I use white noise primarily as the wall upon which my sound shadows appear. This is a development of both Doppler effect explorations (altering sound by shifting its relative position) and convolution resonance (understanding conditions of a space based on the containment of the resonance).

Creating sound shadows is as much about finding the perfect backdrop to feature the movement occurring in the foreground as it is about moving in a way the body can sense. Puckette and Hagan recommend the use of a high-frequency sound, which makes distinct the difference between interference and clear transmission. I have found that blue noise—a type of noise with more high-frequency content than white noise—enables the differentiation between motion and stasis, and it tends to focus attention on the movement once the white noise is established as background noise rather than the main feature of listening.

Through exploration, I have found that sound shadows work best with short, consistent actions that can be transposed onto different body parts. Jumping, swaying, swinging, shaking, and approaching and departing—either from the sound source or the recording device, at varied speeds—were movements that translated well through sound. Here I should clarify that these movements should not be seen as gestures. The concept of the musical gesture brings in theories on body movements related to the shaping of sound with a communicative intention. While using the term gesture to describe musical effects “surpasses the Cartesian divide between physics and the mind” and encourages an embodied understanding, the movement I use is not coded in this way (Jensenius and Wanderley 2009, p. 19). My work attempts to engage the body without triggering an analytic determination of what is transpiring through language or gesture.

The best scene for designing sound shadows is one where the person can move freely within the audible “frame”—the area of clearest reception between the sound source and recording device. This leaves enough space for the background noise to come through. Ideally, the movement crosses at least half of the audible frame, so if the movements are small, such as intricate movements of the fingers, they should be performed close to the recording device. The scale of the effect depends on the distance between the moving body and the recording device rather than between the moving body and the sound source because, again, the recording device represents the perspective of the listening body. Small details should be close so they are in clearer “focus” the way more detail appears in the foreground of an image.

The process I use to create sound shadows recognizes the body as a tool for working with sound. The speed and intricacy that the moving body can express enables non-musicians, like myself, to produce soundscapes that would otherwise require musical training and technical knowledge. My embodied approach works through more intuitive expression and can be thought of as a new mode of sonic design.

8 360-Degree Sound Shadows in *the Presence in My Absence*

In the sound installation *the Presence in my Absence*, I used blue noise as the backdrop in a dark room with five Genelec speakers on stands surrounding the audience. In this installation, audiences were set in total darkness to sense the subtleties in the soundscape. Our ears sense pressure and can detect slight changes (Altman 1992, p. 21). The blue noise surrounding the audience created a consistent noise floor, which the listening body adjusted to. It resembled the rush of air at high pressure, which was most noticeable when momentarily absent in one spot.

I moved my body through the space during the recording process creating sound shadows by momentarily blocking the blue noise between the speaker and the microphone. The sound would periodically drop out of each speaker in the installation, re-enacting the movement and producing the sound shadows. Audience members recalled an initial confusion, questioning their hearing. The sequencing of the sound shadows across the speakers in the room was intended to indicate my moving physical form, similar to what would be felt if my tangible body was moving in the room, passing in front of the speakers and blocking the sound.

The creation of 360-degree sound shadows involved setting up five identical microphones in a circle, surrounding five outward-facing speakers in the centre that were playing blue noise toward the microphones. This is represented below in Fig. 2, with symbols representing the placement of the five speakers in the centre and five microphones on the perimeter of the action. I moved my body around the circle, improvising with pathways and approaching each microphone with the intention of reaching my future audience. Through this process, I both listened for and created the sonic change, responding to and experimenting with the dynamics of different movements. I performed my movements as if the recording equipment were my audience, creating an audible performance for them as they tracked the shadows my body left in the consistent sound. The setup was based on reverse engineering the experience: what acted as playback in the recording process became the audience perspective; the structures recording sound were then *producing* sound in the final piece.

Rick Altman (1992) examines the language we use to describe variances in sound and highlights the pragmatic but limiting ways we define the existence of sound in the world. The language we use fails to differentiate the quality of sound; we name sound based on its production rather than its perception (p. 19). This limits our evaluation of the sonic experience. It was my impression that audience members had difficulty articulating their experience in *the Presence in my Absence.* They seemed to lack the words to describe the change in sound, focusing mainly on the potential implications of what had transpired: they had lost hearing in one ear, something had gone wrong in one speaker, or something was blocking the sound. The installation infers the cause of the interference as the explanation for the experienced disturbance. It is the bodily experience of audible interference in sound shadows that affirms my exploration of embodiment as an important element of sonic design.

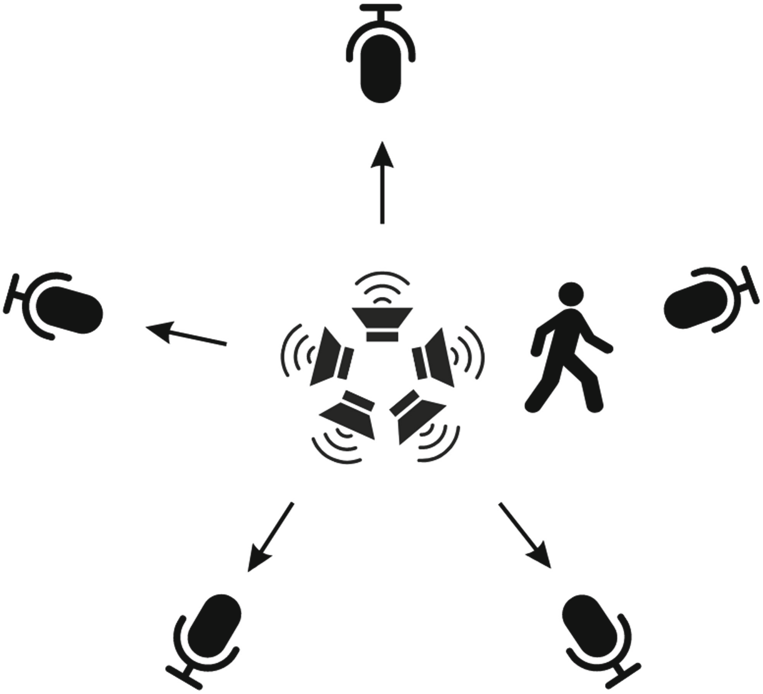


Fig. 2. Diagram representing the recording configuration for 360-Degree Sound Shadows

9 Sound Perceived as a Moving Material Form

As I physically inhabit the recording of sound defined by all the expression and constraints of a human body, I am producing a specific corporeal intention which I hope lands with the listening body and is a more nuanced experience than what could be developed through spatial sound editing.

Sound object theory, put simply, denotes a unit of sound as it is perceived as a whole (Schaeffer et al. 2017, p. 67). Godøy's (2016) *sound-motion objects* represent a sound that, as expressed, may be decipherable as a "meaningful" unit of sound related to the movement of its creation. Godøy's research proposes that "any sound event entails an image and context of a body-motion event" (p. 5). This is important as I position my recorded sound as having perceptible, material form.

The *sonomotiongram* technique developed by Jensenius and Godøy (2013) investigates the translation of information from movement to sound. Such a cross-modal "translation" is based on the universal nature of shape cognition. This research visualizes similarities between sonified movement, movement performed in response to music, and the characteristics of the music itself (p. 75). The instances where traditional music-making and embodied sonic design overlap allow the listening body to ground within an abstract experience. My embodied binaural spatialization technique, which I will explain in the next section, reproduces sound as a moving material form. The more I abstract the moving body in sound, I will offer recognizable sound actions to help reorient the listening body. In my installation, *the third movement,* I do this with re-creations of literal sonic events like footsteps, breath, or by tapping a pen while recording. Based on research for the sonomotiongram, I imagine the tap of the pen, as an action sonified, would somewhat resemble its sound data.

I am proposing that in depicting dynamic material form, sound can perceptually reanimate the moving body in space. The aesthetic specificity of movement and sound holds the potential for auditory creations as performance and a felt sense of bodies remanifested in space. Sofia Dahl's (2022) research on music performance and expression points out that musical cues assist with the ability to track simultaneous action through abstract thinking. This ability to more holistically recognize systems and patterns drives my desire to depict movement using a human body in motion—arguably the most recognizable system of them all. Through my practice-based research, I have made observations about the different modes of understanding that are accessible through an expressive abstraction. The sound object, according to Pierre Schaeffer (2017), is what we hear in the acousmatic performance of sound (the absence of instruments) when we cannot determine the source of the sound (p. 67).

I aim to create the perception of movement in the space around the body and sometimes the sensation of self-motion. *Vection* is the sensation of motion when no actual motion is occurring. This is what I was trying to achieve in the spinning experiment described earlier in this chapter. Sensing movement when you see the train next to you start to move, but your train is stationary, is an example of vection at play (Clarke 2005, p. 75). Vection is typically experienced visually but can be experienced in listening. Since my early explorations, I have found that to create the sensation of self-motion, the sound source must act like a backdrop or scene, less specific than a point in space and more like the space itself. I have only truly been able to produce the sense of self-motion with the aid of tactile, vibrating transducers in a spatial sound studio, and I attribute much of this to the collaboration of the tactile and auditory senses in the near absence of visual information.

10 Embodied Binaural Spatialization

Sound is choreographed as a moving material form to be sensed by the body in my second technique, embodied binaural spatialization, which functions by recording a single, static sound source through in-ear binaural microphones as I move my body around the sound. As a sonic design technique, embodied binaural spatialization produces sound inherently *different*—even just slightly—at different points in space. The soundscape, therefore, is composed through movement, using an embodied process with the intention of enabling an embodied experience of the sound for the listening body. This and the sound shadow technique both encourage the listener to perceive sound and space with proprioception (the perceptual sense that situates the body), which in turn activates the body through kinesthetic empathy. This is how I propose to offer the experience of movement without the listening body having to execute it. *Imagine* being able to feel something without physically experiencing it. In that case, embodiment is a kind of visceral empathy.

Researcher Ana Tajadura-Jiménez is involved in several studies that consider how hearing shapes body representation and posits audition as an important and often-overlooked factor in body awareness and representation (2012). *Predictive coding* deals with the mental schema of our bodies which is formed by information within our environment and is updated continuously based on new sensory information (Burns 2014). It is with this information that we perceive—and predict—how our bodies interact with the world. One such study found that the sound the impact participants' feet made on the ground when walking influenced how they felt in their bodies. Participants reported feeling differently about the size of their bodies while hearing the modified sound of their footsteps in action. It is in listening and embodying the sound of my moving body that I propose to offer a new experience of, and for, the listening body.

Embodied binaural spatialization is a way to generate a sense of movement for the listening body. It makes use of a singular perspective, which is first inhabited by the recording subject and then by the listening body, ideally through headphones. This is a process of reversal similar to that of *the Presence in my Absence.* For example, the listening body experiences a sound in the distance, approaching them from behind, which whips around the right side of their head, circling three times and arching above their head from their right ear to their left before landing in front of their face and fading out. The embodied binaural spatialization process is simply this experience inverted. The sound source—usually a pair of speakers—is a stand-in for the listening body, and I keep it at waist height so that I have room to move easily in any direction. The embodied binaural spatialization technique would use in-ear binaural microphones to create this experience, and I would start walking backward toward the sound source, spinning *myself* three times clockwise and then bowing my head forward as I lean my right ear in towards the speakers and trace the sound across the crown of my head, into my left ear and turning my face to the sound as I back away from the speakers. Even after creating these movement-scapes, I sense a character in the sound. My body then senses the recorded body as an entity outside of itself and responds to its movement. If I had struggled at a certain point, that tension comes through, and it feels real and still somehow new to me, even moments after living it.

11 *The Third Movement*: An Installation in the Kinesphere

The third movement considers the boundaries of the self, musical expectancy, and the tethers of reality through a track composed over seven separate listening stations. In the installation, pictured above in Fig. 3, the headphones were arranged loosely in the track's order and hung on plinths of varying dimensions and orientations throughout the installation. The audience was invited to move through the installation and put on the different headphones, each featuring a looped section of the full twelve-minute installation track with a specific condition for listening. At different heights and facings, the length of the headphone cords helped to indicate the intended position and location of the body, while the cushions encouraged the listening body to resign to the floor, as seen in Fig. 4. The audience populated the space, each occupying a specific location in relation to each other while having an individual experience. I edited drones, samples, and voice recordings to move seemingly autonomously in the 360-degree space using embodied binaural spatialization and developing concepts across listening stations. The sound animated the individual kinesphere of the listening body with an allocentric perspective—all while sharing space with other bodies and keeping track of activity around

Fig. 3. The third movement, installed at Emily Carr University of Art + Design, 2021

them in the virtual-audible and tangible space. The installation created a situation that was constantly permeated.

Fig. 4. Audience member lays down to experience the third movement sound installation at Emily Carr University of Art + Design, 2021

While documenting All Bodies Dance Project (ABDP)'s work *It's Enough (For a Rooftop),* taking place on the rooftop parkade of the Sun-Wah building in Vancouver, I experienced a new sonic illusion. One that re-created a site-specific, past sonic event and inspired a more narrative feature of *the third movement*. I was reviewing an ambisonic (360-degree) recording onsite using a pair of headphones while crouching next to the microphone's recording position as dancers moved to the shade for a break. I started playback, and the voice of the rehearsal director started up behind me, asking the dancers to take their positions. I looked behind me and found the entire group was still quietly occupying the shaded end of the rooftop parkade, including the rehearsal director. I was struck by the distinct sensation of someone behind me, her presence commanding action. Played back, the ambisonic recording convincingly re-created this sonic event because it was being reproduced at the exact site of its recording and occurring as if at the same angle behind me. In the spinning experiment from earlier in this chapter, I noted how the sound of the music in the space helped promote the sense of self-motion based on convolution resonance. The same effect here makes listening to a past sonic event—at the same site of its recording—a more convincing re-enactment of the event. That day, on the rooftop parkade of the Sun-Wah building, the sound was so realistic that I was physically responding to this information before I could fully make sense of it.

The recording I created for the first set of headphones featured one of these past sonic events, created and recorded in the exact same playback position for the listener. I walked up to the microphone slowly, making sure that each step I made was clear and audible. At this point, the hyper-reflective process was heightening my senses. I spoke into the mic intermittently, saying things like "hey" as if trying to get someone's attention. I said, "it's okay, it's supposed to do that," in response to electronic glitch sounds that I later layered onto the track to give the illusion of a bad wire connection. In a hyper-reflection process, the recorded action was layered to manipulate and redistribute the experience of liveness. The sensory confirmation from the environment within which we perceive our bodies generates this sonic illusion using predictive coding. Retaining the specific audible features of the building and installation space made plausible the sound of footsteps and the voice coming from behind the listening body.

The sensory anticipation and deception presented this work as playful trickery. I addressed the space between people—which was no more than a few feet, as pictured in Fig. 5—by occupying the perceived auditory-kinesthetic space with sound, which the listening body entered through headphones. The headphone cord represented a tethering to the site of the present experience, while musical development, dialogue, perceptible movement, and the site-specific past sonic events projected another reality. Portraying sound as a moving material form, *the third movement* brought attention to the conditions of certainty. I wanted the listening body to question what they were hearing and to feel at odds with what they were seeing and what they were sensing.

12 Technology and Access

It is important to note that my work is not *generated* by technology but *mediated* through sound equipment and software. The recording process is essential in communicating nuanced work that could only be produced by a body. Presenting my work using audio technology is the point where the research is embodied by others. One benefit to recognizing the impact that daily media has on the human psyche is that it indicates a unique

Fig. 5. Audience members populating the third movement sound installation, Emily Carr University of Art + Design, 2021

permeability through certain means. I aim to draw up bodily knowledge rather than implanting a message of my own. The technologies I work with are connective, they act as a means of extending my corporeal expression and enable communication from one body to another. Working this way becomes an ever-changing mode of creation through which to conceive novel experiences.

The language we use to understand the world is largely spatial; in fact, a good portion of how we communicate depends on orienting concepts and relating them spatially within our lives, like "diving into a subject" or saying something is "in the realm of possibilities" (Tversky 2019). This is how we qualify and understand the significance of events and circumstances.

Singer-songwriter Imogen Heap is working with a team of engineers, artists, and designers on the creation of gloves that use the artist's hand movements for the live performance of electronic music (Nosowitz 2014). The *Mi.Mu Glove* is a set of gloves comprised of specifically placed sensors that act like a synthesizer. Although the glove is meant to enhance stage presence, it allows the artist to approach the creation of each sound differently, thinking spatially and through the body (2014, para. 4). Based on the theory of multiple intelligences, workflows that enable different ways of understanding sound creation are generative as well as innovative. I am using embodied processes for creation where movement functions as sonic design, doing much of what could be done with software through the moving body. Working this way enables greater access to this type of creation for artists without the tools or education to produce such soundscapes.

Artist jamilah malika abu bakare's interdisciplinary practice prioritizes listening over looking as she seeks justice through art-making (2021). Abu bakare speaks of the gatekeeping of technology not only by gender, race, and class but of the developing art forms themselves, indicating that the expectation for high-level production narrows the scope of voices represented. She emphasizes the power of "art that makes you want to make art," which encourages engagement and inquiry as opposed to upholding an artistic hierarchy

(2021, Sound as freedom [Artist talk]). To abu bakare, art that brings about change is less about merit and more about continuing the conversation. It is the inclusivity of her practice, which involves anti-oppression as well as an emerging artist's initiative, that truly beckons other makers. Her provisional use of technology, purely for recording and playback, places more importance on the act of listening than demonstrating mastery of technique. Abu bakare goes as far as to say that the low-tech appeal of her sound and video works underlines the priorities of her practice, demonstrating a refusal of the commodification of media practices.

13 Where Do We Go from *Hear*?

Abstract art facilitates a heightened awareness that makes sensation more available. Audio, like video, can offer an experience that we may never be able to enact physically by envisioning it, or in the case of sound: by imagining it. Physical embodiment is not reserved for the moving, able body. This point most demonstrates my enthusiasm for creating work for the senses: perception of sensation and actual physical sensation can have the same impact on the body.

It is the availability of research on sound and movement that initiated a line of inquiry on the effect of sound on the body, leading to the development of my artistic research. Ultimately, I am exemplifying an internal, felt experience. This is in an effort not only to be witnessed but to share something as it is felt. Sound installations created through embodied binaural spatialization and sound shadows could one day be reproduced as dance, with intricate movement patterns mapped onto different surfaces, apparatuses, and spatial schemes. This would require more sophisticated technologies, but I am interested in the idea of how much movement the listening body can track simultaneously and at what point the motion amalgamates, amounting potentially to an entirely different form. Can movement be heard in the way a harmony of tones becomes a chord, as a chorus of action?

As I work with embodied sonic design, I challenge the approach that defines music so that the definition and application of sound might expand. The more experimentally I work, the more intently I listen and the more I appreciate what I can hear. I am working in this way to deepen the experience through embodiment. Appealing to the senses is my way of offering an experience that is beyond visual representation, one that is individual but shared. In creating awareness through physical attunement, I endeavour to regulate systems out of balance—the systems within us and those that govern our bodies. Embodied sonic design provides access to alternative states through embodiment: shifting how you feel in a particular setting. Inhabiting the body in a new way can empower us to live differently within conditions we cannot control. Acting as an empathetic intermediary, I create shareable spaces using embodied sonic design, so we can understand beyond our own circumstances. A focus on other ways of knowing, through embodiment, is simply my way of bringing balance to a world that I believe is suffering from its reliance on visual assurance. A sensorium is an escape from rules we did not make yet must adhere to. I offer my creations as a refuge where feelings and ideas can take shape.

References

Abu Bakare, J.M.: Sound as freedom [Artist talk]. Resonance series. Emily Carr University of Art + Design, Vancouver, BC, Canada (2021)

Altman, R. (ed.): Sound theory. Sound Practice. Routledge, Abingdon (1992)

Borgo, D.: Sync or swarm: musical improvisation and the complex dynamics of group creativity. In: Futatsugi, K., Jouannaud, J.-P., Meseguer, J. (eds.) Algebra, Meaning, and Computation. LNCS, vol. 4060, pp. 1–24. Springer, Heidelberg (2006). https://doi.org/10.1007/11780274_1

Burns, C.: Magic shoes: How to hear yourself instantly happy. New Scientist (2014). https://www.newscientist.com/article/dn26524-magic-shoes-how-to-hear-yourself-instantly-happy/

Clarke, E.F.: Ways of Listening. Oxford University Press, Oxford (2005). https://doi.org/10.1093/acprof:oso/9780195151947.001.0001

Dahl, S.: Connecting theories of music performance and expression for sonic design [Conference presentation]. International Seminar on Sonic Design, University of Oslo, Oslo, Norway (2022)

Godøy, R.I., Song, M., Nymoen, K., Haugen, M.R., Jensenius, A.R.: Exploring sound-motion similarity in musical experience. J. New Music Res. **45**(3), 210–222 (2016). https://doi.org/10.1080/09298215.2016.1184689

Herndon, J.: Embodied composition: composing the body with sound. Leonardo **55**(2), 170–174 (2022). https://doi.org/10.1162/leon_a_02137

Jensenius, A.R., Wanderley, M.M.: Musical gestures: concepts and methods in research. In: Godøy, R.I., Leman, M. (eds.) Musical Gestures: Sound, Movement, and Meaning, pp. 12–35. Routledge, Abingdon (2009). https://doi.org/10.4324/9780203863411

Jensenius, A.R., Godøy, R.I.: Sonifying the shape of human body motion using motiongrams. Empir. Musicol. Rev. **8**(2), 73 (2013). https://doi.org/10.18061/emr.v8i2.3926

Kozel, S.: Closer: Performance, Technologies, Phenomenology. The MIT Press, Cambridge (2008). https://doi.org/10.7551/mitpress/9780262113106.001.0001

Lochte, B.C., Guillory, S.A., Richard, C.A. H., Kelley, W.M.: An fMRI investigation of the neural correlates underlying the autonomous sensory meridian response (ASMR). BioImpacts **8**(4), 295–304 (2018). https://doi.org/10.15171/bi.2018.32

Merleau-Ponty, M.: Phenomenology of Perception. Routledge, Abingdon (1989)

Nosowitz, D.: Sculpt Sound With Gestures, Using Imogen Heap's $2,000 Gloves. Fast Company (2014). https://www.fastcompany.com/3028020/sculpt-sound-with-gestures-using-imogen-heaps-2000-gloves

Puckette, M., Hagan, K.: remnant: Exploring Presence and Absence Through Acoustic Disturbances in Space. UCSD Multisensory Perception Lab (2019). http://msp.ucsd.edu/Publications/icmc2019-reprint.pdf

Schaeffer, P., North, C., Dack, J.: Treatise on Musical Objects: An Essay across Disciplines, 1st edn. University of California Press (2017). https://doi.org/10.1525/j.ctt1qv5pqb

Tajadura-Jiménez, A., Väljamäe, A., Toshima, I., Kimura, T., Tsakiris, M., Kitagawa, N.: Action sounds recalibrate perceived tactile distance. Curr. Biol. **22**(13), R516–R517 (2012). https://doi.org/10.1016/j.cub.2012.04.028

Tversky, B.G.: Mind in Motion: How Action Shapes Thought. Basic Books (2019)

Wood, K.: Audience as community: corporeal knowledge and empathetic viewing. Int. J. Screendance **5** (2015). https://doi.org/10.18061/ijsd.v5i0.4518

Sonic Design and Spatial Features

Ulf A. S. Holbrook(✉)

Asplan Viak & RITMO Centre for Interdisciplinary Studies in Rhythm, Time, and Motion,
Department of Musicology, University of Oslo, Oslo, Norway
`post@u-l-v.org`

Abstract. This chapter seeks to contextualise and demonstrate how the core features of sonic design can encompass multidimensional features of space. The Schaefferian sound object is the basis for sonic design, a multidimensional unit which can contain multiple significations and features at the same time. This unit can be described by its main features and be broken down into sub-features and sub-sub-features. Rich and varied attributes from acoustics and psychoacoustics are used to see the sound object not merely in a musical but also in a spatial perspective. My proposed spatial "extension" to sonic design follows a proposal for how the *typomorphology* classification system can be seen in the light of spatial features.

Keywords: Sonic design · sound object · typology · morphology · spatial audio

1 Introduction

When we hear sound, it is all around us; the space this sound occupies is always present in one way or another, or is revealed to us in the listening situation. Space is revealed through the different attributes of the sound matter and is neither empty nor absolute. This chapter looks to map out some of the features and connections between sonic design and spatial features by drawing on Pierre Schaeffer's theories surrounding the sound object and his system of classification and exploration called the *typomorphology*. This was described and laid out in his 1966 book on music theory *Traité des objets musicaux: Essai interdisciplines* (English translation published in 2017). This music theory, with its interdisciplinary approach, was not realised as a how-to guide for composers but instead was the culmination of a research project which sought to bring together issues in musicology, acoustics, philosophy, and psychology (Valiquet 2017).

Through his music theory, Schaeffer did not consider space as necessarily relevant in itself, but rather that *time* is where the object exists. This is despite earlier performance research that utilised spatial technologies like the *potentimètre d'espace*, a device used during the performance of *Symphony pour un homme seul* in 1950 to move sound between three loudspeakers (Holmes 2012). Schaeffer and colleagues did not have access to the technological tools we have today, and it is fruitless to speculate how the *spatial parameter* would be incorporated into his work. However, in the *Outline of a concrete music theory*, with Abraham Moles (Schaeffer 2012), the two authors defined "25 initial words for a vocabulary" (p. 191–194), where words 23–25 are defined as

A. R. Jensenius (Ed.): SSD 2022, CRSM 12, pp. 175–191, 2024.
https://doi.org/10.1007/978-3-031-57892-2_10

spatial music, *static spatialisation*, and *cinematic spatialisation*. Spatial music is any music "that is concerned with the localisation of sound objects in space when works are being projected to an audience;" static spatialisation is defined as static sources in space, locatable to a point; and cinematic spatialisation refers to "projection that makes sound objects move in space at the same time as they move through time" (Schaeffer 2012, p. 194). Building on this, Iannis Xenakis formulated concepts of *stéréophonie statique* and *stéréophonie cinématique*, referring to sounds that are distributed over loudspeakers as points or where the sound sources are mobile and moving, what Maria Anna Harley referred to as *trajectories sonores* (Harley 1998).

The French composer Edgard Varèse iconically referred to his practice of composition as *organised sound* (Varèse and Wen-Chung 1966), where all possible sounds could be of musical interest, expanding the possible scope of composition. This encompassed all ranges of acoustic phenomena as equal in "value" to the perceived limited scope available within the voice and acoustic instruments. Varèse used topological and spatial metaphors to describe his work of "shifting planes, colliding masses, projection, transmutation, repulsion, speeds, angles and zones" (Born 2013, p. 2). Not only has this discourse influenced the subsequent development of spatialisation approaches, but I will argue throughout this text that these metaphors have also paved the way for new approaches to understanding sound, to a practice of expanding sound to space. Varèse argued that with "the liberation of sound," our conventional music notation systems would be inadequate to convey the new music, rather "the new notation will probably be seismographic" (Varèse and Wen-Chung 1966, p. 12).

Sonic design has been proposed as an "interface" for studying musical sound (Godøy 2010); however, from its basis in Schaeffer's theories on the sound object, the contribution of sonic design extends beyond mere musical sound and into spatial features and significations. Schaeffer theorised the sound object as a basic unit of perception where it is "capable of making a rich set of perceptually salient sonic and multimodal features present in our minds" (Godøy 2018, p. 761). This is a proposal of the sound object as a multidimensional unit, meaning that it can contain multiple significations and features simultaneously and is an *ontologically complex* unit. How we perceive these different features depends on our intentional focus and what features we focus on when listening. This multidimensional unit can be described by its main features and broken down into sub-features and sub-sub-features where all the different feature dimensions have various values (Godøy 2021). When perceiving sound objects, we can access the various features of the sound through its mediation to us as a sign. When shifting our intentional focus to the features contained in the sound object, we cannot *see* the source, and this is no longer of any relevance, as we focus on *what* we hear and *how* we hear it. The potential references to an external sound event will still be evident, depending on our intentional focus. Likewise, "smoke is only a sign of fire to the extent that fire is not actually perceived along with the smoke" (Eco 1979, p. 17).

This chapter will contextualise these concerns and discuss how the core feature of sonic design and Schaeffer's theories on the sound object can be extended to multidimensional entities of morphologies of space and as an informing element in spatial audio applications. Before discussing the typomorphological classification system, the central concept of *anamorphosis*, or warping, will be discussed in relation to the physical signal and the subjective perception of sound.

2 Linear and Non-linear Relationships

The relationships between the linear and non-linear are not always easy to define. Still, within the context of sound and audio programming, we can say that in a linear system, we can multiply a signal by a constant for amplification or attenuation of the signal; and in a non-linear system, we can multiply a signal by another signal, as in amplitude modulation (Smith 1999, p. 95). Linear relationships can be plotted in a straight line and divided into modular parts. Linear systems can be taken apart and put back together again, unchanged. Non-linear systems, however, "are not strictly proportional. One can think of them as having internal thresholds; when these thresholds are crossed, they switch into another mode of behavior" (Roads 1996, p. 887).

Non-linear relationships exist between the physical signal and the subjective perception of the same sound. These relationships are always present; we not only listen to "the sound itself" nor perceive it on its own; we also listen from a particular spatial position and perspective. Michel Chion emphasises that it is "not the psychology of the auditor that matters, it is the particular spot where the latter is positioned that does" (Chion 2016, p. 172). This points out that our experiences shift depending on our position, and we should be concerned with exploring the possible correlations between the physical signal and the subjective perception of the sound, specifically the relationship between the signal, the space, and the listener. This is especially pertinent given that the "sonic image emerges, therefore, as a concept that can integrate different listening approaches and provide an understanding of both the intrinsic and the extrinsic aspects of sonic experience" (Barreiro 2010, p. 36).

These piece-wise cumulative images (Godøy 2006), indicate that we piece together a sound and its behaviour in incremental steps. This *becoming* of sound perception can be referred to as what Norbert Schnell called an *action–action* relationship, meaning that any action is not isolated but is always part of an *inter*-action, the result of, belonging to, and becoming an interactive relationship (Schnell 2013). The relationships between the signal, the space and the listener exist between what Dick Raaijmakers described as "from the smallest sound to liquid form" (Raaijmakers 2000, p. 81) as well as the possible transmutations and morphogenesis of objects that can range from "dull matter, hard resonant matter, flowing liquid, bubbling liquid or steam clouds" (Roads 2015, p. 312).

Objects can have multiple significations, and as we shift our intentional focus to attend to different features of these multidimensional objects, we also shift our focus between the object as we hear it spatially and how it is situated in three dimensions:

> The essential aim of spatialization, which is often confused with some strange myth of "spatial music", is to improve the definition of objects through their distribution in space, since it so happens that the ear distinguishes two simultaneous sounds better if one comes from the right and the other from the left. We are not dealing here with a luxury added on to our hearing but something to facilitate it.

> Before even mentioning space and sound architecture, we should talk about the identification of objects and their coexistence. Where they are is of little consequence; it is what this enables that is important: an incomparably clearer, richer, more subtle perception of their contents. In the same way, binocular vision gives the third dimension and by putting things in perspective with each other allows us to judge their properties and relationships better. (Schaeffer 2017, p. 325)

Schaeffer referred to *anamorphosis*, or warping, as the possible non-linear relationship between the physical signal and the sound object, that could be characterised by irregularities that suggest a distortion of physical reality (Schaeffer 2017). This concerns the mapping of correlations between subjective images and the acoustic basis in sound. For example, temporal anamorphosis leads to “time warping” that describes how a “listener’s perception affords conclusions that do not concur with physical reality” (Landy 2007, p. 79).

Anamorphosis is a visual distortion that requires the viewer to be in a specific location to see the correct image(s); it is a technique to create pictures within pictures. One example is Hans Holbein’s painting *The Ambassadors* (1533), a much-cited double portrait of two unknown ambassadors with a still life. The painting features a smeared shape across the front of the painting. This shape reveals itself to be a human skull when viewed at a sharp angle from the right, an example of a *memento mori*.

Another example can be found in the work of Maurits Cornelis Escher, where the pictures within the pictures are accessible for the viewer from one position, depending on where we focus our gaze. His lithograph Waterfall, for example, depicts a waterfall and a waterwheel, where the water seemingly flows downhill after the waterfall, only to return to the waterfall, causing a feedback loop. This warping of an image indicates that it is the subjective perception, from a specific angle, that should be considered significant but not the only thing we should attend to.

The correlations between the physical signal and the heard sound are essential because what we think we hear is not always what we do, in fact, hear. The signal is a carrier of information, but it is not the information itself; it is a representation of information (Garnett 1991)—the physical experience of music is related to the physical vibration propagating through a medium before it reaches our ears. For example, in the fields of sonification and auditory display, the sonification process must be rooted in the data it presents, but what is perceived is still sound, from which we can extract information as we would with listening to any sound. The information contained in the sonification should be perceived by the listeners (Grond and Hermann 2014). The perceptual experience is the psychoacoustic feature attributed to how we make sense of what we hear, and the cognitive features surrounding the listening experience determine the structures we make of what we hear and what it means to us. This approach allows us to explore the deeper facets between physical signal propagation and our subjective perceptions.

The relationships between sounds and their perceptions imply the need for an empirical feature mapping between the percept and the signal (Godøy 2021); that is, our subjective perception of these sounds is considered the most important aspect (Godøy 2019). However, the acoustic features of sounds and their propagation through space are not irrelevant, so we should be concerned with studying the correlations between sound, space and listener. As part of his music theory, Schaeffer presented a framework for the classification and understanding of sound objects, the *typomorphology*, and this system lends itself to the study of both sound and space not merely for the evaluation of sound features.

3 A Brief Outline of the Typomorphology

In Schaeffer's music theory, the typomorphology provides a framework for understanding transitions in sound perception, and likewise, it can provide a framework for understanding spatial transitions. The complexity of Schaeffer's theories should not be underestimated, nor should the rigour in examining the sonic matter. Chion states that a summary of the types, classes, species, and genres of objects can be found in the Summary Diagram of the Theory of Musical Objects), which is a tool for investigation and not simply as a set of results. This, in turn, is further emphasised as: "the general procedure in this music theory is to move forward in a series of approximations rather than in a straight line" (Chion 2009, p. 100). Then, the general idea in this music theory is a series of approximations through a process of analysis and synthesis.

Analysis and synthesis refer to the systematic exploration of features. It is a method to understand the world by breaking it into smaller parts and looking at the possible interactions between the parts and their surroundings. This has been described by Jean-Claude Risset as analysis *by* synthesis (Risset and Wessel 1999). *Analysis* refers to decomposing something of varying degrees of complexity into smaller parts or elements. This also includes interactions and perspectives. *Synthesis* refers to the operations involved in putting these decomposed elements back together as themselves, as new configurations or through the combinations of interactions (Risset 1991; Wright et al. 2000). By drawing on the joint perspectives afforded by both anamorphosis and analysis/synthesis, we can investigate the typomorphology.

The typomorphology is a descriptive inventory that precedes musical activity; it is the initial "phase in the programme of musical research" (Chion 2009, p. 124). The typology is a "first sorting" according to the overall shape of the sound, and the morphology looks at the internal characteristics and features of the sound object. The tasks of the typomorphology are identification, classification and description, and it is divided into three parts (Chion 2009, p. 124):

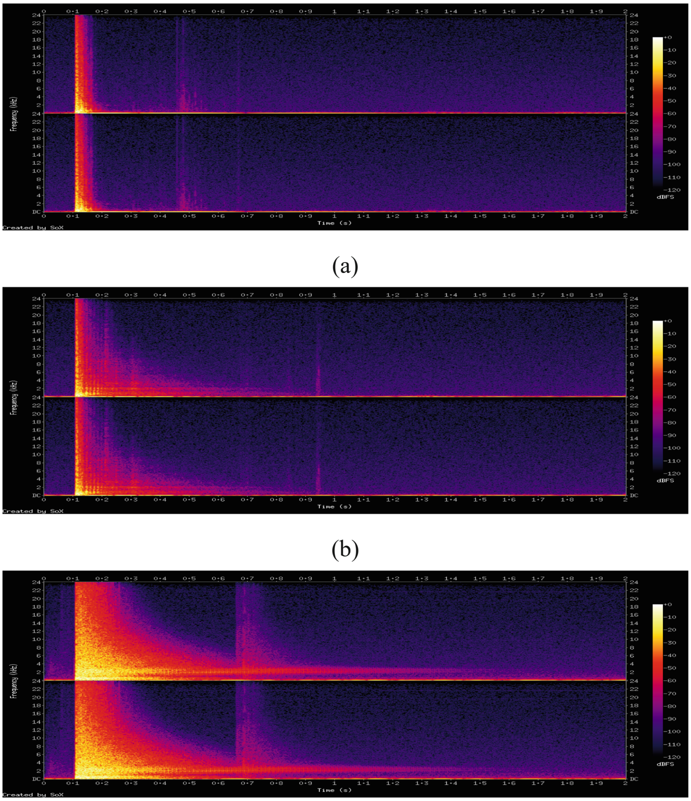
1. Identification of sound objects (typology)
2. Classification by type (typology)
3. Description of characteristics (morphology)

Identification and classification of sound objects is a procedure which consists of isolating and cutting out sound from all possible contexts and then arranging the sound objects by type. This sonic examination is based on subjective judgement. It is done in terms of reduced listening and involves a temporary suspension of our knowledge about the world and the sounds we are listening to in order to access their features. The typology starts by identifying sounds into three different categories based on their dynamic envelope:

1. Impulsive
2. Iterative
3. Sustained

These three categories describe sounds by their dynamic envelope, where an impulsive sound is short and fast produced by striking an object. An iterative sound describes something with a rapid motion, which can be perceived as a stream of impulses. A sustained sound is something prolonged, with a reasonably steady envelope. This perception could be true when hearing an impulsive sound of, for example, a snare drum being hit or a glass breaking in a dampened room like a studio. However, a changing spatial context for different sounds would classify them into different categories. For example, this same impulsive sound played in a reverberant space would cause the sound to be iterative rather than impulsive. As Tor Halmrast demonstrated, the attack of a tone is lengthened due to reverberation, which masks its entrance into another, and, depending on the sound source, this creates an attack of the attack (Halmrast 2018).

Even here, at the very start of the classification of sounds, the spatial *presence* makes evident the sound's relationship to something external to itself. For example, the measure RT_{60} (reverberation time) tells us how long it takes for the sound pressure level to drop by 60 dB (Howard and Angus 2009, p. 301). This is an easily understood parameter, but it says nothing about the materials of the room, the number of reflections, arrival times of these, or their strength. This causes rooms with the same RT_{60} to sound very different (Halmrast 2015). Figure 1 displays three impulses recorded in slightly different spaces. The first impulse was recorded in a heavily dampened space, where soft materials covered the walls, floor, and ceiling. The second impulse was recorded in the same room, with the door open, onto a concrete hallway. The third impulse was recorded in the concrete hallway outside the dampened room, which also has many intersecting corridors. The changes imposed on the sound by the surrounding space are clear. Indeed, when examining concert halls, David Griesinger found that the sonic background of a performance space can have unique timbral and spatial qualities and properties (Griesinger 1997), which introduces different timbral colourations to the sound. This would surely also be true for many other spaces.



(a)

(b)

(c)

Fig. 1. Three impulses recorded in slightly different spaces. a) An impulse recorded in a dampened space with a short reverberation time. b) An impulse recorded in the same space with the door open. c) An impulse recorded in a reflective hallway.

After the initial identification of sounds, they can be classified into *pairs* of typological criteria, where they are used to give approximate distinctions between objects (Chion 2009, pp. 134–137). This includes its *mass/facture*, which describes how a sound occupies the spectrum and how its shape changes over time. *Duration/variation* describes how a sound is subjectively experienced and how it is experienced over time. Finally, *balance/originality* deals with the object's internal structures.

The morphology, then, is divided into seven criteria of *mass*, *dynamic*, *harmonic timbre*, *melodic profile*, *mass profile*, *grain*, and *allure* (often referred to as *gait/oscillation*) (Chion 2009, pp. 158–187). The aim is not to identify abstract values such as pitch classes but to classify and understand sound in its possible diversities. This also extends into spatial features. Of these criteria, the most interesting in this regard are the two last, that of grain and allure.

1. *Grain* is a microstructure in the sound, which can be fine or coarse, and refers to the perceived surface of the object and its tactile texture. It can refer to a rapid gait

or variation or an accelerating iteration. A rapid succession of impulses stops being perceived as impulses but becomes a continuous sound with a characteristic grain.

2. *Gait* (a suitable translation of the French word *allure*, which means to walk, or a way to walk) refers to an undulating movement or fluctuation of sound objects, which can also be described as an oscillation. The oscillation of gait can both be in terms of duration and motion. The gait of a sound can be seen as a "signature" of its source.

These criteria are used as descriptions of sound. Why should we use such a system to describe space when we already have the tools defined by acoustics at our disposal? Indeed, the typomorphology does not afford us a description of space but gives us the means to describe *behaviour in space*. When we hear a sound emitted in an enclosed room, we hear a single fused sound that consists of the direct sound emitted by the source, along with a series of reflections from different surfaces in the room. The ratio of direct-to-reverberant sound is vital in distance perception of a sound source, as well as early reflections, high-frequency attenuation, and air absorption (Moore 2003).

Perception of sound quality, sound colour, or timbre is not solely dependent on the sound "itself," the spaces also contribute. Different rooms sound differently, and "the background can have its own spatial and timbral properties" (Griesinger 1997, p. 725). In a concert hall, the first reflections, primarily through interference by comb filtering, lead to a change in tone colouration and "image shift" (Barron 1971). Likewise, our spatial perception can be influenced by echo disturbances, shifts in the image of the apparent source, shifts in spatial impression, and different modifications of timbre (tone colouration) as functions of differences in intensity and arrival time (Kendall 1995).

The influences of, and changes to, sound given the environment it propagates in shows that the spatial situation is an essential factor in understanding the heard sound, the sound we perceive in the spatial listening situation. This is important for considering spatialisation approaches, which are not merely concerned with panning a source around a set of loudspeakers but should consider the entirety of the listening situation. When examining space from the position of a music theory concerned with manipulating sound materials, it poses a series of problems which are not obvious to resolve. The following section will discuss different approaches to experiencing space from the basis of psychoacoustic perception before examining possible approaches to working with space from a perspective of spatialisation.

4 Perspectives on Space

The classification system represented by the typomorphology provides the listener with a framework for exploring the many features of sound objects. It does not consider space as some abstract entity but analyses sounds for their features, shapes, and motions. As we saw earlier, Schaeffer did not consider space relevant in and of itself, but extending the typomorphology in this way will provide us with a richer set of tools for evaluation and classification. Indeed, the perceptions of spatial environments depend on the listener's accumulated knowledge of the physical and external world:

> When sensing a spatial environment, an individual builds a cognitive map of space using a combination of sensory information and experiences accumulated over a lifetime. The cognitive map of space in our consciousness is subjective,

> distorted and personalized - an active and synthetic creation - rather than a passive reaction to stimuli. (Blesser and Salter 2009, p. 46)

This construction of the cognitive maps we use to sense spatial environments reflects the concerns in emphasising anamorphosis to describe the relationships between sound and signal. It aligns with the message from Schaeffer's musical research that it is through our subjective and attentive perception of the world and the sounds contained within it that we make sense of what we are experiencing. For identifying a sound object, the "identification is done by reference to a higher level of context which includes the identified object, as an object in a structure" (Chion 2009, p. 61). It becomes clearer when examining the different criteria from the typomorphology, that sounds have a relationship to the external world, and it is the sound's morphological criteria that provide us with clues about how it existed spatially and how we can make it exist spatially.

At the opening of this chapter, Varèse was cited by using a series of metaphors to describe the behaviours of sounds in space. These feature categories lend themselves well to the transformation of sound materials but also as space descriptors. These same concerns have been formulated by Jean Petitot, in relation to morphodynamic models and how they unfold as bifurcating, non-linear dynamic systems:

> The phenomenological description of sound images, sound structures and sound organizations is very diverse; it includes forms, figurative salience, clear and fuzzy contours, attacks and fronts, not to mention deformation, stretching, mixing, stability and instability, rupture, discontinuity, harmonic clouds, crumbling and deviation of figures and so on. (Petitot 1999)

The bifurcations Petitot describes are related to both Varèse's topological and spatial metaphors of colliding masses, shifting planes, projection, and transmutation, and to Roads' dull matter, hard resonant matter, flowing liquid, bubbling liquid or steam clouds referred to earlier. A bifurcation is a point where something divides into two parts (or branches) and is used as a model of transition of features (Strogatz 2015). Metaphors are used to describe, experience, and understand something in terms of something else (Lakoff and Johnson 2008).

When we encounter sounds and sound experiences, we use metaphors to describe their features; for example, a sound can be described as "smooth," "shrill," "rough," "boxy" and the like. Metaphors can help composers and listeners to describe something vague as more tangible. The use of metaphors as a language to describe perceptions of sound can be a means of explaining the mental image of a sound (Porcello 2004) and even by adopting metaphors from other fields as a means of sensory evaluation, such as using terminology from the wine industry to describe features of concert hall acoustics (Lokki 2014). In the wine industry, the aroma wheel is a systematic way to discover the various flavours and fragrances found in wine and, looking beyond personal taste, the wine industry has established an overall characteristic of wine. However, with the available terminology and attributes to describe spatial features, this is not found in the same way with sound perception.

We can often refer to objective parameters, as defined in ISO3381-1:2009 as a guideline and standard for room acoustic measurements. However, this guideline does not discuss the subjective perception nor preference of listeners (Lokki 2013). The subjective perceptions of concert halls are difficult to measure, and this highlights the need to go beyond the impulse response measurement and standard criteria (Halmrast 2015) to focus on the perceptual consequences of frequency-dependent phenomena in musical instruments and human spatial hearing (Lokki 2016).

Yet, adopting metaphors for describing sounds and space can lead to inaccurate and conflicting descriptions among listeners. The different spatial attributes of sounds are, as we saw earlier, grounded in real-world experiences, sound perception, and localisation, abstraction of objects, relationships between objects, and the perception of space through the mass and size of objects. This can be described by the following four points:

1. Perception of sound as a whole, through object cognition and smearing in time and space
2. Immersion in the sound, the perception of not only the listening space but also the inherent spatiality of the sounds and their external references
3. The perception of multiple locations and distances and the proximities between sounds, is essential for the understanding of the relationships between sounds in a space
4. The perception of space through the size and mass of objects

"Real-world" indicates that which can be sensed from our surrounding world, either directly through our biological sensory apparatus or through microphones, sensors or other data collection methods. Through "sounds as a whole," we gather some form of impression of the supposed origin of the sound and its spatial context.

To describe the perceptual cues and the mechanisms of human sound localisation, we can use criteria defined through psychoacoustics to aid in the description and classification of sounds. The dimensional features in spatial sound are impressions in terms of *spatial extent* (width, depth and height), *distance* and *direction*, and immersive features such as *presence*, *envelopment*, and *engulfment*. In their normal usage, these attributes describe spatial and musical percepts and how the human mind makes sense of these experiences. These attributes can also provide us with insights into how the identification, classification, and description of sounds can be made through the typormophological framework.

Akin to how a sound object was described at the opening of this chapter, as a multidimensional unit containing multiple significations and features at the same time, experiencing sound and its acoustic correlate is also characterised by an array of multidimensional features, as exemplified above in the wine and food industry. Within acoustics, there is a wealth of terminology for describing space and spatial experiences but no agreement on many features. As an example, *spatial impression* is used to describe whether a space is perceived to be large or small, and *spaciousness* describes whether we are in a large and enveloping space (Griesinger 1999). The terms spaciousness, spatial impression, and envelopment are interpreted variably in the literature, and spatial impression has often been used as a "cover all" term (Rumsey 2002). Several researchers equate spaciousness with *apparent source width* (Griesinger 1997), but spaciousness has no bearing on the perceived size of the source, "a concert hall can be spacious, the reverberation of an oboe can be spacious, but the sonic image of an oboe cannot be spacious" (Griesinger 1997, p. 721). The perceived spatial impression is dependent on lateral reflections between 125 Hz and 1000 Hz. It is a function of the performing level and will be higher with larger ensembles (Barron and Marshall 1981). The combination of early and late arriving energy determines the magnitudes of spatial impression, apparent source width, and listener envelopment (Bradley et al. 2000). If reflected energy arrives within 50 ms of the end of the sound event, it is perceived as a small room (Griesinger 1996). However, to explain spatial impression, both the frequency and level-dependent aspects of the music that arrives at the listener's ears have to be linked.

These differences in terminology and lack of agreement on perceptual attributes can be a source of inaccurate and conflicting descriptions. Still, they are salient features that can be used to further develop the typomorphological framework and its potential spatial features. They are also informing elements in spatial audio applications. The next section discusses two approaches to working with space from a practical perspective and offers methods for thinking about both sonic and spatial design.

5 Approaches to Space

Marije Baalman differentiates between *techniques* and *technologies* (Baalman 2010). Techniques are descriptive of a compositional process, while technologies are descriptive of panning, speaker arrays, encoding/decoding functions, and so forth. When thinking about spatial audio or spatialisation approaches, it is usually limited to panning sounds between multiple loudspeakers. Rather than maintaining this relationship of formal properties to describe sound in space, we can draw on the typomorphology to explore further how sonic design can be used for spatial features.

Drawing on the methods discussed so far, two different but related approaches to *designing* space will be presented. Sound design involves the construction of sound worlds that exist in complement to a visual component, in this instance designing space will refer to the construction of holistic spatial scenes. Recordings made in a forest can evoke a sense of place and particularly a sense of depth (Westerkamp 2002). To gain acoustic knowledge of this space we can record impulse responses to replicate the acoustic presence of a forest and present it in a concert hall. It is common to be mindful of the foreground and background as important components in creating depth in a spatial scene (Lennox et al. 2001). Usually, also time-based processing effects like reverb are used to create larger spaces for the sounds to exist in.

We will yet again return to Varèse's topological and spatial metaphors, which he described as "shifting planes, colliding masses, projection, transmutation, repulsion, speeds, angles and zones" (Varèse and Wen-Chung 1966). In what he termed "zones of intensity," Varèse essentially described spatial experiences and organisation of sound materials, "these zones will be differentiated by various timbres or colors and different loudnesses" and "these zones would appear of different colors and of different magnitude in different perspectives of our perception" (Varèse and Wen-Chung 1966, pp. 11–12). Not only does this description put into perspective the concerns surrounding the realisation of *Pòeme électronique* at the Philips Pavilion in 1958, where the piece was spatialised over 400 loudspeakers, but also prefigures much of the technological advances made in spatialisation technologies for music, sound art, and film.

As with the metaphoric descriptions discussed so far, these features open a multitude of opportunities for how we can explore and analyse space through spatial audio applications. Many modern technologies for spatial audio emphasise a point-source approach, where an individual sound is placed in space as a single point. In the real world, sound does not exist as a point. The sound produced by a source will propagate outward in the surrounding space, and we will experience different frequency reflections and time decays from a series of surfaces in the space surrounding us. All sound sources have complex radiation and directivity patterns, and these complex patterns, combined with a potentially complex set of reflections from the surroundings, illustrate that a single point in space will not suffice as an element of spatial, sonic design.

To overcome this problem, one can design space based on photographs or personal experiences. However, by drawing directly on the physical makeup of the space to be constructed, various 3D scanning techniques can be used. The resulting point-cloud reconstructions can be used to build 3D maps of the space as a basis for developing sonic and spatial design, and through this construction, determine how sound behaves. Point-cloud reconstructions can be made through structure-from-motion, also called photogrammetry or LIDAR scanning. The resulting data collection can be used to build a 3D model which represents the physical constraints of the space. In the instance illustrated in Fig. 2, a point-cloud photogrammetry reconstruction is made of an area in a forest, showing trees extending as vertical lines around a central clearing. This illustrates how point clouds can be used dynamically to navigate a space and to create density maps of sound motions. This approach to spatial audio draws on principles from sonification and its uses of making data available through sound (Hermann et al. 2011). In sonification, the data defines and drives how sound behaves spatially and is concerned with the quality of the sounds and of the contexts to which they belong.

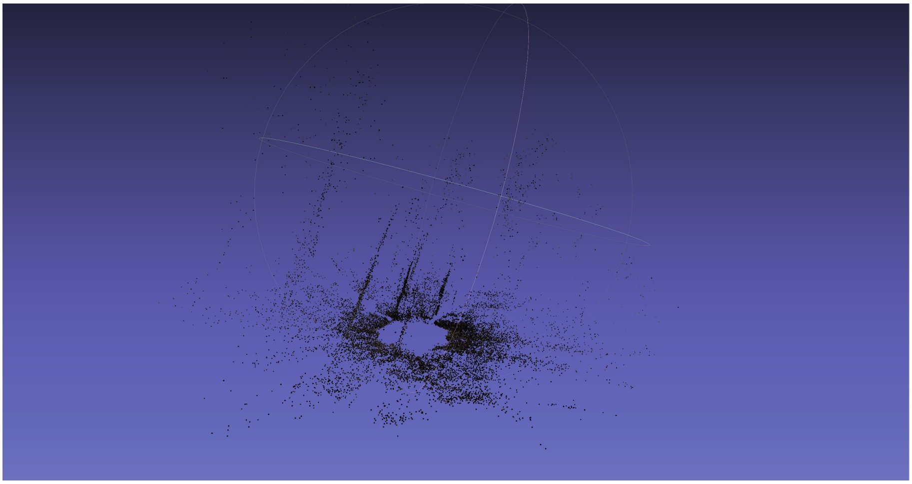


Fig. 2. A sparse, point-cloud reconstruction of a forest clearing, with trees seen as extending vertically around a central point. The example is made using photogrammetry, a technique where patterns are recorded and interpreted through photographic imagery.

By constructing a 3D model of a space, sound behaviour, reflection, and diffraction can be modelled, giving the sonic designer more direct control over constructing a perceptual significance, that is "to describe the 'behaviour' of sounding objects in and through their local environment—this is not just the case of Doppler effect, as it includes the timbral changes due to comb-filter effects as the early reflection patterns change with movement" (Lennox et al. 2001, p. 8). The possible applications to games, virtual reality, sound-field reconstruction, and composition should be clear.

This can also be explored from a "non-real" perspective, where a point cloud can be randomly generated and used to determine densities and motions of sounds in space. In the instance displayed in Fig. 3, the number of points is randomly generated and the distances between them are determined by a rule-based system. This approach is similar to spatial swarm granulation (Wilson 2008), yet the points in this instance are not necessarily bound to granular synthesis processing techniques. For example, each point in the cloud can represent a filter, an impulse response, a sound or simple delay points where a sound is stretched as it moves past. The behaviour of such a point cloud can be determined using the popular Boids algorithm (Reynolds 1987) or as an approach to *timbre spatialization* (Normandeau 2009) or spectral splitting (Wilson and Harrison 2010), where each point represents a spectral bin (Kim-Boyle 2008; Torchia and Lippe 2004).

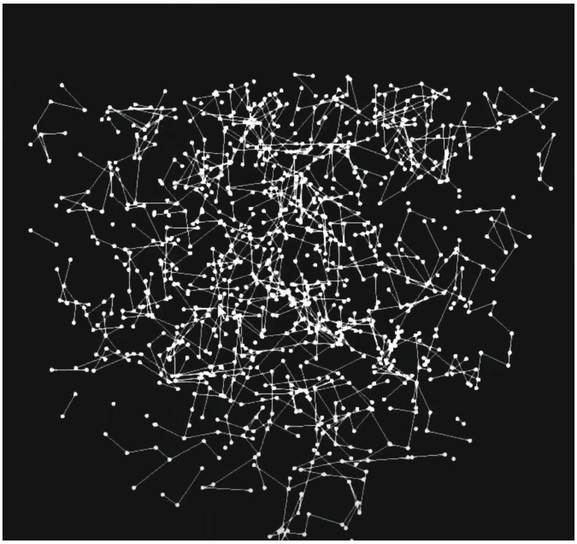


Fig. 3. A randomly generated point cloud which can be used to create density maps and motions with no basis in the real world.

Rather than relying on the panning of individual sources bound to different speakers, these approaches mirror some of the theoretical concerns discussed so far in this chapter. Consider *gait* (allure), the morphological criterion which denotes the fluctuation or undulation of a sound. In Schaeffer's system, this referred to characteristics of the sound itself, but gait can also be used to express how a sound moves through space. In sonic design, the feature space is not fixed, rather it is a field of possibilities (Godøy 1997). The *grain* quality of sound, refers to the perceived surface and tactile perception of a sound and could be extended into describing the perceived surfaces and reflective qualities of a space, as exemplified in the forest point cloud. Indeed, also the typological criteria of mass/facture, which relates how a sound occupies the spectrum and how its shape changes over time, can here be given spatial relevance. The density and distances between the points in the second example can dynamically be changed in response to how a sound changes over time.

Viewing the approaches to space sketched out here in relation to the psychoacoustic attributes discussed in the previous section widens the potential feature space as described through sonic design. Where gait, grain, impulsive, iterative, sustained, facture, duration, and variation describes the inner features of a sound, spatial impression, spatial width, apparent source width, and many other psychoacoustic attributes provide us with a salient framework for extending how we classify sounds and their spatial features.

The morphological visions of both Varèse and Petitot can in this way be given a renewed relevance and context in terms of spatial understandings of sound. Using methods of data extraction and reconstruction can create flexible models of motions and densities of how sounds move and behave. These approaches join Baalman's differentiation between techniques and technologies with that of sonic design, where we can work directly with the different values of the features, sub-features, and sub-sub-features within the multidimensional framework. Through piece-wise cumulative images, we piece together a sound and its behaviour in incremental steps, including its spatial qualities.

6 Conclusion

Sonic design and the system for classifying sounds, the typomorphology, extend our understanding of the interplay between sound and space. The morphological descriptions described by Schaeffer in the *Treatise on Musical Objects* provide a rich tool-set to pursue and understand these perspectives from artistic and scientific perspectives.

Spatialisation designs are often made on purely technical grounds where individual sounds are panned from speaker to speaker, layered, and moved in and out of specific densities. However, these approaches are often considered "after the fact," when the sounds are made; what is left is purely a presentation format. Many current technologies for sound spatialisation emphasise a point source approach, where individual sounds are panned as points with no consideration for the remaining context. The approaches outlined above, with roots in a morphological description of sound motion, density, and presence within a spatial context, sketch out an open-ended approach to building spatial scenes. This approach draws on what was earlier referred to as an *action–action* relationship, where actions are always part of an *inter*-action. This approach reflects the methods of sonic design, where the criteria are subdivided and sub-subdivided in a top–down, subjective exploration of feature categories.

The sound object has been seen by many as a significant part of musical experience, both as a tool for understanding and creating music. In this chapter, by drawing on the methods of analysing and categorising features of sound objects, this has been extended into spatial dimensions and uses. Through this expanded focus, and by drawing on the rich and varied attributes of acoustics and psychoacoustics, there will most likely be many more salient features to identify as we are now not solely considering musical sound but spatial presence as well.

References

Baalman, M.A.: Spatial composition techniques and sound spatialisation technologies. Organ. Sound **15**(3), 209–218 (2010)

Barreiro, D.L.: Sonic image and acousmatic listening. Organ. Sound **15**(1), 35–42 (2010)

Barron, M., Marshall, A.H.: Spatial impression due to early lateral reflections in concert halls: the derivation of a physical measure. J. Sound Vib. **77**(2), 211–232 (1981)

Barron, M.: The subjective effects of first reflections in concert halls–the need for lateral reflections. J. Sound Vib. **15**(4), 475–494 (1971)

Blesser, B., Salter, L.-R.: Spaces Speak, Are You Listening?: Experiencing Aural Architecture. MIT Press, Cambridge (2009)

Born, G. (ed.): Music, Sound and Space: Transformations of Public and Private Experience. Cambridge University Press, Cambridge (2023)

Bradley, J., Reich, R., Norcross, S.: On the combined effects of early-and late-arriving sound on spatial impression in concert halls. J. Acoust. Soc. Am. **108**(2), 651–661 (2000)

Chion, M.: Guide to sound objects: Pierre Schaeffer and musical research. J. Dack & C. North, Trans. (2009). http://www.ears.dmu.ac.uk

Chion, M.: Sound: An Acoulogical Treatise. Duke University Press, Durham (2016)

Eco, U.: A Theory of Semiotics. Indiana University Press, Bloomington (1979)

Garnett, G.E.: Music, signals, and representations: A survey. In: Giovanni De Poli, A.P., Roads, C. (eds.) Representations of Musical Signals, pp. 325–369. MIT Press, Cambridge (1991)

Godøy, R.I.: Formalization and epistemology. Scandinavian University Press, Oslo (1997)

Godøy, R.I.: Gestural-sonorous objects: embodied extensions of Schaeffer's conceptual apparatus. Organ. Sound **11**(2), 149–157 (2006)

Godøy, R.I.: Images of sonic objects. Organ. Sound **15**(1), 54–62 (2010)

Godøy, R.I.: Sonic object cognition. In: Bader, R. (ed.) Springer Handbook of Systematic Musicology, pp. 761–777. Springer, Heidelberg (2018). https://doi.org/10.1007/978-3-662-55004-5_35

Godøy, R.I.: Thinking sound-motion objects. In: Filimowicz, M. (ed.) Foundations in Sound Design for Interactive Media: A Multidisciplinary approach. Routledge, Abingdon (2019)

Godøy, R.I.: Perceiving sound objects in the musique concrète. Front. Psychol. **12**, 1702 (2021)

Griesinger, D.: Spaciousness and envelopment in musical acoustics. Audio Engineering Society Convention 101 (1996)

Griesinger, D.: The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces. Acta Acust. Acust. **83**(4), 721–731 (1997)

Griesinger, D.: Objective measures of spaciousness and envelopment. In: Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction (1999)

Grond, F., Hermann, T.: Interactive sonification for data exploration: how listening modes and display purposes define design guidelines. Organ. Sound **19**(1), 41–51 (2014)

Halmrast, T.: Acoustics in between: perception of sound in rooms beyond standard criteria. Psychomusicol. Music Mind Brain **25**(3), 256 (2015)

Halmrast, T.: Attack the attack: reverberation influences attack and timbre. Proc. Inst. Acoust. Hamburg **40**, 335–345 (2018)

Harley, M.A.: Music of sound and light: Xenakis's polytopes. Leonardo **31**(1), 55–65 (1998)

Hermann, T., Hunt, A., Neuhoff, J.G.: The sonification handbook. Logos Verlag, Berlin (2011)

Holmes, T.: Electronic and Experimental Music: Technology, Music, and Culture. Routledge, Abingdon (2012)

Howard, D.M., Angus, J.: Acoustics and Psychoacoustics. Taylor & Francis, Milton Park (2009)

Kendall, G.S.: A 3-D sound primer: directional hearing and stereo reproduction. Comput. Music. J. **19**(4), 23–46 (1995)

Kim-Boyle, D.: Spectral spatialization - an overview. In: Proceedings of the International Computer Music Conference (2008)

Lakoff, G., Johnson, M.: Metaphors We Live By. University of Chicago Press, Chicago (2008)

Landy, L.: Understanding the Art of Sound Organization. MIT Press, Cambridge (2007)

Lennox, P., Myatt, T., Vaughan, J.: 3D audio as an information environment. In: Audio Engineering Society Conference: 19th International Conference: Surround Sound-Techniques, Technology, and Perception (2001)

Lokki, T.: Throw away that standard and listen: your two ears work better. Build. Acoust. **20**(4), 283–293 (2013)

Lokki, T.: Tasting music like wine: sensory evaluation of concert halls. Phys. Today **67**(1), 27 (2014). https://doi.org/10.1063/PT.3.2242

Lokki, T.: Why is it so hard to design a concert hall with excellent acoustics? Dim **1**, 73–13 (2016)

Moore, B.: An Introduction to the Psychology of Hearing. Academic Press, Cambridge (2003)

Normandeau, R.: Timbre spatialisation: the medium is the space. Organ. Sound **14**(3), 277–285 (2009)

Petitot, J.: Morphological eidetics for a phenomenology of perception. In: Petitot, J., Varela, F., Pachoud, B., Roy, J.-M. (eds.) Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science, pp. 330–371. Stanford University Press, Stanford (1999)

Porcello, T.: Speaking of sound: language and the professionalization of sound-recording engineers. Soc. Stud. Sci. **34**(5), 733–758 (2004)

Raaijmakers, D.: Cahier ”M” – A Brief Morphology of Electric Sound. Leuven University Press, Leuven (2000)

Reynolds, C.W.: Flocks, herds and schools: a distributed behavioral model. ACM SIGGRAPH Comput. Graph. **21**(4), 25–34 (1987)

Risset, J.-C.: Timbre analysis by synthesis: representations, limitations, and variants for musical composition. In: Giovanni De Poli, A.P., Roads, C. (eds.) Representations of Musical Signals, pp. 7–43 (1991)

Risset, J.-C., Wessel, D.L.: Exploration of timbre by analysis and synthesis. In: Deutsch, D. (ed.) The Psychology of Music, pp. 113–169. Academic Press, Cambridge (1999)

Roads, C.: The Computer Music Tutorial. MIT Press, Cambridge (1996)

Roads, C.: Composing Electronic Music: A New Aesthetic. Oxford University Press, Oxford (2015)

Rumsey, F.: Spatial quality evaluation for reproduced sound: terminology, meaning, and a scene-based paradigm. J. Audio Eng. Soc. **50**(9), 651–666 (2002)

Schaeffer, P.: In Search of a Concrete Music. University of California Press, Berkeley (2012)

Schaeffer, P.: Treatise on Musical Objects: An Essay Across Disciplines. University of California Press, Berkeley (2017)

Schnell, N. (2013). *Playing (with) sound: Of the animation of digitized sounds and their reenactment by playful scenarios in the design of interactive audio applications* (Doctoral dissertation). Institute of Electronic Music, Acoustics, University of Music, and Performing Arts Graz, Austria

Smith, S.W.: The scientist and engineer's guide to digital signal processing. California Technical Publishing (1999)

Strogatz, S.H.: Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering. Westview Press, Boulder (2015)

Torchia, R.H., Lippe, C.: Techniques for multi-channel real-time spatial distribution using frequency-domain processing. In: Proceedings of the 2004 Conference on New Interfaces for Musical Expression, pp. 116–119 (2004)

Valiquet, P.: Hearing the music of others: Pierre Schaeffer's humanist interdiscipline. Music. Lett. **98**(2), 255–280 (2017)

Varèse, E., Wen-Chung, C.: The liberation of sound. Perspect. New Music **5**(1), 11–19 (1966)

Westerkamp, H.: Linking soundscape composition and acoustic ecology. Organ. Sound **7**(1), 51–56 (2002)

Wilson, S.: Spatial swarm granulation. In: Proceedings of the International Computer Music Conference (2008)

Wilson, S., Harrison, J.: Rethinking the BEAST: recent developments in multichannel composition at Birmingham ElectroAcoustic Sound Theatre. Organ. Sound **15**(3), 239–250 (2010)

Wright, M., Beauchamp, J., Fitz, K., Rodet, X., Röbel, A., Serra, X., Wakefield, G.: Analysis/synthesis comparison. Organ. Sound **5**(3), 173–189 (2000)

The Sonic Imagery of the Covid-19 Pandemic

Georgios Varoutsos(✉) and John D'Arcy

SARC: Centre for Interdisciplinary Research in Sound and Music, Queen's University Belfast,
Belfast, Northern Ireland
{gvaroutsos01,J.DArcy}@qub.ac.uk

Abstract. The Covid-19 pandemic catalysed disruptions and disturbances in ways of living across the globe. Many of these changes in daily life were felt through stark changes to our soundscapes, particularly those in urban centres. Might we better understand the effects of the Covid-19 lockdowns through sonic analysis? This chapter explores how sound analysis methods, including concepts of the sound-motion object and sonic image, might aid in understanding the environmental soundscapes of the pandemic lockdowns. The discussion focuses on the Sounding Covid-19 project—an initiative involving a series of field recordings carried out during Covid-19 pandemic-related events in the urban environments of Belfast, Northern Ireland (2020–2022) and Montreal, Canada (2020–2021). The project presents the sound archive through various listening experiences, including soundscape compositions, sound mapping and narrative-based radiophonic work. We consider how the pandemic may have invited us to pause and reconsider how we document and archive the present to look back and better understand the future. Sound may be vital in understanding our environment and the socio-cultural shifts over time. This chapter argues that documenting, preserving, and analysing the soundscapes of the pandemic lockdowns may help us reflect on our shared histories in several ways.

Keywords: Pandemic Soundscapes · Covid-19 · Sound-Motion Objects · Sonic Images · Mental Presence

1 Introduction

In a multi-sensory world, sound provides an essential set of temporal and spatial information about the activities occurring in one's surroundings and helps us understand the phenomena of everyday life. The lockdowns of the Covid-19 pandemic significantly altered the activities and patterns in everyday life for many and subsequently transformed how one's sense of normality might be perceived through sound. Kang (2014:43) describes urban soundscapes as perceptions based on societal and environmental conditions, which include aspects such as culture, history, and politics. With this in mind, the soundscapes of pandemic lockdowns indeed reflect the conditions of the time through the fluctuating sonic identities of urban spaces. This chapter considers how we might reflect on our experiences of the pandemic lockdowns via a sonic perspective—listening to sonic activities in urban spaces across the varying stages of lockdowns, exit strategies, and lifted restrictions.

A. R. Jensenius (Ed.): SSD 2022, CRSM 12, pp. 192–210, 2024.
https://doi.org/10.1007/978-3-031-57892-2_11

The chapter aims to reflect on the pandemic through modes of listening and sonic analysis. It is hoped that listening "through" the pandemic may allow us to connect to sonic properties from past and present experiences and offer a method of reflection on the changes that occurred throughout the pandemic. This listening can highlight how restrictions impacted the sonic state of the urban places. Changing patterns in the sounds activated by the elements, animals, machines, and humans may all be read as indicators of the transformative nature of the lockdowns.

The sonic material analysed in this chapter was produced through the *Sounding Covid-19 Repository*—a field recording and soundscape composition project initiated by the first author. This project collected field recordings during pandemic lockdowns in Belfast (2020–2022) and Montreal (2020–2021). Information and documentation on the *Sounding Covid-19 Repository* can be found on the project web page.

(https://georgiosvaroutsos.com/covid-19/) or on Zenodo (https://doi.org/10.5281/zenodo.8245035).

This chapter will analyse these pandemic lockdown soundscapes through several approaches, focusing on Godøy's concepts of sonic images (2010) and sound-motion objects (2019). While Godøy defines these concepts primarily in the realm of musical composition and perception, we consider how these concepts can be applied more broadly to understand soundscapes by introducing Wittmann's theory of mental presence (2011).

In considering the importance of soundscapes in shaping our lived experience, we might look to Udsen and Halskov's (2022) ideas on the soundscape's role in placemaking, and Radicchi et al. (2021), who explain that sound facilitates communication and spatial orientation whilst serving as an emotional source of direction for us, whether consciously or unconsciously.

Through the sonic analysis of the *Sounding Covid-19 Repository*, we attempt to explore the possibilities of understanding the shifting societal changes of the pandemic lockdowns through the medium of sound.

2 An Overview of the *Sounding Covid-19 Repository*

The *Sounding Covid-19 Repository* used soundwalking as a methodology to actively engage with urban spaces, collecting audio material using a handheld field recorder during individual soundwalks throughout various stages of the pandemic. The field recordings were edited and presented in soundscape compositions, which were then published as a website audio archive, online soundmaps, and an interactive location-activated soundwalk experience. The project also involved recorded interviews where participants relayed their personal experiences during pandemic lockdowns. These voices were presented in combination with the field recordings in the radiophonic work *Covid-19 Sound Stories* presented on the project web page.

In early 2020, as countries around the globe introduced lockdowns, the first author was based in Belfast, Northern Ireland. Here, social distancing and Stay-At-Home rules were introduced, with restrictive measures fluctuating at various stages of the lockdown. These changing restrictions palpably transformed the interactions between urban and natural environments and separated people from one another. As travel restrictions were lifted, the project broadened beyond Belfast to include Montreal, Canada. The two places were chosen based on circumstances and opportunities. Most of the attention is focused on Belfast because that is where the first author undertook their PhD during Covid-19 lockdowns. On the other hand, Montreal was where the research was done when the first author was allowed to go home for home visits.

The audio recordings in Belfast and Montreal were created as part of soundwalks, a methodology for actively listening while navigating an environment (Adams et al. 2008; Paquette and McCartney 2012; Drever 2011; Carras 2019). Urban sites were chosen as the locations for active listening and recording exercises. The sites were chosen based on tourist maps of Belfast and Montreal, considering areas that might typically exhibit either familiarity, visitation density, or dynamic social interaction for both locals and tourists. This was also a tool for self-study of how one's emotional, psychological, and physical proximity to the source of the sound affects how one perceives and reacts to it (International Organization for Standardization 2017). Subsequently, the chosen sites were linked together to form soundwalk routes that would comply with government restrictions.

The act of field recording in the *Sounding Covid-19 Repository* aimed to highlight the intrinsic value of recording urban soundscapes and listening back to glean useful information. The recording processes aimed to fulfil a conservational function, a common aspect of many field recording projects (Western 2018; Freeman et al. 2011; Demers, 2009). The project also aimed to document and share the sounds of pandemic lockdowns in ways akin to what Cusack describes as 'sonic-journalism' (2012)—considering all sound activity (not just verbal) to be informative and offering communication and understanding of a moment in time at a specific place.

The recorded soundwalks were carried out each time local government restrictions were changed during the lockdown and exit strategy phases. This reiteration drew on Gorichanaz's (2017) ideas of auto-hermeneutics as a way of embracing and reflecting upon phenomena through repeatable methods. Field recordings were made following the initial restriction guidelines, and to maintain consistency between each lockdown in both cities, used the same constraints to construct a comparative recording framework. Using only accessible equipment, a portable handheld recorder, the Zoom H6 with an X/Y capsule set at 120° recording in stereo, recordings were typically for a duration of five minutes. Subsequently, post-production focused on usable material from the recordings, removing clipping or distorted material. These soundscape compositions were time-compressed to two-minute soundscape compositions following Truax's (2022: 287) concept of focusing on the key features of the recordings. There are 91 pandemic soundscape compositions: 78 for Belfast and 13 for Montreal, with the majority based on two-minute durations and only a few exceptions that are considered other recordings that are not based on urban locations being four minutes (focused on local cultural events such as St. Patrick's Day, The Twelfth, or Christmas Market). This chapter only focuses on the two-minute soundscape compositions that create the comparative nature of the research.

The audio editing processes involved in creating the soundscape composition aimed to produce creative listening experiences, as is often the intention in soundscape composition (Sarwono et al. 2022; Truax 2002; Westerkamp 1999). In this case, the soundscape compositions form a reflective repository to identify particular sounds or sonic activities. The soundscape compositions form an aural chronological overview of the changing urban environments throughout the pandemic lockdowns, exit strategies, and lifted restrictions. Thus, the soundscape compositions somehow document the distinct sound spaces created when government restrictions acted as the invisible agents of change in our urban spaces.

3 Analysing Soundscapes

The sounds of urban environments are a collection of perceptual experiences that create a sensorial link between the listener and space. Soundscapes can be viewed as contextual perceptions, representing many variables occurring within an environment (Brooks et al. 2014). Here we consider varying methods of soundscape analysis that may aid in better understanding the audio archive of the *Sounding Covid-19 Repository*.

Considering urban soundscape analysis, Léobon (1995) and Lebiedowska (2005) describe six types of sonic sources with varied perceptual responses within an urban acoustic environment: i) Background noise, ii) Mechanical, iii) Human activity, iv) Nature, v) Human presence, vi) Speech and communication. Parmar (2022) uses four categories: Earth sounds (wind, water etc.), Human sounds (voice, action), Animal sounds (cries, calls, etc.), and Technological sounds (alarms, motors, etc.). The International Organization for Standardization (2017) distinguishes categories for the urban acoustic environment as either: (1) human activity or facility-generated sounds, including transport, human movement, electromechanical, voice and instrument, other human, and social communal, or (2) non-generated human activity which includes nature and domesticated animals. With a classification system, we can start to quantify the sounds present in the environmental mix and, for instance, observe the changes in human presence over the course of pandemic lockdowns.

Temporal measurements of broadband noise levels can provide amplitude data that might be useful in comparing soundscapes, though spectrograms mapping changing amplitudes of specific parts of the frequency spectrum may be more helpful in identifying the sound level and tonality of specific sounds. However, we might also consider augmenting these quantitative measurements with qualitative descriptions of the individual sounds. These might highlight particular perceptual moments and, as Axelsson et al. (2010) suggest, relate sensations beyond noise levels and speak to broader contexts and considerations such as human well-being. On an unconscious level, we may perceive some sounds as signalling comfort and security, whilst others may trigger anxiety or insecurity.

3.1 Sound-Motion Objects

To comprehend the components of lockdown soundscapes, examining the sonic properties that make up the sonic event using Godøy's theories of sound-motion objects enables the investigation of sonic images. In musical and perceptual contexts, Godøy (2019) explains that sound-motion objects are short durational sound fragments between

0.3 and 3 s that are focused on gesture and have limited perceptual attention. This durational constraint would allow the recognition of prominent dynamic musical features, as well as perceptual motion and feelings towards the sound. Figure 1 shows, for example, a bell from the Albert Memorial Clock that can be heard in Exit Strategies 2020 at 1 m 31 s into the piece, and the chime lasts 2.81 s, providing both a sonic feature of the landmark and awareness of the sound source.

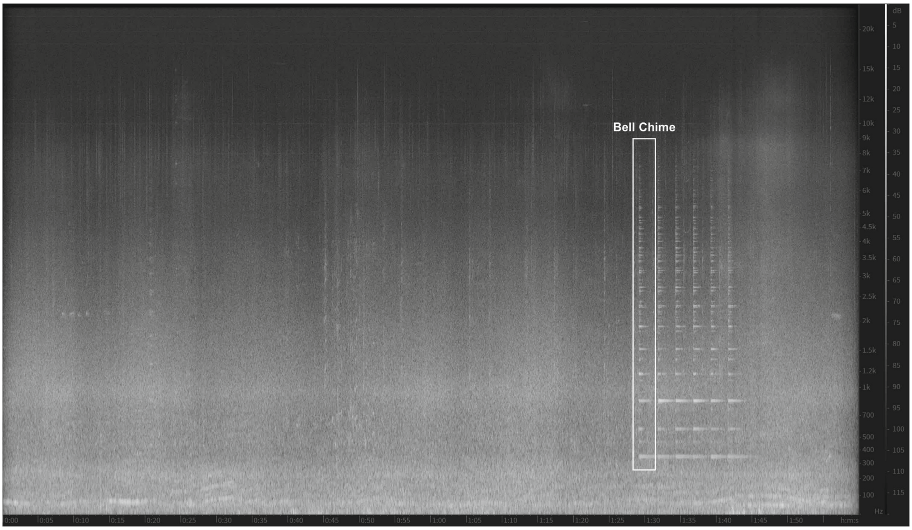


Fig. 1. Spectrogram of Bell Chime of 2020-07-04-Albert Memorial Clock-Exit Strategies 2020.

Much of Godøy's research stems from Schaeffer (1966), who coined the "sound object" and referred to it as the minimal perceptual representation of a sound's features concerning spectral dynamics and focused attention. Organic or artificial sounds provide the ability to hear and perceive the characteristics of one sound from an acousmatic position. Chion (2009) builds on Schaeffer's approach and explains that a sound object is isolated from visual perception, removing context and allowing for a reduced listening approach focused on that sound itself. For example, in 2021-01-02-Place Jaques Cartier-Montreal Lockdown-Part 3, Fig. 2 displays the vocalisation of a yell at 49 s lasting for ~ 1.9 s, depicting the range and a possible response to its sound being heard.

The singular sound listening experience highlights a sound's qualities, not its relationship to a space or place. However, the sound–motion object proposed by Godøy (2019) relays a set of spatial and temporal cues to the listener regarding durational sound fragments, extending from the sound object with consideration of sound and bodily motion.

When we consider sound–motion objects with recorded or listened-to soundscapes, each with its limitations of individualised sound-object listening, we can perceive individual sounds to understand sonic properties and designs without meaning. Gaver (1993) would refer to this approach as "musical listening," a perceptual observation of a sound pattern, quality, and identity. However, sounds within an environment are not always

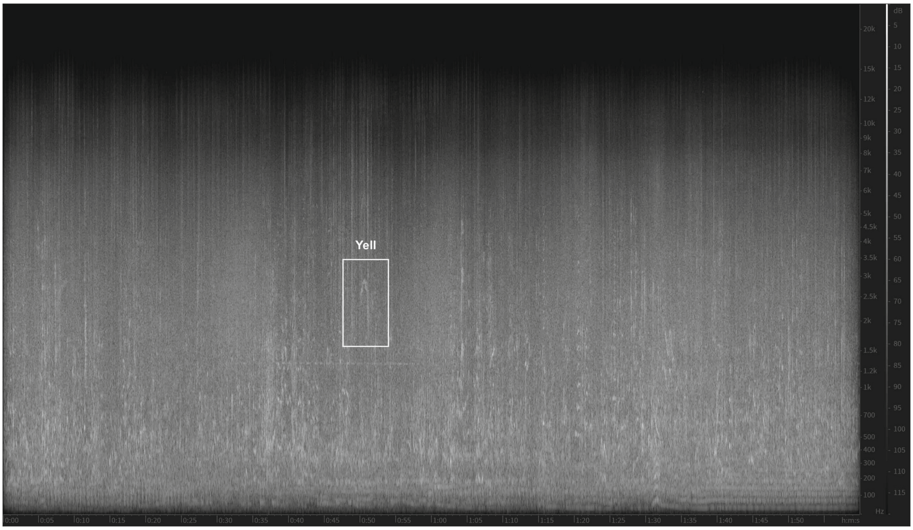


Fig. 2. Spectrogram of 2021-01-02-Place Jaques Cartier-Montreal Lockdown-Part 3.

perceived individually; Gaver also proposes "everyday listening," sounds that create a momentary experience within the environment. Both musical and everyday listening encompasses the sound source, but the interpretive framework determines what sounds can offer the listener. A sound can be an action or an event, depending on the positioning of the listener. One can consider either the sound–motion object, which would identify the features and movement of a particular sound or a sound event, which collects experienced sounds to interpret the environment and allows for reactive decision-making.

3.2 Mental Presence

To embrace a holistic approach with both sound–motion objects and everyday listening to phenomena, Wittmann (2011) proposes the concept of mental presence, a moment of unified experiences of self and presence. Combining spatial-temporal features can encapsulate sound-motion objects within experienced moments beyond the 3-s limit, allowing for a better understanding of sound sequences and their context. Mental presence, while perceptual, aids in the sonic recall by listening to all sonic moments instead of short sonic fragments, where there may be a reduced capability to accurately depict all sonic information. Setti et al. (2022) studied spatial memory and discovered that people could identify the source of an unknown sound within three and a half seconds; however, this was with separate sound playback rather than sequential. Kaplan and Iacoboni (2007) discuss how, in the environment, multimodal representations are better perceived by action sounds than non-active ones. Therefore, to understand sonic changes in an environment and the connection between the perceiver and the lived world, sound in the natural world needs to be understood as a continuous perceptual link to the changing environment.

Soundscape listening can provide a framework for the perceptual interpretation of the self in the present through mental presence and consideration of the sonic characteristics of sound–motion objects. Lähdeoja (2018) reexamines the original ideas surrounding Schaeffer's sound object by introducing contextualisation of sounds to their environment, thereby expanding the concept of gestures and movement. Permitting multiple perceptual understandings that are transferable and applicable to diverse creative forms of soundscape compositions or others.

Truax (2001; 2022) introduces and develops the idea of analytical listening, which involves technological capabilities of re-examining collected sounds for contextual knowledge-making through repeated listening experiences. Regarding the Covid-19 soundscapes, these soundscape recordings and compositions enable a repeatable re-experience and re-examination of the perceived sonic environment over time, which can be compared to sound–motion objects and the expansion of considering mental presence for further analysis.

Godøy (2019) describes how we consider sound production and perception by instrument or body motion for music and sound design. Those same principles apply to listening in on the sonic environment, forming relationships between the features and the perception of heard sounds in a place. Jenison (1997) explains that on an ecological level, we consider physical acoustic properties such as.

- sound intensity: sonic energy with various lengths of decay
- interaural-time-delay: sound heard between the left and right ear
- Doppler effects: sound moving through mediums

All of these inform about audio signals in a space and place. The audio signal provides position, direction, and movement. In contrast, we perceive sound characteristics as salient features to distinguish between a place and space.

We can engage in alternative levels of comprehension regarding the meaning of those listened sounds concerning a place and space by listening to audio signals and forming perceptions of those sounds and events. As listeners, Feld (1996, p. 97) explains that sound can be used as a tool for understanding sonic experiences, coined as an acoustemological framework. Acoustic ecology acknowledges the relationship of the sonic environment to the listener, what sounds mean, and informs us of a place (Devers 2019; Westerkamp 2002; Traux 2001; Schafer 1966). Creating a possibility for knowledge-making through sonic experiences embraces our awareness of sonic presence. During a period known as mental presence, when sensory-motor perception, cognition, and emotion mix to produce a phenomenal experience, a person notices themselves and their surroundings (Wittmann 2011). It extends to being present within the environment and listening to sound to build relationships within a place by attaching meaning to sounds from our perspective.

4 Analysing the *Sounding Covid-19* Soundscapes

Regarding sounds during the pandemic, we need to consider the sonic markers of urban spaces. Sonic markers, or soundmarks, are culturally significant sounds that identify a place and space (Birdsall and Drozdzewski 2018). While a sonic marker agrees with

Godøy's sound–motion object in terms of features, each individual sound marks a particular place yet is not always in line with the context of the space. Interwoven sounds provide aural information to the listener, understanding their position in the place or space and rendering a perceptual understanding of the sonic environment. According to Cuadrado et al. (2020), for listeners to have a complete experience, they need both sound sources and sound events to create meaning, interpretation, and emotional responses. We would define sound sources as the sound–motion objects and the sound events relating to mental presence. Emotional responses to sounds play a crucial role in individuals' considering sounds pleasant or unpleasant, significant or insignificant, which then promotes the sonic identity of a space, either by the individual or community (Liu and Kang 2016; Jeon et al. 2013; Yang and Kang 2013; Cain et al. 2013). Each listener will have an individualised experience, rendering multiple varied perceptions of sonic importance or relevancy from a place, meaning that sound markers can change from person to person and across time.

Hall et al. (2013) examined that the interaction between the individual experience and subjectivity in a physical and a socio-cultural setting is equally significant as the auditory signal. This is not to say that the physical property of sounds in a place should be disregarded, but instead that it is essential to understand that once the listener attaches meanings to sounds, hierarchy, attention, and recall, they create a sonic impression of a place and space. Sound markers are critical to the phonic identity of a place, incorporating the sounds in the environment, associating with the individual or community, and branding the various types of activities that occur from their sources (Rehan 2016). Only listening to sound–motion objects makes a minimal connection to the context of sound events in the place. However, the distinctive characteristics of produced sounds allow one to review the contextualisation and association of that sound with a place.

For instance, certain acoustic characteristics of Belfast's urban spaces were captured on field recordings during lockdown periods. For example, the Albert Memorial Clock is an iconic slanted clock tower structure located within the city that serves as a physical connection between entering and exiting the city. However, its distinctive bell chimes that acoustically complement the landscape are what give it its sonic identity, and these chimes are what make it easily recognisable (Fig. 3).

The clock's chimes are distinguishable based on their acoustic characteristics and the physical materials used in their construction, with each strike travelling horizontally and vertically through space and time. One chime lasts approximately 2.7–2.9 s, making it an acceptable indicator of Godøy's sound–motion object. Nevertheless, if we consider the pattern that develops over time and the accumulation of fifteen-second-long chimes that represent the hours of the day we enter a state of mental presence that enables us to recognise a sound and be aware of our immediate surroundings. This enlarges the specificities of a place by indicating a particular moment only by listening to an extension of the sound–motion objects that are sequentially attached to provide a context.

By analysing the clock's chimes and identifying the specific source of the sound, we can locate them at landmarks in the physical environment. Based on our analysis of the recording's traffic flow, seagull activity, and human-generated sounds, we can accurately determine the level of human activity at the location and time. For instance, Traffic Flow

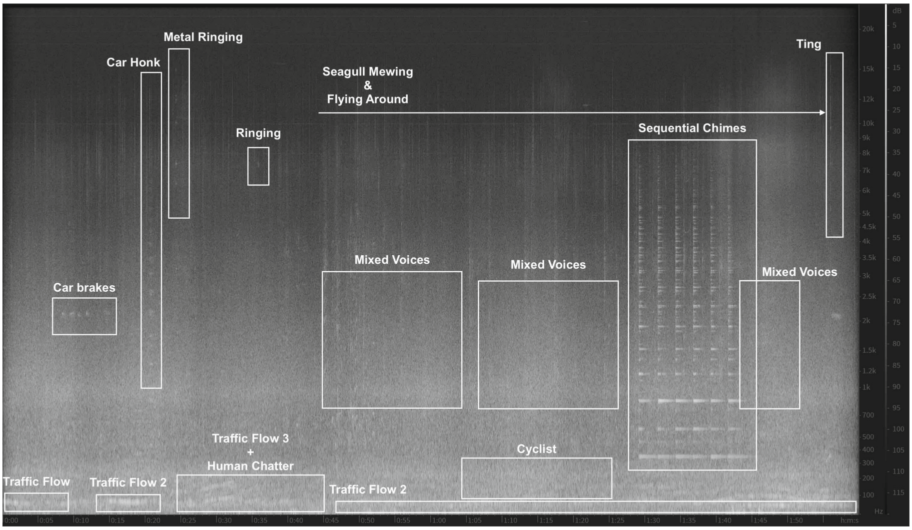


Fig. 3. Spectrogram of 2020–07-04-Albert Memorial Clock-Exit Strategies 2020 - summed channels.

2 lasted 6 s, seagulls for 1 m 12 s, and mixed voices were audible for 20 s. These findings suggest the presence of a group of people and a possible relaxation of restrictions.

With the addition of context information, like metadata descriptions, we can look at how all the sounds in a specific soundscape composition relate to each other and come up with more ways to explain how the sounds interact, form relationships, and consider the state of events related to urban space in the city with either previous experience or collected data. This enables a mode of understanding that these sounds are influenced by an event, such as Covid-19, which has resulted in a decrease in human-generated sounds during lockdowns, including fewer sounds from social spaces, pubs, commutes, vehicles, and other human activities. Due to government regulations, these "regular" urban sounds were absent, which created the impression that the area had been abandoned or isolated. When the restrictions were lifted, the spaces began to flourish with the preconceived sonic properties of these urban spaces. This phenomenon was caused by long periods of lockdown, which encouraged people to reunite, and the architecture of these urban places as platforms for various auditory interactions between humans, urban sounds, and nature sounds.

A second example is the Botanic Gardens, which host a variety of human, natural, and urban sounds but, during the first lockdown, seemed less affected by changes. However, listening solely to each of the sound–motion objects individually, such as a bird call, a spoken voice, a distant or passing vehicle, or a cyclist, would make it difficult to understand the location recording and the context. However, by enlarging the focus with mental presence to combine the mixture of those aforementioned sound-motion objects, it can be perceived that it is in more natural environment settings as there are fewer sounds generated by what ISO (2017) would define as facility activity. Without context, this may be deemed any other park or nature recording and any part of the year. Yet, the metadata provides an additional layer to attach the recordings to specific periods or events, framing the recording to consider that once we recognise certain sound-motion objects, it enlarges our framework for understanding with mental presence. We can consider how sounds fit the context of the place and space, informing us about possible relationships affected or generated by events such as Covid-19.

We design each public urban space for a purpose and function, constantly inviting different interaction sets and sonic relationships from city planners, building architecture and designs, artistic practises, and daily life (Belgiojoso 2014). However, if we only consider sound as a sound–motion object, we may overlook the connections to and from a specific sound, its relationship to the spaces it inhabits, and the listener's perceptual experience.

The state of lockdowns removed humans from inner city life and allowed them to reconnect with the natural world, even while living in a city. Regarding sound–motion objects, it would be difficult to distinguish the Botanic Gardens (an urban city park) from other urban city parks during lockdowns. If we increase our capacity for perceptual awareness by listening for extended durations, we may be able to determine where we are in a given environment. We can distinguish between a city park and a forest based on various urban ambiences and/or sounds, birds, human sounds, and human activity. It is possible that a single sound can aid in the detection of auditory features, but when we listen for longer periods, we gain the ability to differentiate between different environments and identify spaces based on their collective sonic makers.

> Some of the socially important values that are ascribed to soundscapes include creating a sense of place, providing cultural and historical heritage values, interacting with landscape perceptions, and connecting humans to the nature. (Jia et al. 2020)

Reflecting on individual sound markers, the question becomes, what happens when we listen to multiple sounds over time, and how do we derive meaning from those accumulated sounds? Regarding soundscapes and mental presence, we enlarge the listening experience to connect with the surrounding sonic environment as a perceiver and creator of sounds within that space. Sounds produce an ambience or a reflection of space, with various impressions of sounds associated with the pandemic lockdowns. When multiple sounds are active within a period, such as masking, the ability to differentiate each sound becomes muddled unless the context of multiple sounds is considered an event. The sound marker functions as a representative sonic anchor for a particular location and connects the soundscape to the landscape. This is also useful for creating a basis for comparison when repeating listening practises and soundscape recordings. Developing a pattern that requires measurement points, including date, time, recording position, and location. The *Sounding Covid-19 Repository* strictly adhered to the government's guidelines throughout the pandemic. All recordings were limited to a maximum of five minutes and taken at consistent locations and times, with only slight variations due to weather. Moreover, the recordings repeat on the same days each year, except during Lockdown 2, when they focus on the weekly comparisons. This comprehensive two-year study offered a detailed analysis of the impact of pandemic restrictions on urban soundscapes.

Sound's meaning is inextricably linked to the environment in which it is produced, heard, and understood. Our environment influences many of the concepts we use to comprehend the world, and our upbringing unconsciously shapes how we hear. As a result, our understanding of sound is considerably more complex than most individuals think about. Sound is a resource and instrument for constructing the relationship between the listener and their present environment. It is the combination of experiencing the present moment through auditory senses, which bases itself on the situationality of sound.

5 Understanding the *Sounding Covid-19 Repository*

Embracing both sound–motion objects and mental presence, we can listen to recorded sounds to develop an internal visualisation of a recorded place and space by considering spatial and temporal specific sound behaviours in a process Godøy discusses as "sonic images" (2010). A recorded or composed soundscape has the ability to preserve sonic information, allowing individuals to listen to or examine sonic events to imagine the acoustic environment, and therefore has the potential to contribute towards knowledge-making. As previously discussed, a sound–motion object is a short durational perceptual understanding of a sound's features, with the concept of mental presence to extend the time and consider the present awareness of the listener with sound. However, moving from a sound–motion object to a sonic image perspective, we designate a specific time-frame to retain the typology and morphology of a sound object, enabling the visualisation of a sound object's shape, qualities, and movement (Godøy 2019). These essential characteristics enable the imagination of sounds without the need to be present when they occur. Like a soundscape, recorded and listened-to sounds have multiple contexts, such as mobile device listening, designed soundwalk experiences (apps), online via soundmaps or audio players, and other creative playback methods.

The purpose of the soundscape is to generate an auditory understanding of a time and place, which is just as important as visual information for landscape conservation (Brown 2010). Possibilities exist to audibly visualise sounds, ambience, and sonic events from a particular period and expand sensorial knowledge-building and the period's sounds. We create an internal visualisation of sounds, highlighted by sonic markers and other sonic features, from a location to imagine the ambience of a place to be listened to later. According to Kang (2014: 96), the perception of a soundscape is the result of a deliberate design procedure. A sonic imprint develops, which may change over time based on the environmental and urban relationships cultivated or constructed during the urban space's development. Similar to how visual methods such as photography or painting can provide visual information and settings, preserving the soundscape through various recording techniques allows us to return to a sonic environment and place ourselves within it.

Soundscape recordings preserve information such as cultural events, socioeconomic shifts, defining meaning-making moments, and time-stamping a particular occurrence (Dumyahn and Pijanowski 2011). The soundscape recording is a collective representation of dynamic relationships, incorporating our past, learned, and current experiences to render perceptions of the space we occupy. We tend to assume there is a problem in the urban space of a city if there is a lack or absence of human sounds in the urban space because the urban space is designed to reflect these sounds (Ouzounian 2017). This perspective during the Covid-19 lockdown/exit strategies allows people to understand how societal changes affect sound environments and how sound can represent the effects of the pandemic on daily life. These effects can be shown through government, economy, health and well-being, and culture changes, such as business closures, the lack of street performers or musicians, rush hour traffic, and social interaction in different urban settings.

The sound identifies our own knowledge and life experiences from a phenomenological standpoint. Therefore, listening becomes an active moment with urban space and listening to the soundscape while understanding the more significant social impacts intertwined or absent in that space. The lockdown soundscape composition repository contains a two-year timeline of varying sonic environments influenced by government restrictions. It is not necessary for listeners to be present to experience these effects. However, by reviewing the audio material, they can form a sonic image based on the auditory characteristics of city life changes during and after lockdowns. The perceptual experience of isolation and abandonment occurred during lockdowns due to a lack of human movement, voice, and otherly generated human sounds. At the same time, an increase in wildlife was predominant in the foreground of the urban space, despite the presence of motorised vehicles in the distance. In particular, Commercial Court is in the Cathedral Quarter, known for its art and nightlife and an important part of Belfast's identity as a city. It is a specially designed area that would have had a greater amount of human-produced sounds, such as human voice and movement, had there been no lockdown at the time of the recording. Unfortunately, the lockdown rules prohibited certain outdoor activities, preventing businesses from opening and people from occupying designed urban spaces like this, creating an unusual historical period. By comparing the same recording location of Commercial Court in Lockdown 1 and Lifted Restrictions, we can compare the sound-motion objects with the use of mental presence to get an idea of how the area changed between these two times when different outside policies affected it.

Table 1 summarises the recognised individual sound–motion objects heard in the soundscape compositions during a period of mental presence and may help to place a comparative visualisation of the periods through sonic images.

Table 1. Comparison of Commercial Court in Lockdown 1 and Lifted Restrictions. Please refer to the project web page access the compositions or on Zenodo.

2020-03-27-Lockdown 1	2022-03-27-Lifted Restrictions
Seagull's Mewing + Movement	Human chatter
Distant Motorised Transport	Background Music
Electromechanical	Motorised Transport
	Electromechanical

As mentioned earlier, one's perception of a soundscape depends on their surrounding environment and experiences. The first recordings of the lockdown were carried out on March 27, 2020, in Belfast. During the first week of the lockdown, the soundscape was erratic and disorienting, and this was especially noticeable when recordings were carried out on Friday evenings. This is a vastly different recording and imaging of the space compared to Sunday morning, when the area would be much quieter sonically. This area was alive with pedestrians, musicians, shoppers, nightlife, modes of transportation, and other urban or natural sounds before the advent of Covid. *Lifted Restrictions* recordings were taken two years later, on March 27, 2022. Both soundscapes depict how wildlife was or was not present in the areas, indicating a sense of isolation. Within *Lockdown 1 of Commercial Court*, the absence of human-produced sounds was quickly filled by the movement and calling of various birds. A dominant factor that stands out from the local sounds can be an indication of changes taking place.This also provides another sensory experience of isolation, in that there are no human voice sounds near one of the more popular streets in the city centre. The birds' flocking and calling circled overhead, exacerbating the sense of isolation within this space. The only human sounds produced were self-made from recording this moment. Which differs when listening to the *Lifted Restrictions of the Commercial Court*. When lockdowns and restrictions were lifted, people returned to these urban areas, and human interaction sounds once again dominated the listening experience. As a result of the otherly sonic interactions of human sounds, wildlife becomes suppressed and almost nonexistent in this specific urban area. Furthermore, self-isolation is still relevant on a personal level. Human isolation is disrupted as there are forms of gathering and sounds of togetherness from the chatter, laughter, footsteps, and other relatable human sounds in this space once again.

Similarly, upon returning to Montreal, Canada, after months of isolation in Belfast, the research was expanded to specific sites to self-observe and self-reflect on the changes imposed by those local governments. During the Christmas holiday season, the Old Port district of Montreal usually hosts a variety of outdoor celebrations, cultural events, and entertainment shows or performances. Contrasting to the Belfast recording period and conditions, wildlife aids in indicating degrees of isolation, whereas this is not a similar point for Montreal, especially when much wildlife migrates or hibernates during these colder winter periods. However, recordings and compositions from this trip depict a mixture of isolated or less active sonic conditions and varying social encounters mostly indicated by human-generated sounds. People in Montreal attempted to embrace the cultural significance of winter celebrations by continuing to walk through snowy paths after the rule prohibiting them from entering other people's homes was changed just before Christmas and New Year's Eve.

Only in public urban spaces could such celebrations be shared with others. In the *Place Jaques-Cartier-Montréa*l soundscape composition, we can hear forms of speech, individuals purchasing and eating Tire d'érable (maple taffy) from outdoor kiosks—a local culturally traditional dessert—and others continuing to walk through cold conditions around the port area, with some utilising an outdoor light installation in a park square. The yell at 49 s, lasting 1.9 s, was previously mentioned as a sound–motion object in this piece. Still, if we consider mental presence and processing as sonic images to the listening experience, we can start to connect this particular yell with play and expressions of content.

The soundscape composition reflects Covid-19-modified cultural identity and interactions experienced in such urban spaces. By recording and forming these soundscape compositions, there is a process of self-connection to the areas, with another appreciation of the immediate moment. While there is an inability to possibly see everything, listening to the recorded and composed soundscapes allows us to visualise this beyond the single experience. For example, returning to the yell found in the piece, on its own, minimises any form of understanding of the context of the sonic activities present and recording conditions. Having soundscape compositions that encompass a longer durational experience of sounds defined under mental presence parameters, we can formulate the connections beyond the durational limits of sound–motion objects, even considering how the interplay of sounds affects the self of repeated listening experiences. This interacts with Covid-19's real and repeated experiences, ultimately attempting to re-adapt from previous interactions with the space while adhering to current government and health policies. Specifically, at this moment, the 2-m distance had to be continuously reminded of while recording: "Do not be so close to anyone." Having these invisible rules dictate movement in such complex areas allowed for variable sets of sonic interactions, whether formed by one's own bubble or multiple bubbles occupying a space. This mindset must be considered: there are constant considerations not only on how to experience but also on what is brought into or affected by visible or invisible factors.

However, there were significant individual differences based on prior exposure to these urban spaces and the Covid-19 effect during a culturally significant time of the year. The concept of the sonic imprint that will be experienced is generated by the sounds that are heard and specifically listened to, giving rise to a sense of memory as well as a response to a particular location. Extreme changes from the preventative measures taken for Covid-19 radically altered both the past and the present's sonic memories and experiences.

These experiences reinforced that sonic moments can be irretrievably lost and inspired me to record numerous instances that can provide auditory information for others to imagine, experience, and revisit in the future. From an archival standpoint, collecting, gathering, and including sonic information (recordings or other audio material) is a progressive step toward including all ranges to create a broader sense of history (Swain 2003). We can imagine a sound's features, characteristics, location, and relationships. However, this is only true if we experience the sound at the source or via recordings.

Schafer (1977) stated that earwitness reports from persons who were present and who testify or can testify as to what they heard are the only way we may learn about historical soundscapes. Not to imply that every moment should be captured, but historically, sound has not been preserved to the same extent as visual information (photographs, paintings, and text). Smith (2007) explains that visual information alone is insufficient to comprehend complete historical experiences and that various other senses must be preserved. Moreover, a sound's sensory production (replicability) and sensory consumption (contextual relationship) are distinct types of historical review. Sensory consumption focuses on understanding what a sound or sonic event signifies over time, considering sociocultural life and excluding our contemporary ideals or perceptions if we are attempting to place ourselves in the past. We can use soundscapes as a resource and instrument to expand the potential for making historical connections, recalling moments, and imagining spaces.

6 Reflections

This chapter considers how the transformational periods of the Covid-19 pandemic lockdowns might be better understood by listening to and analysing field recordings and soundscape compositions made during these times. By comparing audio recordings made at different points in time through varying lockdown restrictions, we can begin to sonically depict the dynamic shifts in urban spaces caused by lockdowns. This sonic comparison reveals changes in environmental sound markers, acoustics, and social sounds.

The analysis attempts to combine concepts of the sonic image, sound–motion objects, and mental presence to consider the context of environmental sounds and their relationship to the listener. In addition to visualising sounds for historical learning with sonic images and sensory consumption, it is essential to consider the contextualisation of sonic events from a period.

Each city's climate, pedestrian and transportation accessibility, social interaction, and designed spaces are unique. During pandemic lockdowns, the ability to listen to the present sonic environment and identify the changes in social life is possible through considering sound–motion objects and heightening our experience through methods of mental presence. Capturing audio in these urban locations marks a specific period in modern times, and creating a range of lockdown soundscape compositions enables the act of each person to process the sonic information for sonic image association, a way of imagining these changing periods.

Another way to enhance the visualisation of sonic events would be through soundmaps and soundwalking apps to create experiential learning. The Sounding Covid-19 soundscape compositions are featured across various platforms, such as Uno Noll's Radio Aporee (2021), Josh Kopeček's Echoes Soundwalking App (2020), Pete Stollery's COVID–19 Soundmap (2020), Stuart Fowke's Cities and Memories (2020), and others. This variety allows users/listeners to place themselves within the material's listened/recorded/composed experience, either on-site or online. Combining Godøy's concept of sonic images and these experiential tools can explore Smith's sensory consumption. Creating a sense of presence, visualisation, and a new response to the present-day environment (more pronounced if on-site) can be a way of experiencing points in time with an immersed sense of presence in the space and place where the sounds were recorded.

In future sonic preservation work, there may be potential in capturing, documenting, archiving, and analysing sound in varying spatial audio formats, e.g., ambisonic and binaural recordings. Ambisonic recordings depict a 360-degree perspective of the sonic environment and may contribute to developing a stronger sense of presence and contextual meaning-making. Applying similar strategies from the SSID protocol can enable a larger dataset, incorporating audio, video, and survey responses to formulate an extensive comprehension of soundscape investigations (Mitchell et al. 2020). Such a study can contribute to a deeper understanding of the sonic relationships that stereo recordings are limited to capturing or representing.

The *Sounding Covid-19 Repository* serves as a reflective archive, enabling future investigation of soundscapes shaped by the pandemic. As an aural time capsule, the soundscapes preserve a temporal evolution through pandemic lockdowns and exit strategies. Analysis and reflection upon the archive serve to reconnect the listener with these shifting soundscapes and interrogate the broader socio-cultural transformations that shaped them.

References

Adams, M.D., et al.: Soundwalking as a Methodology for Understanding Soundscapes, vol. 30. Reading, U.K. (2008). http://usir.salford.ac.uk/id/eprint/2461/

Axelsson, Ö., Nilsson, M.E., Berglund, B.: A principal components model of soundscape perception. J. Acoust. Soc. Am. **128**(5), 2836–2846 (2010). https://doi.org/10.1121/1.3493436

Belgiojoso, R.: Constructing Urban Space with Sounds and Music. Ashgate Publishing Limited (2014)

Birdsall, C., Drozdzewski, D.: Capturing commemoration: using mobile recordings within memory research. Mobile Media Commun. **6**(2), 266–284 (2018). https://doi.org/10.1177/2050157917730587

Brooks, B.M., Schulte-Fortkamp, B., Voigt, K.S., Case, A.U.: Exploring our sonic environment through soundscape research & theory. Acoust. Today **10**(1), 30–40 (2014). https://doi.org/10.1121/1.4870174

Brown, A.L.: Soundscapes and environmental noise management. Noise Control Eng. J. **58**(5), 493–500 (2010). https://doi.org/10.3397/1.3484178

Cain, R., Jennings, P., Poxon, J.: The development and application of the emotional dimensions of a soundscape. Appl. Acoust. Appl. Soundscapes Recent Adv. Soundscape Res. **74**(2), 232–239 (2013). https://doi.org/10.1016/j.apacoust.2011.11.006

Carras, C.: Soundwalks: an experiential path to new sonic art. Organ. Sound **24**(3), 261–273 (2019). https://doi.org/10.1017/S1355771819000335

Chion, M.: Guide to sound objects. 212 (2009). https://monoskop.org/images/0/01/Chion_Michel_Guide_To_Sound_Objects_Pierre_Schaeffer_and_Musical_Research.pdf

Cusack, P.: Sounds from Dangerous Places. Thornton Heath (Royaume-Uni): ReR Megacorp (2012). https://ualresearchonline.arts.ac.uk/id/eprint/9451/1/Cusack,%20Sonic%20Journalism.pdf

Cuadrado, F., Lopez-Cobo, I., Mateos-Blanco, T., Tajadura-Jiménez, A.: Arousing the sound: a field study on the emotional impact on children of arousing sound design and 3D audio spatialization in an audio story. Front. Psychol. **11** (2020). https://doi.org/10.3389/fpsyg.2020.00737

Cities and Memory. #StayHomeSounds: Sounds from the Global Covid-19 Lockdown (2020). https://citiesandmemory.com/covid19-sounds/. Accessed 12 Feb 2021

Demers, J.: Field recording, sound art and objecthood. Organ. Sound **14**(01), 39–45 (2009). https://doi.org/10.1017/S1355771809000065

Drever, J.L.: Soundscape composition: listening to context and contingency. In: Filimowicz, M. (eds.) Foundations in Sound Design for Linear Media, pp. 358–379. Routledge, New York (2019)

Drever, J.L.: Soundwalking in the city: a socio-spatio-temporal sound practice – CORE. In: 5th International Symposium on Temporal Design. Sheffield, United Kingdom 21–22 July 2011 (2011). https://eprints.gold.ac.uk/id/eprint/7835/

Dumyahn, S.L., Pijanowski, B.C.: Soundscape conservation. Landscape Ecol. **26**(9), 1327 (2011). https://doi.org/10.1007/s10980-011-9635-x

Feld, S.: Waterfalls of Song: An Acoustemology of Place – Resounding in Bosavi, Papua New Guinea. In: Feld, S., Basso, K.H. (eds.) Senses of place, pp. 91–135. School of American Research Press (1996)

Freeman, J., DiSalvo, C., Nitsche, M., Garrett, S.: Soundscape composition and field recording as a platform for collaborative creativity. Organ. Sound **16**(3), 272–2812011. https://www.academia.edu/13672458/Soundscape_Composition_and_Field_Recording_as_a_Platform_for_Collaborative_Creativity

Godøy, R.I.: Thinking sound-motion objects. In: Foundations in Sound Design for Interactive Media. Routledge, Abingdon (2019)

Godøy, R.I.: Images of sonic objects. Organ. Sound **15**(1), 54–62 (2010). https://doi.org/10.1017/S1355771809990264

Gorichanaz, T.: Auto-hermeneutics: a phenomenological approach to information experience. Libr. Inf. Sci. Res. **39**(1), 1–7 (2017). https://doi.org/10.1016/j.lisr.2017.01.001

International Organization for Standardization (2017). ISO/DIS 12913-2: 2017 Acoustics—Soundscape—Part 2: Data Collection and Reporting Requirements. https://www.iso.org/standard/75267.html

Jenison, R.L.: On acoustic information for motion. Ecol. Psychol. **9**(2), 131–151 (1997). https://doi.org/10.1207/s15326969eco0902_2

Jia, Y., Ma, H., Kang, J.: Characteristics and evaluation of urban soundscapes worthy of preservation. J. Environ. Manag. **253**(January), 109722 (2020). https://doi.org/10.1016/j.jenvman.2019.109722

Kang, J.: Urban Sound Environment. CRC Press, London (2014). https://doi.org/10.1201/9781482265613

Kaplan, J.T., Iacoboni, M.: Multimodal action representation in human left ventral premotor cortex. Cogn. Process. **8**(2), 103–113 (2007). https://doi.org/10.1007/s10339-007-0165-z

Kopeček, J.: ECHOES - Geolocated Audio Tours & Experiences (2020). Accessed 9 Aug 2023. https://explore.echoes.xyz/collections/

Lähdeoja, O.: Composing the context: considerations on materially mediated electronic musicianship. Organ. Sound **23**(1), 61–70 (2018). https://doi.org/10.1017/S1355771817000280

Lebiedowska, B.: Acoustic background and transport noise in urbanised areas: a note on the relative classification of the city soundscape. Transp. Res. Part D: Transp. Environ. **10**(4), 341–345 (2005). https://doi.org/10.1016/j.trd.2005.03.001

Léobon, A.: La qualification des ambiances sonores urbaines. Nat. Sci. Sociétés **3**(1), 26–41 (1995). https://doi.org/10.1051/nss/19950301026

Liu, F., Kang, J.: A grounded theory approach to the subjective understanding of urban soundscape in sheffield. Cities **50**(February), 28–39 (2016). https://doi.org/10.1016/j.cities.2015.08.002

Mitchell, A., et al.: The soundscape indices (SSID) protocol: a method for urban soundscape surveys—questionnaires with acoustical and contextual information. Appl. Sci. **10**(7), 2397 (2020). https://doi.org/10.3390/app10072397

Ouzounian, G.: Rethinking acoustic ecology: sound art and environment. Evental Aesthet. **6**(1), 4–23 (2017). https://eventalaesthetics.net/vol-6-no-1-2017-sound-art-environment/

Paquette, D., McCartney, A.: Soundwalking and the bodily exploration of places. Can. J. Commun. **37**(1) (2012). https://www.academia.edu/32681560/Soundwalking_and_the_Bodily_Exploration_of_Places

Parmar, R.: Listening to Places: exercises towards environmental composition. Void Gallery (2022)

Radicchi, A., et al.: Sound and the healthy city. Cities Health **5**(1–2), 1–13 (2021). https://doi.org/10.1080/23748834.2020.1821980

Radio Aporee: Maps - Sounds of the World (2021). Accessed 20 Nov 2022. https://aporee.org/maps/

Rehan, R.M.: The phonic identity of the city urban soundscape for sustainable spaces. HBRC J. **12**(3), 337–349 (2016). https://doi.org/10.1016/j.hbrcj.2014.12.005

Sarwono, J., Sudarsono, A.S., Hapsari, A., Salim, H., Tassia, R.D.: The implementation of soundscape composition to identify the ideal soundscape for various activities. J. Eng. Technol. Sci. **54**(1) (2022). https://doi.org/10.5614/j.eng.technol.sci.2022.54.1.7

Setti, W., Engel, I.A.-M., Cuturi, L.F., Gori, M., Picinali, L.: The audio-corsi: an acoustic virtual reality-based technological solution for evaluating audio-spatial memory abilities. J. Multimodal User Interfaces **16**(2), 207–218 (2022). https://doi.org/10.1007/s12193-021-00383-x

Schaeffer, P.: Traité Des Objets Musicaux : Essai Interdisciplines. Nouv. éd. Editions du Seuil (1966)

Smith, M.M.: Producing sense, comsuming sense, making sense: perils and prospects for sensory history. J. Soc. Hist. **40**(4), 841–858 (2007). https://doi.org/10.1353/jsh.2007.0116

Stollery, P.J.: COVID-19 Sound Map. Sound Scotland (2020). Accessed 23 Feb 2021. https://sound-scotland.co.uk/news/covid-19-sound-map.

Swain, E.: Oral history in the archives: its documentary role in the twenty-first century. Am. Arch. **66**(1), 139–158 (2003). https://doi.org/10.17723/aarc.66.1.9284q6r604858h40

Truax, B.: Speech, music, soundscape and listening: interdisciplinary explorations. Interdisc. Sci. Rev. **47**(2), 279–293 (2022). https://doi.org/10.1080/03080188.2022.2035103

Truax, B.: Genres and techniques of soundscape composition as developed at simon fraser university. Organ. Sound **7**(1), 5–14 (2002). https://doi.org/10.1017/S1355771802001024

Truax, B.: Acoustic Communication, 2 edn. Ablex, Westport (2001). https://encore.qub.ac.uk/iii/encore//record/C__Rb1645991?lang=eng&suite=def

Udsen, A.-S., Halskov, K.: Soundscape design for historical buildings as a sonic place-making process. In: DRS Biennial Conference Series (2022). https://dl.designresearchsociety.org/drs-conference-papers/drs2022/researchpapers/12

Varoutsos, G.: Sounding Covid-19 Repository. Zenodo (2023). https://doi.org/10.5281/zenodo.8245035

Westerkamp, H.: Linking soundscape composition and acoustic ecology. Organ. Sound **7**(1), 51–56 (2002). https://doi.org/10.1017/S1355771802001085

Westerkamp, H.: Soundscape Composition : linking inner and outer worlds, vol. 4 (1999). https://locusonus.org/dropbox/coordination/txts_docs/divers_phonographie_soundscapes/Westercamp/Soundscape%20Composition%20_%20Linking%20Inner%20and%20Outer%20Worlds.pdf

Western, T.: Field recording and the production of place. Critical Approaches to the Production of Music and Sound (2018). https://www.academia.edu/41837883/Field_Recording_and_the_Production_of_Place

Wittmann, M.: Moments in time. Front. Integrat. Neurosci. **5**(October), 66 (2011). https://doi.org/10.3389/fnint.2011.00066. Accessed 8 Aug 2022

Yang, M., Kang, J.: Psychoacoustical evaluation of natural and urban sounds in soundscapes. J. Acoust. Soc. Am. **134**(1), 840–851 (2013). https://doi.org/10.1121/1.4807800

Yong Jeon, J., Young Hong, J., Jik Lee, P.: Soundwalk approach to identify urban soundscapes individually. J. Acoust. Soc. Am. **134**(1), 803–812 (2013). https://doi.org/10.1121/1.4807801

Excitations and Resonances: Misinterpreted Actions in Neon Meditations

Risto Holopainen(✉)

Gråbrødreveien 10, 0377 Oslo, Norway
ebeling@ristoid.net

Abstract. Neon Meditations is a collaborative performance work combining visual art and music, where colours are translated into sound in an electronic instrument controlled by two performers. The sound design follows the principle of excitation and resonance. We use exciters attached to resonating objects that colour and distort the sound. The mapping from gesture to sound, and the fact that this is a multi-agent system, tends to cause confusion about the way the performers shape the sound. Godøy's concept of sound–motion objects is well adapted to acoustic instrumental music, but using Neon Meditations as an example, we will see that it faces many challenges when one tries to extend its application to live electronic music.

Keywords: Live-electronic music · modular synthesizer · sound–motion objects · acoustic processing

1 Introduction

Neon Meditations is an ongoing collaboration with visual artist Per Hess. It is an improvised piece, a recurring performance, and a crossing point between music and visual art. Its background is a curiosity about colour, sound, and their possible relationship. We read the colours of neon light with colour sensors and make music with a modular synthesizer, which both performers control simultaneously in a way that easily causes confusion about action and sound.

The sound design can be described in simple terms as excitation and resonance, or source and filter: the excitatory signals from the analogue modular synthesizer are distributed to exciters attached to vibrating objects which colour and distort the sound. Neon Meditations may be an inconvenient example to illuminate Godøy's concept of sound–motion objects and the motor theory perspective, since these ideas were developed primarily with acoustic music in mind. Nevertheless, it may be revealing to reconsider sound–motion objects from the perspective of live-electronic music.

First, I will describe Neon Meditations from a technical and aesthetic point of view, with a focus on sound, and finally, I will discuss some topics from Godøy's research on sound-related motion as it applies to live-electronic music in general, and illustrate it with the example of Neon Meditations.

A. R. Jensenius (Ed.): SSD 2022, CRSM 12, pp. 211–221, 2024.
https://doi.org/10.1007/978-3-031-57892-2_12

2 Colour and Sound

Per Hess has been particularly occupied with colour in his work as a visual artist. When we first met at one of his exhibitions in 2017, we began discussing a possible collaboration where I would contribute music or sound. As we soon realised, it is by no means obvious how to combine static visual art with music as a form unfolding in time, at least if one is to avoid making one part illustrative and subservient to the other. In Scriabin's landmark work Prometheus, a colour organ dynamically lights the space in different colours following the music. The simultaneous use of coloured light and music as mood-inducing devices has become ubiquitous in concerts and movies, to the point of being barely noticed. Our approach is seemingly less spectacular by the restriction to constant colours, which are unaffected by the music.

Hess has produced a series of neon tubes segmented into fields of different colours. Neon by itself glows with an orange-red, and various fluorescent pigments inside the glass tubes bring out a range of different colours. In fact, what many artists casually refer to as "neon" might also include argon, which produces a bluish colour. Eventually, we decided to use sensors to read colours and control an analogue modular synthesizer with these signals. By doing so, we have turned the neon tubes into a visually appealing part of a musical instrument, although they still can be exhibited on their own.

A long period of experimentation and practice preceded our premier performance. Technical problems had to be solved, such as constructing colour sensors and designing a patch on the modular synthesizer, but we also faced aesthetical choices, such as how to map colour to sound. Actually, "mapping" is a misleading word, since we ended up with a rather complex relation instead of a simple one-to-one correspondence between colour and sound.

The notion of synaesthesia tends to come up in discussions about colour and sound. In the strong form, hearing a sound may induce the visual impression of a particular colour (so-called photisms), or vice versa, in completely idiosyncratic ways. However, there is evidence of more widespread forms of synaesthesia (Marks 1975). Temperature is associated with colours when we describe red and yellow as "warm" or blue and green as "cold." Bright colours are regularly associated with a high pitch and dark colours with a low pitch. Vowels are sometimes associated with colours; in particular, the second formant frequency, and the spread between the first and second formants, appear to be related to colour. According to Marks, synaesthesia is a cross-modal manifestation of meaning in a purely sensory form and is not fundamentally different from non-synaesthetic cross-modal meaning or even abstract verbal meaning.

It may be revealing to consider the various visuo-auditory correspondences from their temporal aspect. Colour doesn't "happen" in time; we typically experience it statically. Hence, for its auditory correlate, we should expect something that can be extended over time. Gestures are completely different; with the closure of their beginning, trajectory, and end, they can be related to musical Gestalts such as tones or short phrases, or musical objects in Pierre Schaeffer's sense (Godøy 2018).

It is commonplace in contemporary art to explain what the work is about. If a trite verbal description exhausts its meaning, one may wonder, what is the point of creating the work? Usually, there is a remainder that resists explanation. Artistic research, in the sense of solving problems posed by the realisation of the artwork, is an intrinsic part of artistic creation. There is also a narrower sense of the term, related to how artistic projects have been adopted in an academic setting with expectations to produce knowledge. However, such knowledge may be specific to the project, highly subjective, or difficult to generalise and share. Peter Osborne has rightly identified some other perils of the academic form of artistic research. Rather than producing critically significant works, it may lead to a new kind of academic art (Osborne 2021). Posing and solving research questions does not necessarily contribute to attaining artistic goals. In Neon Meditations, it is fair to say that the work concerns translating colour into sound. But by no means is it a pedagogical illustration of particular synaesthetic colour-to-sound correspondences, a topic better left to researchers in cognitive psychology. Nor have we been particularly occupied with the relation between gestures and sound. Nonetheless, it turns out that Godøy's research on sound-related motion can shed some additional light on our performance—and perhaps also the other way around.

3 The Patch

To begin with, we had to construct colour sensors, which we did using the simplest means. We use two hand-held colour sensors (Fig. 1), which can point independently in different directions. This design decision already eliminates any straightforward one-to-one correspondence between colour and sound. Instead, there would be a mapping from colour pairs to sound, except that both performers influence the sound.

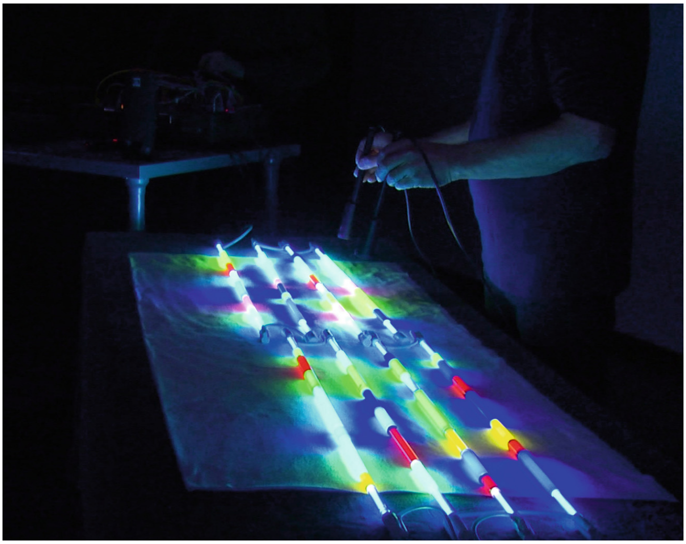


Fig. 1. Neon lights and sensors. Video still, reproduced by permission of Guro Berger.

Each colour sensor receives a constant voltage which passes through a photoresistor. In front of the photoresistor, there are colour filters made of plastic film, which ensure that the sensors actually register hue and not only brightness. The varying control voltage

(CV) is returned to the modular synthesizer in a rather complex patch. First, there is some processing of the CV signals from the colour sensors. Both sensor signals are sent to a comparator which outputs a gate signal when one voltage is higher than the other. The sensor signals also directly control filter cutoff, oscillator frequencies, and amplitude. In parallel, the other performer controls the modular synthesizer through knobs, faders, and an expression pedal. The audio path includes two oscillators, VCAs, filters, and a few other modules with complex cross-modulation. The output is sent to an external mixer via an amplifier, to exciters attached to vibrating objects.

Over the years we have performed the piece, the rough structure of the patch has remained the same, but some modules have been exchanged, causing necessary adjustments. There is no score in a traditional sense, but there is documentation of the patch that I try to follow, which gives the work a certain cohesion and makes it recognisable through its successive iterations.

Two things are crucial to notice here. First, two performers control the same instrument, even with the same parameters. It is a multi-agent system, with predecessors such as the network ensemble The League of Automatic Composers, who connected microcomputers in a network around a kitchen table in the 1970s, even before the internet and the MIDI protocol became available (Rohrhuber 2007). Second, left to its own devices, without our active control, the patch would only produce a monotonous drone. With our aid, it produces a somewhat variable drone.

4 Sound Design

Arguably, sound design may cover the entire process from composition to interpretation. Sometimes it makes sense to distinguish sound design from composition proper. This may be the case if the composition is conceived as an abstract structure, represented by a score or other symbolic information, which is given concrete flesh in an interpretation or realisation. Indeed, the term *musique concrète* itself reflects this focus on the actual sounding music (Schaeffer 1966; Godøy 2021a).

Some of my electroacoustic pieces have been created first as a timbrally crude sketch that I have submitted to *acoustic processing* before mixing a final version (Holopainen 2021a). Acoustic processing here refers to using real acoustic spaces and vibrating objects to colour the sound. I would play sound files through loudspeakers or exciters attached to the resonating bodies of instruments such as guitars, drums, or other objects with a prominent acoustic character and then record it again. This technique is also used in Neon Meditations. Since it is a live performance, the exciters and vibrating bodies are included onstage (Fig. 2).

Schematically, the sound design of Neon Meditations can be considered as separated into excitations and resonances. Excitations can be understood as related to what we do and mental images of actions, whereas resonances can be related to the effects of what we do and images of materials (Godøy 2001). Some acoustic instruments have feedback from resonances to excitation, which complicates their modelling as separate stages, but in our case, the separation is justified.

In the patch for Neon Meditations, excitations come from spectrally rich sawtooth waveforms and white noise, fed into analogue filters with variable cutoff frequencies.

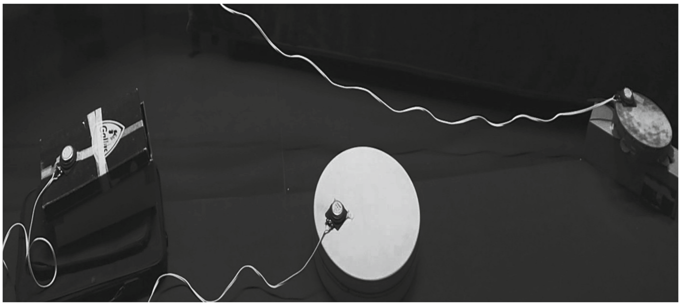


Fig. 2. Exciters and resonators. Video still, reproduced by permission of Guro Berger.

The filters provide a first stage of resonances, but this signal acts like an excitation in the following stage of acoustic processing. The output from the modular synthesizer is fed to the exciters, usually attached to a frame drum, a tambourine, and cardboard boxes (other resonators may be used if available). These resonating bodies strongly colour the sound, and at loud levels, they also distort it by adding rattling or buzzing sounds of their own. By adjusting the position of the exciters on the object, various vibratory modes can be emphasised. The amount of extraneous rattling depends not only on the volume and frequency; it can also be controlled to some degree by how tightly the exciter is fastened to the vibrating object. The use of physical vibrating systems could be taken much further. For example, Daniel Wilson has built feedback systems around exciters and resonators with contact microphones in what he calls a post-electronic modus operandi (Wilson 2012).

The technical and aesthetic ideal of hi-fi treats the loudspeaker as a transparent window into an imaginary world. We are not supposed to notice the presence of the loudspeaker but focus on the music. Michel Chion recalls how cinema sound was once characterised by a cavernous resonance and a wavering sound caused by the uneven speed of the film projector (Chion 1994, p. 99). Modern movie theatres have solved these problems; a powerful deep bass can be produced with little distortion. At home, such deep bass tones would make the furniture or dishes shake. Chion also makes a useful distinction between *fidelity* and *definition* (ibid, p. 98). Fidelity is more of a selling argument than a verifiable notion; it would require hard to arrange comparisons between the original and the reproduction. Definition is a more technical and precise term. High-definition sound covers a broad frequency range, particularly high frequencies that can transmit a sense of acuity and presence, as well as a large dynamic range. In sound distribution by exciters on sounding bodies, the distinction seems apt: clearly, it distorts the sound too much to be considered hi-fi, but on the other hand, the added rattling noise that extends the high-frequency range with a shimmering provides high definition. The exciters and their associated resonators are also point sources, well localised in the spatial field, in contrast to the diffuseness of stereo panning achievable with an ordinary pair of loudspeakers.

In Neon Meditations, the exciters and vibrating objects replace transparent loudspeakers. Their presence onstage is noticeable, both visually and soundwise. Acousmatic listening, another of Schaeffer's famous notions (Schaeffer 1966, ch. 1), refers to any listening situation where we cannot see the source, such as sound played over loudspeakers. Supposedly, it helps the listener focus on the sound as such and be less preoccupied with the causality of the sound source. With the stage presence of rattling tambourines and buzzing shoe boxes, it is far from certain that acousmatic listening is an adequate term, since these objects partly become sound sources of their own while also transmitting sound.

Neon Meditations may be described tentatively as a drone piece, although perhaps not typical of that genre. Our performances typically last about 20 min and consist mostly of a sustained, wavering sound. In Schaeffer's typological classification, it would be called a *thread* ("Trame" in French). The thread type can be encountered not only in natural environments such as waterfalls, but is also common in orchestral music (Chion 1983, p. 134), typically as a background texture. In drone pieces, what otherwise serve as background elements are brought into the foreground.

Drone pieces, in general, are characterised by a total lack of development or any sense of large-scale form. There may be a high density of micro-events or textural and timbral articulations, but very little happens at the phrase level. Many contemporary music genres, including drone pieces, pose certain challenges to the listener's perception and memory (Wanke and Santarcangelo 2021). Retention and protention are intentionally put to the test in these pieces, which have also been described as engaging in a form of"memory sabotage."

Schaeffer's most basic classification of sound objects divides them into three categories: Sustained, iterated, and impulses. Schaeffer also has a category of *objets convenables*, or suitable sound objects, which are of medium length and easy to memorise. Given the drone character of Neon Meditations, we tend to stay away from the *objet convenable* category. For Schaeffer, there was a normative aspect to this category, these sound objects were deemed suitable for music.

5 Interaction

If we ever veer away from the sustained sound type in Neon Meditations, it is by brief passages of iterated sound objects that may become sufficiently separated to be perceived as impulses; it is done simply by turning the frequency knob of an oscillator down to the sub-audio range. As Godøy points out, there are phase transitions between different kinds of bodily motion (Godøy 2021b), directly corresponding to the sustained, iterative, and impulsive types of sound objects. With an electronic instrument, these transitions are producible in one smooth movement, simply by turning a knob, although the resulting sequence of sustained pitched tones, iterations, and separate impulsive sounds remain perceptually distinct categories.

Physiological constraints make it impossible to increase the rate of a bow tremolo to audio frequencies. Electronic instruments do not share our physiological limitations, but should they simulate them?

Here it is important not to confound what *is* and what *ought* to be. Granted, we have distinct perceptual and motoric regions of sustained, iterative, and impulsive types, and electronic instruments afford a seamless transition across the range. From these two facts, someone might suggest that electronic instruments ought to be designed such that these distinct perceptual types can only be produced by dedicated types of gestures, as they are in acoustic instruments. Someone else might instead celebrate the fact that electronic instruments afford a new, non-natural connection between gesture and sound. I believe both views have their merits.

However, there is another aspect of live electronic interaction that needs to be considered. Although much of the writing on the topic specifically addresses interactive computer music, much of it also applies to analogue or hybrid electronic instruments such as modular synthesizers. In any case, interactivity consists of relegating certain tasks to the machine and letting the performer play a role that can be described as that of a supervisor, pilot, or collaborator. In my experience, the interfaces that allow for the most expressive performances are those that permit a detailed control of all aspects of sound and relegate as little as possible to automation. In Sergi Jordà's words:"A good instrument should also be able to produce 'terribly bad' music, either at the player's will or at the player's misuse" (Jordà 2007, p. 104). Such instruments require more practice to master and allow for bad performances, but that is precisely the point.

Virtuoso instruments don't correct the performer's mistakes. In Neon Meditations, there is another reason for not granting the machine too much autonomy. Instruments or systems for generative music can be designed to create a stream of varied output with little to no input, a goal that has been pursued in various media involving feedback, so-called interfaces for self-organising music (Kollias 2018), and in more algorithmic approaches using monolithic systems that merge sound synthesis and slower processes (Holopainen 2021b). In modular synthesizers, self-generative patches can produce endless musical variation with no input. However, in Neon Meditations, where the point is to translate colour readings into sound, such an additional layer would unnecessarily obscure an already complicated gesture-to-sound relation. Furthermore, the multi-agent nature of our system creates a complexity of interaction comparable to what can be achieved by sophisticated interactive digital or analogue computer systems.

6 The Motor Theory Perspective

Godøy (2018) lists four types of music-related motion that may be expected in performances of instrumental music:

- *Excitatory motion*: transfer of energy from musician to instrument
- *Modulatory motion*: dynamically changing pitch, timbre, loudness
- *Ancillary motion*: avoiding strain, etc.
- *Communicative motion*: between performers or toward an audience

Examples of excitatory motion include blowing air into a wind instrument, plucking or bowing a string, or tapping a drum membrane. In live electronic music, where the sound production is already taken care of, it is still possible to simulate excitatory motion, as is commonly done on keyboard instruments where depressing a key produces a sound. In Neon Meditations, on the other hand, we do not even try to simulate such correspondences; it is all about modulatory motion. Both performers modulate timbre, pitch, and loudness. Ancillary actions do not produce sound, at least not on purpose, but are more or less necessary to accommodate the playing. Our performance requires rather static postures and mostly looking down on the instrument. Therefore our communicative motion is minimal while we are engaged in the performance.

In an acousmatic listening situation, when we listen to a recording of musicians, we see none of their motions. Nevertheless, the first two categories of music-related motion are more directly involved in producing and shaping the sound we hear than the last two. We might infer the excitatory and modulatory actions from listening only; at least, we might imagine probable sound-producing actions (see Godøy 2001). However, we are unlikely to guess all sorts of ancillary or communicative motions the musicians were making before the microphone in the studio.

7 Sound-Motion Objects in Live-Electronics

Godøy describes sound–motion objects as multimodal, including sound and corresponding body motion; they typically occur on medium time scales of 0.3–5 s; and they may involve complex motor schemata such as complicated, rapid passages which have to be practised before being performed automatically without conscious control (Godøy 2021b). It is no coincidence that their time range corresponds to Schaeffer's *objet convenable.*

The theory of sound-motion objects suggests that we tend to imagine a plausible physical motion, often a body motion, corresponding to sounds we hear. This is practically unquestionable in the case of singing and acoustic instrumental music but becomes more conjectural in electroacoustic music with less immediate connections between sound production and perception.

In live electronic music, the gesture-to-sound relation may become confused, depending on the mapping from controller to sound production. Some actions correlate with sounds, but there may be ancillary motion with no causal relation to the sound. For an audience without expert knowledge about the controllers and mappings used in the performance, it may be impossible to distinguish ancillary motions from those that modulate or trigger sounds. The antennae of a theremin controller make no distinction between motions with modulatory, communicative, or ancillary intention. If you move at all sufficiently close to the antennae, they register it. Furthermore, some sounds in live electronic music may not correspond directly to performance actions, such as automated sequences or pre-recorded parts that only need to be started and perhaps stopped. In Neon Meditations, two performers at once influence the sound. It is quite unlike playing the piano with four hands, where each pianist knows which part they are playing; in our performance, we may both control filter cutoff or oscillator frequencies, and the audience is likely to have trouble deducing which actions are responsible for the timbral changes that result. Indeed, we also found this confusing at first and had to spend time practising before our first performance.

According to the motor theory perspective, sounds correspond to imagined actions (Godøy 2001). This may hold even in electroacoustic music since we can imagine whatever we want. But it also happens that sounds in live electronic music contradict what is seen. Usually, after a performance, we engage the audience in a dialogue and answer their questions. We have had audience members compare our sound to a car or motorbike. The colour sensors are sometimes mistaken for microphones, which they admittedly resemble. Some audience members, therefore, speculate that they pick up sound directly from the neon tubes, which they do not. The point is not to call out the audience for not getting what we are doing. On the contrary, the active search for causal links may contribute to making the performance an engaging experience.

8 Assessment of the Sound-Motion Object

The notion of sound-motion objects is rooted in the praxis of acoustic instrumental music, which explains some of its biases; it has a certain focus and possibly a few blind spots or limits to its applicability. I will briefly summarise those that come to my mind.

1. Acoustic instruments have been with us for a considerable time. We interact with them as we interact with the rest of our physical surroundings; excitations and resonances inform us about forces that set objects into vibration and about the material properties of these objects. Grounding the theory in ecological perception gives it a general, broad validity.
2. Sound-motion objects, like Schaeffer's *objet convenable*, emphasise the medium duration range of about 0.3–5 s. This makes drone pieces inconvenient examples to illustrate the idea. The focus on gestures and sound–motion objects downplays processes over longer time spans. On the other hand, sound–motion objects are situated at a level above the intermodal concept of texture. Spatial textures of various coarseness should be easy to imagine as timbral qualities of varying roughness. This touch-to-sound correspondence does not seem to necessarily involve motion.
3. Sound–motion objects take acoustic instrumental music as their model; the concept is therefore not a priori equally relevant for electronic music. The freedom to introduce arbitrary mappings from controllers to synthesis parameters may destroy the unity of perception and performance, which can be taken for granted in acoustic music.
4. Even the concept of coarticulation, which is best known from phonetics but is also a reality in vocal and instrumental music, must be reconsidered in live electronic music. Coarticulation involves the fusion of otherwise distinct motions, and prepared actions, such as performers placing their fingers in the correct position on an instrument before playing (Godøy 2021b). This has certain consequences for sound production in vocal and instrumental music. In live electronic music, the role of coarticulation in shaping the performance may be much less important, or at least very different, depending on the specifics of the mappings and interfaces used.
5. The theory of sound-motion objects does have interesting things to say about virtuosity, idiomatic writing and playing (Godøy 2018), but it seems almost overqualified, yet not quite to the point when it comes to motorically less challenging improvised live-electronic performances. Live coding is perhaps the most striking example, where mental effort largely supplants bodily effort; the typing motions, although also involving motor skills, have a most indirect relation to the sound.
6. A single performer is implicitly assumed responsible for the sound production, not two or several performers as in multi-agent systems, nor a hybrid combination of performer and algorithms or other kinds of automation (As Godøy reminded us during the 2022 seminar, the mechanical organ originally needed two performers, one of whom was treading the bellows. The bagpipe also apparently frees the performer

from excitatory motion while playing so that only modulatory motion is required. To these examples, one might add wind chimes, a mechanical instrument that requires no human performer.)

7. The focus is on low-level perception and the physicality or physiology of sound production. This choice of focus is understandable as a complement to, or reaction against, a previously prevailing overly abstract and"disembodied" flavour of music theory. As always, focusing on something is fine as long as it does not replace an old myopia with a new one or pushes other questions worth asking and methods worth pursuing into the background, such as sociological, historical, and aesthetic perspectives on music.

In summary, the concept of sound–motion objects most aptly deals with acoustic instrumental and vocal music. The motor theory perspective offers the plausible view that we experience motion and sound as interconnected, almost synaesthetic aspects of a coherent phenomenon. This is most obvious in acoustic music and may still, to some degree, be true when listening to electronic music, where distinct sound types may be associated with suitable imagined sound-producing actions. In live-electronic music with its arbitrary mappings from gestural controllers to sound, on the other hand, it is a matter of artistic choice whether the motion–sound correspondence should be upset and quite illogical, or follow our expectations by simulating the functioning of acoustic instruments. Maybe live-electronic music deserves its own addendum to the theory of sound-motion objects, wherein we distinguish between the motion we would typically imagine as we hear the music, and the actual, arbitrary mappings from gesture to sound. What complicates it is that these two levels are superimposed and may provide mutually conflicting cues.

As for Neon Meditations, the project has turned out to be surprisingly long-lived. We are less preoccupied with developing the performance than maintaining it and adapting it to new circumstances. For each new performance, we solve the practical matters of sound design in slightly different ways.

Acknowledgements. I wish to thank Per Hess for his support, the reviewers for their helpful comments, and not least, Rolf Inge Godøy, my supervisor for many years, for having been generous with his time and encouragement.

References

Chion, M.: Guide des Objets Sonores. Pierre Schaeffer et la Recherche Musicale. Éditions Buchet/Chastel, Paris (1983)

Chion, M.: Audiovision. Sound on Screen. Columbia University Press (1994)

Godøy, R.I.: Imagined action, excitation, and resonance. In: Godøy, I., Jørgensen, H. (eds.) Musical Imagery. Swets & Zeitlinger (2001)

Godøy, R.I.: Motor constraints shaping musical experience. Music Theory Online **24**(3) (2018)

Godøy, R.I.: Perceiving Sound Objects in the Musique Concrète. Front. Psychol. **12**, 672949 (2021a)

Godøy, R.I.: Constraint-based sound-motion objects in music performance. Front. Psychol. **12**, 732729 (2021b)

Holopainen, R.: Making complex music with simple algorithms, is it even possible?". Revista Vórtex **9**(2), 1–13 (2021a)

Holopainen, R.: Algorithmic Composition by Autonomous Systems with Multiple Time-Scales. Divergence Press (2021b)

Jordà, S.: Interactivity and live computer music. In: Collins, N., d'Escriván, J. (eds.) The Cambridge Companion to Electronic Music, vol. 5, pp. 89–106. Cambridge University Press, Cambridge (2007)

Kollias, P.: Overviewing a field of self-organising music interfaces: autonomous, distributed, environmentally aware, feedback systems. In: ACM Conference on Intelligent User Interfaces, Intelligent Music Interfaces for Listening and Creation, Tokyo, Japan (2018)

Marks, L.: On colored-hearing synesthesia: cross-modal translations of sensory dimensions. Psychol. Bull. **82**(3), 303–331 (1975)

Osborne, P.: No going back – but not forward to there either: once more on art and/as research. In: Slager, H. (ed.) The Postresearch Condition, pp. 5–13. Metropolis M Books, Utrecht (2021)

Rohrhuber, J.: Network music. In: Collins, N., d'Escriván, J. (eds.) The Cambridge Companion to Electronic Music, vol. 8, pp. 140–155. Cambridge University Press, Cambridge (2007)

Schaeffer, P.: Traité des objets musicaux. Éditions du Seuil (1966)

Wanke, R., Santarcangelo, V.: Memory as the aspatial domain for the perception of certain genres of contemporary art music. Music Sci. **4**, 1–18 (2021)

Wilson, D.: Miraculous Agitations: on the uses of chaotic, non-linear and emergent behavior in acoustic vibrating physical systems. Leonardo Music J. **22**, 35–40 (2012)

Technological Innovation

Aspects of Sound Structure in Historic Organs of Europe

Albrecht Schneider(✉)

University of Hamburg, Hamburg, Germany
`aschneid@uni-hamburg.de`

Abstract. This article sketches the development of pipe organs in Europe from Roman times to the Baroque era in order to shed light on concepts of sound structure that apparently guided the design and construction of certain types of organs. For a better understanding of empirical measurements presented in Parts 4 and 5 of this study, Part 2 provides some basic organology. In Part 3, the development from the so-called 'Blockwerk' to organs comprising several wind chests, manuals, and various types of pipes and stops is outlined. In part 5, relations of sound structure to tunings and temperaments are discussed, including actual measurements.

1 Introduction

One of the most ancient and widespread musical instruments in Europe is the organ, as it is used in churches and concert halls as well as in private settings. A few years ago, UNESCO included the organ and organ music in its list of world cultural heritage. The historical record for organs in Europe from the Middle Ages to the Renaissance and Baroque era and also in modern times is very broad and has been studied in great detail, with a focus on archival source material as well as the development of organ building and organ music. Thereby, a number of historical and regional organ types have been identified, and major stages in the development of pipe organs and organ music have been outlined (for comprehensive surveys, see Williams 1966, Klotz 1975, William & Owens 1984, Eberlein 2011). Works on organ building provide information on technical aspects such as the mensuration of pipes and their peculiar geometry in relation to sound generation and the organisation of various pipe ranks within the overall structure of organs (e.g., Adelung 1976). Pipe organs are known for their complex mechanical construction as well as for the amazing variety of 'sound colours' they can produce from pipe ranks of different designs and manufacture. Beginning in the 1930s, characteristics of sound recorded from historical organs have been investigated (e.g., Trendelenburg, Thienhaus & Franz 1936, 1938, Lottermoser 1940, 1983a/b). However, early recordings made for documentation and empirical research on sound properties included only a small number of extant instruments while organs built in the 17th and 18th centuries, respectively, have been used quite frequently for recordings of music (mainly of the Renaissance and Baroque era) within the context of historical performance practice, which involves 'original instruments' of a given period. A milestone, in this respect, was the recordings the organist Helmut Walcha made of Johann Sebastian Bach's complete

A. R. Jensenius (Ed.): SSD 2022, CRSM 12, pp. 225–258, 2024.
https://doi.org/10.1007/978-3-031-57892-2_13

works for organ, where he played instruments of the famous organ builders Friedrich Stellwagen, Arp Schnitger, and Andreas Silbermann. These recordings began in 1947 at Lübeck, where the organ of Friedrich Stellwagen from 1637 was reinstalled in the St. Jakobi church after WW II (see Wölfel 1980, 53ff.). Walcha's recordings of Bach were issued by the 'Archiv Produktion', a special label of the Deutsche Grammophon devoted to 'scientifically documented' recordings of works from various periods of music history. A re-issue of Walcha's recordings of Bach's works for organ appeared in 2000 (Helmut Walcha: J.S. Bach. The Organ works, 12 CDs. Archiv Produktion 2000). Such recordings honoured a number of historical organs in Europe that were esteemed for their excellent craftsmanship and sound quality, which saved them from destruction and removal. Still, quite many historical organs from the Baroque era were badly damaged, to the point of losing original wind chests, part of the mechanical action as well as a significant number of pipe ranks, in the second half of the 19th and the first decades of the 20th century; from various instruments only the organ cases remained (into which a new organ manufactured around the period 1850–1910 was inserted). The reason to abandon old organs was not so much their need for maintenance or repair but a change of concepts when organ builders and organists alike opted for a more 'modern' sound, whereby pipe stops in organs were often designed to emulate orchestral timbres (in particular, strings). Also, devices suited to vary the dynamics of sound were installed in many new organs. This included the 'swell case,' which can be opened or closed in quasi-continuous motion so as to decrease or increase the SPL [dB] of sound radiated into the ambience. Another example is the so-called 'crescendo roller,' a device suited to activate stops in successive order, thereby increasing SPL and spectral density stepwise. Against such 'progressive' inventions, organs from the Baroque era were often viewed as old-fashioned and unworthy of proper maintenance. It was more or less by chance (and due to the fact that not all villages and small towns could afford to order new instruments) that some of the most outstanding organs, as, for instance, the Schnitger organ of Cappel (1679–80; originally built for the St. Johannis monastery church of Hamburg; see Fock 1974, 33f., Vogel, Lade & Keweloh 1997, Edskes & Vogel 2009) survived nearly untouched.

In the 1920s, a group of organ builders, organ experts, musicians, and musicologists unsatisfied with industrial organ manufacture, pneumatic instead of mechanical action, and contemptuous of the gadgetry in contemporary organs (such as "high pressure" stops), discussed the merits of 'classical' organs (the period from ca. 1600–1770) and proposed a program calling for the restoration of historic organs, possibly to their original state. This movement, which in Germany is known as 'historische Orgelbewegung' (historic organ revival, with similar organisations in the Netherlands, France, Italy, and other countries), had practical relevance in that, in the first stage, a survey of surviving organs and their present state was initiated. From inspection of individual instruments as well as from archival studies, their history with all the previous repairs and modifications became evident. Notwithstanding regrettable losses suffered over two or three centuries, there was still a substantial mass of original parts (organ cases, wind chests, more or less complete pipe ranks, ducts, bellows, keyboards, etc.) in place, in various organs, to allow for restoration and/or reconstruction. Working from a comparative basis (one part missing in a certain organ, fortunately, was preserved in another of the same master or one of his contemporaries and could serve as a model for reconstruction), the process of restoration, especially since ca. 1950, has been a continuous and very effective one. Of course, there were severe problems as organ builders had to understand the manufacture of pipes based on pre-industrial techniques of casting organ alloys (from tin, lead, copper), as well as how intonation of flue and reed pipes was facilitated hundreds of years ago. Along with the reconstruction of original pipe ranks and revoicing of pipes, in many historic organs, the tuning has been changed from equal temperament (ET12) back to one of the systems used around 1700 (like meantone or Werckmeister, see below). Though there is no doubt that all the work invested into the restoration of historic organs has brought back an impressive range of instruments diverse in design and 'sound colours,' the question remains of how close the actual sound might come to sound properties an instrument had when it was first installed centuries ago.

What can be said, with some confidence, is that a restoration (of a painting or other work of art as well as of a monument etc.) is always an attempt at finding a convincing solution on the basis of available evidence. The goal is to preserve as much as possible of the original substance and to reconstruct what is missing using appropriate materials and techniques. In regard to organs, the corpus of data gained from restoration projects over several decades is extensive, and so is the experience of organ builders who specialise in restoration. Since their knowledge and craftmanship have been proven in the technical reconstruction of wind chests and actions, re-adjustment of wind supply and revoicing of pipes etc., their efforts likely revived also sound properties as assumed for the original instrument. Of course, this is a process of approximation since almost all historic organs have undergone modifications, usually in the period ca. 1780–1870, which changed their pitch and tuning. The original pitch, referring to some practical usance like the 'Chorton', was generally considerably higher than our standard $a^1 = 440$ Hz. The tuning system applied to organs for more than 200 years had been one of the variants of meantone temperament (see Lindley 1987, Ratte 1991, Schneider & Beurmann 2017). When equal temperament (ET12) came into use ca. 1780–1820, most organs then were tuned to this system (especially in cities wealthy enough to afford such a process that often included an exchange of pipe ranks since certain old stops were not compatible with ET12, see below). Retuning to ET12 and lowering the pitch level were often combined. All these modifications of the past had to be corrected in a process of careful restoration so that a state close to the original construction and voicing was achieved. The expectation of such proceedings is that also the sound of each pipe rank, and of the organ as a complex unit, will come close to what must have been the 'original sound' ('Originalklang') of a Renaissance or Baroque organ. Since we have no recordings from 1600 or 1700, attempts at finding the 'original sound' are demanding and may, to some degree, remain conjectural (that is, they are based on factual evidence yet include inferences). The task to approximate the 'original sound' is by no means restricted to historic organs but exists, in similar ways, for almost all instruments from past centuries. For example, violins and other string instruments of famous makers such as Stradivari, Guarneri, or Stainer were not left untouched over several centuries but have been subjected to repair and re-adjustment, including replacement of strings as well as of bridges and even necks. For appropriate repair of historical violins, specialists had to study in depth the principles of design of those masters and had to become familiar with the materials (wood, glue, lacquer) they had used. Thus, it was possible to restore such instruments in detail, thereby regaining superior playability and excellent sound quality, which, as a huge number of recordings made with historic instruments amply demonstrates, cannot be too far from the original. From all the evidence available, we may conclude that, by skilful and well-informed restoration, a close approximation to the ‘original sound’ of a historic instrument such as a violin, bass viol, flute, oboe, harpsichord, or organ is possible. In this respect, it seems justified to regard sounds recorded from a Baroque organ fully restored to its original state as authentic. Though we cannot relive the past, we can revive its instruments and concepts of sound.

Research directed to characteristics of the sound properties of historic organs gained new momentum when digital recording and signal processing tools became available on a greater scale in the 1990s. As data for such research, it is mandatory to record sounds from each pipe rank on site since the voicing and intonation of pipes receive a final adjustment in the room into which an organ is placed. Also, it is important to record instruments before and after restoration in order to document the previous sound characteristics and to assess the changes that result from the restoration process (cf. Schneider et al. 2006, Ahrens, Braasch & Schmidt 2006). Sound recorded from pipes mounted on their wind chest can be subjected to signal analysis whereby temporal and spectral features suited to describe sound generation in pipes and timbral quality of peculiar pipe ranks can be studied and documented objectively (see Beurmann, Schneider & Lauer 1998, Schneider, von Busch & Schmidt 2001). In this chapter, we continue and expand previous research, which includes actual organ sound as produced with combinations of pipe stops viewed in relation to tuning and temperament.

In the following section, I shall first address some basics of organology, including terminology, as certain concepts and terms will be needed, in Sect. 4, in conjunction with sound analyses of pipe ranks. In Sect. 3, the development and history of some organ types are briefly reviewed since organs of the Baroque era found in parts of Northern Germany and adjacent regions of the Netherlands, preserved certain features known from older types of organs. As in many cultural phenomena, one can observe the interplay of continuity and change also in organ building.

2 Some Basic Organology

A pipe organ is a wind instrument (aerophone) that consists of a system supplying wind to a chest on which one or several ranks of pipes are mounted (for technical aspects, see Adelung 1976 and Williams & Owen 1984). Pipes are distinguished by their mode of operation into flue and reed pipes. In a pipe organ, pressing a certain key on the keyboard will open a valve whereby air streams from the wind chest into a flue pipe or reed pipe, where the airflow will activate a pulse generator coupled to a resonator. Regular sequences of pulses from the generator elicit periodic vibrations in the air column enclosed in each pipe, which acts as the resonator part of the coupled system. Standing waves will be formed in a cylindrical or conical tube of a given length l if the resonance condition $\omega_e = \omega_r$ is met (ω_e = exciting frequency, ω_r = resonance frequency, for $\omega = 2\pi f$). Standing waves and resonance, in turn, is the condition necessary for the production of harmonic sound that is radiated from the open end of a tube (e.g., a diapason pipe).

In aerophones such as flutes and reed instruments, the generator can be described as a nonlinear oscillator, whereas the tube resonator reacts to excitation in a linear response (within certain operation limits).

The oscillator/generator typically interrupts a continuous stream of air by a valve-like mechanism which, in reeds and horns, opens and shuts in a basically periodic motion controlled by, first of all, the pressure and the speed of the air fed into the oscillator. A valve-operated pulse generator can be formed, in real instruments, for instance, by the two lips of a musician pressed into a mouthpiece (as in trumpets and horns). In wind instruments, a single reed (as in a shawm or clarinet) and double reeds beating against each other (as in the oboe and bassoon) can serve as a valve. Instead of a valve, an edge-tone generator can produce a pulse train in a complex cyclic process (of 360°, see Meyer & Bork 1987, 20ff.) controlled by velocity and pressure parameters (cf. Fletcher & Rossing 1991, ch. 16).

In a block-and-duct flute, air passing the duct forms a laminar jet which streams against the edge opposite the duct where the jet bends inwardly into the pipe and outwardly away from the pipe while forming vortices and, consequently, eddies. The periodic change of direction the jet undergoes is brought about by the interplay of velocity and pressure differences. The pulses transmitted to the air column inside the resonator excite vibrations which result in standing waves when resonance is achieved. Since air molecules inside a tube do not undergo shear stress, only longitudinal motion is observed. Eigenmodes and resonance frequencies in an ideal tube open at both ends are in harmonic ratio. In regard to modes of vibration, there are nodes (minima) and antinodes (maxima) for the displacement and pressure amplitude, respectively; in an ideal tube open at both ends, the (alternating) pressure $p_{\sim}$ at each open end must be minimum while displacement x and velocity $\boldsymbol{v}$ of particles must be maximum. Hence, the open end viewed as a boundary condition (see Kalähne 1913, 76ff.) has a pressure node and a displacement antinode while pressure reaches a maximum at $l/2$ for the first mode, and displacement has a node there. The modes of vibration in the air column inside the open tube correspond to frequencies whose ratio is harmonic, that is $f_{\mathrm{n}} = \mathrm{n}f_1$ (n = natural number 1, 2, 3,…), where $f_1 = \frac{c}{2l}$, with l = length of the tube, and c = speed of sound in air (~340 m/s at 15 °C, sea level).

A standing wave fits into a tube if its length l equals 1/2 of the wavelength, λ, or an integer multiple of λ/2, thus: $l = \mathrm{n}\,\frac{\lambda}{2}$ and $\lambda = \frac{2l}{n}$ with $\lambda = \frac{c}{f}$.

Since only half of such a standing wave fits into a tube of length l, it is a λ/2-resonator. For a cylindrical tube closed at one end, it must have a displacement node and a pressure maximum at the rigid wall, while the open end has a pressure node and a displacement antinode. The distance between node and antinode, in this case, is l; a standing wave in the open tube closed at one end requires that l must be 1/4 of the wavelength λ or an odd multiple of λ/4. Hence $l = \frac{(2n+1)\lambda}{4}$ for n = 0, 1, 2, 3, … and $\lambda = \frac{4l}{(2n+1)}$ where $\lambda_1 = 4l$; for the lowest mode of vibration in a λ/4-resonator, the corresponding fundamental frequency is $f_1 = \frac{c}{4l}$ and frequencies of the next higher modes that have a pressure antinode and a displacement node at the closed end are $3f_1$, $5f_1$, $7f_1$ etc. Thus, the cylindrical tube closed at one end yields only odd harmonics. In principle, this is also the case with organ flue pipes closed (stopped) at one end. However, in a real vibrating system, one has to take more parameters into account, such as particle velocity ($\boldsymbol{v}$) and input impedance (Z_{in}), as well as energy losses (D) due to friction and damping (cf. Gough 2014, 635ff.). Though the input impedance to a tube filled with air is small, a certain force from wind pressure is needed to overcome the resistance.

Measurements of impedance in a cylindrical tube as well as in wind instruments yield curves where the maxima approximate harmonic ratios. In organ flue pipes, the actual resonances can deviate somewhat from the exact harmonic ratio so that higher partials increase in their frequency above the $f_{\text{n}} = \text{n}f_1$ ratio. A more general factor is that the pressure nodes of an air column vibrating in a tube do not match its end plane but lie somewhat outside (otherwise, sound radiation, which requires energy transport into open space, would be impossible). Thus, the effective length of a tube is $l + \Delta l$, where Δl is the end-correction which relates the distance a of the pressure node from the end of the tube to its radius (r) or diameter (d). The term a depends on frequency (cf. Meyer 1960) and diminishes with rising frequency; since the effective length of the air column under vibration decreases with frequency, actual resonance frequencies rise accordingly. The quantity a can be calculated (cf. Kalähne 1913, T. II, 221) like $a = \frac{\pi r}{4}$ $= 0.7854\ r$, where r is the radius of the tube's opening. Average values given for the end-correction Δl of open pipes usually are between $0.6r$ and $0.85r$.

In addition to the end correction Δl, another term Δm is required to account for the fact that pressure is not zero at the labium ($l = 0$) but has a positive value (cf. Mühle 1966/1979 for measurements taken from the block-and-duct flute which is comparable to a small flue pipe; see also Fletcher & Rossing 1991, ch. 17.3). Thus, Δm increases the effective length of the air column in the direction of the labium. Slight deviations of resonance frequencies in a tube from harmonic ratios are also caused by friction of particles on the walls of the tube. Friction effects are more marked in tubes or pipes of small diameter (where the wall plane is relatively large as compared to the dimensions of the air column). Further, it should be noted that flue pipes and the resonators of reed pipes develop structural vibrations of their walls (their geometry corresponding to a cylinder, a cone, or to a compound of elements). Wall vibration is relevant in long pipes with a thin wall, as in pipes made of an alloy with a high percentage of tin. Though these effects are measurable (cf. Runnemalm, Zipser & Franke 1999), structural vibrations are very small in amplitude (unless structural eigenmodes and modes of the air column coincide in frequency so that resonance is achieved).

A pipe organ consists of at least one wind chest equipped with one rank of pipes (for details, see Adelung 1976 and Williams & Owen 1984). Most instruments, however, comprise several such ranks of flue and reed pipes, and many organs installed in churches or concert halls combine several so-called 'works' (German: Werke) or 'divisions,' which can be viewed as separate units or even as separate organs. A 'work' generally has its own wind supply (in the era under consideration here, based on bellows), which feeds one or several wind chests via ducts. On each wind chest, there are several rows of pipes representing various organ stops which differ by the type of sound generation (flue and reed pipes; certain pipe models make use of overblowing into the 2nd or 3rd harmonic, see Mahrenholz 1942/1968) as well as by their size and geometry. Figure 1 shows a range of different flue and reed pipes standing on their wind chest from the Oberwerk (OW, upper organ) of the large organ built by Arp Schnitger 1689–93 for the St. Jacobi church of Hamburg. The picture was taken after the restoration of the organ in 1993/94, see Ahrend (1995).

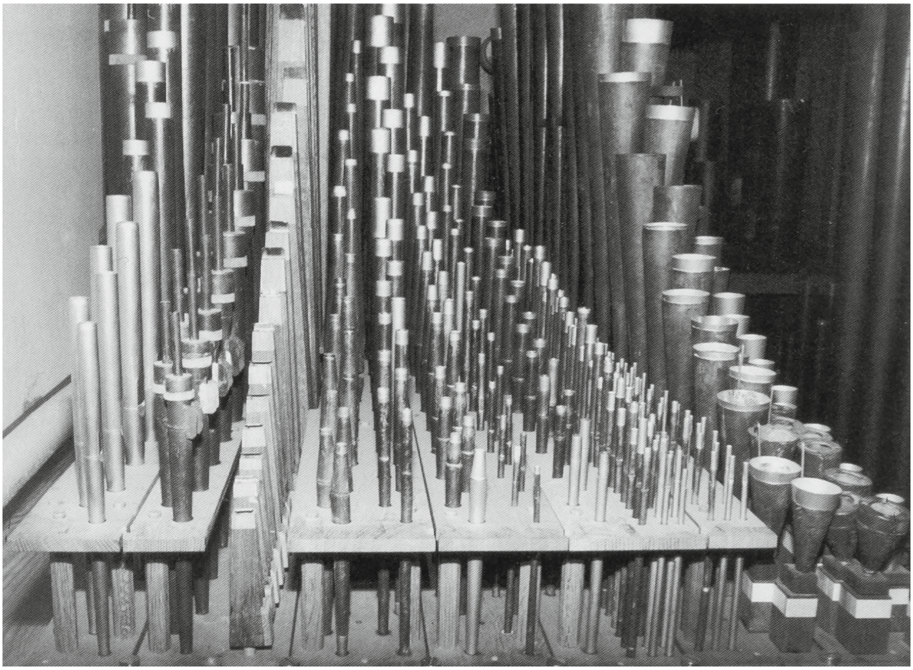


Fig. 1. Pipe ranks of the OW of the Schnitger organ at St. Jacobi, Hamburg. The pipe ranks on the wind chest are (from left to right): Prinzipal 8', Rohrflöht 8', Holtzflöht 8', Spitzflöht 4', Octava 4', Nasat 3', Octava 2', Gemshorn 2', Scharf IV-VI, Cimbel III, Trommet 8', Vox humana 8', Trommet 4' (from Ahrend 1995).

From the picture, it should be clear that pipes vary in regard to their length, diameter and shape. Flue pipes can be cylindrical or conical or may show a combination of cylindrical and conical segments in the resonator. Also, flue pipes can be open at the upper end or stopped, being either completely closed (a $\lambda/4$-resonator, meaning the pitch is about one octave below that of an open pipe of equal length) or partly closed as in the Rohrflöte shown in the picture, where a tube of small diameter is inserted into a disc on top of the pipe. The disc, in turn, is part of a 'hat' which covers the top end of each pipe. The hat is air-tight and can be moved up and down, which changes the effective length of the pipe and, thus, its pitch. A quite simple construction for stopped pipes usually was to close the upper end by soldering a plate on top; for fine-tuning, there may be holes drilled into such a top plate. Wooden flue pipes usually have a quadratic cross-section (as does the 'Holtzflöht' in the picture). In regard to the pitch, timbre, and intensity of the sound emitted from flue pipes, there are several relevant parameters related, most of all, to the geometry of the mouth, which incorporates the lower lip and the upper lip, whose edge acts as a pulse generator. In the process of voicing, an organ builder may make minute changes to the width of the windway (the flue), which alters the thickness of the laminar jet passing through the flue, and at the same time, alters the pressure and the speed of the jet. Among the variable parameters are also the height of the 'cut-up' (German: Aufschnitt) between the lower and upper lip as well as the width of the mouth (for technical details, see graphics in Adelung 1976, Williams & Owen 1984). The result of voicing should be a stable pitch and a harmonic timbre at a sound level as desired. Since the generator is coupled to a resonator, the actual system behaviour also depends on the geometry of the resonator, and in particular, on the relation of the effective length to the diameter of the pipe. With conical or double-conical pipes (e.g., Spitzflöte, see Fig. 1) or with even more complex shapes, numerical calculation of pitch can be quite demanding while actual mensuration is based very much on rule-of-thumb estimates, and even more so on practical experience as it did grow over centuries of organ building.

From medieval treatises on the mensuration of pipes (mensura fistularum; see Sachs 1980), it is evident that theorists considered first the length of different pipes in terms of small integer proportions (analogous to string sections on a monochord). Apparently, just a few theorists recognised that the analogy of strings and pipes did not hold as such, and was insufficient to determine dimensions and pitches for real pipes. There were some considerations where fractions of the diameter of a pipe were added to its length (to account for the factor later understood as the end-correction of the pipe; see Sachs 1980, 65ff.), however, the approach was theoretical rather than empirical. A general aspect inherent in these mensuration problems is that appropriate scaling of organ pipes involves more than one parameter of pipe length since the design of pipes must consider not only their pitch but also the specific timbral quality of a rank. The task to model a row of pipes thus is threefold. First, the sounds emitted from the pipes must realise the steps of a musical scale defined, for flue pipes, in the main by their fundamental frequency. Second, the sequence of sounds from such a row of pipes must bring about an increase in brightness proportional to the increase in pitch per scale step since brightness is a component of pitch and at the same time, a timbral factor (see Schneider 2017). Third, while spectral centroid and sensation of brightness change along the steps of a rising or falling musical scale, spectral energy distribution and spectral envelope, as well as temporal characteristics of sounds for one pipe rank, should follow a certain pattern so as to maintain the timbral quality (by which a rank is identified, by musicians and listeners). This was already a problem in late medieval times when the compass of an organ was restricted to 2–3 octaves, and more so in modern instruments where four or even five octaves in manual keyboards are standard. One of the facts probably experienced in medieval organ building was that continuity in timbral quality cannot be achieved if only the pipe length is varied, with the diameter (and all other parameters) kept constant. With such a design, pipes low in pitch will have a timbre that is too bright, whereas pipes high in pitch will sound dull (see Adelung 1976, 80ff). The lesson learned early from scaling was that several parameters in regard to the geometry and also voicing of pipes must be taken into account, and in doing so, pipes of a certain type (e.g., a cylindrical diapason, an open conical flute, or a trumpet) can be built in different size so as to match pitch levels for a certain octave (32', 16', 8', 4', 2', 1'). In this respect, scaling demands that actual measures must be altered in proportion to each other along relevant dimensions (pipe length, diameter, height and width of cut-up, etc.). Assuming such proportionality, pipes and pipe ranks can be described, first of all, by their pitch level as defined by the pipe length expressed in 'foot.' In modern standard tuning ($a^1 = 440$ Hz), an open flue pipe of approximately 262 cm effective length will produce a sound with the fundamental of 65.4 Hz when the key for C_2 is pressed. The wavelength, in this case, is ca. 525 cm while the pipe of 262 cm approximates 'eight times a foot' (in olden times, spatial extensions were often measured in 'foot', one foot is ca. 32 cm). Thus, the pipe in question will be labelled 8' (eight-foot). To produce a fundamental at 32.7 Hz for the key C_2, the pipe length must be doubled to 16'. Conversely, a 4' pipe at the same key will have a fundamental at 130.8 Hz, a 2' pipe at 261.6 Hz etc. In historical organs of Northern Germany and the Netherlands, some stops with flue and with reed pipes are found in the 32' register, while in most organs, pipe ranks from 16' to 2' are implemented (spanning four octaves); some organs have or had 1' ranks (cf. Praetorius 1619, 162ff., Edskes & Vogel 2009, 167, 169, 173, 198). Mixture stops, as well as special pipe ranks, can incorporate very small pipes (<1'), which add high harmonics and increase spectral brightness (see below).

The diameter of open flue pipes changes in proportion to their pitch. However, while pipe lengths approximate a ratio of 2:1 per octave, the diameters of pipes do not correspond to this ratio and, in fact, can vary considerably to adjust the timbre (number and strength of partials) in a rank of pipes. For example, diameters for the Principal 16' in the HW of St. Jacobi at Hamburg have been measured (cf. Ahrend 1995, 255) as shown in Table 1.

Table 1. Pipe diameters and ratios, Principal 16', HW, St. Jacobi, Hamburg

Pipe	C	F	c	f	c'	f'	c"	f"	c"'
Ø (mm)	232.4	188.8	142.3	114.9	85.9	72.3	50.3	41.1	28.6
Ratio	1	0.812	0.612	0.494	0.37	0.311	0.216	0.177	0.123

Viewed from a historical and geographical perspective, the variety of organ stops in Europe is immense notwithstanding certain standards had been established over time (for detailed accounts, see Mahrenholz 1942/1968, Klotz 1975, Williams 1984, Eberlein 2009). Organ builders experimented a lot to find optimal designs for pipes as well as for the resonators coupled to reed generators. Since the labelling of organ stops was not uniform, one has to take historical and regional traditions of organ building into account. To complicate things further, one and the same name, like the German 'Nachthorn' or the French 'cornet de nuit,' can stand for stops of quite different designs and musical functions.

Most pipe organs from the 15th to the 18th centuries were primarily designed to be used in church services as well as in music events related to religious practice and recreation; this does by no means exclude recitals and concerts where the organ was used as a continuo instrument in an ensemble of strings, woodwinds or brass. Depending on factors such as the size of churches or other rooms chosen to house an organ, financial means available to a community that would order and pay for an instrument as well as the strength of musical activities pursued in certain regions, organs of different sizes and complexity were build that ranged from small instruments (in general, one manual, one wind chest, number of stops < 10, no separate pedal work, or pedal completely missing) to middle-sized and large organs. Middle-sized, in this respect, typically would be a two-manual organ with separate pedal work and an overall number of stops from about 18 to 25 distributed to the three works. Large historical organs of the late Renaissance and Baroque era generally comprise three manuals (in some organs, even four) plus a fully equipped pedal. Thus, there are four or five separate 'works' (divisions) which typically can be distinguished by their location within the overall spatial structure of a pipe organ as well as by different musical functions and sound designs.

Large organs from places like Hamburg, Lübeck, or Stralsund by the time of ca. 1600–1700 typically had three or even four manual works like Hauptwerk (HW, great or main organ), Oberwerk (OW, placed above HW), Brustwerk (BW, right in front of the organist sitting with his face in the direction of the HW and OW), and a Rückpositiv (RP, chair organ in the back of the organist) plus a pedal work (Ped) that had gained special importance by then. One of the peculiarities of northern Germany was that in a number of large organs in use around 1650–1730, four manual works were played from three keyboards; that is, one could either couple two manual works (like OW and BW) to the respective keyboard or use them alternately. In a schematic drawing, the spatial arrangement of a large organ with four manual works plus a pedal work split into two towers flanking the organ to the left and right is shown in Fig. 2.

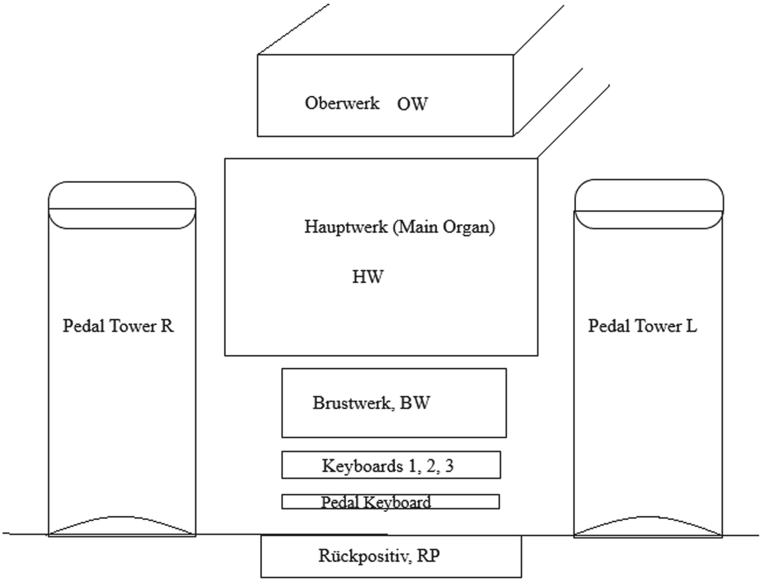


Fig. 2. Scheme of a North German Baroque organ with four manual divisions (HW, OW, BW, RP), three keyboards, and a pedal whose pipes are mounted in two flanking towers.

A real instrument is shown in Fig. 3, a photo taken of the three-manual organ built for St. Nikolai at Altenbruch (near Cuxhaven, close to the North Sea) by Johann H. Klapmeyer in 1727–1730. This instrument (III, Ped, 35 voices) incorporates a considerable number of stops from older instruments that had been built in the 16th century; the RP very likely dates back to 1577 (see Vogel et al. 1997, 218ff.). In the years 1647–1649, Hans Christoph Fritzsche from Hamburg renewed the HW and added stops to the RP. Since organs were expensive in regard to the materials needed for construction (different kinds of wood and metal, such as tin and lead hard to get hold of in those days), it was customary to repair existing stops and wind chests and to keep them in use as part of a new instrument. Thanks to this cost-saving attitude, the organ at Altenbruch and many others of the Baroque era preserved pipe ranks from the 16th and 17th centuries.

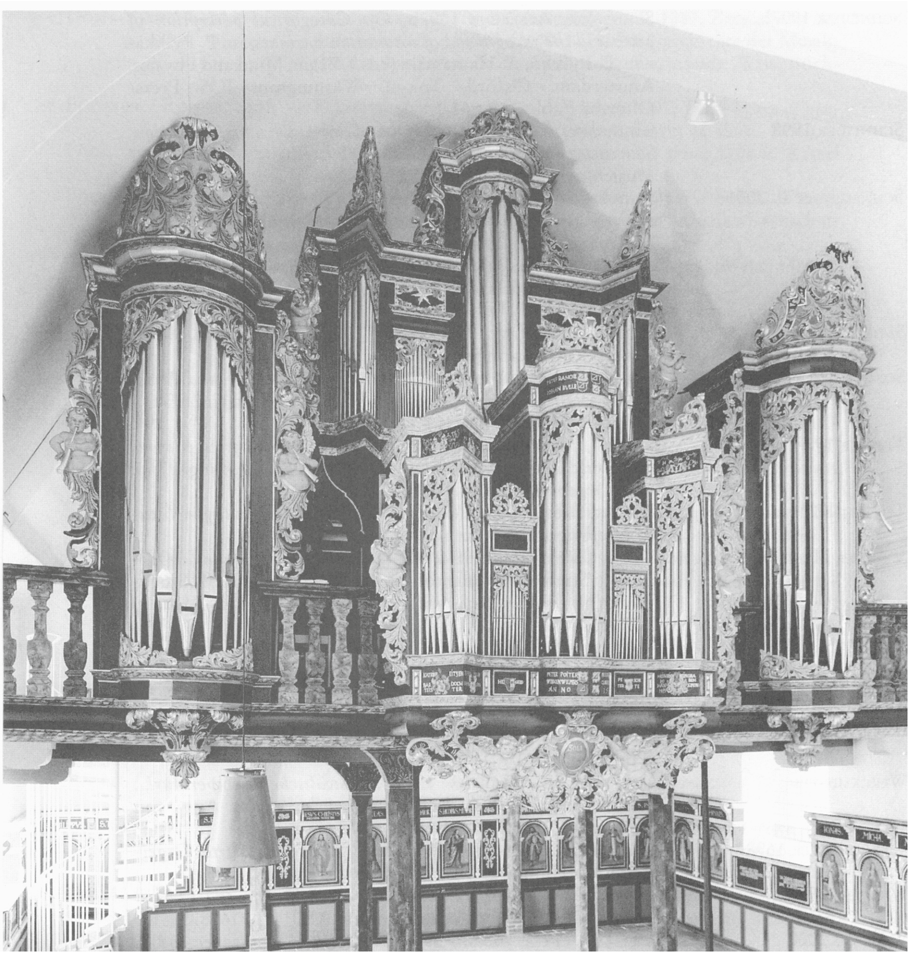


Fig. 3. Klapmeyer organ at St. Nikolai, Altenbruch, with the HW in the back and the Renaissance RP and the two pedal towers integrated into the balustrade of the gallery. The BW is masked by the RP and not visible in this picture.

Splitting the pedal pipes into two towers with a division of C and C#, D and D#, etc., has the advantage that pipes that are just a semitone apart will not interfere with each other in chromatic bass lines (as are found in organ music of the late Renaissance and Baroque era). Flue pipes in an organ can interact in several ways. One factor is that pipes mounted on the same wind chest share the supply of air delivered from the bellows (which was the traditional wind supply before electric ventilators came into use). If the pressure in the wind supply is not strong enough or unsteady, simultaneous use of several large flue pipes can cause a slight but sometimes audible pitch shift relative to the pitch of each single pipe as tuned. Flue pipes standing close to each other moreover influence each other in pitch, an effect known as acoustical coupling ('Mitnahme'; cf. Lottermoster 1983a, 56) that can be explained as a synchronisation of vibration regimes (Abel 2008, Fischer, Bader & Abel 2016).

3 A Few Notes on the Development of Pipe Organs

Though any detailed account of organ history is far beyond the scope of this article (see Williams 1966, Klotz 1975, Williams 1993, Eberlein 2011; for an overview, Williams & Owen 1984), a few remarks concerning some major stages in organ development seem apt to understand aspects of continuity and change in organ design. The very beginnings of the pipe organ lead us back to Roman times. In addition to written and iconographic sources, we have a small number of pipes and other parts from archaeological excavations. Perhaps the most significant find brought remnants of a small pipe organ from the camp at Aquincum (today, a part of Budapest, Hungary) to daylight. This organ, parts of which lay in the rubble after the camp was destroyed by fire, dates to 228 CE; it might have been a *hydraulis*, a type known for its mechanism of hydraulic pressure used to provide wind to the wind chest and further on into the pipes. While the exact mode of wind supply in this instrument is not quite clear, the wind chest survived in relatively good condition and allowed, together with a number of pipes and other parts, a tentative reconstruction of the organ, including inferences as to the dimensions and original tuning of the pipes. These were ordered into four rows, one with open pipes and three with stopped pipes, which yields $13 \times 4 = 52$ pipes (see Kaba 1976 and the article 'Orgel von Aquincum' in the German Wikipedia for pictures). There have been some suggestions in regard to the compass and the scale to which the organ was tuned. Apparently, the four pipe ranks could be activated individually—as in a modern organ—by some mechanism. If used together, pressing a single key (or, rather, pulling a lever) would join the sound from three stopped pipes (of different lengths) plus an open pipe into some sonority that possibly could have included musical intervals (the pitches of a stopped pipe and an open pipe of equal length are an octave apart; see below, Sect. 4).

This aspect is of interest since an important late medieval organ type, the so-called 'Blockwerk,' apparently was designed to produce complex harmonic sounds. In its most basic form, a Blockwerk assembled several or even many rows of pipes ordered according to their length on a single wind chest (see Williams & Owen 1984, Klotz 1975, 10f.). The idea behind this arrangement was that, by pressing a single key (which was broader and longer in size than in a modern keyboard; see Praetorius 1619 and Bormann 1966), a number of pipes responded, which were in harmonic pitch ratios and formed chord-like sonorities. In certain respects, a Blockwerk thus can be viewed as an early large mixture stop which, however, is different from later mixtures in that a Blockwerk typically had only a few pipes for the low keys and a larger number of pipes for high keys, that is, the number of pipes per key increased from the bass to the discant register (see Preaetorius 1619, 94ff; Bormann 1966, Klotz 1975). The next step in organ building was to add ranks of diapason pipes which either were rigidly coupled to the compound of pipes in the Blockwerk or could be switched on and off as desired; thus, an elementary registration was possible (diapason pipes with and without Blockwerk, or the latter alone). Musically, one could use such ranks of diapason pipes to carry a hymn or other melody while the Blockwerk, as a sonorous unit, could support the diapason as well as a group of singers. Since the ranks of diapason pipes were placed in the front of such organs, they often were labelled ‘Praestant’; the Blockwerk mixture set behind these diapason pipes was labelled ‘Hintersatz’ (both terms remained in use over centuries).

The structure of a late medieval organ that was built at Halberstadt around 1360, and revised in 1495, was as follows: this organ had three manuals and a pedal, with 22, 22, 12, and 12 keys, respectively. The compass for the two upper manuals apparently was H–a^1 (B$_2$–A$_4$), with g#^{1} (G#$_4$) missing, and the key for H (B$_2$) probably sounding the tone B (Bb$_2$). The uppermost manual was connected to a compound of pipes of various lengths that made up the discant Blockwerk (see Table 2).

Table 2. The distribution of pipes relative to the keys likely was as follows (cf. Bormann 1966, 44).

Key	16’	8’	5 $^1/_3$’	4’	2 $^2/_3$’	2’	1’	No. choruses
H-f	2	3	4	5	6	6	6	32
f#-c#^{1}	2	4	5	6	7	8	10	42
d^1-a^1	2	5	6	7	10	12	14	56

From the middle keyboard, one could play a double row of 16’ diapason pipes (suited to carry a melodic line); however, the keys of this keyboard also activated the pipes of the discant Hintersatz. The third manual activated the pipes of another Hintersatz, which was an octave lower in pitch than the discant Blockwerk unit, and coupled with the pedal.

The structure of the discant Blockwerk unit reveals that the number of pipes increased towards high pitches and small pipes. The sound pressure level for each of the 16’ and 8’ pipes will be higher, at a given wind pressure, than that of a single 2’ or 1’ pipe. Still, the sheer number of the 2 $^2/_3$’, 2’ and 1’ pipes will reinforce the sound level considerably and, moreover, will shift the spectral centroid upwards in the treble range. Praetorius (1619, p. 100) noted that a Blockwerk such as the old organ of Halberstadt must have produced “ein uberaus starcken schall und laut und gewaltiges geschrey” (an immensely strong sound and enormous screaming).

Praetorius understood the ‘Hintersatz’ as a forerunner to the more modern mixture stop. The difference, however, is that the Blockwerk of late medieval and Renaissance masters, in general, did not use the concept of repetition, which is characteristic of a mixture stop. For example, in the organ built by Berendt Huß and Arp Schnitger for the church St. Cosmae & Damiani at Stade (Altes Land, see Edskes & Vogel 2009), the OW from 1675 has a compass (with a so-called short octave in the bass) C, D, E, F – c’’’ (44 keys spanning four octaves). Among the 11 stops in the OW is a mixture VI (Vogel 1982), as shown in Table 3.

Table 3. Composition of mixture stop (sixfold), Huß-Schnitger organ, Stade

Key																
C									1’		⅔’		½’	½’	⅓’	⅓’
c							1 ⅓’		1’		⅔’	⅔’	½’	½’		
f#					2’		1 ⅓’		1’	1’	⅔’	⅔’				
c’			2 ⅔’		2’		1 ⅓’	1 ⅓’	1’	1’						
f#’		4’	2 ⅔’		2’	2’	1⅓’	1 ⅓’								
f#’’	4’	4’	2 ⅔’	2 ⅔’	2’		1 ⅓’									

Such a mixture stop is called ‘sixfold’ since there are six pipes per key who have interval relations of octaves and fifths. From the structure of the pipes in regard to their length and pitch, it is clear that the mixture serves to ‘brighten up’ sounds of the basic diapason stops (like Principal 16’ or Quintadena 16’ available in this OW). In particular, for tones low in fundamental frequency like C (in Helmholtz designation), which, in this organ (historically tuned to g’ ~442 Hz, a so-called ‘choir tone pitch’) is at ca. 74 Hz. Conversely, tones much higher in fundamental frequency, like c’ (295.4 Hz in this organ), shall have less treble sound added from high-pitched small pipes so as to avoid a sound quality sensed as ‘sharpness’. Instead, for these tones in the upper octaves of the keyboard, the pipes from the mixture stop should reinforce, to some extent, midrange frequencies. Since the pipes of mixture stops internally are tuned to just intonation, that is, harmonic ratios, the acoustical function of such a mixture stop thus is to provide additional partials on top of the sound of pipes from other stops (e.g., diapason- or flute-like stops). This effect can be labelled ‘harmonic spectral enhancement.’ The overall perceptual effect of such a mixture stop is to supplement the sound of diapason- or flute-like stops with a certain amount of spectral brightness that, approximately, should be constant over the whole compass of the keyboard. Therefore, the spectral centroid of sounds from a mixture stop should not change very much over several octaves so that the sensation of spectral brightness from a sequence of sounds played with, for example, Principal 16’ + Octave 8’ + Octave 4’ + Mixtur VI does not surpass a certain range.

The original Blockwerk concept implied that several, if not many, pipes attached to each key would produce a rich, chord-like sonority based on octaves and fifths (see above). The Blockwerk, with its peculiar sound structure, has been linked with the medieval organum as a musical form (cf. Klotz 1975, 9). However, during the 14th and, more so the 15th century, the structure of organs was modified and expanded in line with musical developments. The Halberstadt organ had three manuals and a pedal suited to play some elementary two-part polyphony, perhaps including long-held pedal notes or simple ostinato patterns. In the 15th century, a number of organs already showed separate ‘works’ (like HW, RP, Ped), each equipped with one or several pipe ranks, which could be activated individually and combined at will with one or with several mixture stops. From the treatises on organs by Henri Arnaut de Zwolle from 1447 (facs. ed. 1972; Latin text with commentaries in Bormann 1966, 157ff.), the structure of an organ built by Jehan du Mexe for the cathedral Notre Dame at Dijon (Bormann 1966, 163f., 169f., Klotz 1975, 35, 40ff.) has been inferred like shown in Table 4.

Table 4. Organ of Notre Dame, Dijon (15th century), stop list (reconstructed)

HW, compass $_1$H-f^2, 43 keys	RP, compass F G-f^2, 36 keys	Pedal, F G A B^b, 4 keys
Diapason (8’) II-IV	Diapason (8’) IV-VII	Diapason (16’) II
Mixture (4’), VI-XIV		Mixture (8’) V
Cimbel (1/2’) III		

According to this stop list, the diapason in the HW, RP, and Pedal each consisted of several, i.e., two to seven rows of pipes (marked with Roman capitals). The HW had a substantial mixture with up to fourteen pipes per key (thereby continuing the tradition of the Blockwerk), and the Pedal also had its mixture (fivefold). The important point here is that the HW included a special mixture stop generally known as a threefold cymbal (German: Zimbel, fr. cymbale). Arnaut de Zwolle (fol. 133 v° and 134 r°) gave more information in regard to this stop which offered high-tuned major chords; for instance, pressing the key f^2 would produce sound from a total of 18 pipes, the c^3 key would activate 20 pipes (cf. Bormann 1966, 163) shown in Table 5.

Table 5. Organ of Notre Dame, Dijon, compound of pipes attached to single keys

Key	Diapason (Prinzipal)	Mixture (Hintersatz)	Cymbal	Total
f^2	3 f^2	2 f^2, 6 f^3, 2 c^4, 2 f^4	c^4, f^4, a^4	18
c^3	4 c^3	3 c^3, 7 c^4, 2 g^4, c^5	c^4, e^4, g^4	20

The major-third cymbal (German: Terzzimbel) is of particular interest since, in the course of the 15th century, the major third was accepted as a consonant interval in both music theory and composition, a fact that led to significant changes also in the tuning of organs and other keyboards. By about 1500, the so-called meantone temperament based on just major thirds (see Schneider & Beurmann 2017) had become the predominant tuning system in Europe which was in use, in a number of variants, well into the second half of the 18th century or even later (it was laborious and costly to change the tuning of organs with their multiple pipes).

Splitting the former Blockwerk into separate stops (like mixture and cymbal) as well as dissolving the compound of diapason pipe rows into several independent stops were notable developments in organ building in the 15th and 16th centuries. The dissolution into individual stops is especially clear in Italian organs where the order of diapason stops often was as reported for S. Pietro at Modena (Giovanni B. Facchetti, of Brescia, 1519) and for Santa Maria Rotonda at Brescia, built by Gian Giacomo Antegnati in 1536; see Klotz 1975, 71, 133). The stop list for the Facchetti organ is shown in Table 6.

Table 6. Organ of S. Pietro, Modena, early 16th century, stop list

Manual $_1$F, $_1$G, $_1$A - g^2 a^2 (50 keys)	foot	Flute stops	Pedal
Principale (longest pipe)	8’	flauto (VIII = 4’)	
Ottava (VIII)	4’	flauto (XV = 2’)	
Quintadecima (XV)	2’		
Decima nona (XIX)	1 1/3’		
Vigesima seconda (XXII)	1’		
Vigesima sesta (XXVI)	2/3’		
Vigesima nona (XXIX)	1/2’		

The Italian terms and the Roman capitals relate to the *claves naturales* of the diatonic scale; foot marks are relative and indicate interval relations between stops, not the absolute length of pipes. The organ built by Antegnati had an additional flute stop (XXII, 1’) and an independent pedal (F G A-d^1, 20 keys) with a single stop labelled Contrabassi 16’. The feature that is of interest here (as different from a fixed mixture) is the possibility to select and combine pipe ranks that form harmonic interval ratios; activating those diapason stops one after another means expansion of an additive harmonic synthesis whereby the sound gets brighter with every pipe rank added on top. The concept of additive synthesis of diapason stops, being apparent in many Italian organs, has a modern follow-up in electronic organs of the 20th century, such as the Hammond B3 and the Vox Continental, where the player can mix partials from generators with ‘drawbars’ like pipes of different foot length (16’, 8,’ 5 1/3’, 4’, etc.).

From historical sources, it seems the range of ‘sound colours’ available from those Italian organs was rather small (with a dominance of diapason and flute stops). However, from iconography and written sources, it is well known that late medieval and particularly Renaissance musical practice included many wind instruments (flutes, horns, reeds).. At the beginning of the 17th century, yet much in retrospective, Michael Praetorius, himself a skilled musician and composer, in his ‘Organographia’ gave a detailed account of musical instruments and put special emphasis on trombones, trumpets, the Zinck, several long and cross flutes, and various types of reed instruments such as the Pommer (alto and tenor shawm, Bombart), Schalmey (treble shawm), Dulzian (an early bassoon-like reed), the Krumbhorn (crumhorn), Rankett, etc. (1619, 31–43). In the same work, the chapter on the ‘Historia veterum Organorum (81ff.) elucidates the concepts behind organs of previous centuries, in particular instruments of the Blockwerk style. The chapter on the Historia novorum Organorum (119ff.) discusses the types of organ pipes and the different pipe ranks that came into use, mostly in the 16th century, and which Praetorius knew from first-hand experience. He describes the more common diapason and flute-like ranks, followed by flue pipes with more complex geometry (like the Gemshorn), the stopped pipes and the various reed pipes. He also adds chapters on the tuning of reed pipes and on the suitable design of organs and presents a comprehensive survey of stop lists (Dispositionen) from various organs that had been recently built for churches at Danzig, Lübeck, Hamburg, and other places. More information is condensed into a catalogue of pipe ranks. Finally, his 'Sciagraphia oder Theatrum Instrumentorum' offers many figures that illustrate the instrument types addressed in the text.

From Praetorius and other sources, we understand how diversified pipe ranks had become between roughly 1450–1600. One of the reasons already mentioned was the dissolution of huge compounds of pipes (Blockwerk and Hintersatz) into separate ranks, another was that organ builders strived to emulate the broad range of flutes, horns and reed instruments that played an important role in Renaissance music. These instruments all had a peculiar sound quality (some came close to the human voice, some reeds had a nasal sound, etc.), which made them distinct and identifiable in an ensemble. Organ builders must have recognised the benefit they could have for organs devised as multi-timbral instruments. If several distinct 'sound colours' were available, the organist could play a cantus firmus or characteristic melodic line with a reed stop against other voices, for which a soft sound (from stopped pipes like a Gedackt) might be appropriate. Such a concept would work easily in a two-manual plus pedal organ with different pipe ranks available in each department. In small instruments (one manual, no separate pedal), parallel usage of two 'sound colours' was possible if some stops could be assigned to either the bass or the discant half of the manual (which was thus divided into two registers).

From a historical perspective, the diversity of pipe ranks and an increase in the number of stops is evident from many organs of the 16th century that were built in France as well as in a large region comprising the Low Countries (understood as a geographical term) and parts of Germany (for a detailed account, see Klotz 1975, ch. X-XVI). The division of pipe stops into distinct groups according to sound properties and musical function, in general, followed a scheme like:

A. Diapason stops (flue pipes from 16' to 2' with relatively narrow diameters like Prinzipal 16' or Praestant 8', Oktave 4', Oktave 2'); mixtures (usually III to IV) and related aliquot stops (like Zimbel or Sesquialter) composed of rather small and narrow flue pipes;
B. Open and stopped flue pipes with a wider diameter (like Hohlpfeife, Quintaden, Nachthorn); the sound quality in theses pipe ranks is more mellow or flute-like, in stopped flue pipes it can be hollow (like a voiced syllable 'hu') due to the prevalence of low odd partials;
C. Reed pipe stops like Posaune 16', Trompete 8', Krummhorn 8', Schalmei 4'.

In a well-designed, middle-sized or even large organ, one could expect a selection of stops from all three groups in every department (HW and/or OW, RP and/or BW, Ped). Perhaps the largest organ in use before 1600 was built 1583–1585 by Julius Antonii (from Bergues-Saint-Vinoque; Flemish: Sint Winoksbergen) for St. Marien at Danzig (56 voices on HW, RP, BW, Ped plus 3 tremolo units and a wind-operated drum). This instrument offered an enormous range of pipe ranks (as listed by Praetorius 1619, 162f.), including various diapason and mixture stops, many different flutes (from 16' to 1'), and no less than 11 reed stops such as a trombone 16' (Ped), two trumpet 8' (RP, Ped), two Krummhorn 8' (RP, Ped), two Schalmei 4' (RP, Ped), two Zink 4' (RP, BW), a Regal 8' (BW), and a Kornett 2' in the pedal. Praetorius (1619) presented the stoplists of some other large organs as were built at, for example, Lübeck (St. Petri, Gottschalk Johannsen, 1587–91, 45 stops on OW, BW, RP, Ped), Stralsund (St. Marien, Nicolaus Maaß ca. 1592, 43 Stops on OW, RP, BW, Ped), Hamburg (St. Jacobi, 53 stops on OW, BW, RP, Ped). In 1680–87, Arp Schnitger succeeded in building an even more complex organ for the St. Nikolai church of Hamburg (67 stops on HW, OW, RP, BW, Ped; four manuals with 'short octave' in the bass register, C to c^3, 47 keys; see Fock 1974, 46ff.). This marvellous instrument, the result of a long tradition of organ building first developed in the Low Countries and continued in Northern Germany, unfortunately was destroyed, in a disastrous fire, on May 5th 1842. Another famous organ that fell victim to this fire was the organ built by Henrick Niehoff for St. Petri of Hamburg (ca. 1550, 42 stops, see Praetorius 1619, 169f. and Fock 1939, 298ff.). Niehoff, who worked for many years from s'Hertogenbosch in the province of Brabant, is understood as a foremost organ builder of his era as he pursued a concept of contrasting sound colours produced from various flue and reed stops. In particular, he recommended the Terzzimbel (also labelled 'klingende Zimbel' and 'rauschende Zimbel'), a special type of mixture which, due to its composition, added high harmonics including major thirds to the sound of other stops, thereby amplifying both spectral fusion and brilliance. In the historic organ of Altenbruch (see above), there is an original Zimbel in the HW (probably built by Hans Christoph Fritzsche in 1649) that produces significant spectral energy in high-frequency bands (up to and even beyond 10 kHz; see Schneider et al. 2006).

4 Sound Generation: Empirical Observations

One obvious feature in the speech of flue pipes is the noisy transient in the onset of sounds which has been studied extensively (cf. Fletcher 1976, Nolle & Finch 1992, Castellengo 1998). The main reason for this phenomenon is that alternating pressure $p_{\sim}$ needs time to build up from the pulse train passing from the edge tone generator into the resonator tube and that, after reflection at the open end, a stable regime of periodic vibration needs to be established resulting in standing waves. The pipe viewed as a cylinder filled with a mass of air has a certain input impedance Z_{in} which is quite small, but so is the wind pressure in most historic organs as measured in a duct or chest (usually 50–80 mm water column depending on the size of the organ and the room in which it stands). As a mass of air enclosed in a large flue pipe has some inertia, the onset in 16' and 8' pipes can last quite long ($t_{\text{on}} > 50$ ms, for some pipes even $t_{\text{on}} > 100$ ms; see examples in Beurmann et al. 1998, Schneider et al. 2001, 2006). Quite often, the second mode of vibration (the octave in an open flue pipe, the twelfth in a stopped pipe) is activated before the fundamental sets in. The higher partial kind of 'signals' the onset of such a tone to the listener '(see Fig. 4)':

Another characteristic feature of the onset of many flue pipes is the 'spitting' noise ("Chiff") preceding periodic vibration. The noisy transient, together with the attack of

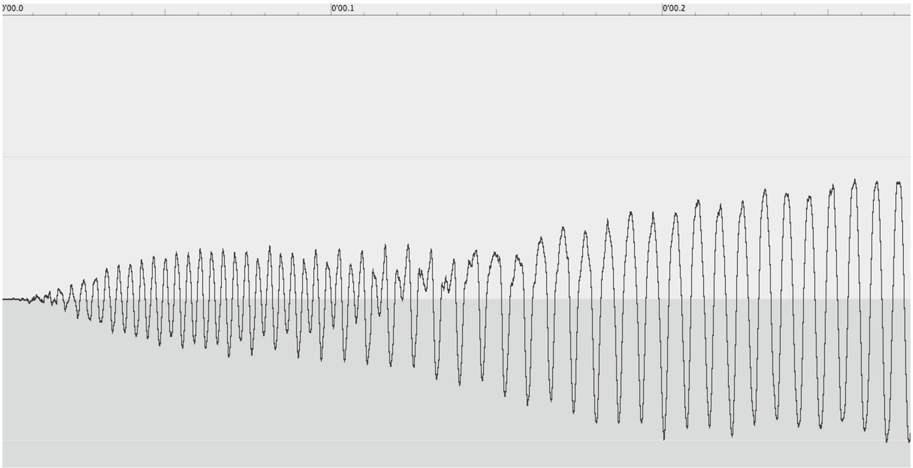


Fig. 4. Stade, St. Cosmae & Damiani, the organ built by Berendt Huß and Arp Schnitger 1668–1673; RP, Oktave 4’, key/tone C; onset begins with the 2^{nd} harmonic (Oscillogram).

the partials, has a sound quality of its own that helps listeners sense the onset of single tones. While the length of the pipe that produced the transient shown in Fig. 4 was ca. 1.20 m, transients appear even in very small flue pipes. In Fig. 5, the onset for the tone/key c in the Nachthorn 1’ of the pedal in the organ of St. Cosmae at Stade is shown. The pipe length for this tone is actually 1/2 foot; the fundamental frequency consequently is high at ca. 1177 Hz.

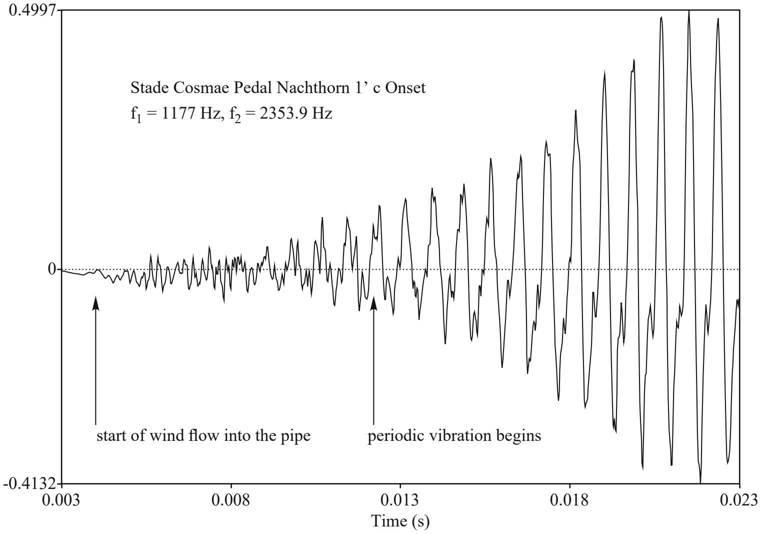


Fig. 5. Nachthorn 1’, tone/key c, Onset with noisy transient, periodic vibration.

In flute-like aerophones, one can observe, in a series of overlapping short-time spectra, the process whereby relatively broad and flat components carrying energy turn into harmonic partials with marked peaks (see Schneider 1998). The physical process thus covered is the transition from relatively broad resonance zones to definite resonance frequencies in the tube as the standing wave regime becomes stable. The time needed to establish standing waves in the tube is dependent, among other parameters, on the length and width of the tube. In a small pipe like the Nachthorn 1', a periodic vibration pattern appears after ca. 10 ms (Fig. 5). In the spectrum of this sound taken shortly after onset, four partials are prominent.

In general, reed pipes differ from flue pipes in that there is a fast and hard attack in their sound with less noise involved in the transient. The periodic regime of vibration is often established almost instantly, even in large generator plus resonator systems, as is demonstrated in Fig. 6, which shows the onset of sound radiated from the c pipe of the Dulzian 16' in the RP of the Huß/Schnitger organ at Stade.

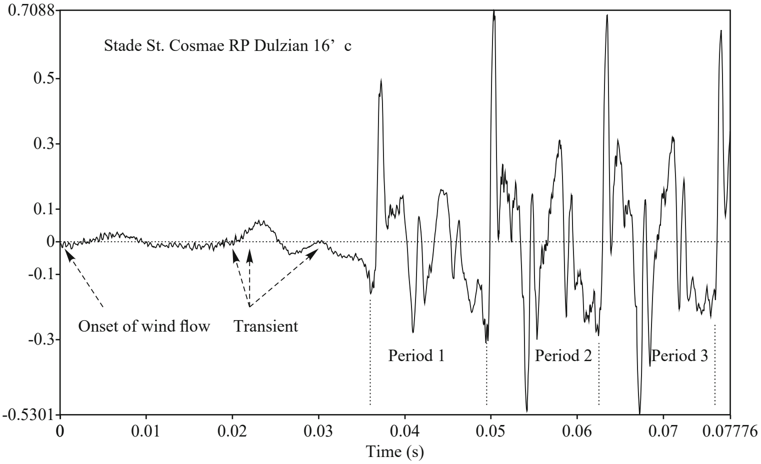


Fig. 6. Dulzian 16', tone/key c (f_1 ~ 77 Hz). The section to the left marks the time from the start of wind supply to this pipe and the transient (ca. 34 ms) which is immediately followed by full periods of vibration (T ~ 13 ms).

Sounds from reed pipes typically have a rich spectrum with dozens (in some cases more than 100) harmonics. The spectra, in general, show a cyclic structure where spectral amplitudes and the spectral envelope are similar to the envelope of a Sin[x]/x function, and certain harmonic partials are more or less suppressed due to the duty cycle of the valve defined by τ/T (τ = pulse width, T = length of the period in ms; for examples from the large Schnitger organ of St. Jacobi at Hamburg see Beurmann et al. 1999, 159ff.). Numerous reed pipe spectra show considerable energy in frequency bands known from phonetics as 'formant zones'. Such a concentration of spectral energy lends sounds a vocal quality. The spectrum of a sound of the Regal 8' in the RP of the organ at St. Jakobi of Lüdingworth (see Edskes & Vogel 2009) illustrates this peculiar aspect (Figs. 7 and 11). The Regal 8' was built by Antonius Wilde in 1598/99 for an organ expanded by Arp

Schnitger in 1692. In this stop, the tongue, the shallot, the tuning wire and the resonator of each pipe are made from brass which gives the pipes a distinct ‘sound colour.’ As is evident from Fig. 7, one of the energy concentrations is around 3.4 kHz (the region of the so-called ‘singing formant’ for male opera singers is at ca. 3 kHz; see Sundberg 1987).

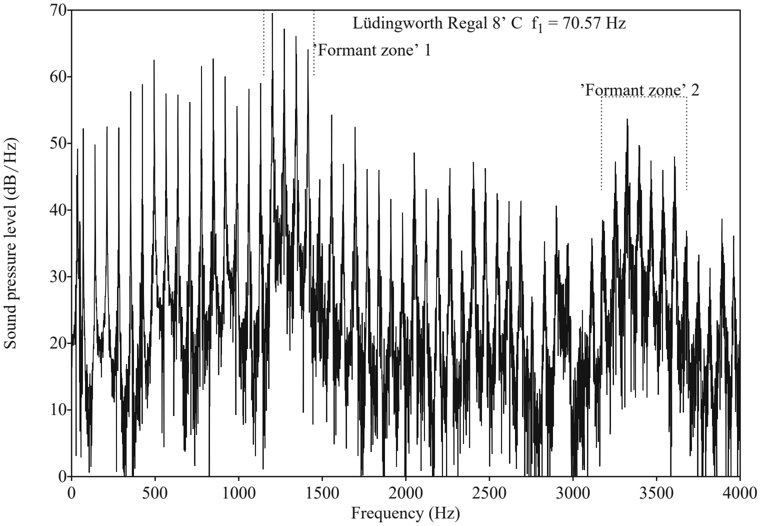


Fig. 7. Regal 8’, tone/key C, formant-like spectral energy concentrations at 1.3 and 3.4 kHz.

A closer inspection of the same sound with the Burg algorithm (see Marple 1987, ch. 8) reveals that, in fact, two concentrations of spectral energy can be identified as formants, with centres at about 1.3 and 3.4 kHz, respectively.

The temporal and spectral composition of such sounds from reed stops is of perceptual and musical significance. First of all, their pitch is clearly defined both from the fundamental f_1 and the periodicity pitch $f_0 = 1/T$ resulting from the joint effect of numerous harmonics. Second, the presence of formant-like energy concentrations in the spectrum gives such sounds a vowel-like quality (which is also observed in the tone of Italian master violins, see Mores 2017). In effect, reed stop sounds, such as those produced by the Regal 8’ of Lüdingworth provide the listener with ample information in regard to pitch structure and timbre. With reed stops available in each division of the organ, one could emphasise prominent voices in a polyphonic setting or could give consecutive sections of a musical work alternating ‘sound colours.’

5 Sound Structure and Tuning

The design of pipe organs from the medieval Blockwerk and the Italian instruments of the 16th and 17th century up to a range of organs with multiple stops built in the Low Countries and Northern Germany (with some extensions also to Denmark and Sweden) between ca. 1600 and 1750 converges in one fundamental aspect which can be described as 'massive additive sound synthesis'. To understand this concept, one has to remember that each pipe in an organ is a sound generator of its own which produces a more or less complex sound with a periodic time function and a harmonic spectrum. In a large organ of the Baroque era, such as the (extant, fully restored) Schnitger organ of St. Jacobi at Hamburg, there are more than 4,000 pipes (see Ahrend 1995), and even in smaller instruments, there are several hundred sound generators each tuned to a certain pitch. Thus, any combination of pipes will produce a complex harmonic sound where many spectral components carry energy and reinforce the tonal structure (e.g., in a major chord played with *organo pleno* registration).

As has been pointed out, the early Blockwerk organ (without any facility for registration) must have had a complex sonority for every key (in historic treatises, the expression indeed is "die Orgel schlagen"). From medieval treatises on music theory and organology, we may assume that the early Blockwerk was tuned to a chain of pure fifths. Thus, the tuning was 'Pythagorean' based on ratios such as 3/2, 4/3, 9/8, 81/64, etc. To be sure, these ratios concern the horizontal dimension of tuning (i.e., the distances between fundamental frequencies of the tones within a scale which, in modern times, can be expressed in Hz or in cents calculated therefrom). The vertical dimension of tuning, in a Blockwerk as well as in various mixture stops and special pipe ranks consisting of several rows of pipes (e.g., the Rauschquinte or the Sesquialter) concerns the relations of fundamental frequencies of the pipes within that pipe rank and in particular, the interval and frequency relations of pipes activated by each key of the manual. The pitch intervals in the vertical direction were always (and still are) tuned to small integer ratios, that is, in just intonation, in order to produce a high degree of spectral fusion and harmonicity.

As we know from Arnaut de Zwolle (see above), the Terzzimbel, which comprises pipes tuned to sound as just major thirds 5/4 and major chords, was introduced quite early into organ building. This did not cause problems as long as the music played on a Blockwerk may have been restricted to hymns or other melodic formations. In this case, every note played on the keyboard would produce a rich harmonic sound from the group of pipes assigned to a particular key (similar to a modern synthesiser with several harmonic oscillators where a complex sound can be activated from a single key). The problems began when separate ranks of diapason pipes and extra manuals were added to the Blockwerk and when musical settings or improvisations played on those organs included polyphony (already beginning in the 14th century and clearly so in works of the 15th century; see Klotz 1975). Though simultaneous intervals according to music-theoretical rules of that era were restricted to perfect consonances (the octave 2/1, the fifth 3/2, the fourth 4/3), in the 15th century, major thirds appeared in various sources (like the well-known Buxheimer Orgelbuch). Given that the horizontal tuning in the keyboard was still Pythagorean, most of the major thirds in a twelve-note scale would be of the size of a 'ditonus,' comprising two whole tones 9/8, which results in the 'Pythagorean major third' of 81/64 (of 408 cents). The Terzzimbel, however, had just major thirds 5/4 (of 386 cents). Playing the interval of a major third on the keyboard and at the same time activating the pipes of a Terzzimbel would inevitably bring about a controversy of two major thirds differing in interval size by a so-called comma of ca. 22 cents. The sound of two groups of pipes which differ in their tuning by ca. 22 cents, results in severe amplitude modulation, which listeners sense as roughness.

The problem was solved, in one essential aspect, when the so-called meantone temperament was introduced into keyboard tuning ca. 1500 (or shortly after). The term ‘meantone’ (which was coined in the 19th century) refers to the fact that, in this tuning, an interval of a just major third like c–e (386 cents) is divided, by the tone d, into two intervals of equal size (193 cents). However, the so-called quarter-comma meantone temperament (see Schneider & Beurmann 2017) was derived from practical tuning based on four pairs of major thirds like $b_\flat$–d–$f^\#$, f–a–$c^\#$, c–e–$g^\#$ and $e_\flat$–g–b as is evident from the tone lattice A shown in Table 7.

Table 7. Tone lattice for A: Meantone tuning and B: Just intonation pitches

	A: Meantone tuning	B: Just intonation pitches
-2	$f^\#$ ↔ $c^\#$ ↔ $g^\#$	$c^\#$ — $g^\#$ — $d^\#$ — $a^\#$ ....
	∣ ∣ ∣	∣ ∣ ∣ ∣
-1	d ↔ a ↔ e ↔ b	a — e — b — $f^\#$ — $c^\#$ — $g^\#$ ....
	∣ ∣ ∣ ∣	∣ ∣ ∣ ∣ ∣ ∣
0	$b_\flat$ ↔ f ↔ c ↔ g	f — c — g — d — a — e — b ....
	∣	∣ ∣ ∣ ∣ ∣ ∣ ∣
+1	$e_\flat$	$d_\flat$ — $a_\flat$ — $e_\flat$ — $b_\flat$ — f — c — g ....

In the left tone lattice, the sign|denotes a just major third 5/4, and the sign ↔ denotes a fifth narrowed by one-quarter of a comma, that is by ca. 5.5 cents (the ‘meantone fifth’ thereby is ca. 696.5 cents). In the right lattice, the|also denotes a major third 5/4, and – denotes a pure fifth 3/2. Tones in the -1 row are flat by one comma (22 cents) relative to the pitch of a tone of the same name in the row marked 0 (which represents the basic chain of just fifths). Tones in the -2 row are two commas flat, and tones in the $+1$ row are one comma sharp relative to their equivalents in the 0 row (the lattice of just intonation tones can be extended in the horizontal and in the vertical as needed but is restricted here to the section shown for demonstration). From the two schemes, it is evident that they share the formative interval of the just major third in the vertical and that they differ in the size of the fifths. The main reason is that tuning an instrument to just intonation ratios requires more than 12 pitches and tones per octave since it distinguishes between sharps and flats; for example, an E-major chord needs the $g^\#$ as major third, a f-minor chord needs an $a_\flat$ as minor third, etc. With but twelve tones and pitches to be tuned per octave and the decision for implementing as many just major thirds as possible (which, in addition, would bring about just minor sixths 8/5 of 814 cents), so-called ‘quarter-comma meantone temperament’ with no less than eight just major thirds was the best possible solution. This tuning soon became widespread and was reflected also in works of Renaissance and Baroque keyboard music which indeed feature the 'sweetness' of just major thirds and minor sixths. A piece that clearly demonstrates such features is John Dowland's 'Lachrimae Pavan' (originally for lute) which was set, with variations, for organ and other keyboard instruments by composers like William Byrd, Jan Pieterszoon Sweelinck, Peter Philips, Melchior Schildt, and Heinrich Scheidemann. These variations sound great when played on an organ or other keyboard instrument tuned to 1/4-comma meantone temperament because of the high degree of fusion in simultaneous major thirds and minor sixths. However, while there are eight just major thirds in 1/4-comma meantone tuning, the other four are far too wide (at 428 cents), and some of the minor thirds are far too narrow. Besides the slightly narrowed 'meantone fifths' of 696.5 cents, there is one very poor fifth at $g^{\#}$ – e_b of 738.5 cents (blamed as the 'howling wolf'). In effect, 1/4-comma meantone offers a number of highly consonant major chords (B_b, F, C, D, A, E, E_b) as well as a number of harmonious minor chords (c, d, a, a, e, $f^{\#}$, $c^{\#}$). In contrast, some major and minor chords sound rather harsh, particularly those where the wide major thirds $c^{\#}$ – f, $f^{\#}$ – b_b, $g^{\#}$ – c, b – e_b of 428 cents are involved. To overcome these deficiencies, one had to avoid the poor intervals and chords (or could use them, in certain settings, as an expression of grief and pain as was appropriate in the context of the musical 'Affektenlehre').

A technical and musical remedy suited to overcome the limits of 1/4-comma meantone was to increase the number of pitches per octave beyond twelve. A practical solution for harpsichords and organs was to provide extra strings and pipes so as to split $g^{\#}$ and a_b, e_b and $d^{\#}$, thereby eliminating not only the 'wolf' but also improving the compass of chords that can be played in acceptable quality (that is, with a sufficient degree of harmonicity and the absence of unbearable roughness). With 1/4-comma meantone tuning as background, the compass of keys and chords used in musical works, in general, was from A-Major to B_b-minor. Inserting the two tones/pitches a_b and $d^{\#}$ into lattice A above shows that the B-major chord and the A_b-major chord, the f-minor chord and the $g^{\#}$-minor chord are now at hand. The development of keyboard instruments with more than twelve keys/pitches per octave seems to have started in Italy, in the 16th century, in attempts at reviving classical Greek chromatic and enharmonic scale models to be used in contemporary music. In 1548, Zarlino had a harpsichord with 19 keys/pitches per octave, and Vicentino expanded the number of keys and pitches per octave to 31 (see Schneider & Beurmann 2017, 415ff.). Two such enharmonic instruments with 31 keys, the 'Clavemusicum omnitonum' built by Vitus de Transuntino in 1606 (see Barbieri 2008, 25f.), and a Hammerklavier from the late 18th century (Johann Jakob Könnicke, Vienna; see Barbieri 2005, 463ff.) have survived.

Though instruments with 17, 19, or even 31 keys per octave were rare since their construction was far from easy, the concept of adding two extra keys and pitches per octave into the keyboard of harpsichords as well as of organs to improve on the meantone tuning must have been more common. Werckmeister (1698, 79, 81) complained that one finds keyboards with three or more subsemitonia implemented; in his opinion, this was an obstacle to musical performance. At Hamburg, Gottfried Fritzsche (also Frietsch, he originally came from Meissen in Saxonia) built a new organ for St. Maria Magdalena in 1629, which, according to Mattheson (1721, 180f.), had several subsemitonia in each octave. In 1633/34, Fritzsche expanded the organ of the St. Petri church, where he supplied the HW with a new chest and added several stops; he implemented subsemitonia for $d^{\#}/e_{b}$, $g^{\#}/a_{b}$ and $a^{\#}/b_{b}$ in the HW as well as in the newly built BW (see Schröder 2006, 32). In 1635, he also must have implemented split keys for $d^{\#}/e_{b}$, $g^{\#}/a_{b}$ and $a^{\#}/b_{b}$ in the RP of the organ in St. Jacobi (see Fock 1939, 350). Organs with 14 or even 16 keys per octave were still found in England in the 19th century (and even new instruments with split keys were built there; see Williams 1968, 62f.). When, in the 1990s, a new organ was planned for Örgryte Nya Kyrka at Gotenburg that should emulate the large North German Baroque organs as were built at Hamburg by the Scherer family, Gottfried Fritzsche, and Arp Schnitger (as well as by Friedrich Stellwagen at Lübeck and Stralsund), a decision was made to incorporate split keys in this four-manual organ analogous to the extra keys Fritzsche had provided (see Speerstra 2003). The Gotenburg instrument also has split keys for $e_{b}/d^{\#}$ and $g^{\#}/a_{b}$ in the pedal.

Since split keys in organs meant extra costs for additional pipes and mechanics, and as organists perhaps found it difficult to master keyboards with 14 or even 15 keys, a less arduous way to deal with tuning problems was to try various temperaments (the Latin word ‘temperari’ means to balance) such as proposed by Werckmeister, Neidhardt, and other theorists and practitioners of music around 1700 (see Lindley 1987, Ratte 1991). A general tendency of most such proposals is to enlarge the just major thirds, which made up the core of 1/4-comma meantone tuning, and to widen the narrowed fifths of this system so as to approximate the 3/2 ratio. In the tuning scheme known as Werckmeister III (from 1691, see Rasch 1983), there are four narrowed fifths (similar to 1/4-comma meantone) while all other fifths are pure. Major thirds in this system vary from 390.2 to 407.8 cents, and minor thirds from 294.1 to 311.7 cents. Werckmeister III thus was a step back from a tuning based on just major thirds to a tuning based on pure fifths (like the Pythagorean system). While Werckmeister still maintained some grading within intervals and chords in regard to harmonicity vs. roughness, ET12 later levelled such differences. The advantage of a tuning like Werckmeister III understood as a ‘Wohltemperierung’ (making all major and minor chords acceptable though by no means equal like ET12) was that it allowed using most keys around the circle of fifths. As some impressive works of German organ music of the Baroque era are in E-major (a Praeludium and fugue by Buxtehude, BuxW 141, and a similarly complex work by Vincent Lübeck, LübWV 7), their rendition in 1/4-comma meantone is problematic since the B-dominant chord needed in E major suffers, with the pitches actually available in this tuning (see tone lattice A), from the false major third b–e_{b} (of 428 cents) and the narrowed fifth. It has, therefore, been suggested that those works (as well as works by other composers of the Baroque era, including J. S. Bach) would require a ‘Wohltemperierung’ like Werckmeister’s, as an adequate tuning system.

To assess the quality of a certain tuning or temperament like 1/4-comma or 1/5-comma meantone, Werckmeister III, ET12 etc. objectively, sound analysis directed to temporal and spectral parameters seems appropriate (see Schneider et al. 2004, Schneider & Beurmann 2017). In particular, measurements of all major and minor chords as played on an organ or harpsichord provide data for a comparative evaluation. The harmonic-to-noise ratio (HNR, see Boersma 1993), calculated from the periodicity of a signal as measured by autocorrelation or cross-correlation, shows differences between the twelve major and twelve minor chords of a chromatic scale for a given tuning as well as differences between several tunings (e.g., variants of meantone temperament, Werckmeister, ET12; see Schneider & von Busch 2015, Schneider, von Busch & Adam 2017, Schneider & Beurmann 2017). Empirical data from such measurements thus allow us to classify major and minor chords for each tuning in regard to spectral harmonicity where low HNR readings (quantified in dB) indicate a poor degree of harmonicity (the respective sounds are likely to give rise to a sensation of roughness). Higher HNR readings indicate that spectral components of the three complex harmonic sounds making up a chord are less divergent in frequency and amplitude, thereby enhancing the periodicity of the signal and that such chords are sensed as more consonant by listeners.

A comparison of tunings in use on pipe organs poses a problem, in one respect, since it is not possible, under realistic conditions, to change the tuning of a certain historic organ so that one could record sounds from pipes in meantone tuning on one day, and repeat the process a short time later after retuning the same organ to some other system. Thus, one has to compare sound data recorded from different organs. This, however, seems justified if the recordings can be done under nearly identical conditions from instruments of the same period, which have been restored recently according to the same criteria. For an actual comparison, we made recordings of the organ Arp Schnitger had built in 1688–90 for St. Mauritius at Hollern (Altes Land, close to Hamburg; see Edskes & Vogel 2009) and of the organ at St. Wilhadi in Stade (Altes Land), built by Erasmus Bielfeldt 1732–36 (see Vogel, Lade & Keweloh 1997). The Schnitger organ at Hollern is tuned to 1/4-comma meantone, the Bielfeldt organ at Stade to Werckmeister III. A comparison of the HNR data demonstrates that, for 12 major chords, meantone yields a number of relatively high readings (for C, D, E_b, E, F, G, A, B_b) in contrast to some poor ($C^{\#}$, $F^{\#}$, $G^{\#}$, B). In Werckmeister, differences between the 12 major chords are still present but not as large as in meantone. For the 12 minor chords, meantone again shows a clear pattern of higher vs. low HNR readings, and also Werckmeister exhibits an uneven pattern. However, the differences (expressed in dB) between individual minor chords are not as big as in meantone. In conclusion, a comparative evaluation of HNR data suggests that Werckmeister, on average, is more balanced than meantone in regard to major chords and, to a lesser degree, also minor. HNR data thus confirm the concept of Werckmeister III as a tuning suited to compose and perform organ music within a wider compass of keys. This advantage, however, is not without problems.

One has to remember that the pipe ranks in a Blockwerk and then in mixture stops were tuned to just intonation, typically in octaves, pure fifths and just major thirds. Stops like a Terzzimbel, Sesquialter or Terzian (see Mahrenholz 1942/1968, 228ff.) work very well in 1/4-comma meantone for those chords which incorporate just major thirds, but must be avoided in remote keys where the poor major thirds of 428 cents of the horizontal keyboard tuning would create roughness against the just major thirds of 386 cents from the Zimbel. As the contrast between good and poor chords in Werckmeister is less marked, adding a Zimbel to a standard registration like Prinzipal 8’, Oktave 4’ + 2’ perhaps would yield tolerable or even fair results in regard to HNR readings (which can be related to the psychoacoustic parameters of harmonicity vs. roughness). In Fig. 8, HNR readings for twelve major chords played on the HW of the Bielfeldt-organ at Stade with three stops (Prinzipal 8’ + Oktave 4’ + 2’) are shown in Fig. 8.

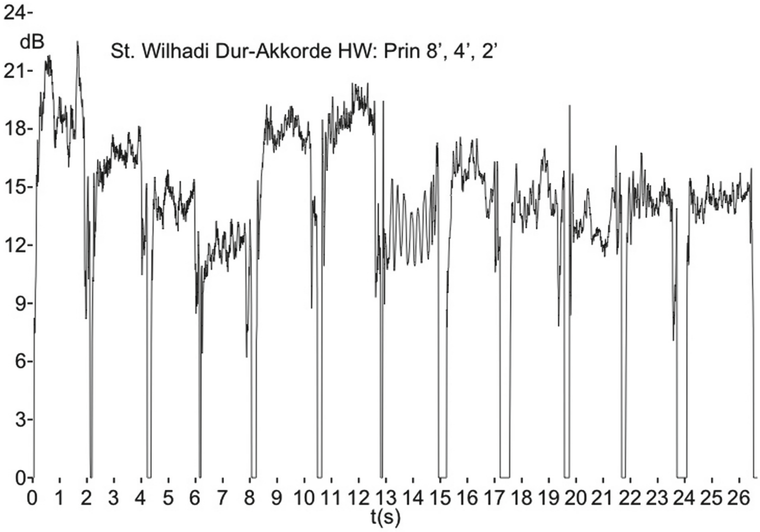


Fig. 8. Major chords, Werckmeister III, St. Wilhadi, Stade, HW, Prinzipal 8', Oktave 4' and 2'. The very short and high HNR at the onset of the A-major chord appears because the sound at the onset starts with a single harmonic partial corresponding to the 2nd mode of vibration in one of the pipes.

If the Cimbel (threefold) available in the same HW is added, the effect is significant, as Fig. 9 demonstrates (all sounds had been normalised to −3 dB before analysis). For all twelve chords, HNR readings are markedly lower, indicating the overall level of periodicity and harmonicity is reduced. Moreover, the flux in HNR over time, which indicates modulation effects such as AM and roughness, is already visible in Fig. 8 and increases significantly with the Cimbel added (Fig. 9). Though Werckmeister III tuning allows for a greater compass of usable keys, on the one hand, it does not work well with mixture stops which incorporate just major thirds, on the other.

The discrepancy between horizontal and vertical tuning encountered in Werckmeister III is of a more general nature. In a fundamental way, tunings based on pure or slightly tempered fifths (such as Werckmeister and ET12) differ from tunings based on just major thirds, such as 1/4-comma meantone and its expansions on keyboard instruments with 17 or more keys (see Barbieri 2008, Schneider & Beurmann 2017). The reason is that powers of one prime number do not equal powers of another prime number (e.g., $3^n \neq 5^m$), to the consequence that a Pythagorean major third 81/64, derived from four pure fifths 3/2 like c – g – d – a – e, differs from a just major third 5/4 by 21.5 cents (the so-called syntonic comma). This discrepancy, well-known from Greek musical theory, must have become a problem for organ builders in the 15th century when horizontal keyboard tunings most likely were still based on chains of pure fifths, while musical works demanded just major thirds as consonant intervals. Moreover, the Terzzimbel was invented as an organ stop complementing the usual mixture (based on pure fifths, see above), and thus actual sounds produced from a combination of horizontal and vertical tuning could have employed two different major thirds at the same time. Since the just

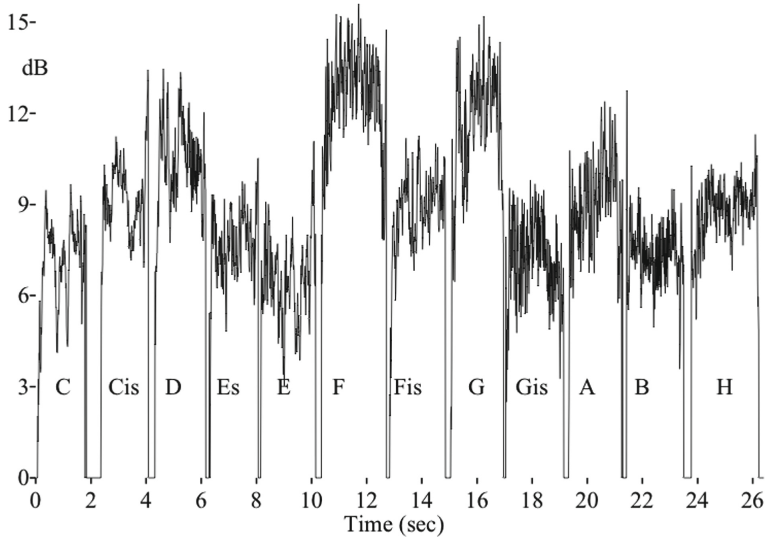


Fig. 9. HNR readings, Major chords, Werckmeister III, St. Wilhadi, Stade, HW, Prinzipal 8', Oktave 4'and 2' plus Cimbel (threefold).

major third 5/4 was accepted as a consonant interval of structural importance (most clearly by Zarlino 1558, 1573), keyboard tuning had to adapt to this situation and did so by inventing 1/4-comma meantone (which was outlined, in a practical way, by the organist and organ expert, Arnold Schlick in 1511). This tuning system saw a number of variants (like 1/5-comma) and extensions to more than 12 tones/pitches per octave in order to expand the compass of usable keys and chords but, in general, was implemented in its original form (see above). For example, it was found in the course of restoration that long flue pipes of the Schnitger organ at St. Jacobi of Hamburg (1689–93) had been left untouched in their 1/4-comma meantone tuning implemented by Schnitger (see Ahrend 1995).

Demands on tuning and temperament began to change around 1680–1720, when organists, many of them active also as composers (like Dietrich Buxtehude, Nikolaus Bruhns, Vincent Lübeck, Georg Böhm, and of course J. S. Bach) ventured into more remote keys (in the circle of fifths) and also used hitherto unknown chord progressions and modulations. As a parallel process, one has to note the transition from a predominantly modal organisation of music in the 16th and still in the 17th century to modern concepts of major and minor tonalities. While organ music preserved modal structures even in advanced compositions like Fischer's *Ariadne musica* (1702, 1710) and J. S. Bach's organ chorales as well as the Duetto BWV 802 from the third part of the *Clavier-Übung* (published 1739), elements of major/minor tonality are also ingredients of those works. It has been suggested that a performance of Bach's Fantasia in g-minor (BWV 542, coupled with a fugue) would need an organ "to be well-tempered, though not necessarily equal-tempered" (Williams 1980, Vol. 1, p. 120). In fact, the modulations found in this work are far-reaching (from D-major to $D_\flat$-major in bars 31ff.) and involve no less than 25 different pitches if one would intend to play this section in just intonation. If played on a ‘well-tempered’ organ tuned to Werckmeister III like St. Wilhadi at Stade, with a conventional registration (HW: Prinzipal 8, Oktave 8’, 4’, 2’; Pedal: Subbass 16’, Oktave 8’, 4’), HNR measurements yield the pattern shown in Fig. 10.

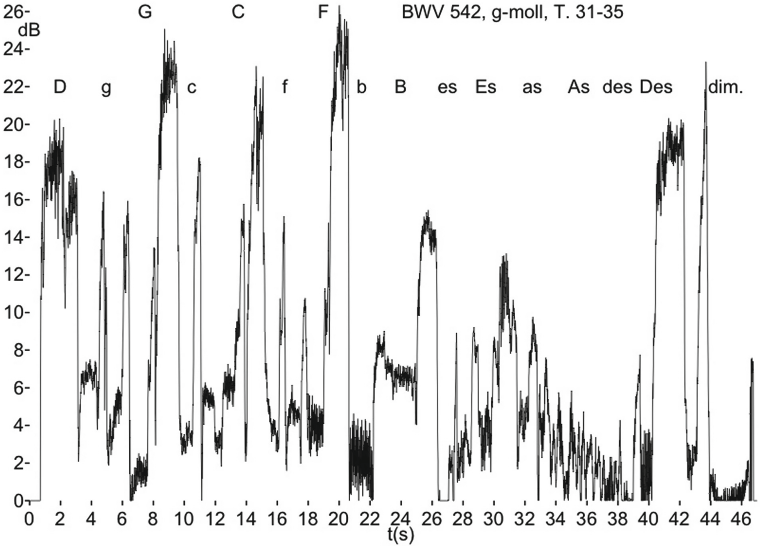


Fig. 10. J.S. Bach, Fantasia g-minor (BWV 542), bars 31–35, Werckmeister III tuning.

The modulation in this section proceeds around the circle of fifths from sharps to flats; the relevant chords in bars 31–36 are D – g – G – c – C – f – F – b$_\flat$ – B$_\flat$ – e$_\flat$ – E$_\flat$ – a$_\flat$ – A$_\flat$ – d$_\flat$ – D$_\flat$ – dim – e (Uppercase: Major, lowercase: minor). Figure 14 shows this modulation up to the first diminished chord (dim) in bar 35; in the Fantasia, it is followed by two more sonorities which resolve to an e-minor chord in bar 36. From Fig. 14, it is obvious that the Werckmeister tuning yields quite good HNR readings for a number of major (D, G, C, F, Des [D$_\flat$]) and minor chords (g, c, f) while some others are acceptable (B [B$_\flat$], Es [E$_\flat$], As [A$_\flat$], as [a$_\flat$], des [d$_\flat$]), and the remaining b [b$_\flat$] and es (e$_\flat$) are rather poor. Thus, there is a grading with respect to spectral fusion versus roughness, as one would expect from a temperament such as Werckmeister III. The problem that becomes obvious is that with advanced compositions employing far more than 12 tones (as identified from the notation), a tuning system suited to realise harmonic and melodic structures with precise intonation also would need more than 12 tones and pitches per octave. In this respect, common temperaments like Werckmeister III or ET12 fall short of providing adequate acoustical means for the performance of music that makes use of advanced harmony. After various temperaments had been explored, from ca. 1500–1800, in tuning organs and other keyboards (see Lindley 1987, Ratte 1991), ET12 finally was accepted as a standard mainly for practical reasons. In this process, the newly invented pianoforte played a “key” role because the vast number of instruments manufactured in Europe per year required certain conventions in regard to compass, tuning, and standard pitch (of $a^1 = 435$ Hz, which came as late as 1858 in France and 1885 as international agreement).

As a tuning and pitch system, ET12 incorporates slightly narrowed fifths and markedly enlarged major thirds as well as narrowed minor thirds. In effect, ET12 is closer to Pythagorean tuning than to just intonation. Since the octave is divided into 12 steps of equal size (100 cents), ET12 allows modulation from arbitrary starting points to whatever target (key/chord) is chosen. However, the difference between sharps and flats is levelled, and none of the intervals besides the octave is just. In regard to historic organs, their mixtures and stops including just major thirds like the Sesquialter and especially the Terzzimbel did not fit, in their sound structure of harmonic partials, to the horizontal ET12 tuning, which, unlike Werckmeister III or a similar 'Wohltemperierung', does not provide for some keys and chords with major thirds closer to the 5/4 than the 81/64 interval. As a matter of fact, when ET12 was implemented in historic organs all over Germany and adjacent regions, ca. 1780–1870, in particular, mixture stops with pipes tuned to major thirds were altered or completely removed. Quite many old reed-pipe stops (like Trechterregal, Bärpfeife, Dulzian) met the same fate and were dismissed for their rich harmonic sounds (that is, for the quality that once had made reed pipes so attractive for organ builders and musicians alike). Instead of such stops, pipe ranks with a more mellow sound emulating bowed strings and other 'orchestral colours' were installed on wind chests to accommodate a much different concept of organ music inspired by predominantly homophonic genres.

6 Concluding Remarks

The design and construction of pipe organs from medieval times to the Baroque era show remarkable achievements in regard to mensuration and technical manufacture of pipes, formation of pipe ranks and stops as well as setting up a disposition for each organ where stops combine into an overall sonic unit. At the same time, they maintain a characteristic timbral sound quality. Such a concept of 'diversity within unity' became evident in particular in the 16th century when organs were built, perhaps first in the Rhineland and the Low Countries (see Klotz 1975, 93ff.), with a growing number and diversity of stops. It is from this era that the typical tripartite organisation of stops results, that is, there are (a) diapason pipes of different foot lengths with relatively narrow diameters as well as mixtures and the occasional Zimbel; there are (b) flute-like stops with pipes of a wider diameter; and there are (c) various reed stops emulating reed instruments and horns of the Renaissance. In the course of the 16th and further, in the 17th century, the divisions of larger organs became well-equipped with stops from these three groups, whereby in particular, the pedal chest gained in volume and gravity. This was a condition prerequisite to assigning voices to the pedal for the performance of polyphonic settings such as bicinia, canons, or fugues. One can see a clear interdependency between developments in organ design and construction, on the one hand, and compositional practice, on the other. Organ building reached a zenith already around 1600, with a number of large three-manual organs (plus pedal) as listed by Praetorius (1619). It seems that Gottfried Fritzsche was the first to build a fourth division equipped with its own clavier as he expanded the BW in St. Jacobi, Hamburg (1635/36; Fock 1974, 55f.). This organ was enlarged and improved by Arp Schnitger (1689–93) and is fully restored (except that 1/5-comma meantone tuning has been substituted for the original 1/4-comma to allow for a wider range of usable keys). Another organ built by Schnitger with four divisions and four manuals for St. Ludgeri at Norden (1686–88, 1692, IV, Ped, 46 voices; see Vogel et al. 1997, Edskes & Vogel 2009) is of interest since it had to be fitted into a church of unusual architecture, where the nave is much lower in height than the transept and the choir. Schnitger chose to place the organ on a balcony on a side wall of the choir, which extends 'round the corner' into the transept. While HW, OW, BW and RP radiate their sound into the choir, most of the pipes in a single huge bass tower (which contains all pedal stops) 'speak' into the transept. The organ at St. Ludgeri proves masters like Schnitger could solve complex mechanical and even acoustical problems.

Significant developments in organ building between ca. 1500 and 1700 gave organists, many of them composers as well, ample opportunity to create a wealth of works written to be performed on 'the queen of all instruments', the pipe organ. As Praetorius (1619, 85) remarked, the organ should incorporate all other instruments by emulating their peculiar sound characteristics. The great variety of organ stops found in late Renaissance and Baroque organs and the diversity of sounds they produce must be regarded as an important part of our sonic and musical heritage.

Acknowledgements. The author made recordings of more than a dozen Baroque organs in Hamburg and Northern Germany from 1997 to 2017. He thankfully acknowledges the support of organists and colleagues who took part in these recordings on site, in particular Richard von Busch, Ingo Duwensee, Eberhard Lauer, Lüder Schmidt, Dorothea Schröder, Andreas Beurmann†, and Christof Adam. I'd like to thank Jürgen Ahrend and Wolfgang Weidenbach for permission to reproduce their photos (Figs. 1 and 3 of this article).

References

Abel, M.: Synchronization of organ pipes. In: Schneider, A. (ed.) Systematic and Comparative Musicology: Concepts, Methods, Findings. Frankfurt/M.: P. Lang, pp. 231–240 (2008)

Adelung, W.: Einführung in den Orgelbau, 3rd edn. VEB Breitkopf & Härtel, Leipzig 1976

Ahrend, J.: Die Restaurierung der Arp Schnitker-Orgel von St. Jacobi in Hamburg. Bestandsaufnahme und Restaurierungsbericht zusammengestellt von G. Lade, mit einem Beitrag von C. Edskes. In: Reinitzer, H. (Hrsg.). Die Arp Schnitger-Orgel der Hauptkirche St. Jacobi in Hamburg, pp. 127–265. Christians, Hamburg (1995)

Ahrens, Chr., Braasch, J., Schmidt, S.: Die Silbermann-Orgel in der Dresdener Hofkirche nach der Restaurierung: wie nahe sind wir dem Klang von 1755? Rubin. Wissenschaftsmagazin der Ruhr-Universität Bochum, Jg. 2006, no. 1, pp. 6–13 (2006)

Arnaut de Zwolle, H.: Les traités d'Henri-Arnault de Zwolle et de divers anonymes (Paris, Bibl. Nat. ms. latin 7295). Facs. ed. par G. Le Cerf & R. Labande, Paris 1932. Reprint Kassel: Bärenreiter 1972 (for the Latin text with translation and commentary see also Bormann 1966) (1447/1972)

Barbieri, P.: Enharmonic Instruments and Music 1470–1900. Il Levante, Latina (2008)

Beauchamp, J.: Analysis and synthesis of musical instruments sounds. In: Beauchamp, J. (ed.) Analysis, Synthesis and Perception of Musical Sounds, pp. 1–89. Springer, New York (2007). https://doi.org/10.1007/978-0-387-32576-7_1

Beurmann, E., Schneider, A., Lauer, E.: Klanguntersuchungen an der Arp-Schnitger-Orgel zu St. Jacobi, Hamburg. Systematische Musikwissenschaft/Systematic. Musicology **6**, 151–188 (1999)

Boersma, P.: Accurate short-term analysis of the fundamental frequency and the harmonic-to-noise ratio of a sampled sound. Inst. Phon. Sciences Univ. Amsterdam Proc. **17**, 97–110 (1993)

Bormann, K.: Die gotische Orgel von Halberstadt. Eine Studie über den mittelalterlichen Orgelbau. Merseburger, Berlin (1966)

Castellengo, M.: Acoustical analysis of initial transients in flute-like instruments. Acustica acta acustica **85**, 387–400 (1999)

Eberlein, R.: Orgelregister. Ihr Name und ihre Geschichte, 2nd edn. Siebenquart, Köln (2009)

Eberlein, R.: Die Geschichte der Orgel. Siebenquart, Köln (2011)

Edskes, C., Vogel, H.: Arp Schnitger und sein Werk. Bildband mit den erhaltenen Orgeln und Prospekten Arp Schnitgers. Bremen: Hauschild (Engl. edition: Arp Schnitger and his works. Documentation of the organs and façades built by Arp Schnitger in Germany, the Netherlands, Portugal, and Brazil. Bremen: Edition Falkenberg 2016) (2009)

Fischer, L., Bader, R., Abel, M.: Aeroacoustical coupling and synchronization of organ pipes. J. Acoust. Soc. Am. **140**, 2344 (2016)

Fletcher, N.: Transients in the speech of organ flue pipes – a theoretical study. Acustica **34**, 224–233 (1976)

Fletcher, N., Rossing, T.: The Physics of Musical Instruments. Springer, New York (1991). https://doi.org/10.1007/978-1-4612-2980-3

Fock, G.: Hamburgs Anteil am Orgelbau im niederdeutschen Kulturgebiet. Zeitschrift des Vereins für Hamburgische Geschichte **38**, 289–373 (1939)

Fock, G.: Der historische Orgelbau im Küstengebiet zwischen Hamburg und Groningen. Acta Organologica **1**, 11–20 (1967)

Fock, G.: Arp Schnitger und seine Schule. Ein Beitrag zur Geschichte des Orgelbaues im Nord- und Ostseeküstengebiet. Bärenreiter, Kassel (1974)

Gough, C.: Musical acoustics. In: Rossing, Th. (ed.) Springer Handbook of Acoustics, 2nd edn., pp. 567–701. Springer, Heidelberg (2014). https://doi.org/10.1007/978-1-4939-0755-7_15

Kaba, M.: Die römische Orgel von Aquincum (3. Jahrhundert). Bärenreiter & Akadémiai Kiadó, Budapest, Kassel (1976)

Kalähne, A.: Grundzüge der mathematisch-physikalischen Akustik. T. II, Leipzig, Berlin: Teubner (1913)

Klotz, H.: Über die Orgelkunst der Gotik, der Renaissance und des Barock, 2nd edn. Bärenreiter, Kassel (1975)

Lindley, M.: Stimmung und Temperatur. In Hören, Messen und Rechnen in der frühen Neuzeit (= Geschichte der Musiktheorie, Bd 6). Darmstadt: Wiss. Buchgesellschaft, pp. 109–331 (1987)

Lottermoser, W.: Klangspektren einer Silbermannorgel. Akustische Zeitschrift **5**, 324–330 (1940)

Lottermoser, W.: Orgeln, Kirchen und Akustik. Vol. 1: Die akustischen Grundlagen der Orgel. Frankfurt/M.: Das Musikinstrument (1983a)

Lottermoser, W.: Orgeln, Kirchen und Akustik. Vol. 2: Orgelakustik in Einzeldarstellungen, 2nd edn. Frankfurt/M.: Das Musikinstrument (1983b)

Mahrenholz, Chr.: Die Orgelregister, ihre Geschichte und ihr Bau, 2nd edn. Kassel: Bärenreiter (Reprint 1968) (1942/1968)

Marple, S.: Digital Spectral Analysis with Applications. Prentice-Hall, Englewood Cliffs (1987)

Mattheson, J.: Anhang von mehr als 60 Orgel-Wercken. In: Fr. Niedt. Musikalische Handleitung, T. 2, 2nd edn. Schillers Witwe & Kissner, Hamburg (1721)

Meyer, J.: Resonanzeigenschaften und Einschwingvorgänge von labialen Orgelpfeifen. Dr. Ing.-thesis, Technical Univ. Braunschweig. (1960)

Meyer, J., Bork, I.: Akustische Untersuchung der physikalischen Qualitätskriterien von Querflöten. (Technical Report, Physikalisch-Technische Bundesanstalt). Braunschweig: PTB (1987)

Mores, R.: Vowel quality in violin sounds – a timbre analysis of Italian masterpieces. In: Schneider, A. (ed.) Studies in Musical Acoustics and Psychoacoustics, pp. 223–245. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-47292-8_8

Mühle, Chr.: Untersuchungen über die Resonanzeigenschaften der Blockflöte. Dr. Ing.-thesis, Technical Univ. Braunschweig (Frankfurt/M: Das Musikinstrument 1979) (1966/1979)

Nolle, A., Finch, T.: Starting transients of flue organ pipes in relation to pressure rise time. J. Acoust. Soc. Am. **91**, 2190–2002 (1992)

Praetorius, M.: De Organographia. (=Syntagma musicum, T. II). Wolfenbüttel: Holwein (Repr. Kassel: Bärenreiter 1929) (1619)

Rasch, R.: Introduction. In: Werckmeister, A. (ed.) 1691, repr. Utrecht 1983, pp. 8–51 (1983)

Ratte, F.J.: Die Temperatur der Clavierinstrumente. Bärenreiter, Kassel (1991)

Runnemalm, A., Zipser, L., Ranke, H.: Structural vibration modes of a blown open pipe. Acustica acta acustica **85**, 877–882 (1999)

Sachs, K.-J.: Mensura fistularum. Die Mensurierung der Orgelpfeifen im Mittelalter. Bd 2, Murrhardt: Musikwiss. Verlagsges (1980)

Schlick, A.: Spiegel der Orgelmacher und Organisten (originally published by the author; reprint Mainz: Schott 1974) (1511)

Schneider, A.: Klanganalysen an Aerophonen der Volksmusik. In: Födermayr, Fr., Burlas, L. (eds.) Ethnologische, Historische und Systematische Musikwissenschaft. Oskár Elschek zum 65. Geburtstag. Bratislava: ASCO, pp. 51–80 (1998)

Schneider, A.: Psychophysics and psychoacoustics. In: Bader, R. (ed.) Springer Handbook of Systematic Musicology, pp. 555–741. Springer, Heidelberg (2017)

Schneider, A., von Busch, R., Schmidt, L.: Klanganalysen an Arp-Schnitger-Orgeln. In: Ristow, N., Sandberger, W., Schröder, D. (Hrsg.) »Critica musica«. Studien zum 17. und 18. Jahrhundert. Festschrift Hans Joachim Marx zum 65. Geburtstag. Stuttgart, Weimar: Metzler, pp. 247–270 (2001)

Schneider, A., Beurmann, A., von Busch, R., Schmidt, L., Lauer, E.: Über Beziehungen von Orgelstimmungen zu klanglichen Merkmalen. Ein Beitrag zur Diskussion über Orgelstimmungen aus akustischer Sicht. Acta Organologica **28**, 323–346 (2004)

Schneider, A., von Busch, R., Schröder, D., Schmidt, L.: Zur klanglichen Dokumentation von Denkmalsorgeln. Acta Organologica **29**, 405–426 (2006)

Schneider, A., von Busch, R.: Zur Verwendung der Viertelkomma-Mittelton-Stimmung auf historischen Orgeln: einige empirische Daten und musikalisch-akustische Kriterien. Acta Organologica **34**, 437–454 (2015)

Schneider, A., Beurmann, A.: Explorations in keyboard temperaments. Some empirical observations. In: Schneider, A. (ed.) Studies in Musical Acoustics and Psychoacoustics, pp. 405–444. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-47292-8_13

Schneider, A., von Busch, R., Chr. Adam,: Mitteltonstimmung und Wohltemperierung im Vergleich. Ars Organi **65**, 163–172 (2017)

Schröder, D.: Die Orgelgeschichte der Hauptkirche St. Petri vom Mittelalter bis 1885. In: Schröder, D. (Hrsg.) Gloria in excelsis deo. Eine Geschichte der Orgeln in der Hauptkirche St. Petri, Hamburg. Neumünster: Wachholtz, pp. 23–42 (2006)

Sundberg, J.: The Science of the singing voice. DeKalb, Il.: Northern Ill. U. Pr (1987)

Trendelenburg, F., Thienhaus, E., Franz, E.: Klangeinsätze an der Orgel. Akustische Zeitschrift **1**, 59–76 (1936)

Trendelenburg, F., Thienhaus, E., Franz, E.: Klangübergänge bei der Orgel. Akustische Zeitschrift **3**, 7–20 (1938)

Vogel, H.: Supplementary information (Booklet) to the LP 'Die Norddeutsche Orgelkunst I (Works of Heinrich Scheidemann and Melchior Schildt recorded on the Huß/Schnitger organ at St. Cosmae & Damiani of Stade). Hamburg (1982)

Vogel, H., Lade, G., Borger Keweloh, N.: Orgeln in Niedersachsen. Hauschild, Bremen (1997)
Werckmeister, A.: Musicalische Temperatur. Frankfurt, Leipzig: Calvisius (repr. Utrecht: Diapason Pr. 1983) (1691)
Werckmeister, A.: Erweiterte und vermehrte Orgel-Probe. Quedlinburg: Calvisius (1698)
Williams, P.: The European Organ 1450–1850. Batsford, London (1966)
Williams, P.: Equal Temperament and the English organ, 1675–1825. Acta musicologica **40**, 53–65 (1968)
Williams, P.: The Organ Music of J.S. Bach. 3 Vols. Cambridge U. Pr (1980–1984)
Williams, P.: The Organ in Western Culture, 750–1250. Cambridge U. Pr, Cambridge (1993)
Williams, P., Owen, B.: Organ. In: Sadie, S. (ed.) The New Grove Dictionary of Musical Instruments, vol. 2, pp. 838–916. Macmillan, London, New York (1984)
Wölfel, D.: Die wunderbare Welt der Orgeln. Lübeck als Orgelstadt. Schmidt-Römhild, Lübeck (1980)
Zarlino, G.: Le Istitutioni harmoniche. In: Venezia (3rd ed. 1573) (1558)

Exploring the Electroacoustic Music History Through Interactive Sonic Design

Anna-Maria Christodoulou(✉)

Department of Musicology, RITMO Centre for Interdisciplinary Studies in Rhythm,
Time and Motion, University of Oslo, Oslo, Norway
a.m.christodoulou@imv.uio.no

Abstract. Algorithms and related technologies are widely used for many musicology-related tasks, such as music analysis or even music composition. The use of algorithms in music analysis may be crucial for a deeper understanding of music theory and history, yet experiential knowledge is proposed here as a more interactive way to take a journey through music history and, more specifically, the evolution of electroacoustic music. From acousmatic music to serialism and from Musique Concrète and Elektronische music to Post-Schaefferian Electronica, numerous techniques have been developed for sound generation and manipulation. In this chapter, SuperCollider is used as a tool to create an interactive composition and to provide a walkthrough of electroacoustic music through live coding. The musicological aspects of the different composition techniques of this music style are explored through their integration into the algorithmic composition.

Keywords: Musicology · Music Technology · Electroacoustic music · Live Coding · SuperCollider

1 Introduction

Music creation and performance have faced a tremendous evolution with new technology tools. In music styles, such as electroacoustic and electronic music, methods like digital signal processing and algorithmic composition (creating music through a computer program) have become the core of the composition process, forming the future of music creation. Various tools are used to support such creative attempts, one of which is SuperCollider, an open-source interface and programming language created in 1996 by James McCartney (McCartney, 2002). SuperCollider is widely used by artists for algorithmic composition and live coding (live-scripting an algorithmic composition). Still, it also provides various libraries for researchers interested in manipulating and/or analyzing sound (Collins N., 2011). Therefore, this tool is helpful for the creation and study of musical sound. The syntax of SuperCollider is based on C++ programming language but has its own unique commands, adapted to the needs of sound manipulation and design. This tool is used to create the interactive script presented in this chapter: *SonicDesignHistory* (Christodoulou, 2023). The reason behind selecting this tool, apart from the open-source nature and the number of sound control possibilities, is the large

A. R. Jensenius (Ed.): SSD 2022, CRSM 12, pp. 259–270, 2024.
https://doi.org/10.1007/978-3-031-57892-2_14

community of artists and academics behind it. By attempting to create an interactive script in such a community, there is a clear aim to start a conversation about exploring music history through algorithmic compositions and providing useful tools that will inspire and assist many SuperCollider users.

Music can be used to convey information (Shelemay, 2006). More specifically, in this chapter, there is an assumption that music–more specifically, an interactive music script–can become an effective way to present the historical evolution of a music genre. The exploration begins with a focus on electroacoustic music history, using it as a starting point. To gain a deeper understanding of the electroacoustic music style, music theory and analysis are incorporated. In this endeavor, music technology is employed as the core method, with the script development being facilitated by the use of SuperCollider. The outcome of this attempt could also be characterized as a "lecture-recital" since it combines music history presentation through a music composition. More specifically, during the International Seminar of Sonic Design (2022), *SonicDesignHistory* was presented, where a selection of composers and techniques from the electroacoustic music scene was displayed in a historical sonic design walkthrough, which was scripted live. It is worth mentioning that interactive music notebooks have been created before for educational purposes (Horn, Banerjee, & Brucker, 2022) and data science (Hermann & Reinsch, 2021), but this was the first known creative sonic design attempt of a music history overview. It should be mentioned that a notebook here means an interactive script that contains code and text.

It is important to understand the primary intention of this composition and the reasoning behind selecting an algorithm for the presentation of a music history summary. First, getting a deeper apprehension of the algorithmic composition techniques is possible by investigating multiple sonic outcomes and testing different ways to implement a particular strategy. Also, through this investigation, it is clear that even though there can be a large amount of computer automation in the composition process, it is still a human-controlled music structure. Furthermore, interaction and experience are expected to be more effective in understanding a concept and maintaining the audience's attention. So, it is assumed that through such a presentation, the audience or the *SonicDesignHistory* users will get a clearer understanding of what electroacoustic music consists of and how it evolved over time. For me, the creator of the script and composer, this attempt is helpful to understand how the techniques work and distinguish outstanding elements of particular composers while carrying elements from their past.

The selected composition strategies that are presented in the *SonicDesignHistory* are assessed through their exploration of various pieces, taking into consideration their musicological importance and contribution. There is also an attempt to detect the elements of the various techniques that distinguish and unite the musical styles. This chapter describes the electroacoustic music composition process and the techniques I have used to achieve a historically faithful result. There is also a statement on the various components of the electroacoustic techniques and how each one inspired the creation of their legatees. Furthermore, there is a discussion about the various challenges faced in such an attempt and the prospective applications that can be developed further.

2 Methodology

I have chosen to work with sonic design and sonification to get a different grasp on music history than what is achieved through historiography with a verbal presentation of music history and music examples. The emphasis is not on the historical evolution of music as such, but mostly on the techniques and the artists that formed this history. The main method adopted for this project is the combination of various composition and sound processing techniques, taking advantage of multiple SuperCollider functions (as developed by James McCartney), useful examples from scientific sources (Karamanlis, 2021), and related work (LaFleur, 2020). *SonicDesignHistory* is based on the concept of unfolding musical sounds, which are created by borrowing elements from their previous sounds or sonic events. After integrating these elements, they develop their own unique sonic result, essentially a new musical event. In this case, a music event is an example of a specific electroacoustic music technique, and the music elements are the sound components of this technique.

SonicDesignHistory was created using the SuperCollider IDE. During execution, blocks of code are scripted and/or activated live, while explanatory comments guide the audience, providing information about the processes and techniques that are displayed during the performance, as seen in Picture 1. The composition is divided into three main categories, based on the historical era presented each time and with a special focus on the music of Europe and the USA. There have been multiple attempts to divide Electroacoustic Music into categories before, mostly based on the musical techniques (Manning, 2004) and the major computer developments that accompany the music evolution (Holmes & Pender, 1985). Based on these division attempts, I decided to categorize the historical eras as follows: Early Electroacoustic Music (1948–1960), Electroacoustic Music Evolution (1960–1990), and Digital Age (1990- today).

The main reason behind selecting composers originating from certain parts of Europe and the USA was my familiarity with the literature and the composing techniques, as well as previous interaction with the related compositions. The selection of composers was based on their influence on Western electroacoustic music evolution and their originality. On the other hand, many techniques were integrated into this script due to being influenced by previous composition norms (such as Stockhausen's integration of Schoenberg's 12-tone technique). In terms of the coding implementation of the techniques, it would be possible to reuse some commands from one technique to the other, making it possible to create a live coding manipulation of sound that encompasses the concept of a sound being "born" from its previous one (Fig. 1).

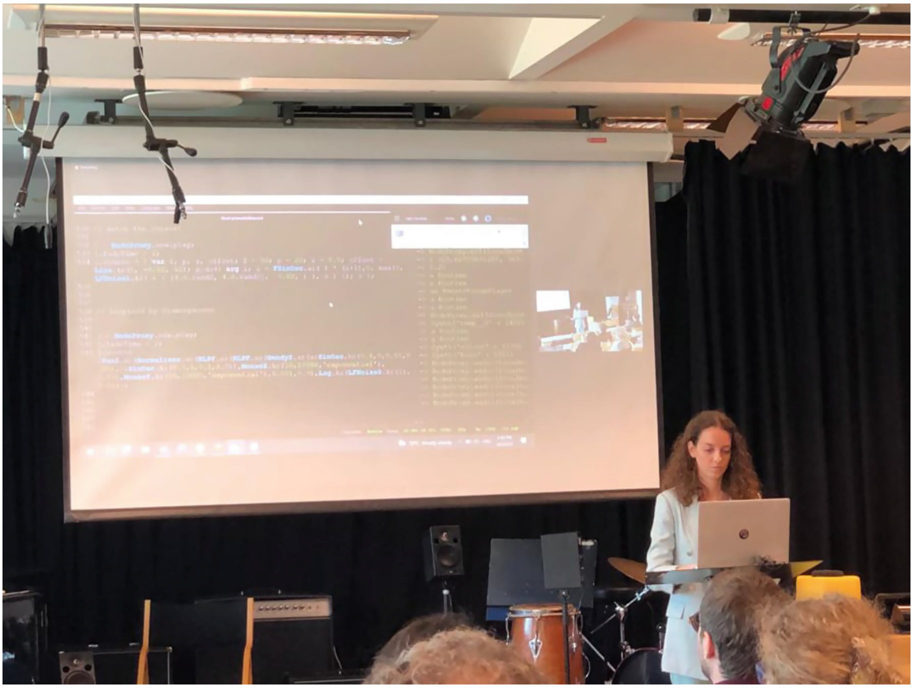


Fig. 1. Live presentation of SonicDesignHistory during the International Seminar of Sonic Design, 2022. The picture shows the script being executed live in the SuperCollider IDE, with comments on the screen guiding the audience through the process (Photo: Léo Migotti).

3 The Early Electroacoustic Music Era (1948–1960)

In the live version of *SonicDesignHistory,* an introductory sound is the first element of the composition. This sound does not add to the overall electroacoustic music history display; it is only produced to accompany the SynthDef activation, making this process more interesting for the audience. SynthDefs are synthesizer definitions used as sound-producing units. These classes are widely used in the script, and—as will be mentioned later on—they do not produce sound before activated. This first element of the algorithmic composition consists of a simple sound resulting from a fast sine oscillator (FSinOsc). The sound output was assigned to a NodeProxy, a placeholder for sound playing in the SuperCollider server. NodeProxies are chosen multiple times in the composition and let the user smoothly activate and deactivate the sound output while offering the possibility to change the sonic result in real-time. This initial sound is the base for the first technique, additive synthesis, one of the oldest and most studied composition strategies of this music genre (Karamanlis, 2021). The main goal of additive synthesis is the formation of a complex waveform by multiple simple – usually sinusoidal – waveforms (Karamanlis, 2021). Its concept originates from pipe organs, and their multiple register stops (Roads, 1995), while the actual idea comes from the Fourier Transform, which allows a complex waveform to be divided into multiple simple periodic waveforms (Karamanlis, 2021). The imitation of additive synthesis was achieved by using a class (Mix.ar) which mixed an array of four channels into one, creating a complex signal consisting of four different sinewave oscillators. To create a faster and easier-to-follow presentation, the SystemClock function was selected, which creates an automatic playback of complex signals in a certain number of seconds.

The first music era was dedicated to Europe and, more specifically, to the Studio for Electronic Music (WDR) in Germany. During the Early Electroacoustic Music era, Elektronische Musik was born in Cologne, introducing a new set of composition techniques, all of which included electronics. The first music event selected for this era is Herbert Eimert's approach to serialism, based on Schoenberg's 12-tone technique. This music segment was inspired by Eimert's "Klangstudie II" (1952). More precisely, one of the musical elements that is introduced is a set of delayed "bubbly" sounds created by combining delayed non-band-limited sawtooth and sinewave oscillators. A wall of reverberated, low-frequency sounds is created in the background, consisting of manipulated noise signals. Furthermore, a SynthDef class was created to resemble the sound of a piano instrument that could play 12 specific notes.

It is important to mention that in the live version of the composition, all SynthDefs were activated at the beginning of the script since they don't make any sound until they are activated. This is a common practice in live coding; it saves time and provides the audience with a simpler, more comprehensive algorithm to watch during the performance. The piece of code that contains the SynthDefs could be completely hidden from the audience during the live presentation. On the one hand, there is the goal to be able to present this script to an audience coming from backgrounds irrelevant to live coding or interfaces like SuperCollider, so it is important not to spend a lot of time showcasing these functions to avoid confusion. Furthermore, there is a clear aim that the interactive notebook is openly available to the audience, so it is important not to require any domain-specific knowledge to interact with it. On the other hand, a brief presentation of the SynthDefs was necessary for those curious to take a quick look into the composition's elements.

The second music event created for the Early Electroacoustic Music era in Cologne was based on Karlheinz Stockhausen's aleatory techniques. Therefore, this part of the script was inspired by the concept that some musical aspects are left to chance. Stockhausen used aleatoric techniques to provide the performers with freedom in sequence regarding the musical fragments, and one of the most notable examples of such an attempt is "Klavierstück XI" (1956). Here, only the concept of sequence freedom is borrowed, and the twelve tones that were selected for the previous music event are played randomly, using a pattern object (Prand) that would randomly select an item from a defined list. Here, there was a list consisting of twelve tones (frequencies) and another list consisting of three SynthDefs (piano, string, and bell instruments). This enabled the creation of a chaotic, random-sounding event that characterized some compositions of Elektronische Musik.

The next part of the sonic design was another important European city: Paris, France. Here, Musique Concrète was born, often considered the 'polar opposite' of Elektronische Musik, mostly because the artists of Musique Concrète used recorded sounds as their input for the compositions. This was opposed to Elektronische Musik, in which the artists would create their own sounds electronically. Musique Concrète also refers to how the composers would work and manipulate directly the so-called "sound objects" (Schaeffer, 1966). Therefore, it was important to create a music event that did not emerge from the previous one but contrasted it in an obvious way. Musique Concrète was developed first in the history of electroacoustic music, but for aesthetic reasons, in my composition, it was presented after Elektronische Musik.

The only composer who was selected to represent Musique Concrète in this attempt was one of the most important figures: Pierre Schaeffer. The technique that was selected for this music event was tape manipulation, and the piece of code that was created as an example was inspired by his composition "Symphonie pour un homme seul" (1949). Here a pre-recorded sound file of female vocals was used, and random slices of that file were selected using the pseudo-random generators (TWChoose) of SuperCollider. At the same time, Schaefferian typology was taken into consideration for the introduction of three main fracture types (the energy envelopes of the sound objects): impulsive sounds (fast and short), sustained sounds (prolonged with steady energy), and iterative sounds (stream of impulses) (Godøy, 2021).

As mentioned earlier, composers were selected from certain regions of Europe and the USA. After exploring some main aspects of Musique Concrète and Elektronische Musik, the next part displays certain elements of the Early Electroacoustic Music era in the USA. One composer selected for this music was Steve Reich and his phase-shifting composition technique. More specifically, a modified, pre-existing attempt to algorithmically recreate Steve Reich's "Piano Shift" (LaFleur, 2020) was implemented in *SonicDesignHistory*. Here a SynthDef was created to resemble a piano sound, different from the one that was used in (LaFleur, 2020), while the note playback strategy remained the same. Two global variables were defined, one that stored the MIDI values for the notes and one that stored their timing. These notes are played every second by the SuperCollider routine ~steady (LaFleur, 2020). In this composition, one pianist is speeding up to put the second pianist out of phase until they are synced. To recreate this technique computationally, the instrument is enclosed in a routine called ~phasing (LaFleur, 2020).

The final music event of the Electroacoustic Music era was dedicated to John Cage. There was a reference to his work "4:33" (1952), which was also used as a creative transition between the two eras. Therefore, all the previous sounds and sonic events were gradually silenced by using fading-out and release functions. This passage was not four minutes and thirty-three seconds as in the original form, but thirty seconds, allowing the audience and the room to take part in the composition by letting their "unintended" sounds be heard, according to the original concept of Cage's work (Davies, 1997). This transition was also practically useful for creating this script since it was challenging to find common elements that would assist the smooth transition from the complicated wall of sound of the Early Electroacoustic Music era to the simple, low-frequency sound that would initiate its Evolution era.

4 The Electroacoustic Music Evolution Era (1960–1990)

The Electroacoustic Music Evolution era is characterized by integrating computers into the composition process, occasionally giving these computers the freedom to make decisions for the music creation and production (Serra, 1993). Performers and composers of this time would take advantage of the digital processes, both for creating modern instruments and for sound transformation, providing endless opportunities for novelty and original creation (Emmerson, 2001). Electroacoustic music performances gradually integrated digital devices that allowed the performers to encode and process note information in real-time, resulting in the so-called “interactive compositions” (Emmerson, 2001). One of the composers who lived during this era and took advantage of the new technologies in his work is Ioannis Xenakis. Xenakis is selected as a representative of the Electroacoustic Music Evolution era. I believe that he signifies this era by his original way of composing and integrating natural sciences into his work, while his stochastic music approach can be used as an illustration of the musical advancements of this time.

Stochastic music was created with the composer’s aim to arrange the music structure using probability calculus (Manning, 2004). Xenakis’ composition “Diamorphoses” (1957-8) was the main inspiration for this era. First, a wall of sound was built by combining a fast sinewave oscillator (FSinOsc) and a low-frequency noise with random frequency values in a loop. In SuperCollider, there are three different dynamic stochastic synthesis generators named after Xenakis’ GENDY model: Gendy 1, 2, and 3. These allow the user to initialize a set of memory with X number of points that are modified one by one with each new period. Dynamic stochastic synthesis is a process during which probabilistic waveforms are generated after being stochastically calculated. Here, a dynamic stochastic synthesis Generator (Gendy2) was used, accepting a sinewave oscillator as a parameter for a random number generator (Lehmer) and another sinewave oscillator as a parameter for another random number generator, both perturbed by Xenakis.

5 The Digital Age (1990-Today)

After a brief display of Electroacoustic Music Evolution, the last era in my script is the Digital Age. The integration of noise as a major part of the composition process is a well-known element of this era, although it is not an innovative practice. Russolo (1885–1947) had already worked on mechanical noise-producing instruments decades prior to the Digital Age (Holmes & Pender, 1985). Noise is one of the elements that are shared among all of the eras. In the Early Electroacoustic Music era, Schaeffer aspired to combine music and noise in his work, while ambient noise was a common substance of Elektronische Musik (Holmes & Pender, 1985). In fact, in the NWDR studio, white-noise generators were quite common. A good example of a music art integrating noisy fragments would be Eimert’s “Klangstudie I,” where “noises appear into washes of echo frizz” (Holmes & Pender, 1985). Cage also had obvious influences of noise blending into his work (such as “Fontana Mix”). It is important to mention that until the Digital Age and the general evolution of recording with digital means, the existence of noise was sometimes inevitable. This could be one of the reasons behind various attempts at creative noise integration into music art. It should be noted, though, that the amount of noise was mostly controlled by the composers.

With the digitalization of music recordings, noise manipulation remained an important composition technique. In the script, a white noise (WhiteNoise.ar) is generated, instantiating the first music event of the Digital Age. This signal is transformed into a hi-hat sound by implementing a high-pass filter (RHPF.ar) (Karamanlis, 2021). Using this as a beat, the next music event is the introduction of microsounds and glitches. More specifically, a SynthDef (Rumush, 2015) designed to imitate a chaotic wall of glitch sounds was manipulated in a way that would match the aesthetics of the current work. This instrument consisted of one bass-like sound from a sinewave oscillator, three tone-like sounds generated from sinewave oscillators, one of which is placed in the stereo field, one pink noise generator (PinkNoise.ar), and an impulse oscillator (Impulse.ar), on which a resonant low-pass filter was applied.

The next music event of the Digital Age covers the creation of ambient sounds. Ambient sounds were introduced early in the 1960s with artists such as De Maria and Varèse, as well as later during the Electroacoustic Music Evolution with Brian Eno and Harold Budd (Holmes & Pender, 1985). Ambient music was correlated to atmospheric soundscapes and background decoration of public spaces, such as airports (Holmes & Pender, 1985). The main idea was to create a musical piece that would make the audience pay attention to everyday sounds that are otherwise ignored (Manning, 2004). It was fused with music styles such as jazz and electronic, but it gradually got a unique identity as a separate music style, known as "ambient" or "space music." With the technological developments in music creation, other genres, such as pop or rock, became more popular. Aphex Twin made the music style relevant again during the 90s with the publication of "Selected Ambient Works'85 - '92" and "Classics" (1995) (Manning, 2004). In this interactive algorithmic composition, the ambiance is presented using an instrument designed by (Karamanlis, 2021) to imitate a "relaxed" sound environment. The main element of this instrument was the use of a bank of fixed-frequency resonators (Klank) which can be used to simulate the resonant modes of an object.

The final example of this era—and this interactive script in general—was the creation of a soundscape. A sinewave oscillator and a dynamic stochastic synthesis generator consisting of three other sinewave oscillators were selected and placed in the stereo field, creating a windy soundscape that would gradually become the only element of the composition. As this music event was activated using a NodeProxy, the rest of the instruments and players faded out. It is now clear that the structure began with simple sounds that gradually led to a chaotic burst of sound, abruptly cut by silence, and gradually building up to yet another chaotic wall of sounds that faded out with the creation of an ambient environment, which was then reduced to a windy soundscape and led naturally to the termination of *SonicDesignHistory*.

6 Similarities and Differences

The presentation of methods and techniques used in the interactive system shows that there are plenty of elements that unite and distinguish the different music styles and eras. To create something unique and innovative, many composers borrowed previous techniques and advanced them with respect. One example is the use of filters, which were used for sound manipulation throughout the whole electroacoustic music history, in the earlier time as delays and reverbs and later as low-pass filters for low-frequency and noise signals manipulation. Also, between the Early Electroacoustic Music era and the Electroacoustic Music Evolution, there is a common element of randomness. Stockhausen let performers select their own sequence in which they would play a music part, restrained to a predefined circle (Manning, 2004). Schaeffer would select random pieces of an audio file to achieve the desired sound montage. Xenakis used Lehmer's random number generator to control the music structure based on mathematical sequences. None of them used randomness in a way that would create chaos or absolute freedom.

Despite their commonalities, many new strategies resulted in the creation of new eras characterized by novel sound identities. For example, Schaeffer's tape manipulation was a unique practice, as well as Reich's phase shift and Cage's emphasis on silence. Of course, it is worth mentioning that even between artists of the same music era, there were contradictions and completely opposite composition directions, as happened with Musique Concrète and Elektronische Musik. However, it is worth mentioning that there was also the common element of electronic sound, either "pure" or with integrated acoustic elements.

My exploration aimed to achieve a historically faithful result by reading up on multiple sources, such as books and journals, and consulting historic musicologists. I believe that it was indeed a successful attempt, having a clear walkthrough of some of the main parts of electroacoustic music history. The presentation received positive comments during the Sonic Design Seminar, but in the future, it would be highly interesting to gather user feedback and perform a scientific evaluation of the script. The composers of electroacoustic music are not restricted to the ones selected here, and neither are the composition methods, but the goal was to present a brief overview of some of the selected strategies in Europe and the USA. Also, the relationship between sound and algorithms became clear through investigating techniques that define the sound by manipulating it in a certain way. It was also clear how complex this form of synthesis can be and how close it is to human-made art.

There were plenty of challenges throughout the conception of this creative attempt. When the purpose is only to introduce creative activity, one can justify their choices mainly on their vision and personal expression. Here, this wasn't the case, since apart from the creativity aspect that was crucial, it was also important to present a meaningful structure and convey substantive information regarding the evolution of electroacoustic music history. An important challenge was the selection of composers and techniques, as well as a meaningful categorization of these in a clear algorithmic structure. Furthermore, even though, in theory, many of the composition techniques seem to share some commonalities, it is not always the case when it comes to their algorithmic implementation, and the same applies to the resulting sound. Therefore, important decisions had to be made regarding the placement of several techniques within the script. Another important challenge was the simulation of analogue techniques in a digital environment. There were sometimes differences in sound, and it was difficult to achieve certain results. However, the resulting sound was very close to the desired outcome, and it seemed like it was feasible to imitate both the techniques and the sounds of the electroacoustic music scene using algorithmic methods.

As far as the live coding was concerned, interaction was an important factor. The script has been made for audiences unfamiliar with algorithmic composition and live coding, so all the information should be presented concisely. For this, I have developed an oral presentation preceding the live coding session to guide the audience's attention and help them understand the script's syntax.

7 Conclusion

This project has led to a deeper understanding of various electroacoustic music composition strategies, including their historical development and implementation. Familiarization with the techniques was essential to provide the comfort to interact with them creatively. The interactive script was structured to be conveyed easily to the general public. It is also important to address the aesthetics that this particular form of music technology provides for music creation. Therefore, the techniques and styles that were imitated were put together in a way that would be aesthetically pleasing for the audience.

I believe there is high musicological value in creating such interactive scripts. It is a new way to present musical information, and the interactivity makes it both playful and helpful. While my project was mainly analytical in nature, it could expand music creativity while combining music theory and history at the same time.

In the future, it would be interesting to explore more aspects of electroacoustic music history. For example, more composers and techniques could be included, such as Stockhausen's envelope design and techniques like ring modulation. It would be interesting to include music from other countries, beyond the current European and North American examples. Also, it would be intriguing to present multiple ways to algorithmically implement the same technique and examine all the possibilities and sound outcomes. The notebook is hopefully a source of inspiration to other SuperCollider users and live coders, and it could potentially start a discussion and evolution of technologically-mediated music history studies by enabling an open collaboration and sharing of ideas.

Finally, future work includes creating an online interactive application to assist education. More specifically, developing a way to teach music history, algorithmic composition, and interactive sonic design with descriptive comments and valuable sources will be relevant. This could be a new way to teach music history and the theory behind the multiple techniques. Many of the techniques are theoretically complex, so practice and interaction are helpful for overall comprehension.

Acknowledgments. I would like to express my warmest gratitude to Prof. Orestis Karamanlis from the University of Athens for providing me with useful knowledge on SuperCollider and for all the feedback and support throughout this work, Prof. Rolf-Inge and Dr. Olivier Lartillot from the University of Oslo for providing me with useful sources and constructive feedback, and Mrs. Marcia M. Spence for providing me with useful tips on writing this chapter.

References

Christodoulou, A.: Exploring the electroacoustic music history through interactive sonic design. Zenodo (2023). https://doi.org/10.5281/zenodo.8296794

Collins, N.: SCMIR: a SuperCollider music information retrieval library. In: ICMC 2011, pp. 499–502. Michigan Publishing, Huddersfield (2011). https://composerprogrammer.com/research/scmir.pdf

Davies, S.: John Cage's 4′33″: is it music? Aust. J. Philos. **75**(4), 448–462 (1997). tandfonline.com

Emmerson, E., Smalley, D.: Electro-acoustic music (2001). *Electro-acoustic music | Grove Music (ionio.gr)

Godøy, R.I.: Perceiving sound objects in the musique concrète. Front. Psychol. (2021). https://www.frontiersin.org/articles/10.3389/fpsyg.2021.672949/full

Hermann, T., Reinsch, D.: sc3nb: a python-supercollider interface for auditory data science. In: AM 2021: Audio Mostly 2021, pp. 208–215 (2021). https://dl.acm.org/doi/10.1145/3478384.3478401

Holmes, T., Pender, T.M.: Electronic and Experimental Music. Charles Scribner's Sons, New York (1985)

Horn, M., Banerjee, A., Brucker, M.: TunePad Playbooks: designing computational notebooks for creative music coding. In Barbosa, S., et al (eds.) CHI 2022: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, pp. 1–12. Association for Computing Machinery, New Orleans (2022). https://dl.acm.org/doi/pdf/10.1145/3491102.3502021

International Seminar on Sonic Design. Oslo, Norway (2022). https://www.hf.uio.no/imv/english/research/news-and-events/events/conferences/2022/sonic-design/index.html

Karamanlis, O.: Σύνθεση και Επεξεργασία Ήχου με το SuperCollider (2021). https://synthesis.orestiskaramanlis.net/

LaFleur, E.: Programming Steve Reich's piano phase in SuperCollider (2020). www.ezralafleur.com, https://www.ezralafleur.com/programming-steve-reichs-piano-phase-in-supercollider/

Manning, P.: Electronic and Computer Music. Oxford University Press (2004)

McCartney, J.: Rethinking the computer music language: super collider. Comput. Music J. **26**(4), 61–68 (2002). https://www.jstor.org/stable/3681770?casa_token=jBJFIKOk19MAAAAA%3A-EUaP37_iDYujFEtoSF7NKpLHnuW3qNr6cD1b5zII1PHXoEEvfqPPu0fvQDmL1x4v0rGjnj3UuHwekN72BDw-GzMA5sa5y5a-XCSip0nqHMv72M1zw#metadata_info_tab_contents. Languages and Environments for Computer Music

McCartney, J.: SuperCollider 3.12.2 (n.d.). https://doc.sccode.org/

Roads, C.: The Computer Music Tutorial. MIT Press, Cambridge (1995)

Rumush. Chaos Glitch SynthDef» by rumush (2015). https://sccode.org/1-50l

Schaeffer, P.: Traité des objets musicaux. Éditions du seuil, Paris (1966)

Serra, M.: Stochastic composition and stochastic timbre: GENDY3 by Iannis Xenakis. Perspect. New Music **31**(1), 236–257 (1993). https://doi.org/10.2307/833052

Shelemay, K.K.: Music, memory and history: in memory of stuart feder. Ethnomusicol. Forum **15**(1), 17–37 (2006). https://www.jstor.org/stable/20184538?casa_token=Gv0shD3TuVsAAAAA%3Aw0tAUgRcpBGjmVb3Du4IEYoXjxYzGsj68UOK3Yugjz2po3XpUlez3wxFh0a5MxHZmTPnr1EnV1jWQQETtZziQOte9yAE_6bomEQdfN_YsGTGMRqLWA&seq=17#metadata_info_tab_contents

Wilson, S., Cottle, D., Collins, N.: The SuperCollider book (2011). https://dl.acm.org/doi/book/10.5555/2016692

Musicological and Technological Perspectives on Computational Analysis of Electroacoustic Music

Olivier Lartillot(✉)

RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion,
Department of Musicology, University of Oslo, Oslo, Norway
`o.s.g.lartillot@imv.uio.no`

Abstract. Analysing electroacoustic music remains challenging, leaving this artistic treasure somewhat out of reach of mainstream musicology and many music lovers. This chapter examines electroacoustic music analysis, covering musicological investigations and desires and technological challenges and potentials. The aim is to develop new technologies to overcome the current limitations. The compositional and musicological foundations of electroacoustic music analysis are based on Pierre Schaeffer's *Traité des objects musicaux*. The chapter presents an overview of core analytical principles underpinning more recent musicological approaches, including R. Murray Schafer's soundscape analysis, Denis Smalley's spectro-morphology, and Lasse Thoresen's graphical formalisation. Then the state of the art in computational analysis of electroacoustic music is compiled and organised along broad themes, from detecting sound objects to estimating dynamics, facture and grain, mass, motions, space, timbre and rhythm. Finally, I sketch the principles of what could be a *Toolbox des objets sonores*.

Keywords: Electroacoustic music · Music analysis · Music Information Retrieval · Musicology · Computational analysis

1 Introduction

Sound design has been elevated as a sound art through the rise of "*musique concrète*" (*concrete music*), introduced by Pierre Schaeffer around 1948. This music composition technique uses recorded sounds as raw material and, through the later addition of electronic sound production, has been generally known under the term *electroacoustic music*. Today, electroacoustic music encompasses many styles—from purely concrete or electronic to hybrid forms that include instrumental performances—in both "academic" and "popular" styles. In a project to establish a corpus of historical electronic music [1], the authors found it hard to trace a categorical boundary between art music and popular music. The present study does not focus on a particular type of music but rather on the sound qualities of the music, whatever it is, leaving aside more traditional

A. R. Jensenius (Ed.): SSD 2022, CRSM 12, pp. 271–297, 2024.
https://doi.org/10.1007/978-3-031-57892-2_15

musical aspects (mainly related to pitch and tonality) already addressed by traditional musicology. The focus will be on electroacoustic music but with an eye to possible applications of the methodologies for other types of music.

This chapter's topic of interest is related to the *analysis* of the sound qualities of electroacoustic music. This has remained a marginal academic activity compared to the analysis of more traditional types of music, "classical" music in particular. For instance, the reference articles about music analysis, cf., [2–4], do not even mention electroacoustic music [5]. Furthermore, many electroacoustic music composers do not consider analysis a necessary activity. Some even consider it as potentially hazardous [5].

Music analysis encompasses many approaches, from understanding the creative composition process on one side to the variability of listeners' reception, understanding, and appreciation on the other. The latter, the esthesic perspective, can be considered in various ways, from surveys of the broad feelings experienced by listeners to more systematic investigations of aspects of the music that could impact the listener's experience [6]. The focus of this chapter lies in the latter approach. Guided by previous musicological systematic studies on the topic and the state of the art in computational sound and music analysis, we investigate whether computational tools can offer new ways to go beyond the current limitations of musicological analyses. We will see a gap between, on one side, the overarching analytical methodologies and ideals developed by musicologists and, on the other side, the modest contributions of today's computational systems. The complexity and infinite richness of the electroacoustic sound universe make it challenging to design computational analytical approaches and, for the musicologists, even to formalise and systematise their modus operandi. Once we provide the machine with the capability to analyse electroacoustic music, the resulting tool could metamorphose the paradigms framed by musicology.

Interestingly, musique concrète was still in its infancy when an extensive and seminal theorisation of its compositional process was published in Pierre Schaeffer's *Traité des objets musicaux* [7,8]. Despite its deep influence on later musicological works, the treatise was not aimed at analysis. Rather, Schaeffer described it as "first and foremost a treatise on listening" ([8], p. 539). It was oriented towards a particular music aesthetics based on clearly separated sound objects, alluding to the limited music technology at that time.

Since the *Traité*, a few important analytical frameworks have been developed, as will be discussed in later sections. This overview of musicological methodologies of electroacoustic music analysis enables us to highlight the most important points of the *Traité* and augment it with a large area of descriptors that can be structured along various categories. We can then use this categorisation as a reference grid to compile an overview (presented in Sect. 4) of the state-of-the-art computational music analysis suitable for electroacoustic music. We will see that what today's technologies can offer is of great interest for the analytical investigation, but still, a lot of progress needs to be made. Thanks to this deep and synthetic understanding of musicology's needs, Sect. 5 sketches a proposed answer to those needs, with the objective in the longer term to establish a *Toolbox des objets sonores*.

2 Pierre Schaeffer's *Traité des objets musicaux*

The *Traité des objets musicaux* has played a monumental role in the establishment of a theoretical framework for electroacoustic music, and musique concrète in particular, both for music analysis but also composition. Pierre Schaeffer should not be considered only as an artistic trailblazer for his vision of a new musique concrète and the foundation of the French school (around the INA-GRM in particular), but also as the proponent of a highly multidisciplinary scientific endeavour to theorize the musical activity of sound listening. Schaeffer saw the limitations of psychoacoustics and music psychology, which were at the time restricted to individual parameters. The activity of listening to sound, which he considered to be studied in a domain called acoulogy, would require attention to the multidimensionality of highly interdependent dimensions.

From the start, the *Traité* was conceived by Schaeffer as a first step towards a more complete treatise of musical organisation. This first step is mainly based on the articulation of two parts. First, a *typology*, to identify sound objects based on criteria of articulation (related to discontinuities in the sound) and prolongation, a dichotomy taken from the opposition between consonant and vowels in linguistics ([8], Chaps. 24–26). Second, a *morphology*, to qualify the sound objects within their contexture (Chaps. 28–34).

2.1 Sound Objects and Reductive Listening

The core notion in the *Traité* is, as indicated in its name, musical objects, or sound objects. These sound objects are supposed to be detected through a phenomenological approach called *reductive listening*, meaning that the focus should be on the sound material itself, without reference to the origin (production) or signification (context) of the sounds (Chap. 15). For one listener, the same sound object, when listened to repeatedly, is fluctuating and unstable due to the variability of the listener's intentions, trying each time to focus on some particular aspects of the sound. This is not considered subjective but a bundle of complementary perspectives around the same object, leading to a set of unified traits.

2.2 Typology

The first step in Schaeffer's approach, the typology, aims at identifying the sound objects through segmentation and classification. The typology is decomposed into two dimensions (Chap. 24):

- *Facture* addresses the overall shape of sound objects under three different characteristics. The main one is related to how the energy of the sound production of a given sound object is maintained over time: either sustained (continuous energy over some duration), iterative (discontinuous energy production, leading to succession of sound and silence) or impulsive (significantly short duration). For sustained and iterative objects, there is also a distinction between moderate and immoderate duration, with respect to a duration threshold of around seven seconds. Finally, in the case of immoderate duration, there is a further distinction between homogeneous and unpredictable facture depending on whether the dynamic evolution is either stable or predictable, or variable and unstable.

- *Mass* relates to the sound's inner material, pitch, and spectral distribution. Mass is considered an elementary morphological characterisation further developed in the morphology step. The first distinction is whether the mass is fixed or varies over the temporal duration of the object. Fixed mass is further distinguished, whether a clear pitch or an inharmonic sound. Varying mass is distinguished by whether it evolves simply and predictably or, on the contrary, in a somewhat random fashion.

A third underlying dimension considered in the typology is related to the capacity of objects to be integrated into music structures, distinguishing between balanced, redundant and eccentric objects.

2.3 Morphology

The morphology describes the internal features of the sound objects, i.e., textural, spectral, timbral, and pitch-related. The treatise presents seven distinct morphological criteria (chapter 34):

- *Mass* is related to sound spectrum characterisation and is decomposed into seven classes: pure sound (one single fundamental), tonic sound (harmonic sound), tonic group (a chord), dystonic sound ("son cannelé", where pitch becomes ambiguous due to inharmonic partials), nodal sound (noise occupying a specific spectral range), nodal group (several ranges), and white or coloured noise (occupying complete spectrum).
- *Harmonic timbre* qualifies the spectral envelope of the series of partials based on sub-dimensions such as full/hollow/narrow, rich/poor, and bright/matt.
- *Dynamics* takes into consideration seven nuances of intensity and eight dynamic profiles, as well as eight attack classes. It also formalises the decomposition of sound objects into three phrases: attack, body and decay.
- *Grain* (or granularity) relates to the sound's rugosity due, for instance, to very fast oscillations of dynamics, or very rapid iteration of short sounds. This is studied along three main parameters: oscillation amplitude (the dynamic range of the oscillation), rate (how fast the oscillation is) and type (resonance, friction, and iteration).
- *Gait* ("allure" in French) relates to the slower fluctuations in harmonic content, pitch, loudness, etc. It can be either a continuous oscillation (for instance, a regular and continuous oscillation in the pitch curve or a regularly recurring continuous variability of the spectrum, of a slight and continuous oscillation in dynamics, etc.) or a more discontinuous succession of events of more clearly distinguishable sub-objects. Two parameters associated with gait are agent (mechanical, living, and natural) and form (order, fluctuation, and disorder).

- *Melodic profile* describes a non-periodic (or else very slow) variation. It relates to the profile in pitch height of the pitch component(s). Four melodic profiles are distinguished: podatus, torculus, clivis and porrectus.
- *Mass profile* relates to the evolution of the temporal harmonic or inharmonic spectrum, with four classes of profiles: swelled, delta, thinned and hollow.

For both melodic and mass profiles and for gait and dynamics, each object is characterised using a combination of two dimensions (amplitude and variation rate), with three classes in each dimension, leading to a 3×3 table with nine different categories.

2.4 Beyond the *Traité*

As acknowledged by Pierre Schaeffer himself, the ambitious programme of music research that was heralded by the *Traité* was cut short and restricted to its first part. It mainly focused on the taxonomy and morphology of individual sound objects, as presented above, without the initially planned additional morphology of the constructions of elementary objects into structures. However, the conceptual and methodological framework developed in the *Traité* is still valuable for analysing electroacoustic music.

Schaeffer conceptualised and wrote the *Traité* at a time when musique concrète, at its early stages, was characterised by structural simplicity due to technological limitations. For that reason, objects are identified in the typology through articulation and stress ("appui") (Chap. 21), which would be suitable solely for objects with apparent attack and sustain phases. The framework does not handle the "polyphony" of superimposed sound objects nor more complex music productions featuring continuously evolving elements. This is also problematic in the morphology, for instance, for the characterisation of dynamics, due to the limitations of dividing all sound objects into three phases: attack, body, and release [9].

The typology is aimed at identifying sound objects through segmentation but also at placing these objects within a classification, to organise the collection of objects. The objective seems to be to select objects of sufficient quality for integration into the composition. Thus, there seems to be an underlying poietic perspective motivating its conception. As discussed below, this aesthetically normative connotation of the *Traité* has been criticised.

The proposed morphology is very rich, conceptually and methodologically speaking, and had a substantial impact on later research. There is, however, a belief in the possibility of a very systematic and highly articulated method in the *Traité*, which did not actualise effectively. Nonetheless, this is inspiring, from a scientific point of view, for the design of a systematic and comprehensive framework, as discussed in the rest of this chapter.

3 More Recent Analytical Approaches

Most, if not all, musicological research on electroacoustic music has been highly influenced by Schaeffer's theoretical and analytical accomplishments, while opening new perspectives. This section presents a chronology of the major developments.

3.1 R. Murray Schafer's Soundscape Analysis

As part of his analytical study of soundscapes, R. Murray Schafer classified sound objects into two main branches related to physical characteristics and referential aspects [10]. Concerning the physical characteristics, he builds on characteristics by Schaeffer but separates the analysis of each sound object into its three successive phases: attack, body and decay. For each, the following characteristics are evaluated:

- *Relative duration*, for attacks: sudden, moderate, slow, multiple; for the body: non-existent, brief, moderate, long, continuous; for the decay: rapid, moderate, slow, multiple
- *Frequency and mass*, five degrees from very low to very high
- *Fluctuation and grain*, steady state, transient, multiple transients, rapid warble, medium pulsation, slow throb
- *Dynamics*, five degrees from very soft to very loud, plus the transitions from loud to soft and from soft to loud

The referential categorisation is divided into natural sounds (the four elements, animals, seasons), human sounds (voice, body, clothing), society-related sounds (rural, urban, maritime, domestic, activities), mechanical sounds, silence and indicators (alarms, etc.).

3.2 Denis Smalley's Spectro-Morphology

The composer Denis Smalley observes that still, at the end of the 20th century, there is a lack of shared terminology for describing sound materials and their relationships in electroacoustic music. He proposes an approach founded on a spectral typology, a morphology, and a study of motions, structuring processes and space [11,12].

Articulating between a typology and a morphology seems reminiscent of Schaeffer's typo-morphology. However, there is a contradiction between the two scholars concerning what should be part of typology and what should be in the morphology. Indeed, Schaeffer considered the dynamic shape of sounds (the facture) as one core element of the typology. In contrast, its other element, the mass, was integrated into the typology in a compact form and the morphology in a more extended form. For Smalley, this is quite the contrary: the typology is founded on Schaeffer's idea of mass—here developed through an exciting reflection about the possible states along the note-to-noise continuum—while the morphology studies how "spectral types are formed into basic temporal shapes" ([11], p. 65).

The morphology discerns three morphological archetypes as the source of traditional instrumental sounds (Fig. 1). The first describes impulsive attacks. The second includes attacks with a decay, either closed (with a quick decay that is strongly attack-determined) or open (including an intermediary continuous sound). The third covers graduated continuants, modelled on sustained sounds, with a graduated onset and a graduated termination. The three archetypes further contain one or several temporal phases, among the trilogy onset/constituant/termination, itself echoing the trilogy attack/body/decay.

Departing from these traditional reference points, Smalley extends the archetypes into a broader listing of morphological models by manipulating the duration and spectral energy of the three phases [12]. Morphologies can be linked or merged to "create hybrids" ([11], p. 71), in the form of morphological stringing, where correspondences can be merged within open constituents through cross-fading, or as a consequence of reversed onset-termination.

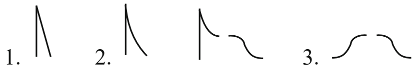


Fig. 1. Morphological archetypes: 1. attack-impulse, 2. closed attack-decay and open attack-decay, 3. graduated continuant, based on [11, p. 69]

Smalley also developed a refined motion typology, related to real and imagined motions created by spectro-morphological design. Here a motion category can be defined as "the external contouring of a *gesture*, or the internal behaviour of a *texture*" ([11], p. 73). He develops an additional typology related to the internal motion style of spectral texture, with four modes (streaming, flocking, convolution and turbulence), either continuous or discontinuous, and with an additional axis (iterative/granular/sustained) and three additional characteristics: periodicity, accelerating vs. decelerating and grouping patterns.

Smalley also discusses the variable scales of significant units in electroacoustic music. He argues that a unit "is often difficult or impossible to perceive, particularly in continuous musical contexts which thrive on closely interlocked morphologies and motions" ([11], p. 80). This exposes the limitations of the *Traité*, which focuses on isolated sound objects. In Smalley's theoretical framework, there can be a multi-levelled structure, with possibly permanent or temporally fractured hierarchies of various temporal dimensions. Finally, a detailed spatiomorphology is detailed [12,13].

3.3 Stephane Roy's Hierarchical and Functional Analysis

Stephane Roy demonstrates an impressive ability to carry out detailed analyses of electroacoustic music. His approach is based on producing visual representations (which he calls a "transcription") of the pieces, based on depicting sound objects using a large palette of graphical styles [5]. Then he develops a hierarchical analysis based on "units" of multiple hierarchical levels. The graphical representation is completed with a detailed textual description.

Can such eloquent analyses be systematised into a reproducible methodology? The approach is developed within the music semiology of Jean-Jacques Nattiez [14], who supervised Roy's doctoral thesis, and is based on the "neutral-level analysis." The idea is that the analysis could be conducted by systematically applying a limited list of objective rules to the music, following an approach initially developed by Nicolas Ruwet [15]. In my view, a close study of Ruwet's argumentation proves the scientific invalidity of the approach [16]. The whole analysis is founded on subjective decisions, contrary to what is claimed. However, despite the epistemological failure, the discussion about the possible mechanisms underlying music analysis (including Gestalt rules and auditory scene analysis through stream segmentation) is of high interest and will be discussed later.

The originality of Roy's approach is the functional taxonomy that can be associated to units based on their inter-relationships. It is based on the inner characteristics of each unit, the relationships of these characteristics among units, and the overall context of the development of those units throughout the piece. The functions are structured into four categories:

- *Orientation:* introduction, trigger, interruption, conclusion, suspension, appoggiatura, generation, extension, prolongation, transition
- *Stratification:* figure, support, foreground, accompaniment, tonic and complex polarising axis, movement, background
- *Process:* accumulation vs. dispersion, acceleration vs. deceleration, intensification vs. attenuation, spatial progression
- *Rhetorical:*
  - *Relational:* call and response, announcement and reminder, theme and variation, anticipation, affirmation, reiteration, imitation, simultaneous and successive antagonism
  - *Rupture:* deviation, parenthesis, indication, articulation, retention, rupture, spatialisation

This functional typology is also translated into a set of graphical symbols that are added to the visual analyses.

Roy also experimented with the adaptation of other notated music analyses to his "transcription" of electroacoustic music: Nicolas Ruwet's paradigmatic analysis, as mentioned above, as well as Lerdahl and Jackendoff's General Theory of Tonal Music [17] and Leonard Meyer's implicative analysis [18].

3.4 Lasse Thoresen's Graphical Formalisation

In addition to creating a phenomenological perspective on Schaeffer's framework, Lasse Thoresen has adapted Schaeffer's typomorphology augmented with a graphical formalisation [19,20]. The typomorphology is simplified by removing the normative concepts of object suitability, originality and redundancy, the distinction between “facture” and “entretien,” as well as the duration threshold (although very long notes are formalised in the form of ambient notes). This leads to a simpler typology, where long sustained notes with unpredictable dynamics are called vacillating, while long iterative notes with unpredictable iteration are called accumulated. In between those extremes are the concepts of stratified and composite objects.

The core contribution of Thoresen is the design of a graphical formalisation of Schaeffer’s theory, representing each sound object in the time/frequency space with its typological characteristics, as illustrated in Fig. 2. It enables us to go into more detail, localising over time the spectral particularities (position and characteristic of each spectral subgroup) and indicating their individual facture. Additional global graphical characterisation of objects is made available, such as the distinction between flutter and ripple notes, with an indication of their inner pulse regularity as well as of possible accelerando and ritardando. While Schaeffer kept a single structural level for the successive objects, lacking, therefore, an actual structural analysis, Thoresen takes benefit of this representation to show how objects are made of sub-objects, which can be characterised as well. For instance, the individual components of an accumulation, the construction of a sound web (“trame”), a large note, an ostinato, a cell, an incident or accident (special cases of, respectively, composite and stratified objects), a chord.

This formalisation enables us to address Schaeffer’s morphology, representing the mass of each object, its evolution over time (expanding, bulging, receding, concave, etc.), and its dynamic profile. A few graphical conventions have been added to indicate particular aspects of the morphology:

- *Mass:* saturated spectrum and white noise
- *Dynamic profile:* categorisation of onset (brusque, sharp, marked, flat, swelled, gradual, inexistent) and ending (abrupt, sharp, marked, flat, soft, resonating, interrupted)
- *Pitch, dynamic and spectral gait:* characterising both deviation and pulse velocity
- *Granularity:* characterising the coarseness and the velocity of the grains, as well as sound spectrum location, weight (or importance) and spectral placement of the grains.

It is also possible to indicate the brightness level of each sound, as well as its gradual change.

Lasse Thoresen also identifies “time-fields,” describing the segmentation of form sections, and “dynamic forms” tracing the perceived directions of energy flow [20,24]. He also pursued the application of two other central terms in Schaffer’s analytical work [7,8], namely “caractère” (character) and “valeur” (value). Whereas “sound-character” refers only to a timbral constant that supports pertinent values, a form-building entity, termed integral sound-character, consists of a union of sound-character and its temporal behaviour [20,23].

3.5 Ecrins Audio-Content Description

The Ecrins project, a collaboration between IRCAM and INA-GRM, was aimed at offering tools for the classification of online sound samples, based in particular on Schaeffer's typomorphology [25]. It also contributed to the theoretical establishment of a taxonomy of analytical descriptors, introducing audio content features such as duration, dynamic profile (flat, increasing, decreasing), melodic profile (flat, up or down), attack (long, medium, sharp), pitch (note pitch or area), spectral distribution (dark, medium, strident), space (position and movement), and texture (vibrato, tremolo, grain).

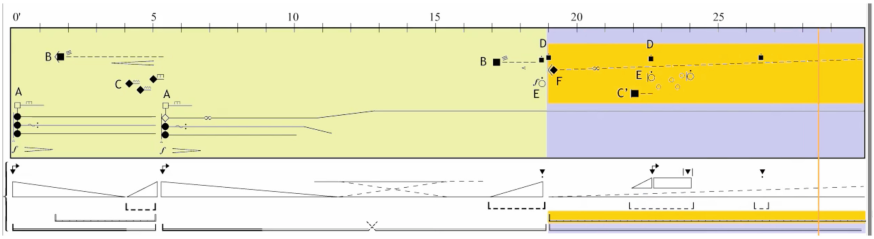


Fig. 2. Graphical analysis by Lasse Thoresen of the beginning of Åke Parmerud's *Les Objets Obscurs*. Screenshot of an animated version [21] of [20, Figure 11.4], from the companion website [22]. The orange rectangles highlight the section being heard, as indicated by the orange vertical playhead. (Color figure online)

3.6 Structural and Functional Analyses

Some aspects of structural and functional analyses have been mentioned above, but there exists also a large range of works related to the establishment of units or sections—possibly along multiple hierarchical levels, and not necessarily following a strict hierarchy—and in assigning various functions or categories to the units or sections [19,26]. This is not a research question specific to electroacoustic music, so it can be investigated for traditional instrumental music, on score or audio recordings of performances. The main question of interest for a computational implementation of approaches of this type is whether they are systematic and can be formalised with explicit discovery methods. This is a question that exceeds the scope of the present study.

3.7 Pierre Couprie's Morphology

Pierre Couprie proposes a methodology for morphological analysis of electroacoustic music offering a comprehensive synthesis of previous approaches [9]. The internal morphology focuses on what is inherent to the sound and does not depend on any external factor:

- *Spectrum:* type (related to Schaeffer's facture), density (compact, normal, transparent), movement type (stationary, linear, breaking, oscillating), movement cycle, amplitude and rate, acceleration or deceleration (related to Shaeffer's gait)
- *Dynamics:* attack profile (related to Schaeffer's categorisation, but simplified a bit), movement type (same as for spectrum), movement cycle, amplitude and rate, acceleration or deceleration (also related to Shaeffer's gait)
- *Grain:* number of sounds, spectral positions, characterisation (with respect to type and amplitude), speed
- *Internal space:* considered in 2 dimensions in a 3 × 3 grid.

The referential morphology links the considered object to other elements in the work (indicating whether the citation is exact, transformed, or an evocation) or external to the work itself (from another work, or from a general concept). The references are analysed according to these categories:

- *Causality*, based on Schafer's categorization
- *Voice description:* type, text, rhythm, speed, pitch variation, colour, cadenza, silence, density, alliteration
- *Effects:* temporal modification, internal spectrum, dynamic envelope, external element
- *Emotions and sentiments.*

The structural morphology is based on analytical tools to reveal the structures of the work along all levels.

4 Computational Electroacoustic Music Analysis

Analysing electroacoustic music can be challenging. One reason is that there are fewer formalised rules than in many other genres. Another reason is the absence of a written music representation provided by the composer. On the other hand, the composer might provide detailed sketches describing a piece, and related computer code for processing and synthesis may sometimes be made available for analysis. The present study does not consider such poietic information, instead focusing on analysing a piece simply from available audio.

Can some of the analytical frameworks introduced above be formalised, systematised and automated with the help of computer implementations? First, we need to clarify and formalise the analytical principles. In the following, we will discuss, first, the detection of basic objects in the music, and second, how to address various musical dimensions.

4.1 Sound Object Detection

Much research has been dedicated to automated score transcription of music performance recordings, particularly detecting individual notes, characterising their temporal positions, pitch, instrumentation, playing style, etc. [27]. This is a complicated problem which has been tackled using two different methods.

The first is purely based on signal processing, computing representations based on mathematical equations to decide the location of the note events and their characteristics. These equations are based on general music acoustics and psychoacoustics principles, particularly those related to sound scene analysis and Gestalt rules.

An alternative approach, which has largely superseded the first one this last decade, is based on machine learning and especially deep learning [27]. Here, an artificial neural network is trained using large audio collections, often indicating where notes should be found and their characteristics. The main difficulty in this approach is the need for such an extensive training dataset. This is particularly problematic for electroacoustic music because there do not exist many detailed analyses. More problematic is the lack of consensus about such analyses, and one analyst might struggle to decide what constitutes a sound object and what does not. Even if we could create an extensive dataset, the large variety of electroacoustic music styles is such that the machine could not generalise well to styles or, for instance, synthesis techniques not included in the dataset.

A possible solution might be to rely on unsupervised learning, where the model is not trained on examples given beforehand but through an automated search for regularities. To my knowledge, these approaches have been used to broadly segment audio recordings into distinct parts, but not yet for more detailed detection of individual sound events. All in all, the problem of automated detection of sound objects in electroacoustic music remains unsolved.

4.2 Dynamics

Once a sound object has been segmented, with a set of partials and/or wider energy bands evolving within a specific time and frequency region, the characterisation of its dynamics might look at first sight somewhat straight-forward, measuring the amplitude of the signal on a relatively slow temporal scale. However, perceived dynamics are not directly correlated with the linear amplitude of the sound or even a more subtle logarithmic relationship. It requires taking into account more subtle properties of the auditory system, for instance, related to the variable impact of the different frequency regions, the effects of critical bands, and the presence of masking effects.

Even more complicated is the fact that listeners' assessment of the dynamics of a given sound event is not simply related to the mere properties of the sound itself but also to their experience of how, when listening to live sound production (such as an instrumental music performance, but not only), the spectral quality of the sound changes depending on the actual loudness of the sound. For example, if the spectral quality of a recording corresponds to the production of a loud sound but is played back with low loudness, the sound dynamics would generally be perceived as loud. As Smalley mentioned:

> During execution of a note, energy input is translated into changes in spectral richness or complexity. When listening to the note we reverse this cause and effect by deducing energy phenomena from the changes in spectral richness. ([11], p. 68)

For these reasons, predicting the perceived dynamics of each sound object is challenging, which has been addressed, for instance, using machine learning approaches [28].

In opposition to the problematic concept of perceived dynamics, it can be valuable to simply estimate the dynamic evolution of the loudness throughout the sound object. As mentioned in the literature review above, a sound object can be decomposed into three main phases: attack, body and decay. Another typical pattern is the attack-decay-sustain-release (ADSR) concept used in many synthesisers to generate natural-sounding sounds. But real-life sounds—and even more complex, artificial electro-acoustic sounds—may have dynamic curves that do not easily fit ABD or ADSR patterns. The detection of attack and (final) decay phases can be done by computation of temporal derivatives and detecting when they reach particular thresholds. This will work on simple examples, but more refined heuristics may be needed for more complex temporal envelopes. Characterisations such as attack time, attack slope, etc., play an essential role in timbre characterisation; as we will see later, these can be directly measured from the extracted attack and decay phases.

Dynamics can be assessed not only for individual sound objects but also for the resulting mix. This results in a single dynamic curve indicating the overall profile. The traditional method for dynamic curve estimation consists of discarding the fast-evolving part of the signal to focus solely on the slowly evolving part using signal processing methods such as low-pass filtering or windowed analysis. I developed a new method that can adequately represent a sudden increase of dynamics while discarding micro-silences (shorter than one second), while simultaneously attempting to model saturation effects taking place within separate frequency registers [29].

Figure 3 shows an example of the dynamic curve I developed, computed here for the analysis of Pierre Schaeffer's fourth of the Five Studies of Noise (*Cinq études de bruits*), initially called "Composée, ou étude au piano", composed from piano sounds recorded for Schaeffer by Pierre Boulez. The dynamic curves are compared with simple RMS computation. We can notice in particular that some parts in the piece—for instance, 100 s after the start—have rather low RMS values but a larger value in the dynamics curve. In other places—such as between 170 and 180 s—RMS values oscillate rapidly, while the dynamics curve indicates a more progressive evolution. The dynamics curve is obtained through a decomposition of the energy into Mel bands, filtering of each band separately via an original filtering model, and concluded with a summation along bands.

4.3 Facture

Schaeffer's notion of facture—which, as described above, corresponds to the characterisation of sound objects as impulsive, sustained or iterative—can be approximated using relatively simple signal processing approaches. Once extracting the dynamic curve, as discussed in the previous paragraph, we can qualitatively differentiate between sound objects that are either clearly impulsive or sustained through observation of the duration of the attack, body and decay phases. But what are the actual thresholds governing the limit between those categories, and what is the impact of the three successive phases in the appreciation of a sustained sound? There does not seem to be any published study on that matter.

Detecting whether a sound is iterative can be computed, for instance, by extracting the envelope curve or by computing the spectral flux over time. But iterativity is not only about a dynamic oscillation; there should also be some invariance of what is supposed to be repeated at each successive iteration.

4.4 Mass, Harmonicity and Pitch

Schaeffer's concept of mass is related to physical characterisations commonly studied in signal processing. First, a time-frequency image of the sound object is computed through, for instance, a spectrogram, showing the evolution of the spectral distribution of the sound for successive short instants (or window frames). Noisy parts of the sound can be detected as regions of high energy with relatively large frequency widths. Partials are characterised by regions with narrow widths and can form harmonic series of one or several pitches. In other words, each pitch comprises a series of partials around multiples of the fundamental frequency. The possible deviation of the partials to the ideal series indicates the inharmonicity of the sound, often found in complex percussion instruments like bells.

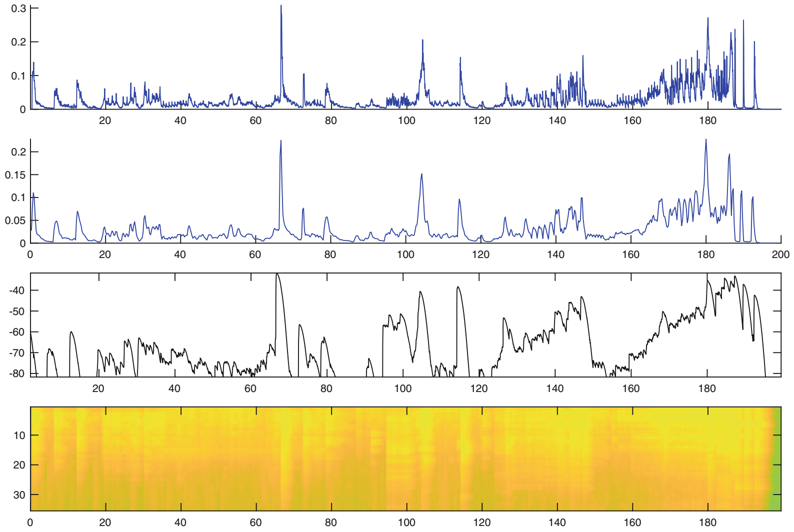


Fig. 3. Analysis of dynamics in Pierre Schaeffer's fourth *Étude de bruits* . Top-down: 1. Root Mean Square (RMS) computed on 0.1 s frames with half-overlapping, 2. RMS on 0.5 s frames, 3. proposed dynamics curve and 4. decomposition of that dynamics along 35 Mel bands (higher amplitude shown with brighter colour) with subsequent filtering within each band. (Color figure online)

"Pitch salience" indicates the relative prominence of a series of partials corresponding to one or several pitches. It can be estimated by first computing the *autocorrelation function* to detect pitch-related periodicities; pitchness is then estimated as the ratio of the magnitude of the highest autocorrelation peak to the magnitude of the 0-lag peak [30,31].

Similarly, the qualification of harmonic timbre (full/hollow/narrow, rich/poor, and bright/matt) can be based on a statistical description of the partials' distribution. Hollowness is related to the ratio of amplitudes of even and odd harmonics [32], while fullness and narrowness denote the width of the spectral distribution. Brightness, in the context of harmonic timbre, could correspond to the ratio of high-frequency partials or to the frequency centroid of the partials.

4.5 Temporal Motions

By estimating dynamics, pitch, harmonicity, and spectrum on successive time frames of a sound object, we obtain a temporal evolution of those different characteristics. One particular interest of the Ecrins project (cf. Sect. 3.5) is its detailed study of the categorisation of dynamics profiles, derived from Schaeffer's classification (dynamic, melodic and mass profile, and gait). The dynamics profile is estimated through an envelope extraction, followed by low-pass filtering, B-spline approximation, thresholding and peak picking [33], as illustrated in Fig. 4. This allows us to estimate the temporal ratio of the ascending and descending phases as well as their slopes. Simpler estimation and classification of the dynamic profile is proposed in [31], where a series of features computed from an estimation of dynamic curves (flatness coefficient, number of onsets, maximum amplitude time, derivative before and after the maximum, and temporal centroid) were used as predictors for a machine learning classification into five classes: ascending, descending, ascending/descending, stable and impulsive.

Concerning periodic motions, grain and gait are considered in the Ecrin project under one single concept called "grain/iteration." Dynamic periodicity is estimated through auto-correlation, while timbre and pitch periodicity is estimated using a similar method based on the similarity matrix [33]. Then, the amount of repetition and the cycle period are measured, and the repeated element is characterised. There was also the intention to classify melodic profiles, but this has not been implemented due to the complexity of the task. Some simple classification strategies are proposed too [31].

Estimating two parameters associated with Schaeffer's gait—agent (mechanical, living, natural) and form (order, fluctuation, disorder)—and of the categorical classes associated with the profiles, remains to be studied. The characterisation of spectromorphological design into imagined motions, based, for instance, on the taxonomy proposed by Smalley, is an even more challenging topic, and its computational systematisation has not been addressed either.

4.6 Spatial Analysis

Very few studies exist in computational music analysis of audio recordings addressing the spatiality of the sound production [34]. Two features have been designed for electroacoustic music analysis, focusing mainly on stereo mix [1]:

- Stereo spatial ebb is a measure of spectral movement comparing left and right channels
- Two channel loudness difference is the absolute difference in perceptual loudness between the left and right channels

There has been relatively little focus on analysing more advanced spatialisation techniques from audio recordings. The spatialisation can be represented relatively straightforwardly if one starts from multichannel audio with a specification of the spatial localisation related to each track. However, there is a lack of analytical approaches and related tools for performing spatial analysis.

4.7 Other Timbral Aspects

In Schaeffer's theory, timbre is studied through mass, harmonic timbre, attack characterisation and granularity. What about other aspects of timbre? Some timbral aspects might be implicitly indicated in Smalley's typology of the internal motion style of spectral textures. In music psychology research, timbre has been conceptualised as a three-dimensional space, with spectral centroid (the distribution of energy along frequency), spectral flux (related to contrast in the temporal evolution of the spectrum), and attack characterisation [32]. Harmonic brightness, as part of Schaeffer's harmonic timbre, could be related to the spectral centroid in the case of harmonic sound. But more generally, a simpler estimation of the brightness of the whole spectral distribution can be carried out by estimating the energy ratio above a given threshold [35] or by computing the spectral centroid. More generally, it has also been suggested to measure the distribution of energy along frequency bands [1,29]. The second dimension in the timbre space, spectral flux, can be related to the study of fluctuation, or granularity. The third dimension, attack characterisation, was discussed in Sect. 4.2.

Sensory dissonance [1,36], spectral entropy and flatness [1] are other relevant descriptors. Since they do not require a given harmonic series, they can also be computed on a general mix of sound objects. Another timbral description is transientness [1]. In Music Information Retrieval (MIR) research, timbre has been very often described in the form of Mel Frequency Cepstral Coefficients (MFCC), which is a technical representation of the spectral shape of the sound. It offers a particular interest for structural analysis, as discussed later.

One aspect of timbre that is central to everyday listening, and also to traditional music listening, is related to the recognition of sound categories based on the type of sound production and the association to the typical family of sound production classes and the underlying contexts (especially for non-musical sounds). This identification of sound class is the opposite of what Pierre Schaeffer aimed at achieving with the concept of reductive listening, but at the same time,

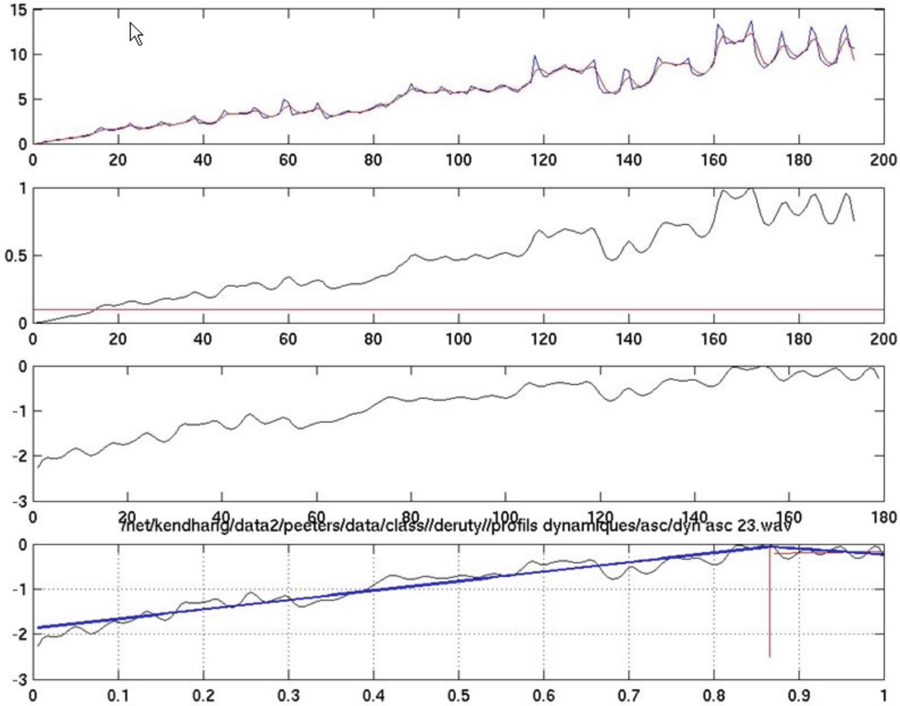


Fig. 4. Estimation of dynamic profile parameters: a) loudness (blue) and smoothed loudness over time (red), b) 10% threshold applied to smoothed loudness, c) smoothed loudness in log-scale, d) Maximum value (vertical red bar) and B-spline modelling. From [33] (Color figure online)

a critical aspect of more modern sound analysis methods such as Schafer's referential categorization. Current machine learning technologies enable the classification of each successive instant of an audio recording according to the detected sound categories, with a taxonomy that nicely resembles Schafer's referential categorisation. For instance, the Sound Analysis framework released by Apple can recognise over 300 sound classes in four categories: Sounds of things (train, car horn, ...), Animals (cow moo, duck quack, ...), Human sounds (singing, laughter, ...) and Music (along various instrument classes). However, since individual sound events are not yet clearly detected from complex pieces, the referential categorisation of these individual sound events remains an open challenge.

4.8 Rhythm

Rhythm is considered in Schaeffer's morphology solely in terms of the possible internal iterativity and gait within one sound object. No other aspect of rhythm is represented in more recent musicological approaches, except that Thoresen's graphical formalisation enables the representation of the cyclic repetition of sequences of short events and conceptualises the pulse velocity and its possible change over time.

The near absence of rhythmical representation is due to the aesthetics of musique concrete and electroacoustic music, especially in the early decades. However, although not explicitly discussed, musique concrète and electroacoustic music feature interesting rhythmic elements. The "curated corpus of historical electronic music" [1], introduces rhythmical features for the computational analysis of electronic music. A first set of features is based on statistics related to the temporal position (or "onsets") of sound objects. Another set is related to statistics concerning beats. Computational methods exist to describe rhythmical pulsation from audio without detecting an actual beat sequence. The autocorrelation function of the dynamic curve can be used to detect pulsations and their hierarchies and estimate metrical clarity and centroid [37].

Figure 5 shows this type of rhythmical analysis for Pierre Schaeffer's fourth *Étude de bruits*. We notice a prominent and regular periodicity of period 0.75 s because those early studies by Pierre Schaeffer were highly based on using special phonograph discs with a "sillon fermé" (closed groove), thus with a fixed period. But other periodicities can be seen, such as a period of 1 s at the beginning of the piece, or very fast repetitions here and there. A bit before 100 s, we see a 0.75 s loop divided into 6 regular subbeats. Between 160 and 180 s, the subdivision of the loop is a bit more complex, with a seeming decomposition into 8 sub beats, but also containing other internal patterns.

4.9 Structural and Semantic Analysis

There exists a large range of research on the topic of computational formalisation and automation of structural analysis of recorded music. One common technique

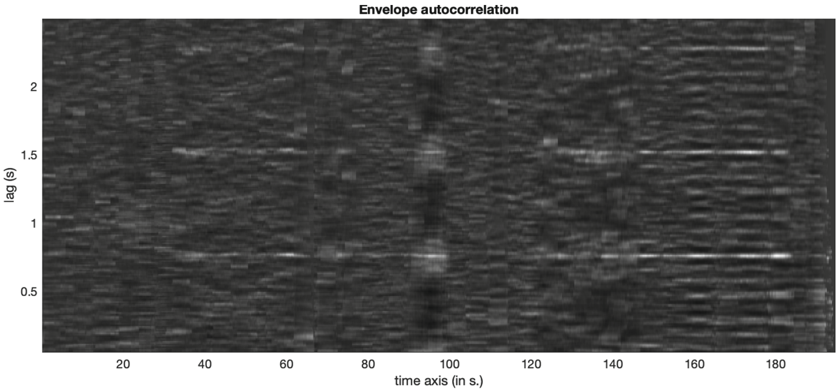


Fig. 5. Rhythmic periodicities (shown in white) throughout Pierre Schaeffer's fourth *Étude de bruits*, with time from left to right, and periods (i.e., duration between successive beats, indicated in seconds on the left) in ascending order from bottom to top.

is based on computing a similarity matrix along a given audio or musical features computed on successive window frames on a given audio recording [38]. Figure 6 shows an example of a similarity matrix for Pierre Schaeffer's fourth *Étude de bruits*, here focusing on simple timbral aspects related to MFCCs. From that matrix can be detected sharp transitions between successive segments (the succession of squares of various sizes along the diagonal) and repetition of sequential patterns (little white lines parallel to the diagonal).

An overview of computational approaches in structural and semantic analysis of recorded music [39] is beyond the scope of this chapter. But concerning electroacoustic music in particular, one particular system has been developed that allows detecting the repetition of samples in a given piece of music, even in the case of "polyphonic" superposition of samples [40]. The system was designed to be partially automatic, requiring an interaction with the user. It remained in the form of a prototype, demonstrated with artificial musical examples made of concatenation and juxtaposition of pre-selected samples.

4.10 Software

A large panoply of software can be of interest for analysing electroacoustic music. Basic representations of the sound, such as waveform or spectrogram, can be computed using free or commercial software. Audio and music features can be

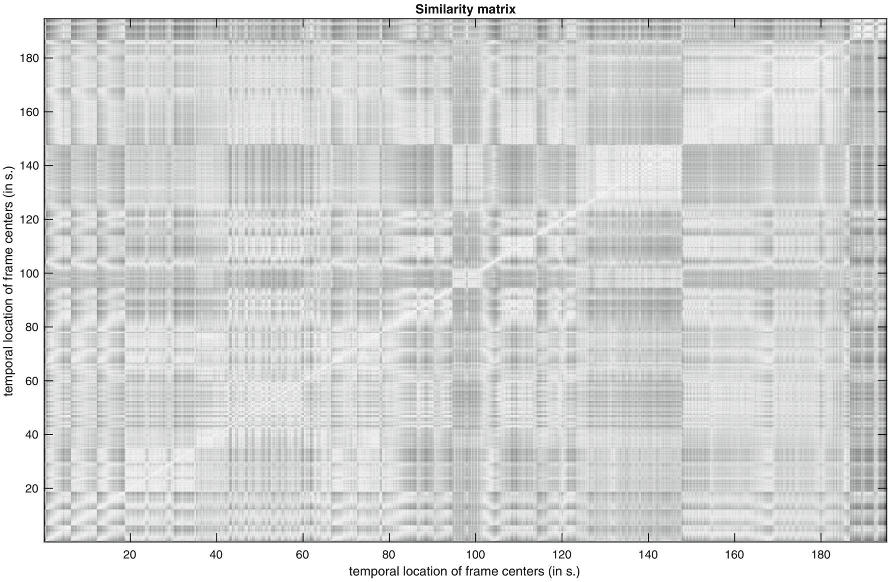


Fig. 6. Similarity matrix related to the MFCC computed on 0.1 s overlapping frames throughout Pierre Schaeffer's fourth *Étude de bruits*, where the frames are compared using the Euclidean distance. The high similarity is shown in white.

computed using software such as Sonic Visualiser with Vamp plug-ins, PRAAT, MIRtoolbox or AudioSculpt.

One common way to manually analyse electroacoustic music is to annotate the spectrogram by adding forms related to particular sound objects. The most common software for visual annotation of electroacoustic music are the following:

- *iAnalyse* developed by Pierre Couprie since 2006, is aimed at displaying music representations in pedagogical settings for musicians, teachers and musicologists [41]. The music timeline is decomposed into successive pages, to which can be added graphic annotations, such as annotations based on Lasse Thoresen's conceptual framework (cf. Sect. 3.4). This enables the user to illustrate music analyses, produce listening guides from annotated scores and help musicologists in their analyses. A playhead can be synced to the visualisation, the graphical annotations can be animated, and audio descriptors computed from other software can be integrated into the display.
- *EAnalysis* also developed by Pierre Couprie, this time in the context of the project "New multimedia tools for electroacoustic music analysis" hosted at the Music, Technology and Innovation - Institute for Sonic Creativity at De Montfort University in Leicester (UK), funded between 2010 and 2013. EAnalysis allows the integration of various types of representations (acoustical, mathematical, musical), for music analysis purposes, as illustrated in Fig. 7.
- The *Acousmographe* is a software developed by INA-GRM for the annotation of general audio representations such as waveforms and spectrograms with graphical and textual representations. The *Aural Sonology Plug-In* is inspired by the compositional procedures and the theoretical reflection of

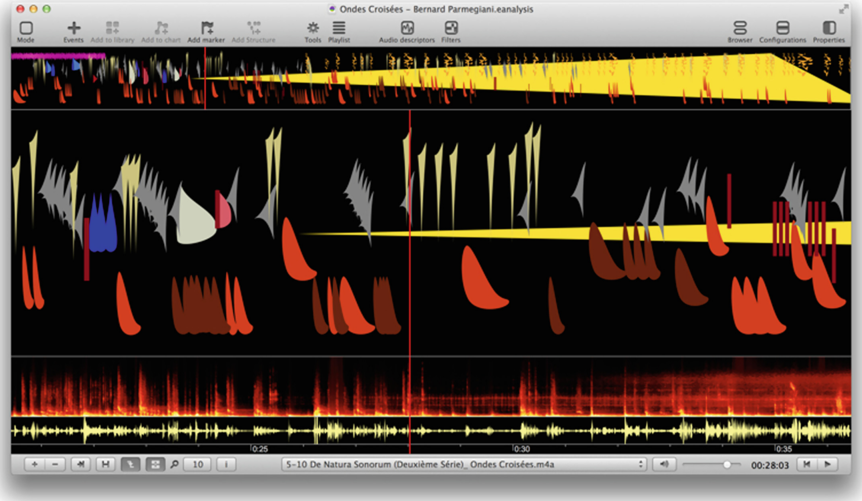


Fig. 7. Screenshot of the *EAnalysis* software.

Lasse Thoresen (cf. Sect. 3.4) to help the listener to conceptualise and write down sound objects heard. The plug-in is equipped with a library for spectromorphological and form analysis, which includes time fields (the temporal segmentation of the musical discourse), layers (the synchronous segmentation of the musical discourse), dynamic form (time directions and energetic shape), thematic form (recurrence, variation, and contrast) and form-building transformations (simple and complex gestalts, transformations between them, e.g., proliferation/collection, fission/fusion; liquidation/crystallisation).
- The *Acousmoscribe* is another annotation tool developed by the SCRIME team at the University of Bordeaux, this time based on the theoretical framework developed by Jean-Louis Di Santo [42].
- *TIAALS* (Tools for Interactive Aural Analysis) a toolbox developed as part of the Interactive Research in Music as Sound (IRiMaS) at the University of Huddersfield, for musicologists to use in conducting and presenting research in which audio and video are fully integrated into the research process and its dissemination. TIAALS focuses on sound material analysis and the realisation of typological, paradigmatical or other analytical charts.
- Other annotation software built on top of audio feature extractor tools are *CLAM Annotator* (on top of the CLAM framework) and *ASAnnotation* (on top of AudioSculpt).
- The *EASY* (Electro-Acoustic muSic analYsis) Toolbox providing a 3D visualization environment for sonic exploration and interaction [43] (cf. Fig. 8). The temporal evolution of timbre is represented as a curve in the 3D timbre space. 26 signal processing features can be computed. Automated segmentation of audio recordings is carried out, mainly based on k-means clustering.

The software above offer various ways to display basic visual representations of the music and to manually annotate them with more advanced analytical representations.

Interviews with three musicologists [40], revealed that they wanted some automated sound object segmentation to correct and enrich manual annotations. They also wanted the possibility to detect all repetitions of the same sample to retrieve isolated voices from a mix. There have also been suggestions of automated high-level structural analysis, for instance, with the possibility to detect the repetition of sequential patterns of sound objects.

5 Towards a *Toolbox des Objets Sonores*

There remains a large gap between, on one side, the overarching analytical methodologies and ideals developed by musicologists and, on the other side, the relatively modest contribution of what computational automation can offer today. The complexity and infinite richness of the electroacoustic sound universe make it challenging to design computational analytical approaches and for musicologists to even formalise and systematise their modus operandi. Once we provide the machine with the capability to analyse electroacoustic music, the resulting tool could metamorphose the paradigms framed by musicology.

Based on the panorama outlined above, I would like to emphasise the following capabilities:

1. to detect and precisely describe and characterise the components constituting the piece of music, from basic objects to groups of objects to structural segments
2. to reveal intra- and intertextuality, concerning the repetitions (with possible transformation) of those components
3. to reveal this rich information in the form of visualisations
4. to allow the analysts to modify those analyses

One overarching aim of my work is to develop technologies automating the analysis of music of all kinds, with a high level of richness and on many different musical dimensions. These technologies are aimed at being made available in the form of toolboxes for analysts (such as MIRtoolbox) as well as interactive music visualisations. In this context, one ambition here, in collaboration with Rolf Inge Godøy, is to develop technologies in line with Schaeffer's "programme de recherche musicale," hence a "Toolbox des objets sonores."

The main difficulty concerns detecting the more or less "elementary" components of the piece of music. This corresponds to Schaeffer's sound objects, but as discussed above, this notion is somewhat limited and should also include

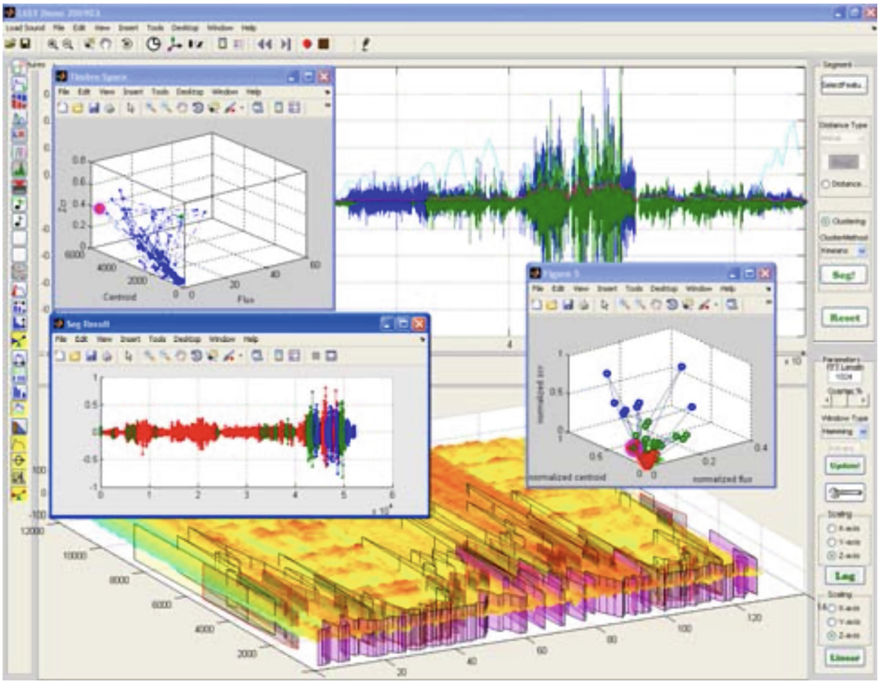


Fig. 8. A screenshot of the EASY software, from [43].

the possibility of "polyphonic" superposition of objects, horizontally and vertically. Here computational formalisation can help the theoretical development: whereas manual analyses require the theory to simplify the general organisation, defining one level of organisation, or else a rather strict structural hierarchy, computational formalisation can be based on less stringent rules, allowing the emergence of a richer variety of structures. Hence, various object candidates can be suggested in parallel on a given piece of music, and there is no need to make decisions at that stage. The computer tool can work in dialogue with the musicologist, who can correct the computational predictions, and the computer might also learn from those mistakes or the musicologist's preferences. The detection of such components is based on auditory scene analysis and inspired by Gestalt rules and cognitive morphodynamics [44,45]. Allowing components to contain smaller subcomponents, implicitly enables the detection, tracking and formalising of iterative and granular objects. Iteration can sometimes appear only at some parts of the super-object. Besides, the successive sub-objects do not need to be iterations of the same pattern. In this way, many descriptions can be related to the succession of sub-objects: the similarity between successive sub-objects, the contrast between them, etc.

A large range of descriptors (such as mass) can be computed, both on the overall mix and on each isolated component. The available list of signal processing features (such as harmonicity) needs to be more closely articulated with the dimensions and the corresponding categorisations proposed by the musicological works since Schaeffer's *Traité*. But here also, the simplicity of strict classifications in those works can be replaced with multidimensional parametric spaces, in which particular regions define the theoretical concepts. A bit like *phase diagrams*. The closer a given position in the diagram is to the paradigmatic centre or border of a region, the more clearly the concept is associated with the corresponding sound object. For instance, the concepts of being impulsive or sustained can be considered as two phases defined in a multidimensional parametric space, including dimensions of attack, sustain and decay times.

The intratextual connections between components in a piece of music can be drawn by detecting similarities along particular parameters or even detecting repetition of the same or similar samples or synthesis types. Iterativeness can be considered as a particular type of succession of sub-components featuring such similarity. Sequential patterns of sounds can also be detected, as well as the iteration of such sequential patterns, as formalised in Thoresen's theory.

The richness of this analysis needs to be made accessible to both musicologists and the public. In particular through the design of visualisation strategies. One visualisation follows the traditional unfolding of time from left to right, like scores, spectrograms and acousmographs, and shows the various constituents (or sound objects) with the depiction of their particularities through forms and colours. Interactivity allows one to browse through the various types of information to be displayed and to highlight the intratextual connections. The overall structure and form of the piece can be shown as well.

Such a "rolling" representation can be compared to another representation in which the elements currently being played are visible anywhere on the screen, and then simply disappear. “This method of presentation is much more natural and makes the display experiential rather than simply informative” [46]. For this second, “experiential” type of representation, the mapping strategy between music and visuals has so far been based on the display of specific simple forms or colours related to elementary musical aspects. The objective here is to make a more immersive visualisation, depicting the music as it unfolds in time with more richness.

Another application is to show a whole corpus of music in the form of a 2D or 3D interactive space where each piece is represented by one point. Intertextual analysis shows the relationships between pieces of music based on similar configurations. The pieces of music are distributed according to their features and can be clustered based on similarities and commonalities.

6 Conclusions

As I have tried to show through this overview, the dream of establishing a systematic, formalised and computerised analysis of electroacoustic music is on its way to becoming a reality. Considerable challenges remain, in particular, related to detecting sound objects and other basic constituents of the pieces of music to be analysed. Fortunately, much progress has been made concerning descriptions of the overall sound along various sound and music dimensions. Gathering a range of the state of the art in computational music analysis within a toolbox would make all the separate research accessible to a larger community. Offering the possibility to perform some approximate segmentation at the more basic levels, and to carry out all those analyses on the different individual objects, would interest musicologists.

This technological progress could enable, in the longer term, to automate analyses along Schafer’s physical morphology as well as Smalley’s spectromorphology, and could also allow automation of graphical representations such as those proposed by Thoresen. On the other hand, any attempt at automation of Smalley’s motion typology or his functional or spatial approach, or any higher-level structural or functional analysis, would require much more work.

Through developing the “Toolbox des objets sonores,” accompanied by interactive interfaces for visualising and browsing music pieces and music catalogues, we hope to stimulate musicological interest in electroacoustic music. We have experienced that the visualisation of such music offers the general public new ways to enjoy the richness of this art. In addition, this would allow further scientific research around this topic in the domains of music psychology and music cognition in particular.

Acknowledgements. The idea of developing a toolbox for analysing electroacoustic music results from numerous discussions with Rolf Inge Godøy. This work is supported by the Research Council of Norway through its Centers of Excellence scheme, project number 262762 and the MIRAGE project, grant number 287152. The author would like to thank the reviewers for their fruitful help.

References

1. Collins, N., Manning, P., Tarsitani, S.: A new curated corpus of historical electronic music: collation, data and research findings. Trans. Int. Soc. Music Inf. Retrieval **1**(1), 34–43 (2018)
2. Bent, I.: “Analysis”. In: The New Grove Dictionary of Music and Musicians (2001)
3. Dunsby, J., Whitall, A.: Music Analysis in Theory and Practice. Yale University Press, New Haven (1988)
4. Cook, N.: A Guide to Musical Analysis. George Braziller, New York (1987)
5. Roy, S.: L’analyse des musiques électroacoustiques : modèles et propositions. Éditions L’Harmattan (2004)
6. Delalande, F.: Music analysis and reception behaviours: Sommeil by Pierre Henry. J. New Music Res. **27**(1–2), 13–66 (1998)
7. Schaeffer, P.: Traité des objets musicaux. Éditions du Seuil, Paris (1966)
8. Schaeffer, P.: Treatise on Musical Objects: An Essay Across Disciplines. University of California Press, Berkeley (2017)
9. Couprie, P.: La musique électroacoustique: Analyse morphologique et représentation analytique. PhD, Paris IV, Sorbonne (2003)
10. Schafer, R.M.: The Soundscape: Our Sonic Environment and the Tuning of the World. Knopf, New York (1977)
11. Smalley, D.: Spectro-morphology and structuring processes. In: Emmerson, S. (ed.) The Language of Electroacoustic Music, pp. 61–93. MacMillan Press, London (1986)
12. Smalley, D.: Spectromorphology: explaining sound-shapes. Organised Sound **2**(2), 107–126 (1997)
13. Smalley, D.: Space-form and the acousmatic image. Organised Sound **12**(1): 35–58 (2007)
14. Nattiez, J.-J.: Music and Discourse: Toward a Semiology of Music (Musicologie générale et sémiologue, 1987). Princeton University Press, Princeton (1990). Translated by Carolyn Abbate
15. Ruwet, N.: Méthodes d’analyse en musicologie. Rev. Belge Musicologie/Belgisch Tijdschrift voor Muziekwetenschap **20**(1/4), 65–90 (1966)
16. Lartillot, O.: For a Copernican revolution in the understanding of universality of structural analysis of music. In: Leroy, J.-L. (ed.) Topicality of Musical Universals / Actualité des universaux musicaux. Éditions des archives contemporaines (2013)
17. Lerdahl, F., Jackendoff, R.: A Generative Theory of Tonal Music. MIT Press, Cambridge (1983)
18. Meyer, L.B.: Explaining Music. University of California Press, Berkeley (1973)
19. Thoresen, L., Hedman, A.: Spectromorphological analysis of sound objects: an adaptation of Pierre Schaeffer’s typomorphology. Organised Sound **12–2**, 129–141 (2007)
20. Thoresen, L.: Emergent Musical Forms: Aural Explorations. Studies in Music from the University of Western Ontario, vol. 24 (2015)
21. Thoresen, L., Hedman, A.: Video animation of Les objets obscurs, Part III, by Åke Parmerud. https://www.auralsonology.com/analyses/chapter-11-analysis-of-les-objets-obscurs-part-iii-by-ake-parmerud
22. Thoresen, L., Hedman, A.: Aural sonology. https://www.auralsonology.com
23. Thoresen, L., Hedman, A.: Sound-objects, values and characters in Åke Parmerud’s *Les objets obscurs*, 3rd Section. Organised Sound **14**(3), 310–320 (2009)

24. Thoresen, L., Hedman, A.: Form-building patterns and metaphorical meaning. Organised Sound **15**(2), 82-95 (2010)
25. Geslin, Y., Mullon, P., Jacob, M.: Ecrins: an audio-content description environment for sound samples. In: International Computer Music Conference (2020)
26. Delalande, F., et al.: Les Unités Sémiotiques Temporelles: Éléments nouveaux d'analyse musicale. Marseille: MIM/Documents Musurgia (1996)
27. Benetos, E., Dixon, S., Duan, Z., Ewert, S.: Automatic music transcription: an overview. IEEE Signal Process. Mag. **36**(1), 20–30 (2019)
28. Elowsson, A., Friberg, A.: Predicting the perception of performed dynamics in music audio with ensemble learning. J. Acoust. Soc. Am. (JASA) **141**(3), 2224–2242 (2017)
29. Juslin, P., Sakka, L., Barradas, G., Lartillot, O.: Emotions, mechanisms, and individual differences in music listening: A stratified random sampling approach. Music. Percept. **40**(1), 55–86 (2022)
30. Slaney, M.: Auditory toolbox: a Matlab toolbox for auditory modelling work. Apple Computer Technical report, vol. 45 (1994)
31. Ricard, J., Herrera, P.: Morphological sound description: computational model and usability evaluation. In: 116th Convention of the Audio Engineering Society (AES) (2004)
32. McAdams, S.: The perceptual representation of timbre. In: Siedenburg, K., Saitis, C., McAdams, S., Popper, A.N., Fay, R.R. (eds.) Timbre: Acoustics, Perception, and Cognition. SHAR, vol. 69, pp. 23–57. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-14832-4_2
33. Peeters, G., Deruty, E.: Automatic morphological description of sounds. In: Acoustics '08, Paris (2008)
34. Tzanetakis, G., Martins, L.G., McNally, K., Jones, R.: Stereo panning information for music information retrieval tasks. J. Audio Eng. Soc. **58**(5), 409-417 (2010)
35. Juslin, P.: Cue utilization in communication of emotion in music performance: relating performance to perception. J. Exp. Psychol. Hum. Percept. Perform. **26**(6), 1797–813 (2000)
36. Park, T.H., Hyman, D., Leonard, P., Hermans, P.: Towards a comprehensive framework for electro-acoustic music analysis. In: Proceedings of the International Computer Music Conference (2011)
37. Lartillot, O., Grandjean, D.: Tempo and metrical analysis by tracking multiple metrical levels using autocorrelation. Appl. Sci. **9**(23), 5121 (2019)
38. Foote, J.T., Cooper, M.L.: Media segmentation using self-similarity decomposition. SPIE Storage Retrieval Multimedia Databases **5021**, 167–75 (2003)
39. Klien, V., Grill, T., Flexer, A.: On automated annotation of acousmatic music. J. New Music Res. **41**(2), 1–21 (2012)
40. Gulluni, S., Buisson, O., Essid, S., Richard, G.: An interactive system for electro-acoustic music analysis. In: 12th International Society for Music Information Retrieval Conference (2011)
41. Couprie, P.: Designing sound representations for musicology. In: Sound and Music Computing Conference (SMC), pp. 563–569. Université de Saint-Etienne/INRIA/GRAME, Saint-Etienne (2022)
42. Desainte-Catherine, M., Santo, J.L.D.: L'acousmoscribe, un éditeur de partitions acousmatiques. In: EMS Conference (2009)
43. Park, T.H., Li, Z., Wu, W.: EASY does it: the electro-acoustic music analysis toolbox. In: 10th International Society for Music Information Retrieval Conference (2009)

44. Godøy, R.I.: Musical shape cognition. In: Grimshaw, M., Walther-Hansen, M., Knakkergaard, M. (eds.) The Oxford Handbook of Sound and Imagination. Oxford University Press, New York (2019)
45. Godøy, R.I.: Perceiving sound objects in the Musique Concrète. Front. Psycholo. **12**, 672949 (2021)
46. Nanayakkara, S.C., Taylor, E., Wyse, L., Ong, S.H.: Towards building an experiential music visualizer. In: 6th International Conference on Information, Communications & Signal Processing (2007)

Designing Musical Instruments and Room Acoustics with Acoustic Metamaterials

Rolf Bader(✉) and Patrick Kontopidis

Institute of Systematic Musicology, University of Hamburg, Neue Rabenstr. 13,
20354 Hamburg, Germany
r_bader@t-online.de

Abstract. Mechanical musical instruments have less timbre variability than electronic instruments. Extended playing techniques and more sophisticated acoustic instrument designs have recently appeared. We suggest acoustic metamaterials as a new way to extend the timbre of mechanical instruments beyond their present sound capabilities. In this chapter, we present three examples of acoustic metamaterials: (1) a one-dimensional string, (2) a labyrinth sphere, and (3) a two-dimensional membrane. The string is covered with additional masses, which leads to a dispersion relation of the harmonic overtones in the sound spectrum. The resulting sound still has a detectable pitch but is very different from a regular string on a mechanical instrument. The labyrinth sphere has a clear band-gap damping and can be used in loudspeakers, musical instruments, or room acoustics due to its small size. A circle of masses is attached to the membrane, leading to a cloaking behaviour of vibrations from within the circle to outside and vice versa. Again, the resulting sound is considerably different from a regular drum and leads to increased variability of musical articulations. Using a microphone array, laser interferometry, impedance tube, and high-speed video recordings with subpixel tracking, the vibrations on the string and the membrane are investigated and discussed in relation to new instrument designs.

1 Introduction

Designing sound is the aim of musical instrument builders, composers, musicians, and music software engineers. Several composers have suggested categorizations of sounds, such as Pierre Schaeffer's spectromorphology [23]. At the core of his thinking was the *sonic object*, a "chunk" of sound perceived as a *gestalt* or single object [22]. There are clear relations between sound objects and physical objects, such as the scratching of violin bows and strokes on percussion instruments [21]. Proposed categories for sound descriptions, such as Schaeffer's impulsive, sustained, and iterative categories, are closely related to sounds from everyday environments. Still, composers and musicians have always searched for new, unheard-of, or unexpected sounds. They might come from mechanical instruments, like the friction instrument Terpodium, the armonica glass instrument, the longitudinal waves used in Clavicylinder, or, more recently, electronic instruments [24]. The electronic music studios, starting from the 1950s, had technical

A. R. Jensenius (Ed.): SSD 2022, CRSM 12, pp. 298–320, 2024.
https://doi.org/10.1007/978-3-031-57892-2_16

possibilities to produce sounds perceived as ‘bigger than life.’ One early example is the out-of-body experiences reported when playing back voices on headphones with a time delay at the Studio für Elektronische Musik of the WDR in Cologne, which started a psychedelic music experience in the 1960s [24].

Acoustic metamaterials are promising further to enlargen the sonic objects available in mechanical musical instruments. These have acoustic properties not found in traditional materials [15,17]. Traditional materials used in mechanical instruments include wood (stringed instruments) and metal or bronze (gongs, bells, brass instruments). It has also been common to use artificial materials: nylon (in strings), mylar (drum heads), carbon (violin top plates), or other polymers or hybrid materials like sandwich plates (e.g. plywood used for jazz or classical guitars), textile structures (found, e.g. in combination with carbon), or plates coated with lacquer (stringed musical instruments).

Metamaterials are a new kind of material with novel acoustic properties, such as extreme damping, frequency band gaps, or acoustic cloaking. Metamaterials are not complicated to build, as their properties arise from being constructed with a complex geometry rather than from new polymers or the inclusion of nanoparticles. We believe that such metamaterials will dramatically increase the sonic capabilities of musical instruments in the coming decades.

This chapter first briefly introduces the way metamaterials work, focusing on the sonic qualities of acoustic metamaterials. It then discusses ways of sonically designing instruments, if and to which extent traditional musical instruments might already show metamaterial behaviour, and how such alternations influence music composition, performance, and musical instrument building. Finally, some examples of metamaterials used in musical instruments and for room acoustics and noise cancellation are given.

1.1 What Is an Acoustic Metamaterial?

Acoustic metamaterials have two fundamental properties:

- Complex, often periodic geometries acting on the sub-wavelength level of the frequencies to be manipulated.
- Properties like negative Young’s modulus, negative density or negative refraction, band gaps, acoustic lens effects, cloaking, or extreme damping.

The first property is astonishing at first, as the traditional view is that geometries smaller than the wavelength of an interacting wave do not alter the wave considerably. Still, when many of these small geometries, all of the size much smaller than an incoming wave (in the subwavelength domain), are placed next to each other (often in a complex way), the resulting geometry can manipulate an incoming wave tremendously.

The second property seems unphysical at first. Young’s modulus is the proportional constant between an applied stress on an object and the resulting strain. So, if one is pressing an object, it shrinks and compresses. If one needs a lot of force to compress the object, like with wood, the Young’s modulus is high;

if only little force is needed to make it shrink a lot, like with rubber, the Young's modulus is low. It never gets negative, which would be unphysical in a static case. A negative Young's modulus would imply that an object expands when pressed. However, waves are much more complicated, and metamaterials could lead to a negative Young's modulus. The same holds for density and mass per volume. In the static case, a negative mass is unphysical. However, with waves, a negative density is possible. This allows for negative refraction. If a wave hits an obstacle, e.g., a pillar in a room, the wave moves forward and spreads slightly in the side directions behind the pillar. This spreading is called refraction, and the amount of refraction is always positive, with zero refraction if no spreading occurs. A negative refraction means that a wave hitting a pillar would not spread but shrink in space. If so, the wave would condense into one single point and therefore, an acoustic lens is built. This scenario would benefit room acoustics, where the audience sitting behind a pillar would hear the music as if there were no pillars.

These fundamental metamaterial features allow many types of sonic design. We can distinguish between two manipulation approaches: the frequency domain (acting as a filter) and the spatial domain (acting in a room). The first is attractive for sonic design, and the second is relevant for room acoustics. We will look at some examples of frequency manipulation later. Room acoustics examples include Metamaterial Wall [27] for extreme low-frequency damping in recording studios and rehearsal rooms.

Like semiconductor materials, band gaps are a solution to a wave equation with single masses. So is a complicated dispersion relation, a frequency-dependent wave-speed, leading to a deviation of the strings spectrum away from a harmonic relation of 1:2:3:... of the spectral partials. Such broadband sounds are not known from musical instruments or natural materials.

Regarding spatial audio, metamaterials can be used in two forms. In the drum examples discussed below, a spatial cloaking on the drum head leads to a broadband band gap. This band gap is the sound part of the manipulation. Still, the complex wave distribution on the drum head leads to a tremendously altered sound radiation from the drum head. The apparent source width (ASW), binaural Interaural Cross Correlation or generally, the spaciousness of the sound will change strongly. So, although the spatial distribution of the waves on the membrane cannot be seen when playing too much, they still play a part in aural perception.

The cloaking behaviour found with the membrane was first introduced as a single-frequency suggestion with electromagnetic waves [18] and was also shown for plates [19]. A possible neat application is hiding an object from outside inspection of electromagnetic waves, like when passing the security luggage control of an airport. As inspection is performed with a single electromagnetic wave frequency, placing an object in a metamaterial will hide it from such inspection. Another single-frequency example is that of an acoustic lens. If a wave hits an obstacle, it is diffracted with a positive diffraction angle, making it broader spatially. When using a metamaterial with a negative Young's modulus and negative density, which is impossible with natural materials, the diffraction index gets negative, and the spatial distribution of the wave behind the obstacle gets smaller. Such a wave will condense behind the obstacle into a single point, like the focus of an acoustic lens [17]. Still, until today, these applications have only worked for single frequencies and are, therefore, unsuitable for musical applications. An example of the application is a metamaterial covering a pillar in a concert room such that a person sitting behind the pillar gets the sound right the way as if the pillar was not there, as mentioned above. There is no application yet for broadband signals.

In this chapter, three examples of applying metamaterial behaviour to musical instruments are demonstrated with a membrane, a device for loudspeakers, and a string. All result in interesting new sounds or increased articulatory ability for players. The chapter first discusses the measurement setup and techniques applied for the instruments, mainly a microphone array, laser interferometry, impedance tube measurements, and high-speed camera recordings with subpixel tracking. The results section discusses the instruments and their possibilities.

1.2 Are Musical Instruments Already Metamaterials?

Traditional musical instruments are not built to be metamaterials. Still, the many existing complex instrument geometries might lead to reconsidering them as such. Although the fan bracing of guitars or the bracing of piano soundboards is built mainly for stability, such regular substructures might lead to behaviour similar to metamaterials. Indeed, the pitch glides of Chinese gongs [9] and the brassiness of crash cymbals [10] or tam-tams [11] are caused by complex geometries.

Membranes used in rock or jazz drum kits, as well as with *tablas* of Indian music, or the *pat wain* or the Myanmar *hsain wain* orchestra, often show additional masses attached to them. They are used for different purposes. Jazz drummers use tape and other dampening materials, especially on the snare drum. Also, tom-toms are taped to reduce the loudness and duration of their tone. Here, the detuning of these drums plays a minor role as they are tuned by tuning pegs at the drum head rim. *Tabla* [12,13] and *pat wain* [14] drums are tuned by adding a plate or a tuning paste. The aim is twofold: the drum is tuned to a pitch, and the overtone spectrum is changed to a more harmonic overtone spectrum of the fundamentally inharmonic spectrum. This increased pitch perception makes them more usable in melodic performance.

Strings with regularly attached masses are phononic crystals [15]. They show a dispersion relation with a non-constant slope and may have band gaps. Musical instruments with such strings are very seldom. One example is the ancient *m'na'anim*, a Jewish instrument dated around 1000 BC and shown in the Encyclopedia of Diderot [16]. It consists of a wooden box, similar to a *cajon*, with one string on which wooden balls are attached. As it is not played today, the resulting sound is unknown. In contemporary classical guitar music, attaching an additional mass on the guitar strings next to the bridge is standard to produce a sound similar to that of the *mbira*, a West African thumb piano. The instrument still sounds pitched, but inharmonic frequency components add a percussive timbre. The *mbira* has a similar sound; the metal plates have a strong fundamental frequency, resulting in clear pitch perception. At the same time, the additional inharmonic frequencies of the rod are responsible for the percussive part.

2 Methods

2.1 Frame Drum

Metamaterials have been used with membranes to achieve damping over a large bandwidth [6,7]; for a review, see [8]. With massive rings attached concentric on the membrane, one or only a few resonance frequencies exist up to 1 kHz, which leads to strong damping of the membrane within this range with prominent peaks at the resonance frequencies. Such applications differ from the concept proposed in this paper. There, a strong overall damping is aimed for, whereas with musical applications, only a partial damping is needed to maintain an audible sound. Also, with such heavy masses, the membrane between the mass and the membrane boundary and the membrane between two rings can mainly be considered like a spring. Then, no additional vibrations are expected on the membrane. This is similar to the case of a one-dimensional sonic crystal, where masses are attached to springs, only that the membrane is two-dimensional. As with concentric rings, the distance between the ring's outer and membrane boundaries is constant for all angles; only one spring length and strength are present. So, these applications differ in principle from the construction and the aims of the dot masses attached asymmetrically on a membrane present in this study.

As discussed above, circular, linear or complex shaped geometries might result in a cloaking behaviour, where a travelling incoming wave looks the same in both cases: with and without the structure in its way. Therefore, for an observer behind the structure, this structure is invisible [17]. Such geometries can also act as cages, where waves cannot travel out and vice versa. This has been found in optics [18] and applied in acoustics [19,20]. This behaviour is frequency dependent and, therefore, is also a way to build a musical metamaterial, enhancing the articulatory ability of a musical instrument.

A frame drum with a mylar drum membrane and a diameter of 40 cm was used [28]. At the drumhead, a circle-shaped area (m) with a diameter of 10 cm is separated using a set of 2×10 neodymium magnets sticking at the front and the back of the membrane. The magnets are circular, with a diameter of 5 mm and a height of 5 mm.

The area separated by the magnets is assumed to act as cloaking, separating vibrations inside and outside this area.

2.1.1 Laser Interferometry

A Verdi Single FAP (fibre array package) diode-pumped solid-state frequency-doubled Neodynium Vanadate ($(\mathrm{Nd} : \mathrm{YPO}_4)$ laser (LSR) source radiates a beam of wavelength $532\,nm$ and beam diameter of $d_{LSR} = 2.25 \pm 10\%$ mm. The beam

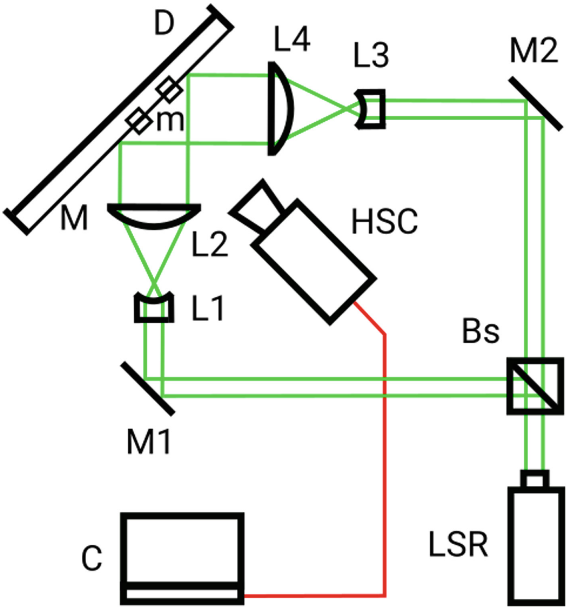


Fig. 1. The experimental set-up. (LSR) laser, (Bs) Beam splitter, (M1, M2) planar mirrors, (L1, L3) semi-concave lenses ($f_{L1,L3}$=−16 mm, $d_{L1,L3}$=10 mm), (L2, L3) semi-convex lenses ($f_{L2,L4}$=300 mm, $d_{L1,L3}$ = 100 mm), (M) drumhead of the frame drum (D), (m) circle-shaped part of the drumhead, separated utilizing a set of 2 × 10 neodynium magnets. (HSC) high-speed camera, (C) analysis using a PC. Green lines mark the beam paths. (Color figure online)

is split by a beam splitter (Bs). The split beams are directed to planar mirrors (M1) and (M2). Subsequently, the beams are expanded via an optical lens system consisting of a semi-concave lens with a focal distance of $f_{L1,L3} = -16$ mm and a diameter of $d_{L1,L3} = 10$ mm and a semi-convex lens with focal distance of $f_{L2,L4} = 300$ mm and a diameter $d_{L2,L4} = 100$ mm.

For the laser interferometry experiment, the drumhead was manually excited by an impulse hammer. The excitation was performed outside as well as inside the separated area of the drumhead. The split and widened beams were directed to the drumhead (M) of a frame drum (D). The impulse response leads to a characteristic interference pattern at the drumhead. The pattern was recorded using a high-speed camera (HSC) with a frame rate solution of 10,000 frames per second. The received data were analyzed utilizing Mathematica on a PC by subtracting adjacent recorded frames [5].

The drum head was also excited by a Brüel & Kjaer Vibration Exciter 4809 in the middle and outside the circle. Both low (65 Hz) and high (918 Hz) frequencies were used, two eigenfrequencies of the drum head with magnets on. The described experimental set-up is depicted in Fig. 1.

2.1.2 Microphone Array

The sound pressure fields of the frame drum were recorded with a microphone array in the near-field, 3 cm in front of the membrane. The grid constants of the array are 5 cm in the x-direction and 4 cm in the y-direction. The microphone array records sound fields with up to 128 microphones with a sampling frequency of 48 kHz and a sample depth of 24 bits simultaneously (Fig. 2).

Fig. 2. Modified frame drum positioned in front of the microphone array.

The recorded sound fields are back-propagated to the surface of the membrane using the Minimum Energy Method (MEM) [1], a multipole method assuming as many radiation sources as microphones. It has successfully been used to measure the vibrations of musical instruments [3,4] (for a review on microphone arrays and back-propagation methods, see [2]).

For the recordings with the microphone array, the drum was struck at three positions: within the circle, at the circle rim between two magnets, and outside the circle at the position opposite to the circle. Each recording resulted in 120 sound files at the microphone positions. The frequency spectra were calculated from these, and all peaks up to 1 kHz were determined. For each of these frequencies, the recorded sound field was back-propagated to the surface of the drum.

2.2 Impedance Tube

It is common to use an impedance tube to obtain the acoustic properties of a material or geometric structure. The reflection and absorption coefficients and the characteristic specific impedance can be found during this measurement process. The tube is designed with a speaker on one and a sound-hard boundary on the opposite end. A sample material can be placed before the sound-hard boundary to measure sinusoidal or noise inputs inside the tube. Several standards can be used to realize the method, such as the standing wave method [25] or the transfer function method [26]. The latter has some advantages regarding the measurement effort and the tube size. Still, the standing wave method guarantees a better signal-to-noise ratio due to the possibility of band-pass filtering. Consequently, it was decided to use it here.

2.3 Modified String

A steal string of length 74.5 cm and diameter 0.25 mm was modified by attaching 74 lead masses along its length with a mass-to-mass distance a = 1 cm. Two different masses are added adjacently, a lighter mass of $m_1 = 0.008$ g and a heavier mass of $m_2 = 0.08$ g. The string is attached to a wooden plate over two bridges. It is typically used as a monochord, an instrument used since ancient times to discuss the relation between string length and musical intervals.

The string was displaced, and the sound was recorded using a piezo attached to the bridge and a microphone near the radiating soundboard. The string with attached masses was tuned to two fundamental frequencies, 100 Hz and 200 Hz.

Analytically, such a string has a dispersion relation with two frequency bands separated by a band gap like [15]

$$\omega^2 = \beta\left(\frac{1}{m_1} + \frac{1}{m_2}\right) \pm \sqrt{\beta^2\left(\frac{1}{m_1} + \frac{1}{m_2}\right)^2 - \frac{4\beta^2}{m_1 m_2} sin^2 ka}\ , \tag{1}$$

where β is the string tension, m_1 and m_2 are the two masses, k is the wave number, and $a = 1$ cm is the grid constant. The graph of this plot is shown in the results section, together with the measurements.

Both the lower and the higher frequency bands contain discrete eigenfrequencies of the manipulated string, which are no longer harmonic, as they would be with a simple string. Compared to a regular harmonic spectrum, the lower band shows a compressed harmonic spectrum, while in the higher band, the spectrum is both compressed and stretched, again compared to a harmonic spectrum. As the number of eigenfrequencies in the lower band equals the number of masses, 74 in our case, and when tuning the lowest to 100 Hz, considering the compression of the spectrum, a band gap is expected starting from about 7 kHz. Additionally, a second band gap should appear at about 10 kHz.

2.3.1 High-Speed Camera and Subpixel Tracking

Additionally, the movement of the string was recorded using a high-speed camera (Vision Research Phantom V711) with a frame rate of 10,000 fps. Using the subpixel tracking software MaxTraq, the motion of all 74 masses was extracted from the high-speed video, resulting in a 74-time series. The time series were Fourier analyzed, and the string displacement could be shown for the string eigenmodes.

3 Results

3.1 Metamaterial Drum

The drum was struck at three different positions since it was expected that the circle would act as a cloaking to the sound. If so, striking within the circle should keep most of the vibrations on the inside, while striking outside the circle would lead to a much-lowered energy in the circle. In Fig. 3 and Fig. 4, the results of the microphone array recordings and back-propagation are shown considering this point.

Figure 3 was calculated by first detecting the maximum absolute amplitude of each mode. Then, only these maximum amplitude positions were accumulated on the membrane for each strike case. Then, all points on the membrane showing more than 20% of accumulated maximum points are displayed.

The case of striking in the circle is shown at the top of Fig. 3. Most maximum points are within the circle. When striking at the circle rim, shown in the middle graph, the distribution of maximum amplitudes is more widespread over the membrane. Finally, in the case of striking outside the circle, shown as the bottom plot in the figure, no considerable maxima are within the circle.

To differentiate this finding with respect to frequency, the amount of absolute amplitude within the circle is shown in Fig. 4 as a fraction of the whole absolute amplitude on the drum. The curves show the three cases of striking in the circle, at the rim and outside the circle. Again, striking in the circle leads to a strong increase of amplitudes within the circle, compared to the cases of striking at the circle rim and outside the circle. Still, this increase only appears above about 400 Hz. As the fundamental frequency of the drum is 34 Hz, we can conclude that the low frequencies are not much affected by the circle, while the higher ones are.

The relatively high fraction of amplitudes in the circle at very low frequencies is remarkable. The lowest peak detected at 7 Hz is not audible and most likely refers to the motion of the drum as a whole, including the wooden frame. This motion is unavoidable as frame drums only sound when the wooden frame is

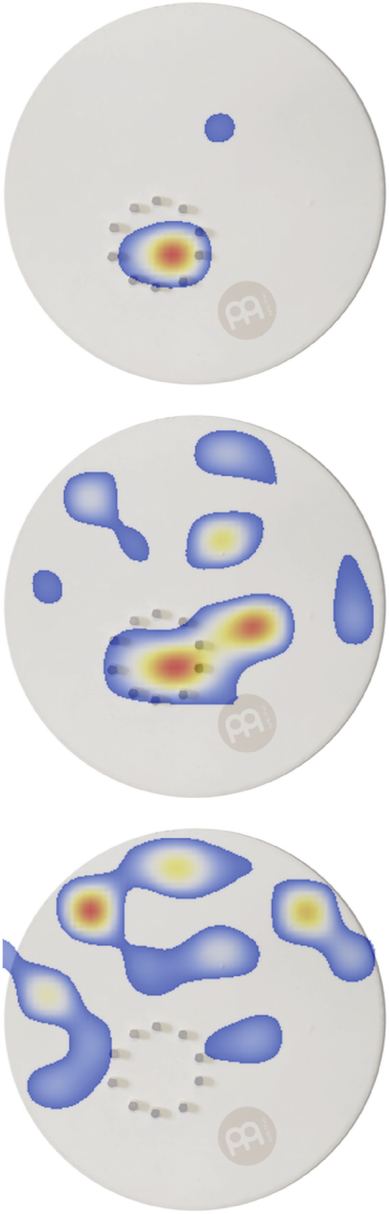


Fig. 3. Density distribution of maximum amplitude values of modes on the drum up to 1 kHz for three hammer strike positions, showing densities above 20%. Top: strike in the circle, Middle: strike at circle rim, Bottom: strike outside the circle at the opposite side of the circle. While most maximum values for the strike are in the circle, very few are within the circle when the drum is struck outside the circle. A medium case is found when striking at the circle rim.

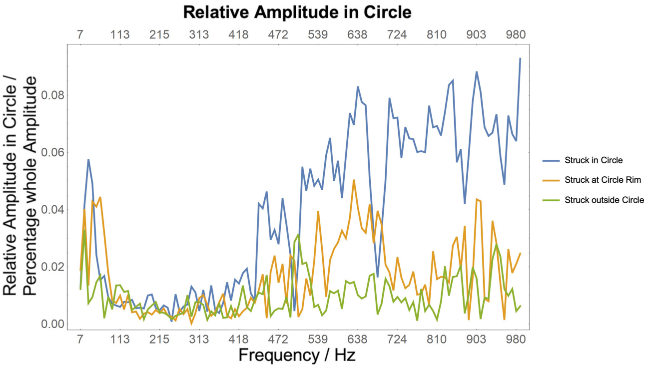


Fig. 4. Frequency-dependent absolute amplitude within the circle compared to total absolute amplitude on the whole drum for three strike cases: a) in the circle (blue), b) at the circle rim (orange), and c) outside the circle (green). For frequencies below around 400 Hz, the amplitude strength of the three cases is about the same; above about 400 Hz, the amplitude strength within the circle strongly depends on the strike position. Strikes in the circle have stronger amplitudes there than strikes at the rim, with strikes outside the circle showing the least amplitudes in the circle. (Color figure online)

free. Fixing it firmly, which would avoid this low vibration, would lead to a very much damped sound and can not be implemented in an experimental setup.

Still, the very low frequencies at 34 Hz (the ‘monopole’ vibration) and around 65 Hz (the ‘dipole’ vibration), show much more relative amplitude within the circle in the cases of striking in the circle and striking at its rim, compared to the case of striking outside the circle. The reason for this behaviour can be found when examining the modes more closely. The very low modes need to make the circle region move, too, as the anti-node regions are large. With 34 Hz it basically covers the whole membrane. With the 65 Hz ‘dipole’ case the membrane is two-split with two regions about half the membrane each. Of course, no monopole and dipole modes exist due to the circle, as in the case of an isotropic membrane.

The higher modes above the dipole, quadrupole, octopole and many other more complex modes with an integer number of axial and circular nodal lines, these modes can deform in such a way as to avoid the motion of the circle region nearly completely. This holds for all three strike cases. It seems that even when striking in the circle, the circle cannot maintain a vibration of these frequencies. The small leakage of vibrations leaving the circle is then taken over by the rest of the membrane, leading to a similar motion when striking outside the membrane.

To confirm these findings in Fig. 5, laser interferometry measurements for the case of striking in the circle are shown. The transient strike is displayed as six snapshots at 0 ms, 0.2 ms, 0.6 ms, 1 ms, 3 ms, and 6 ms. Each black/white line indicates an amplitude increase of one wavelength of the used laser light. Therefore, many circles do not indicate an amplitude ripple but a steep amplitude slope.

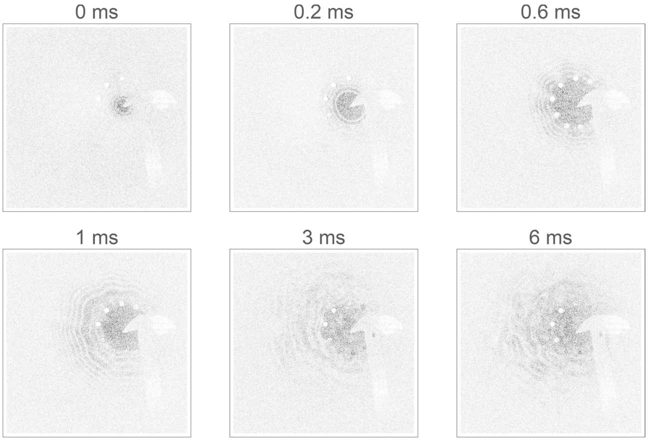


Fig. 5. Laser interferometry time-dependent measurement of the initial transient of a hammer strike on a drum with a separated circular area for several time steps. At 0 ms, a circular wave leaves the strike point, which meets the circle boundary at about 0.2 ms. The boundary elements lead to a split of the circle and the appearance of Huygens wavefronts outside the circle beyond 0.6 ms. At 3 ms, the reflected waves on the membrane lead to complex vibrations.

Starting at 0 ms, the strike leads to a circular wavefront, leaving the strike point, which is shown at 0.2 ms. At about 0.6 ms, this circular wavefront meets the circle rim. Here, it is scattered, and new wavefronts start at the open rim positions, as expected. At 1 ms, these wavefronts form another wavefront outside the circle, slightly ripped, as this wavefront is formed from a finite number of elementary waves according to the Huygens principle. Two cases at 1 ms and 3 ms show the wavefront outside the circle becoming more and more complex as the wavefront is then already reflected at the drum boundaries and leads to a complex waveform.

It can be seen at 1 ms that the circle still has a strong amplitude, much stronger than that leaving the circle. This picture continues at 3 ms and 6 ms, supporting the findings above that most amplitudes stay within the circle when striking.

The same transient time development when striking outside the circle is shown in Fig. 6. Again at 0 ms a circular wave leaves the impact point which arrives at the circle at about 0.6 ms. At 1 ms, the strong amplitude is still present outside the circle while only a small fraction enters the circle. This continues at 3 ms. At 6 ms, there is also some energy in the circle, which is expected from the above findings. The circle region also moves with some amplitude for very low frequencies at 34 Hz and 65 Hz. Still again, overall, most vibrations keep out of the circle when striking outside.

To differentiate the low/high-frequency difference further, the drum was driven by a shaker in and outside the circle at two frequencies, 65 Hz and 918 Hz.

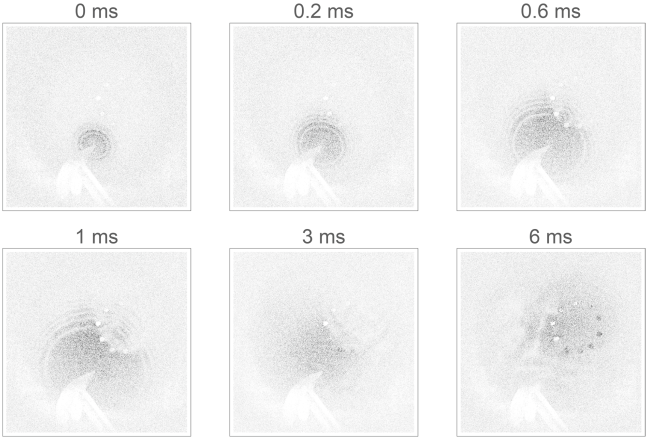


Fig. 6. Laser interferometry measurement of a hammer strike on a membrane with a separated circle area, striking outside the circle. A circular wavefront leaves the strike position and reaches the circle boundary at 0.2 ms. The boundary leads to the formation of a Huygens wavefront inside the circle from about 0.6 ms. From 1 ms on, the vibrations inside the circle are much less than those outside. Still, after about 6 ms, there is motion inside the circle, at small wave vectors and, therefore, at low frequencies only.

In Fig. 7, snapshots of the vibrations are shown at maximum amplitudes of the sinusoidal vibrations. On the top row, the 65 Hz cases are shown on the left of the case when driving in the circle and on the right when driving outside the circle. Broad vibrations can be seen in both cases, indicating a distorted dipole motion. Although the amplitude is stronger inside the circle when driving inside than outside, some amplitude is still outside. When driving outside, the amplitude is about equally distributed. This follows the microphone array's findings, especially with that of Fig. 4. There, energy was present in the circle in all striking cases and even stronger when striking in the circle.

The two lower plots in Fig. 7 show the laser interferometry measurements for the 918 Hz driven sinusoidal, again driven inside the circle on the left and outside on the right. When driving inside the circle, nearly all amplitudes are within the circle, while when driving outside, nearly all amplitudes are outside the circle. At the same time, the circle boundary cloaks the inner circle area.

The circle is cloaking vibrations in both directions, from within the circle to outside of it and vice versa, for frequencies above about 400 Hz. For frequencies below 400 Hz, it is cloaking such that vibrations from outside do not enter the circle. Still, some vibrations escape the circle and form modes outside when driving the circle. But also, in this case, the circle is not taking part in the vibrations considerably. The cloaking no longer works for very low frequencies caused by large anti-nodal areas on the membrane.

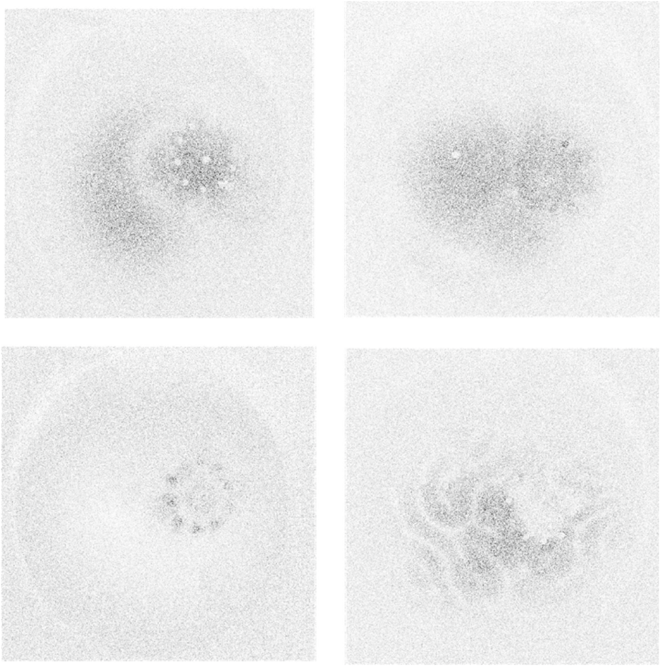


Fig. 7. Snapshots of forced oscillations at 65 Hz (top row) and 918 Hz (bottom row) inside (left column) and outside (right column) the circle. At the low frequency of 65 Hz, the vibrations are strong both inside and outside the circle. At the high frequency of 918 Hz the driving of the membrane inside the circle only leads to a vibration inside, while driving the membrane outside the circle the movement is only outside the circle and very low amplitudes are present in the circle. Therefore, the circle at this frequency of 918 Hz acts as a cloaking of waves in both directions. Comparing with Fig. 4 allows the conclusion that above about 400 Hz, the circle acts as a cloaking element.

3.2 Spherical Labyrinth Structure

The spherical labyrinth structure metamaterial was 3D printed using polylactic acid (PLA), a material suitable for such additive manufacturing. Since an experiment on a plane PLA plate revealed its sound-hard properties, it is assumed that the kind of material does not play a crucial role as the air cavities within the structure cause the metamaterial behaviour. The spherical structure was placed in the mid-point of the impedance tube, with the round side facing the driving speaker. This round front side has small holes where the sound travels into tubes separated by walls. In Fig. 8, the structure is displayed with its front and back sides.

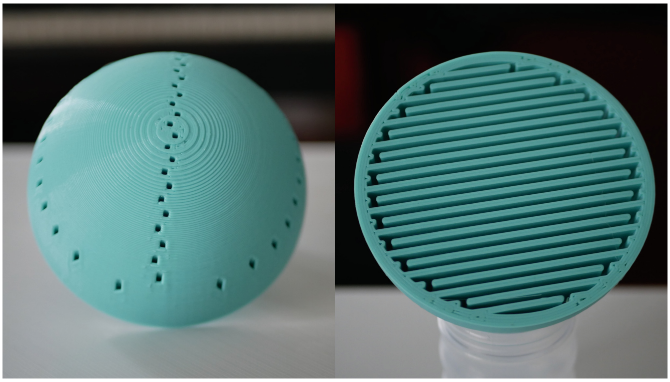


Fig. 8. Spherical metamaterial with holes at the front side and a labyrinth structure inside.

Different boundary conditions at the back of the spherical structure were applied. At first, open boundary conditions were used with no reflective wall behind the sphere, as shown in Fig. 9. The spectrum of reflected waves shows strong reflection throughout the whole frequency range with only about 20% transmission through the sphere.

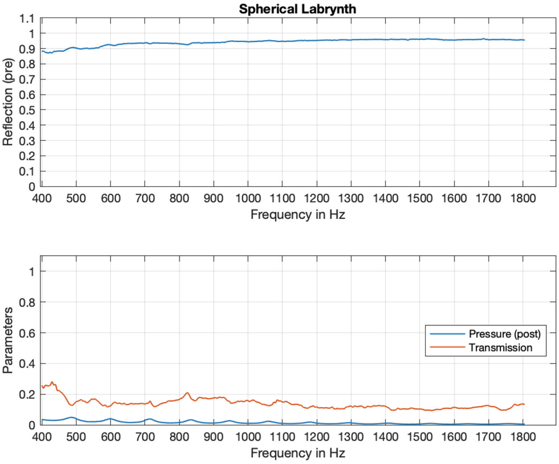


Fig. 9. Reflection, transmission, and pressure curves of the metamaterial without a reflective boundary fixed at its back. The upper plot shows the reflected sound in front of the metamaterial. The plot below shows the detected transmission and sound pressure after travelling through the geometry.

The reflection behaviour strongly changes when a PLA plate of 3 mm thickness is applied to the back of the sphere. A band gap at 770 Hz appears with about 60% absorption, shown in Fig. 10. As 20% are still transmitted, the reflection at the band gap is reduced from 80% to 20% due to the change in boundary conditions.

The band gap is typical for metamaterial behaviour. It only appears when a plate is applied behind the sphere. Still, the plate is no metamaterial on its own. The sphere only develops metamaterial behaviour with such boundary conditions. This is unexpected, as an open tube theoretically has a boundary condition of zero pressure, so there is no transmission. It is, therefore, a reflective boundary condition, like a wall boundary. Still, the wall boundary allows arbitrary and much higher pressure inside the labyrinth than the open boundary condition.

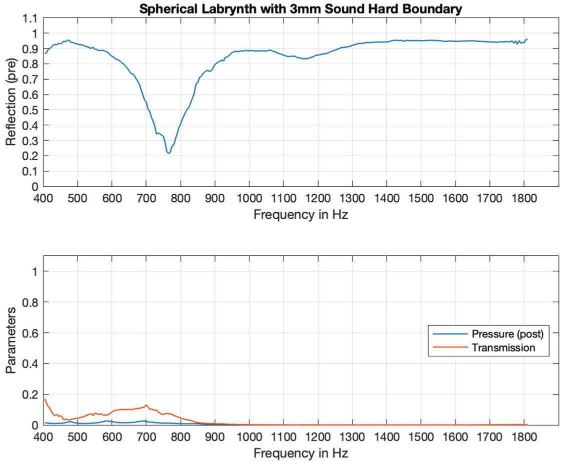


Fig. 10. Same as Fig. 9 but now with an attached back plate to the sphere. The upper plot shows a band gap in the sound reflection caused by the combination of the labyrinth sphere and the back plate. The lower plot shows about 15% of transmission at the band gap.

This structural change provides an excellent opportunity to test the initial hypothesis that the material does not affect the metamaterial behaviour and that only the air volume causes the band gap. Therefore, the back plate of 3 mm thickness is now replaced by one with 10 mm, again manufactured using PLA and placed in the same position as before.

Again, a band gap appears, but now at 680 Hz, lowering the 770 Hz of the 3 mm plate by 90 Hz, so quite considerable, as shown in Fig. 11. Additionally, the transition is nearly zero.

So, the labyrinth sphere only acts as a metamaterial with a closed back. The low transmission in the case of no back plate means to be caused by the boundary conditions of zero pressure at the back of the sphere with no back

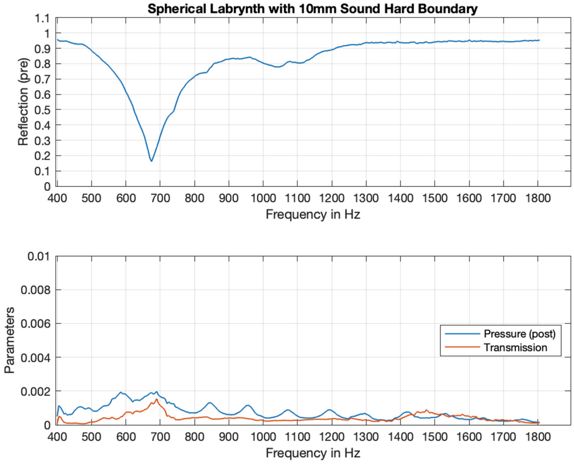


Fig. 11. Same as Fig. 10 but now with a back plate of 10 mm thickness. The upper plot shows a frequency shift of the band gap by 90 Hz from 770 Hz to 680 Hz. The transmission is nearly zero.

plate attached. This zero pressure strongly reduces the sound pressure inside the labyrinth, not allowing sound pressure to enter the sphere considerably, making it highly reflective.

When attaching a back plate, arbitrary pressures are allowed on, and therefore also inside, the labyrinth boundary. Then, the fundamental mode of the labyrinth resonates at 770 Hz and leads to a strong damping around this frequency. A band gap appears.

When adding a thicker back plate, the reflection at the band gap decreases to near zero. This points to the thinner back plate vibrating with the band gap frequency and radiating energy further down the tube. This only appears at the band gap. With the 3 mm plate, transmission outside the band gap is nearly zero. When testing the 10 mm plate alone in the impedance tube, it shows nearly perfect hard boundary conditions, so nearly total reflection.

The decrease of the band gap frequency with the 10 mm plate compared to the 3 mm plate can only be caused by a change in the effective length of the small holes at the front of the sphere and their end-correction. The acoustic length of a tube differs from its geometrical length due to the air at the tube ends moving outside the tube a bit. This leads to an end-correction of the tube, making the acoustic length larger than the geometrical one. This behaviour is frequency-dependent. Lower mode frequencies have a larger end-correction than higher ones. This end-correction is also expected to increase with increased sound pressure inside the labyrinth. This leads to a decrease in the resonance frequency of the labyrinth. Indeed, the strength of this decrease by 90 Hz is unexpected.

The spherical labyrinth has strong damping behaviour in a band-gap manner, where the band-gap frequency can be altered by changing the back plate thickness. Further experiments with a different-sized geometry could underpin these results and show if the effect is also scalable by varying the dimensions of the labyrinth itself.

3.3 Manipulated String

The spectrum of the manipulated string, as measured by the piezo and the microphones, shows peaks in two bands, one band from 100 Hz to about 7 kHz and the other from 9 kHz to 12 kHz. The peaks are in a quasi-harmonic spacing starting from 100 Hz. Many peaks are double, pointing to degenerated modes. These can easily occur in such a system if the masses are not perfectly equally spaced, which is nearly impossible. Double peaks lead to a beating, an interesting musical effect often artificially created, e.g., with the piano where three strings represent one key in the middle range and are detuned slightly from one to another to achieve a beating.

To be sure that 100 Hz is the lowest mode, the highspeed camera recordings of the vibrating string were analyzed using subpixel tracking. Due to the small amplitude of the string in combination with the large amount of 74 masses, the time series of the masses, as taken from the subpixel tracking algorithm, were quite noisy. Fourier-analyzing each time series and taking the amplitudes and phases of the lowest three frequencies could reconstruct the mode shape. Due to the poor signal-to-noise ratio (SNR), the modes are very noisy. Still, this analysis aimed to identify the modes as the fundamental and the next two higher modes. In Fig. 12, the first three modes are shown. They are clearly the lowest three modes of a vibrating string. We are, therefore, sure that the lowest frequency of the string is really at 100 Hz.

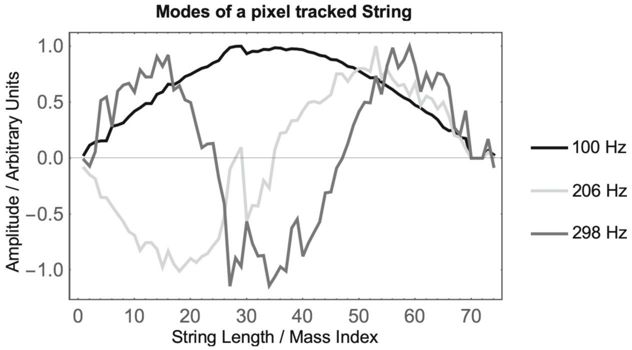


Fig. 12. Mode shapes of the first three modes of the manipulated string as analyzed using a high-speed camera and subpixel-tracking each mass. The resulting 74 time series were analyzed, and the amplitudes at the three lowest frequencies were plotted. The modes are distorted due to background noise, still it is possible to identify them with their respective frequencies. These frequencies are no longer in a harmonic ratio of 1:2:3.

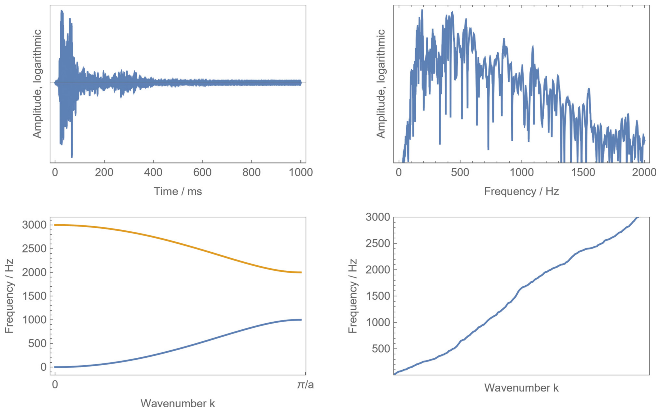


Fig. 13. Time series (top left) and spectrum (top right) of a string with 74 masses attached, showing no longer regular harmonic overtone spectrum and sound time decay. On the bottom left is a theoretical dispersion relation between frequency and wave number in units of grid (mass) distance, having two branches with a frequency gap between them. The lower branch is expected for a string with regular masses attached, all having the same weight. On the bottom right, a rough dispersion relation of the measured string is shown, coming close to theoretical expectations.

Figure 13 shows the time series of a string picked near one end. The sound is very inharmonic, shows a complex time development, and sounds very different from what one would expect from a mechanical string. The spectrum at the top right shows blurred peaks, still not perfectly irregular. On the bottom left, frequency vs. wave number shows a theoretical dispersion relation of a string with adjacent larger and smaller masses. A perfectly harmonic overtone spectrum would show a straight line with a constant positive slope, as wave speed $c = \omega/k$ holds for a perfect string. The shown dispersion relation has two branches with a band gap between them. Both branches are curved. An automatic peak tracking algorithm was implemented in Mathematica to estimate the dispersion relation of the string. Peaks up to 3 kHz are displayed at the bottom right of the plot. They show a similar behaviour to the theoretical lower branch of the figure on the bottom left. Still, no band gap was found, indicating that the string's parts between the masses are not perfectly rigid and move to some extent, making higher frequencies possible. Still, in this case, the resulting sound is not expected from a mechanical musical instrument due to the metamaterial effect, the sonic crystal.

4 Conclusions

The use of metamaterials can change a musical instrument's sound considerably. Changing existing instrument geometries can lead to added band gaps in their spectra, and using several such band gaps will lead to a designed sound, as shown in the examples in this chapter. With percussion instruments, musical articulation is realized by striking or knocking at different positions on, e.g. drums or cymbals. By adding metamaterial structures to them, the variability of such sounds can be increased considerably. The dispersion relation of sonic crystals leads to the stretching or contraction of harmonic overtone structures. Such shifting of frequencies in a harmonic spectrum makes it pseudo-harmonic, an interesting effect of a combined harmonic/inharmonic sound.

The cloaking of a membrane circle is frequency-dependent because the circle is not in a free field but on a membrane with boundaries leading to eigenmodes of the whole system. For high frequencies, the eigenmode shapes outside the circle are complex enough that the membrane acts like a free field, and therefore, the regular cloaking behaviour appears. For lower frequencies (here below 400 Hz), cloaking still works in one direction: waves from outside do not enter the circle to a large extent, but waves can leave the outside area when driving within the circle. For very low frequencies, the cloaking then nearly vanishes.

The transient laser interferometry measurements also showed that some energy leaves the circle when striking the drum in the circle at the very beginning of the sound. These vibrations trigger the modes between about 100 Hz and 400 Hz outside the circle. Therefore, deciding which frequency range to drive when striking in or outside the circle is possible.

It also appears that when striking at the rim of the circle, a mixture of the two extremes, striking outside the circle or at the very centre of it, can be achieved. This holds for both the frequency range up to about 400 Hz and that above this range.

Furthermore, the cloaking of the circle leads to a different radiation behaviour of the drum than when struck outside the circle. At higher frequencies, only the circle area vibrates; it acts like a monopole and radiates sound from a clearly defined point. When striking outside the circle, complex modes appear with an entirely different radiation behaviour. Therefore, depending on the driving point, the same frequency might have two different radiation patterns. A monopole radiation is perceived as a loudspeaker-like source, while a complex radiation pattern is perceived as a live musical instrument. Thus, a musician gets new articulation possibilities with such a manipulated drum.

The drum shows much higher timbre variability than a regular drum. With regular drums, the drummer can only vary the sound by striking at different positions. Striking in the middle leads to a sound dominated by low frequencies, and striking more to the edge increases the energy at higher frequencies, making the sound brighter. Striking outside the circle with the presented manipulated drum, these articulations are still possible. Additionally, the drummer can produce entirely new sounds when striking the membrane at different positions within the circle.

When striking at the very centre, even very strongly, the sound has only energy in the low frequencies. Still, it differs from a regular drum struck in its middle due to the transient behaviour of such a strike; higher partials are more present than a regular drum struck at the drum centre. Such a sound is not possible to produce for drummers with regular drums.

The labyrinth sphere is suitable for damping single frequencies with a small geometry. Furthermore, it can be tuned by manipulating the labyrinth boundaries and altering the damping frequency. Adding several such spheres will lead to a sonic design of a broadband spectrum for airborne sounds. Such structures can be used in loudspeakers to design a desired sound and in any other structure for sound manipulation, like in guitars, drums, or other instruments with cavities. Applications in room acoustics are also possible. When tuning the damping spectrum by mechanically altering the boundary conditions of the single spheres, a musician, composer, or sound designer can alter the instrument's overall sound or radiating body.

The manipulated string is much harder to take control of. Theoretically, a band gap should have appeared in this sonic crystal, although such a band gap could not be found. The reason seems to be the assumption that the string between two masses is not moving within itself but only acts as a single spring. This is an oversimplification; therefore, such a band-gap does not appear. Still, the dispersion relation is complex, leading to a spectrum that is no longer harmonic and far from random. A clear pitch can be heard, making an instrument with such strings still usable for musical performance. The overall sound appears between a string and a percussion instrument. The slight differences in the attached masses lead to degenerated modes and, therefore, additional beating in the sound. Also, the temporal development of a played tone is considerably different from that of a regular string with about exponential decay. Maybe this decay is even more unheard of than the spectrum, as it is not that of a string but cannot come from a percussion instrument, which would also decay exponentially. Further investigations into the instrument's decay are necessary to understand such a behaviour. Nevertheless, the sound is impressive due to its unnaturalness.

The search for new sounds using acoustic metamaterials has only begun. The amount of possible applications is tremendous and is present in all parts of mechanical sound production and room acoustics. Furthermore, these ideas can also be explored in electronic and algorithmic music production, adding metamaterial behaviour to filters, physical modelling, or other electronic music production techniques.

5 Supplementary Material

Nine sound examples of a metamaterial drum and two of a metamaterial string are available in this repository: https://zenodo.org/records/10512430.

Acknowledgements. We thank Paul Testa and Jost Fischer for helping with the experiments.

References

1. Bader, R.: Reconstruction of radiating sound fields using minimum energy method. J. Acoust. Soc. Am. **127**(1), 300–308 (2010)
2. Bader, R.: Microphone Array. In: T. Rossing (ed.) Springer Handbook of Acoustics, pp. 1179-1207 (2014)
3. Bader, R.: Radiation characteristics of multiple and single sound hole vihuelas and a classical guitar. J. Acoust. Soc. Am. **131**(1), 819–828 (2012)
4. Bader, R.: Characterizing classical guitars using top plate radiation patterns measured by a microphone array. Acta Acust. Acust. **97**, 830–839 (2011)
5. Moore, T.: Measurement Techniques. In: Bader, R. (ed.) Springer Handbook of Systematic Musicology. Springer Handbooks, pp. 81–103. Springer, Heidelberg (2018). https://doi.org/10.1007/978-3-662-55004-5_5
6. Naify, C.J., Chang, C.-M., McKnight, G., Null, S.: Transmission loss of membrane-type acoustics metamaterials with coaxial ring masses. J. Appl. Phys.**110**, 124903-1-8 (2011)
7. Tian, H., Wang, X., Zhou, Y.-H.: "Theoretical model and analytical approach for a circular membrane-ring structure of locally resonant acoustic metamaterial," Applied Physics A. Mater. Sci. Process. **114**, 985–990 (2014)
8. Huang, T.-Y., Shen, Ch., Jing, Y.: Membrane- and plate-type acoustic metamaterials. J. Acoust. Soc. Am. **139**, 3240–50 (2016)
9. Jossic, M., Thomas, O., Denis, V., Chomette, B., Manmou-Mani, A., Roze, D.: Effects of internal resonances in the pitch glide of Chinese gonsg. J. Acoust. Soc. Am. **114**, 431–442 (2018)
10. Touzé, C., Chaigne, A.: Lyapunov exponents from experimental time series: application to cymbal vibrations. Acustica **86**, 557–567 (2000)
11. Rossing, T.: The Science of Percussion Instruments. World Scientific, Singapore (2001)
12. Rossing, T.D.: Science of Percussion Instruments. World Scientific, Singapore (2008)
13. Sathej, G., Adhikari, R.: The eigenspectra of Indian musical drums. J. Acoust. Soc. Am. **126**(2), 831–838 (2009)
14. Bader, R.: Finite-difference model of mode shape changes of the Myanmar pat wain drum circle using tuning paste. Proc. Mtgs. Acoust. **29**, 035004 (2016)
15. Deymier, P.A. (ed.): Acoustic Metamaterials and Phononic Crystals. Springer, Heidelberg (2013)
16. Diderot, D.: Encyclopédie, Ou Dictionnaire Raisonné, vol. 66: Suite Du Recueil De Planches. Cap.: Luthier, Planche I, Figure 8 (1779)
17. Haberman, M.R., Norris, A.N.: Acoustic metamaterials. Acoust. Today **12**(3), 31–39 (2016)
18. Mirzaei, A., Miroshnichenko, A.E., Shadrivov, I.V. Kivshar, Y.S.: Optical Metacages. Phys. Rev. Lett.**115**, 215501-1-5 (2015)
19. Colombi, A., Roux, P., Guenneau, S., Rupin, M.: Directional cloaking of flexural waves in a plate with a locally resonant metamaterial. J. Acoust. Soc. Am. **137**(4), 1783–89 (2015)
20. Dubois, M., Shi, C., Wang, Y., Zhang, X.: A thin and conformal metasurface for illusion acoustics of rapidly changing profiles. Appl. Phys. Lett. **110** 151902-1-5 (2017)
21. Halmrast, T., Guettler, K., Bader, R., Godoy, R.I.: Gesture and Timbre. In: Godoy, R.I., Leman, M (ed.) Musical Gestures. Sound, Movement, and Meanig. Routledge (2010)

22. Godoy, R.I.: Images of Sonic Objects. Organized Sound **51**(1), 54–62 (2010)
23. Godoy, R.I.: Coarticulated Gestural-sonic objects in music. In: Gritten, A., King, E (eds.): New Perspectives on Music and Gesture. Ashgate (2011)
24. Bader, R.: How Music Works - A Physical Culture Theory. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-67155-6
25. ISO 10534-1 (1996). Acoustics - Determination of sound absorption coefficient and impedance in impedance tubes - Part 1: Method using standing wave ratio
26. ISO 10534-2 (1998). Acoustics-Determination of Sound Absorption Coefficient and Impedance in Impedance Tubes-Part 2: Transfer-Function Method
27. Bader, R., Gernert, S., Mischo, J.: Metamaterial labyrinth wall for very low and broad-band sound absorption. Preprint (2020). https://doi.org/10.5281/zenodo.3870554
28. Bader, R., Fischer, J.L., Münster, M., Kontopidis, P.: Metamaterials in musical acoustics: a modified frame drum. JASA **145**(5), 3086–94 (2019)

Exploring Musical Agents with Embodied Perspectives

Çağrı Erdem[1,2,3](✉)

[1] RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion,
University of Oslo, Oslo, Norway
`cagrie@uio.no`
[2] Department of Informatics, University of Oslo, Oslo, Norway
[3] Department of Musicology, University of Oslo, Oslo, Norway

Abstract. This chapter presents a retrospective of five interactive systems I have developed focusing on how machines can respond to body movement in music performance. In particular, I have been interested in understanding more about how humans and non-human entities can share musical control and agency. First, I give an overview of my musical and aesthetic background in experimental music practice and a less conventional approach to sound and music control. Then follows a presentation of embodiment and music cognition theories that informed the techniques and methods I employed while developing these systems. Then comes the retrospective section structured around five projects. *Biostomp* explores the unintentionality of body signals when used for music interaction. *Vrengt* demonstrates musical possibilities of sonic microinteraction and shared control. *RAW* seeks unconventional control through chaos and automation. *Playing in the "air"* employs deep learning to map muscle exertions to the sound of an "air" instrument. The audiovisual instrument *CAVI* uses generative modeling to automate live sound processing and investigates the varying sense of agency. These projects show how an artistic–scientific approach can diversify artistic repertoires of musical artificial intelligence through embodied cognition.

Keywords: Musical Artificial Intelligence · Multi-Agent Systems · Embodied Cognition · Human-Computer Interaction

1 Introduction

Artificial intelligence (AI) and multi-agent systems (MAS) can already accomplish highly complex musical tasks, such as modeling instrumental acoustics (Damskägg et al., 2019), synthesizing raw audio (Caillon and Esling, 2021), symbolic music generation (Briot et al., 2020), and generating music from text prompts (Agostinelli et al., 2023). However, real-time musical interaction with AI and MAS is still in its infancy. Music performance is a highly embodied phenomenon, and less is known about how machines can perceive humans as

A. R. Jensenius (Ed.): SSD 2022, CRSM 12, pp. 321–341, 2024.
https://doi.org/10.1007/978-3-031-57892-2_17

embodied entities and how humans can communicate with machines with multiple modalities. This chapter presents a retrospective of five interactive systems I have developed with these questions in mind and focuses on how machines can respond to body movement. The chapter provides an overview of a multiyear artistic–scientific exploration, its iterative methodology, and how theories and methods from the performing arts, computer science, and music cognition informed each other.

I have been particularly interested in exploring human and non-human entities controlling sound and music together, which I call *shared control*. What are the benefits of shared performance control? Following brief introductions of the key terms, I will begin with an overview of my musical and aesthetic background in experimental music practice. This is important to understand where these projects come from. Next is a presentation of embodiment and music cognition theories that informed the techniques and methods I employed while developing the systems, clarifying the emphasis on "embodied perspectives" and reflecting on the interdisciplinarity of my entwined artistic–scientific research model. The retrospective and discussion of the interactive systems I developed based on five shared control strategies will follow: *Biostomp*, *Vrengt*, *RAW*, *Playing in the "air"*, and *CAVI*. Together, these projects show how applying embodied cognition theories can help diversify artistic repertoires of musical AI and MAS.

1.1 Musical Agents

In the field of New interfaces for musical expression (NIME), it has been common to use a variety of machine learning (ML) techniques for action–sound mappings since the early 1990s (Lee et al., 2021, Jensenius and Lyons, 2017). Over the last decades, there has been a growing interest in researching musical agents within the broader field of artificial intelligence (AI) and music (Miranda, 2021). *Agent* comes from the Latin *agere*, meaning "to do" (Russell, 2010). Essentially, anyone or anything that can act with a purpose can be seen as an agent. For example, an agent's sole task might be to recognize the music's particular rhythm while others track simple musical patterns, such as repeating pitch intervals (Minsky, 1981). Such artificial agents are concerned with tackling musical tasks and are what I call *musical agents*. They are artificial entities that can perceive a human performer through sensors, process that information, and act upon their environment by generating sounds and visuals.

1.2 Embodied Perspective

Musical embodiment is concerned with how the body shapes human musical experiences. For example, the effort a musician and a listener exert often depends on the uncertainty of some musical situations, such as technically challenging tasks. Then, one can use the body to communicate, such as nodding to signal their bandmate to return to the tune's main melody. From an enactive perspective, human perception is shaped by our actions (Schiavio, 2015). The enactivist approach asserts the living body as the cognitive system. In other words, the regulation and control of cognition as a homeostatic system are determined by its biological structure (Schiavio and Jaegher, 2017). Thus, cognition can be seen as the action Varela et al. (1991, p. 172):

> By using the term *action* we mean to emphasize once again that sensory and motor processes, perception and action, are fundamentally inseparable in lived cognition. Indeed, the two are not merely contingently linked in individuals; they have also evolved together.

Since cognition emerges not just through information processing but mainly from the dynamic interaction between the agent and the environment, the embodied perspective is concerned with an agent's percept of receiving input and processing abilities. More concretely, it questions an agent's ability to perceive the human body and map percept sequences to actions. Although numerous examples of interactive AI and MAS exist in the literature, only a few have dealt with such embodied perspectives.

1.3 Musical Control

In my work, I question the sound and music control—or the lack thereof—in many interactive music systems. As a noise music artist and improviser, my practice focuses on techniques and approaches that foster unconventional expression in music performance. In particular, I have been inspired by John Cage's (1991) exploration of nonintention, which led me to ask how machines could be given more initiative. How can I share the performance control with another musical agent? An analogy can be two persons playing the same guitar, one exciting the string while the other modifying the pitch on the fretboard. Technically, these two entities are agents, regardless of whether they are human or not. If they practice, they can have reasonable control over the system, which, however, can be possible if they lower their expectations of what to expect from their actions. The outcome will always be contingent on the other entity's influx. One may not even be able to make a sound if the other does not allow it. That is inherently different than two agents improvising on their instruments.

2 Artistic Foundation

It is common for experimental musicians to use electronic hardware in unusual ways. Some tutorials, such as that of Collins and Lonergan (2020), teach, for example, how to hack household electrical appliances. Still, shorting a handheld radio's circuit board to make wizard sounds can be considered "wrong" by many people. One such "wrong" instrument that could spark off a niche performance tradition within the experimental music scene is the "no-input mixing board". The principle is the same as creating loops between a speaker and a microphone. It does not require specialized equipment, and any mixing board can be used. Albeit rare examples of meticulously controlled performances with elaborate rigs, such as Marko Ciciliani's composition *Mask* (2001), no-input mixing is known for its emergent peculiarities (Charrieras and Hochherz, 2016). Performing on a mixer involves sharing musical initiatives with the tool, hence waving the control and being dependent on it. According to Locke (1959), actions are performed in a two-stage temporal sequence. First, possibilities randomly blossom. Then, we choose one action possibility in the next phase: de-liberation. When we act, what was previously out of control is now a determined action. In playing instruments like a no-input mixer, the thought and action processes, hence the decision-making, are distributed between the player and the tool's internal dynamics. Toshimaru Nakamura, states (Paul, 2009):

> You shape the feedback into music. It's very hard to control it. The slightest thing can change the sound. It's unpredictable and uncontrollable, which makes it challenging. It's because of the challenges that I play it. I'm not interested in playing music that has no risk.

The "risk" that Nakamura remarks here implies a preferred uncertainty rooted in a lack of control. That is unconventional in most traditions of playing a musical instrument. Artistically, however, it enables new approaches to performance techniques and music technology innovation.

2.1 Feedback

To better understand the concept of feedback, we can develop an analogy between playing music and driving a boat. The helm of a ship can be seen as analogous to the control interface of a musical instrument, such as a no-input mixer mentioned above. The sea is the electrical current circulating in the components and becoming sound waves through the speakers. As the captain, you shift the steering according to the feedback from the environment concerning waves, winds, and so on. In other words, you continuously evaluate the possibilities, introduce a move, and validate the result before restarting the "loop".

We see such information-feedback paths in all living systems adapting to their environment (Kline, 2015), which can be described as an *autopoietic* organization Maturana and Varela (1980). Poiesis is Greek for "creation" while auto denotes "self". Thus, autopoietic systems consist of self-creating processes (Straussfogel and von Schilling, 2009), which refers to the recursive interactions between the components of living organisms, such as proteins, nucleic acids, lipids, etc. That is a basic understanding of *cybernetics* (Wiener, 1948), which comes from the Greek word *kubernetes*, meaning the helmsman.

The idea of feedback can be traced at least as far back as the beginning of humankind's written record. The first premise of today's rule-based systems is based on the *if...then* condition, which can be found in *modus ponens* of antiquity. Ctesibius' water clock (*clepsydra* c. 250 BC) is considered the first machine to operate under its control. Fast-forward to the 20th century, Nicolas Schöffer created *CYSP 0 & 1* in 1956, human-scale robotic sculptures responsive to changing sound, light, and movement, premiered in a performance with the Maurice Bejart dance company (Shanken et al., 2012). "We are no longer creating a work; we are creating creation," remarked Schöffer (Whitelaw, 2004), signaling the artistic paradigm shift. John Cage, Eliane Radigue, Steve Reich, and David Rosenboom were some of the composers who incorporated feedback into their music. David Tudor's *Bandoneon! (A Combine)* was one of the first pieces that transformed an entire physical space into a self-oscillating instrument via acoustic feedback loops (Goldman, 2012). A milestone was the *Cybernetic Serendipity* exhibition (1968), which happened with 130 contributors, from composers, artists, and poets, to engineers, scientists, and philosophers (Reichardt, 1968).

2.2 Biofeedback

In cybernetics, a particular topic called *biofeedback* emerged as a medical technique that uses electronic devices to measure the physiological processes (Moss, 1999) in the form of visualization or sonification. In the arts, Alvin Lucier's 1965 piece, *Music for Solo Performer, for enormously amplified brain waves and percussion*, was the first to use electroencephalography (EEG) electrodes on a performer's scalp to capture the alpha rhythm of the brain (typically 8–12 Hz). Following an amplification apparatus created by Edmond Dewan, the amplified alpha rhythms excited the sounding body of percussion instruments (Straebel and Thoben, 2014).

In the following years, several other pieces employed biofeedback techniques, such as John Cage's *Variations V* (1965) (Miller, 2001), David Rosenboom's *Ecology of The Skin* (1970) (Rosenboom, 1972), and Stelarc's *Third Hand* (Dixon, 2019). Eventually, the biofeedback paradigm shifted into a new paradigm of *biocontrol* in the 1990s (Tanaka and Donnarumma, 2018). One of the first pieces here was Atau Tanaka's *Kagami*, featuring *The BioMuse* (Lusted and Knapp, 1988), a "biocontroller" that monitors the electrical activity in the body in the form of both EEG and electromyography (EMG) (Tanaka, 1993). The main difference between biofeedback and biocontrol is that the former focuses on measuring bodily processes regardless of the level of intention or willfulness. At the same time, the latter aims at deliberate control.

2.3 Biocontrol

Easier access to fast computers allowed a widespread interest in using the human body as part of musical instruments at the turn of the 21st century. The Myo sensor was particularly important in making bio signals available to larger groups of people through its wireless 8-channel EMG armband with a built-in inertial measurement unit (IMU). Ata Tanaka's *Myogram* (2015) is a piece composed for two Myo armbands and an octophonic sound system, described as "spatial sound trajectories of neuron spikes projected in the height and depth of the space, with lateral space divided in the symmetry of the body" (Tanaka and Donnarumma, 2018, p. 13).

In addition to bioelectric signals, muscle contractions also produce mechanical vibration, which can be captured as acoustic signals through *mechanomyograms* (MMG) (Caramiaux et al., 2015). Donnarumma (2011) pioneered "biophysical music" using his custom device Xth Sense, which uses an electret microphone-based armband to capture "muscle sounds". Donnarumma describes his experience using such bio-interface as "a relationship of configuration, where specific properties of the performer's body and the instrument are interlaced, reciprocally affecting one another" (Tanaka and Donnarumma, 2018, p. 15).

2.4 Coadaptation

Artist-scholars, such as David Borgo David Borgo Borgo and Kaiser (2010) and Marco Donnarumma (2016) suggest a mutual configuration with the (technological) practice and the environment. The latter actively co-constitutes music with living bodies and their activities. If your microphone faces the speaker too closely on a concert stage, creating audible acoustic feedback, you will most likely be triggered to change the microphone direction spontaneously. This could be seen as similar to reaching out the hands while falling. According to Chi et al. (2000), we execute several physiological and biological processes for a single, deliberate task, most of which are often not deliberate or intentional. In that regard, the biological signals produced by muscles reflect the in-betweenness of the human body's voluntary and autonomic functions.

Over the years, I have performed with several different muscle interfaces. This includes the MMG- and EMG-based devices I have developed myself, as well as various commercial products, such as the consumer-grade Myo armband and the medical-grade Delsys Trigno system (some of these works will be introduced in later sections). My experience is that using muscle signals for precise control is challenging. I agree with Tanaka (2000) describing biosignals as "truly living signals," which reflect the in-betweenness of the human body's voluntary and autonomic functions. The causality flows in one direction when we move toward a specific goal. Simultaneously, the dynamic interaction with the environment bestowing the body can flow back via the body's autonomic responses. In other words, the bodily experience of the environment feeds back into one's actions. Starting from these perspectives, I wanted to explore embodied strategies and approaches for interacting with non-human musical agents in artistic settings.

2.5 Musical AI and MAS

Embodied perspectives are scarce in the literature on (musical) human–computer interaction. Literature reviews of artificial intelligence and multi-agent systems for music, such as those made by Collins (2006) and, more recently, Tatar and Pasquier (2019), highlight that musical AI & MAS prioritize interaction based on symbolic audio (e.g., *M & Jam Factory* by Joel Chadabe and David Zicarelli (Zicarelli, 1987), *Cypher* by Rowe (1992), or *Band-out-of-a-Box* by Thom (2000)); audio (e.g., *Voyager* of Lewis (2000), and (*FILTER*) system of Nort et al. (2013)); or cognitive/affective systems (e.g., *OMax* by Dubnov and Assayag (2005), or *MASOM* by Tatar and Pasquier (2017)). However, body movement is also integral to musical interaction and a focal point in developing and performing with new interfaces for musical expression. What is relatively underexplored is how musical agents can interact with embodied entities, e.g., humans, other than merely listening to the sounds of their actions. Rare examples include *Robotic Drumming Prosthesis* by Bretan et al. (2016), *RoboJam* by (Martin and Torresen, 2018), the multimodal agent architecture proposed by Camurri and Coglio (1998), and the musical robot swarm of Krzyżaniak (2021).

3 Embodiment

Embodiment in music interaction essentially refers to actions originating in the body (Leman et al., 2018). As such, the body is the prime medium for interaction. *Gesture* is a commonly used term to describe meaning-bearing human actions and has attracted growing attention in music research (Gritten and King, 2006 Godøy and Leman, 2010, Gritten and King, 2011), spanning new musical interactions (Cadoz and Wanderley, 2000, Jensenius et al., 2010, Tanaka, 2011). However, the term gesture is overwhelmingly multifaceted and differently used in the literature (Jensenius, 2014). In the following, I will clarify the term by dividing it into different levels of body movement, for which using a single term—gesture—is confusing (see Jensenius and Erdem (2022) for more details).

3.1 Low Level

Using a bottom-up approach, I start with low-level body movement, which refers to physical phenomena. Such as *force*, a biomechanical phenomenon that sets the object in *motion*, which refers to the physical displacement of the object. Humans and animals generate voluntary and passive muscular forces to process energy while interacting with the environment (Uliam et al., 2012). When playing a musical instrument, all its different parts transmit forces, motion, and energy from one to another. The experience of playing a musical instrument originates in the sum of the material properties of the instrument and the features of interactive human motion. See, for instance, how the upper harmonics vary by alternating the bow pressure (Motl, 2013), or the amplitude modulation (AM) in a vibrato effect (Dromey et al., 2009). Physical phenomena like force and motion and their variations' influence on the resultant sound can be objectively measured via several motion capture technologies (Jensenius, 2018).

3.2 Middle Level

Differently from force and motion, (embodied) *actions* denote intentionally executed motion fragments, which are subjective phenomena. Godøy and Leman (2010) refer to "cognitive units" to describe such *chunking* of continuous motion and force. Thus, one can think of the action as mental imagery (Godøy, 2009a). As long as an action is not communicated intentionally, it does not necessarily

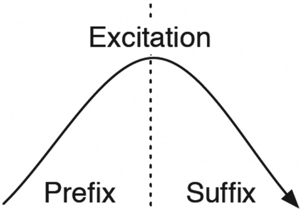


Fig. 1. An action, such as hitting a guitar string, is realized through an excitation phase, which incorporates a prefix and a suffix (Jensenius, 2007, p. 24).

bear a meaning. Hence, I place it in the middle level, between low-level physical signals and high-level communicative actions. Since this middle level is subjective, it is impossible to precisely define, for example, the start and endpoints of an action. Consider the case of hitting a guitar string once. As Godøy (2009b) suggests, the attack has an *excitation phase* having a *prefix* (lifting the arm) and *suffix* (moving down) as illustrated in Fig. 1. *Fidgeting* are the motion parts not directed by a goal nor intentional or conscious.

Since motion and sound are temporal phenomena, we perceive different features in different timescales (Godøy, 2009a). That is a necessity of our cognitive apparatus, for example, in chunking the action segments. Godøy suggests a three-level grouping:

- *Sub-chunk level*: The *micro* timescale for pitch, loudness and timbral features (<0.5 s)
- *Chunk level*: The *meso* timescale as well as the timescale for sound-producing actions (0.5–5 s)—short-term memory
- *Supra-chunk level*: The *macro* timescale for longer contexts (>5 s)—long-term memory

There are many types of music-related body motion (see Jensenius et al. (2010), for an overview), but in the following, I will primarily focus on *sound-producing* actions. Cadoz (1988) suggested that these can be subdivided into *excitation* actions, such as right-hand guitar fingering, and *modification* actions, such as left-hand pitch modifications. As depicted in Fig. 2, excitation actions can be divided further into the three main categories proposed by Schaeffer (1966) and presented by Godøy (2006):

- *Impulsive*: A fast attack resulting from a discontinuous energy transfer (e.g., percussion or plucked instruments).
- *Sustained*: A more gradual onset and continuously evolving sound due to a continuous energy transfer (e.g., bowed instruments).
- *Iterative*: Successive attacks resulting from a series of discontinuous energy transfers.

Identifying the excitation phase can be relatively straightforward when dealing with a single impulsive action but becomes highly complex when combining

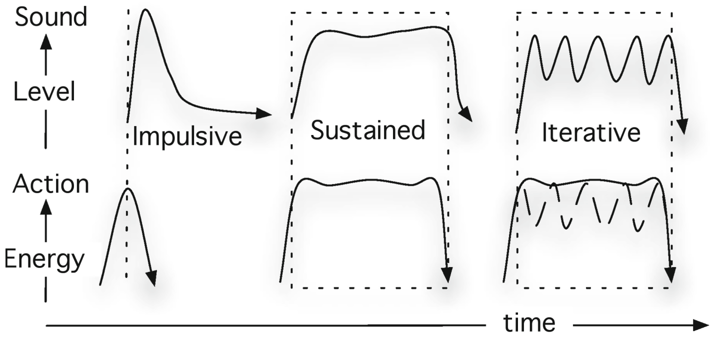


Fig. 2. An illustration of three categories for the main action and sound energy envelopes resulting from different sound-producing action types (Jensenius, 2007, p. 26). The dotted lines correspond to the duration of contact during the excitation phase.

multiple actions. Such action series can be seen as a form of *coarticulation*, the merging of individual actions into larger shapes (Godøy, 2013). Analyzing such action shapes can be challenging from an empirical point of view, particularly segmentation of motion capture recordings for motion–sound analysis.

3.3 High Level

Gestures are actions with an associated *high-level* communicative meaning. The meaning-bearing aspect of gestures has been studied in linguistics: "Gestures exhibit images that cannot always be expressed in speech [...] With these kinds of gestures, people unwittingly display their inner thoughts" according to McNeill (1992, p. 12), emphasizing that bodily gestures are essential to communication.

In music, the word gesture is often used synonymously with both motion and action. However, the challenge is to define the *musical gesture* in a way that covers both motion-related definitions and sonic properties, such as the sound shapes presented by Smalley (1997). The threefold grouping presented in this section provides an embodied perspective on such different levels and definitions of musical gesture.

4 Retrospective

In this section, I present an overview of some of my interactive music systems:

1. *Biostomp*: a muscle-based motorized audio effects controller that explores the boundaries between control and the lack thereof (Erdem et al., 2017)
2. *Vrengt*: an interactive dance piece in which two performers share the control of the system (Erdem et al., 2019)

3. *RAW*: a muscle-based instrument exploring a chaotic behavior in control and automatized ensemble interaction (Erdem and Jensenius, 2020)
4. *Playing in the "air"*: a predictive action–sound model using deep learning based on a custom dataset collected throughout a series of laboratory experiments (Erdem et al., 2020)
5. *CAVI*: an agent-based interactive system using a generative model trained on the data collected in the previous study (Erdem et al., 2022)

Since each system has been described elsewhere, I will breeze through their implementations and focus on details about control structures and sonic design.

4.1 Biostomp

Biostomp is an interface that lets the performer use muscle contractions to control audio effects parameters in live performance situations (a video playlist is available at https://youtu.be/cgnns9z-Nl4). Unlike wearable integrated motion units (IMUs) that measure three-dimensional motion, muscle contractions do not always happen intentionally, which is typical of most biological processes. That can be challenging when using muscles for control. On the other hand, biological idiosyncrasies can also be used creatively in music, similar to how musicians benefit from nature's indeterminacy (Borgo, 2005, Cantrell, 2007).

Biostomp relies on the mechanomyogram (MMG), which denotes low-frequency mechano-acoustic signals generated by contractions in muscle fibers (Watakabe et al., 2001). MMG is the signal resulting from contracting a muscle and can be captured via electret condenser microphones worn on the body part, such as limbs, in the case of Biostomp. When recording audio signals from "inside" of the body, these recordings include multiple bodily "sounds," such as blood flow and heart rate.

Direct transmission of biologically-occurring muscle signals was the primary design consideration for *Biostomp*. It was designed as a self-contained system and avoided any complex mapping and sound design. Instead, it is based on a one-to-one mapping between the MMG amplitude and a motorized headpiece designed to be hooked on potentiometers. The performer then decides which audio effects to control.

The variety of playing modes of *Biostomp* depends on the effects type and the variable signal intensity ("predictability"). In the user study, I observed how different users reacted to different combinations of control and effects. For example, there is a drastic difference in controllability between dynamic (e.g., overdrive) and time-based (e.g., delay) effects. Several users were positive about the system's surprising and less controllable aspects. Nevertheless, most reported that it became more predictable after practicing for some time, which may or may not be favorable. I will return to this aspect later since predictability and user reactions are fundamental commonalities among the five interactive systems being presented.

4.2 Vrengt

Vrengt (Norwegian for "inside-out") is an interactive system that allows a dancer and a musician to control the same sound and music parameters in the interactive system (a video teaser is available at https://youtu.be/vXJ0l9Q68nc). It was designed through a recursive process: capturing and sonifying the dancer's (micro)motion and the shared control of the sonification parameters, which, in turn, affected the dancer's motion. The idea was to work on *sonic microinteraction*, an interaction mode common in acoustic instruments but rarely found in interactive systems (Jensenius, 2017).

In *Vrengt*, I used muscle sensing through electromyograms (EMG), the signal that puts the muscles in motion. Tanaka (2015) describes EMG as capturing the intention to move. It is a bioelectric signal that captures human micromotion indirectly as this level of interaction does not always result in overt body movements (Tanaka, 2015, Jensenius et al., 2017). EMG often reports small or non-visible motion akin to consciously executed actions and automatic body processes (Ortiz et al., 2011). As for the specific sensor device, we chose to work with the (at the time) commercially available Myo armband.

The second interaction method employed in *Vrengt* was capturing the dancer's breathing through a wireless audio signal. Breathing is fascinating in that it is mostly involuntary and unconscious but can also be voluntary and conscious. We preferred using audio over a wireless headset microphone so that the dancer could create acoustic feedback loops by changing her proximity to the speakers on the stage. In doing so, breathing was also used as an aesthetic element. Since the dancer's position on stage influenced the produced sound, the physical space became an integral part of the performance. This was particularly effective in the piece's opening when the dancer was blindfolded for artistic purposes. Then, she had to rely on the auditory feedback from the system to orient herself.

Sonification was a core method used in the sound design of *Vrengt*, which gave the dancer a direct and immediate sonic response. Sonification is often seen as an objective approach to representing data through sound (Hermann and Hunt, 2011). However, in our context, sonification was not the end goal. Instead, we used sonification as part of the creative process.

Drawing on our perceptual and cognitive capacity regarding the link between sounds and sound resources, what Godøy (2001) describe as *mental imagery*, we focused on two techniques in the sound design: (1) Physics-based synthesis of everyday sounds and (2) abstract techniques. In doing so, we could also explore the dancer's sensations concerning the sound synthesis techniques' sonic imagery and mappings. As for abstract techniques, we explored waveshape distortion, ring modulation, and exponential frequency modulation. According to the dancer, while physics-based sounds evoked more straightforward imagery, abstract techniques for sound synthesis resembled shapes that she could "fill with any image you want".

We decided to work with fixed mappings in Vrengt. This was decided early on to accommodate that two human performers would share the control of the system. The dancer's incoming sensor and audio data were processed and interpreted in real-time by the musician, who used knobs and faders on a MIDI controller (Fig. 3). This way, both performers could experience the other's agency. Both performers perceived this as inspiring and fuelled further implementation of artificial agents.

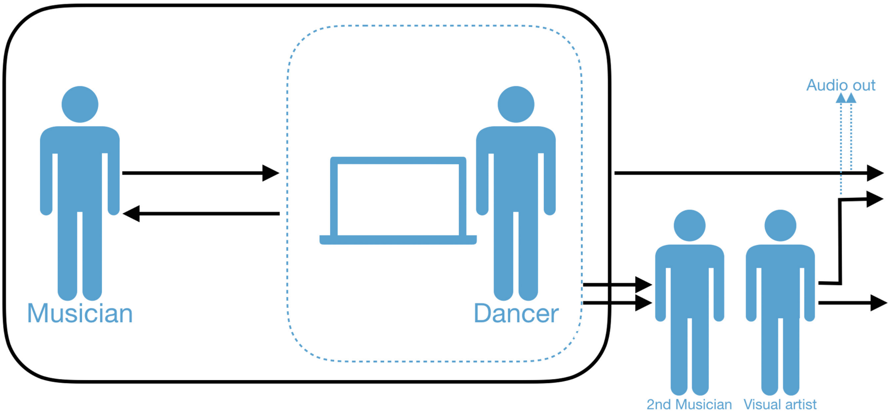


Fig. 3. The setup for the final collaborative performance, showing the levels of connection between performers and instruments (Erdem et al., 2019).

4.3 RAW

The name of *RAW* comes from the system's primary distinctive property, using raw bioelectric muscle signals (EMG) at the audio rate (a video teaser is available at https://youtu.be/_--dzA5pl9k). This was inspired by *Myogram* by Tanaka (Tanaka and Donnarumma, 2018), which uses a direct audification of EMG signals. *RAW* uses two Myo armbands, one on each forearm. Four EMG channels (two per forearm) are buffered every quarter of a second, which is then converted to an audible level by increasing the frequency via a time-scaled sawtooth signal. In doing so, the inherent noise of the raw signal is also frequency-shifted, thus creating a quite noisy high-frequency layer in the audible spectra, requiring filtering. This is where the performer can start being creative as a composer. For example, speeding up the signal to extreme values introduces glitches reminding of well-known electronic music textures, similar to those of Ryoji Ikeda (Emmerson and Landy, 2016).

Two channels of EMG per forearm are sonified, corresponding to extensor and flexor muscle groups. This provides four drone sound channels, controlled by each wrist's extension and flexion. Other poses, such as ulnar or radial deviation, open or closed hands, and neutral poses, create different combinations. One can imagine such a scenario as mixing four audio channels using faders on a mixing board. This approach can be awe-inspiring, but requiring a multi-channel sound system limits its applicability in different ensemble settings. Therefore, I explored several algorithmic approaches for generating control signals.

In the control part of *RAW*, I used multiple feature extractors simultaneously. First, amplitude envelopes were extracted as the continuous EMG signal's root mean square (RMS). For more precise actions, such as event triggering, I used the IMU data, particularly the *jerk*, the rate of change of the acceleration. In air performance, where the performer can move in any direction, the relativity of jerk-based excitation may not always be favorable. Therefore, I trained a support vector machine (SVM) classifier to recognize pinch grips, which I use for triggering purposes. Such gesture recognition helps when performing based on muscle signals for more precision-requiring actions.

A second control part was based on chaotic attractors, such as Hénon-and-Heiles or Lorenz systems, to create melodic motives. The EMG was pitch-shifted at the audio sample rate using additional oscillators. When using a pinch grip, the SVM model can recognize and draw a new set of points on the orbit, where each point refers to a frequency. Although the new frequency may sound random compared to the previous one, it converges into a melodic line. In practice, that does not always work as expected. For example, if the interval between two points is too long, it never converges to a globally familiar pattern. However, the interval can become too repetitive if it is too short.

A third control part was based on two multi-layer perceptron (MLP) artificial neural networks (ANNs). They can be used both in pre-trained mode or in online training mode. The networks were used with a simple gamification strategy. Each ANN mapped eight EMG channels of one armband to a point in an XY plane, of which both axes were mapped into an oscillator parameter. The goal of the "game" is to make two points meet so that a new random event is triggered. As a performer, this is one of the fascinating features of the system.

RAW is based on real-time audio analyses for automated ensemble interaction. Real-time audio analysis is challenging at many levels, particularly in free improvisation settings. The solution was to use an adaptive algorithm and limit the system's scope to rhythm-related tracking using mainly spectral flux and dynamics-tracking using envelope-following. The system also incorporates an effects outboard with a selection of time-based processing modules. These can be employed for live sound processing, producing highly efficient duo performance results. However, in bigger ensembles, such processing can introduce too much ambiguity.

4.4 Playing in the "Air"

Later versions of *RAW* inspired a new project on guitar *ergomimesis*. Magnusson (2019, p. 36) suggests this term for mimicking the *ergon*, Greek for work or function. Thus, ergomimesis denotes carrying out the function and the incorporated working memory, *ergogenetic* memory, from one context or domain into another. I began from an "air guitar" perspective, although the aim was never to mimic the guitar in the air. Instead, I wanted to employ the embodied knowledge of playing the guitar and use these possibilities and constraints in constructing a new instrument.

The first part of the project involved a controlled experiment in a laboratory context. A total of 36 participants performed tasks based on guitar-like versions of each of the three basic sound-producing action types proposed by Schaeffer (1966): impulsive, iterative, and sustained. Analyses of the motion capture, EMG, and sound data from the experiment showed explicit action–sound correspondences compatible with theories of embodied music cognition (Erdem et al., 2020, p. 15).

Following the empirical exploration of how biomechanical energy transforms into sound, we used these transformations as part of a machine learning framework based on Long Short-Term Memory (LSTM) networks and compared nine model configurations. The aim was to determine how much latency these models would be subject to when used as a musical instrument (Erdem et al., 2020, p. 30). Our results showed that the models could predict audio energy features of free improvisations on the guitar, relying on an EMG dataset of three distinct motion types (a video is available at http://bit.ly/air_guitar_smc). Our modeling approach provided empirical support for the embodied music cognition theory.

4.5 CAVI

The inspiration for *CAVI* came from the concepts of emergent coordination (Knoblich et al., 2011), collaborative emergence (Sawyer and DeZutter, 2009), and temporal (un)predictability (Haggard et al., 2002). Following the considerable latency of the trained models, I focused on generative modeling. Instead of a discriminative supervised model, I used a recurrent neural network (RNN) combined with a mixture density network (MDN) layer (Bishop, 1994). This MDRNN model continuously tracked the data streamed from a Myo armband worn on the right forearm of the performer and generated new electromyogram (EMG) and acceleration data.

One interesting question is whether coordination or joint action can emerge between a performer and a musical agent that somewhat simulates the performer's likely actions using generative predictions. To explore that, *CAVI* continuously tracks the performer's motion input, consisting of 4-channel EMG and 3-channel ACC signals, and generates what will likely come next. In brief, *CAVI* generates control signals solely based on the performer's excitation actions. The generated data were used as control signals mapped to digital audio effects module parameters. This could be seen as playing the electric guitar through some effects pedals while someone else is tweaking the knobs of the devices.

CAVI's effects modules rely on time-based sound manipulation, such as delay, time-stretch, and stutter. The jerk of the generated acceleration data triggers the sequencer steps, functioning as a matrix that routes the effects and sends & returns. The generated EMG data (corresponding to the same flexion and extension muscle groups similar to previous projects) is mapped to effects parameters. The real-time analysis modules track the musician's dry audio input and adjust the parameters according to pre-defined thresholds. These machine listening agents include trackers of onsets and spectral flux. For example, if the performer plays impulsive notes, *CAVI* increases the reverb time drastically, becoming a drone-like continuous sound. If the performer plays loudly, the system decides about its dynamics based on the particular action type of the performer (A video is available at https://youtu.be/kmYEEEnjm0s).

CAVI is an audiovisual instrument not only for aesthetic reasons but also to avoid potential causality ambiguities. The design presents CAVI as an uncompleted, creepy but cute creature with only legs that are too small for its body, no arms, a tiny mouth, and a big eye (Fig. 4). In real-time animation, the body contracts but does not make full-body gestures. Instead, the eye blinks from time to time when *CAVI* triggers a new event, opens wide when the density of low frequencies increases or stays calm according to the overall energy levels of sound.

Fig. 4. A still image from the performance piece "Me & My Musical AI 'Toddler", recorded for the online NIME 2022 conference. The performance setup comprised the author, *CAVI*, and, in addition, six self-playing guitars (Photo: Adrian Axel).

5 Discussion

From playing acoustic instruments to performing with computers, my journey illuminated a gap: the intimate, embodied experience of the former seemed absent in the latter. The intrigue around translating the sensation of effort—an inherent yet elusive aspect of human experience—to computational systems drove me to explore embodied music cognition theories. Rolf Inge Godøy's decades-long work on shape cognition (Godøy, 2019) grounded my approach, enabling systematic analysis and fostering innovation in music technologies.

Employing muscle sensing as a motion capture method revealed the intriguing complement that motion-based interfaces could bring to existing interaction paradigms. While biological processes might be challenging for direct control due to their involuntary nature, their unpredictability can be harnessed for improvisational musicking.

My work then expanded on the concept of "air performance", where, unlike acoustic instruments, there is no tangible feedback. Explorations into Godøy's gestural-sonic objects and his idea of chunking on varying timescales informed my work's evolution from biofeedback to biocontrol. These ideas and conceptions inspired me to think and design in terms of dynamic sound shapes. For example, *RAW* is heavily based on responding to a sustained chunk with an impulsive action. Similarly, mental imagery became instrumental in *Vrengt*, where sonic design and dance interplayed through metaphoric mappings. Mental imagery can serve as a shared language, bridging the communication gap between musicians and dancers.

The culmination of these investigations led to the development of systems for coadaptation. By embracing biological unpredictability, I aimed for shared control structures rooted in the embodied human experience. This was not about using machines as tools but promoting more initiative in musical interactions, adapting mutually, and shifting the narrative.

While much has been achieved, the journey is ongoing. As an artist–researcher, I stand at the confluence of embodiment, artificial intelligence, and multi-agent systems. The challenge ahead is not merely about integrating human complexity with machines but envisioning a harmonious coexistence and diversifying the known ways of musicking. As we continue to develop human-in-the-loop technologies, there are many unanswered questions: How do we strike a balance between the urge to take over musical control and the serendipity in waving it? How can we employ our communicative skills and human understanding in musical human–machine interactions? How do we ensure that as we innovate, we foster creativity and expression? I will aim to answer some of these in the years to come.

References

Agostinelli, A., et al.: MusicLM: Generating music from text. arXiv:2301.11325 [cs, eess] (2023)

Bishop, C.M.: Mixture density networks. Technical report NCRG/97/004, Neural Computing Research Group, Aston University, Birmingham. Aston University (1994)

Borgo, D.: Rivers of consciousness: the nonlinear dynamics of free jazz. In: Jazz Research Proceedings Yearbook, pp. 46–58 (2005)

Borgo, D., Kaiser, J.: Configurin(g) KaiBorg: interactivity, ideology, and agency in electro-acoustic improvised music (2010)

Bretan, M., Gopinath, D., Mullins, P., Weinberg, G.: A robotic prosthesis for an amputee drummer. arXiv:1612.04391 (2016)

Briot, J.-P., Hadjeres, G., Pachet, F.-D.: Deep Learning Techniques for Music Generation. Computational Synthesis and Creative Systems, Springer, Cham (2020). https://doi.org/10.1007/978-3-319-70163-9

Cadoz, C.: Instrumental gesture and musical composition. In: ICMC 1988 - International Computer Music Conference, Cologne, Germany, pp. 1–12 (1988)

Cadoz, C., Wanderley, M.M.: Gesture - music. In: Wanderley, M., Marc Battier, I.-C.P. (eds.) Trends in Gestural Control of Music (2000)

Cage, J.: An autobiographical statement. Southwest Rev. **76**(1), 59–76 (1991)

Caillon, A., Esling, P.: RAVE: a variational autoencoder for fast and high-quality neural audio synthesis. arXiv:2111.05011 [cs, eess] (2021)

Camurri, A., Coglio, A.: An architecture for emotional agents. IEEE Multimedia **5**(4), 24–33 (1998)

Cantrell, M.: Enactive reading: John Cage, chance, and poethical experience. Genre **40**(1–2), 131–156 (2007)

Caramiaux, B., Donnarumma, M., Tanaka, A.: Understanding gesture expressivity through muscle sensing. ACM Trans. Comput.-Hum. Interact. **21**(6), 1–26 (2015)

Charrieras, D., Hochherz, O.: Chasing after the mixer, pp. 253–254 (2016)

Chi, D., Costa, M., Zhao, L., Badler, N.: The EMOTE model for effort and shape. In: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 2000, pp. 173–182. ACM Press/Addison-Wesley Publishing Co., USA (2000)

Collins, N., Lonergan, S. (eds.): Handmade Electronic Music: The Art of Hardware Hacking, 3rd edn. Routledge, New York (2020)

Collins, N.M.: Towards autonomous agents for live computer music: realtime machine listening and interactive music systems. Ph.D. thesis, University of Cambridge (2006)

Damskägg, E., Juvela, L., Thuillier, E., Välimäki, V.: Deep learning for tube amplifier emulation. In: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 471–475 (2019)

Dixon, S.: Cybernetic-existentialism in performance art. Leonardo **52**(3), 247–254 (2019)

Donnarumma, M.: Xth sense: researching muscle sounds for an experimental paradigm of musical performance. p. 9 (2011)

Donnarumma, M.: Configuring corporeality: performing bodies, vibrations and new musical instruments. Ph.D. thesis, Goldsmiths, University of London, London, UK (2016)

Dromey, C., Reese, L., Hopkin, J.A.: Laryngeal-level amplitude modulation in vibrato. J. Voice **23**(2), 156–163 (2009)

Dubnov, S., Assayag, G.: Improvisation planning and jam session design using concepts of sequence variation and flow experience. Zenodo, Salerno (2005)

Emmerson, S., Landy, L.: The analysis of electroacoustic music: the differing needs of its genres and categories. In: Landy, L., Emmerson, S. (eds.) Expanding the Horizon of Electroacoustic Music Analysis, pp. 8–28. Cambridge University Press, Cambridge (2016)

Erdem, C., Camci, A., Forbes, A.: Biostomp: a biocontrol system for embodied performance using mechanomyography. In: Proceedings of the International Conference on New Interfaces for Musical Expression, pp. 65–70. Zenodo, Copenhagen (2017)

Erdem, C., Jensenius, A.R.: RAW: exploring control structures for muscle-based interaction in collective improvisation. In: Proceedings of the International Conference on New Interfaces for Musical Expression, pp. 477–482. Zenodo, Birmingham (2020)

Erdem, C., Lan, Q., Jensenius, A.R.: Exploring relationships between effort, motion, and sound in new musical instruments. Hum. Technol.: Interdisc. J. Hum. ICT Environ. **16**(3), 310–347 (2020)

Erdem, C., Schia, K.H., Jensenius, A.R.: Vrengt: a shared body-machine instrument for music-dance performance. In: Proceedings of the International Conference on New Interfaces for Musical Expression, pp. 186–191. Zenodo, Porto Alegre (2019)

Erdem, C., Wallace, B., Jensenius, A.R.: CAVI: a coadaptive audiovisual instrument-composition. PubPub (2022)

Godøy, R.I.: Imagined action, excitation, and resonance. In: Musical Imagery, vol. 13, pp. 237–250. Swets & Zeitlinger Publ., Lisse (2001)

Godøy, R.I.: Gestural-sonorous objects: embodied extensions of Schaeffer's conceptual apparatus. Organ. Sound **11**(2), 149–157 (2006)

Godøy, R.I.: Chunking sound for musical analysis. In: Ystad, S., Kronland-Martinet, R., Jensen, K. (eds.) CMMR 2008. LNCS, vol. 5493, pp. 67–80. Springer, Heidelberg (2009a). https://doi.org/10.1007/978-3-642-02518-1_4

Godøy, R.I.: Gestural affordances of musical sound. In: Musical Gestures, pp. 1–23. Routledge (2009b)

Godøy, R.I.: Thinking sound and body-motion shapes in music: public peer review of "gesture and the sonic event in Karnatak music" by Lara Pearson. Empir. Musicol. Rev. **8**(1), 15–18 (2013)

Godøy, R.I.: Musical shape cognition. In: Grimshaw-Aagaard, M., Walther-Hansen, M., Knakkergaard, M. (eds.) The Oxford Handbook of Sound and Imagination, vol. 2. Oxford University Press, Oxford (2019)

Godøy, R.I., Leman, M.: Musical Gestures: Sound, Movement, and Meaning. Routledge, London (2010)

Goldman, J.: The buttons on Pandora's box: David Tudor and the bandoneon. Am. Music. **30**(1), 30–60 (2012)

Gritten, A., King, E.: Music and Gesture. Ashgate, Aldershot (2006)

Gritten, A., King, E. (eds.): New Perspectives on Music and Gesture. SEMPRE Studies in The Psychology of Music. Ashgate, Routledge, Farnham (2011)

Haggard, P., Clark, S., Kalogeras, J.: Voluntary action and conscious awareness. Nat. Neurosci. **5**(4), 382–385 (2002)

Hermann, T., Hunt, A.: Interactive sonification. In: Hermann, T., Hunt, A., Neuhoff, J.G. (eds.) The Sonification Handbook, pp. 273–298. Logos Verlag, Berlin (2011). OCLC: ocn771999159

Jensenius, A.R.: Action-sound: developing methods and tools to study music-related body movement (2007)

Jensenius, A.R.: To gesture or not? An analysis of terminology in NIME proceedings 2001–2013. In: Proceedings of the International Conference on New Interfaces for Musical Expression. Zenodo, London (2014)

Jensenius, A.R.: Sonic Microinteraction in "the air". In: Lesaffre, M., Maes, P.-J., Leman, M. (eds.) The Routledge Companion to Embodied Music Interaction, 1st edn., pp. 429–437. Routledge, New York, London (2017)

Jensenius, A.R.: Methods for studying music-related body motion. In: Bader, R. (ed.) Springer Handbook of Systematic Musicology. SH, pp. 805–818. Springer, Heidelberg (2018). https://doi.org/10.1007/978-3-662-55004-5_38

Jensenius, A.R., Erdem, C.: Gestures in ensemble performance. In: Timmers, R., Bailes, F., Daffern, H. (eds.) Together in Music: Coordination, Expression, Participation, pp. 109–118. Oxford University Press, Oxford (2022)

Jensenius, A.R., Lyons, M.J. (eds.): A NIME Reader: Fifteen Years of New Interfaces for Musical Expression. Current Research in Systematic Musicology, Springer, Cham (2017). https://doi.org/10.1007/978-3-319-47214-0

Jensenius, A.R., Sanchez, V.G., Zelechowska, A., Bjerkestrand, K.A.V.: Exploring the Myo controller for sonic microinteraction. In: Proceedings of the International Conference on New Interfaces for Musical Expression New Interfaces for Musical Expression, pp. 442–445. Zenodo, Porto Alegre (2019)

Jensenius, A.R., Wanderley, M.M., Godøy, R.I., Leman, M.: Musical Gestures: concepts and methods in research, pp. 12–35 (2010). 978-0-415-99887-1

Kline, R.R.: The cybernetics moment: or why we call our age the information age (2015). OCLC: 890127838

Knoblich, G., Butterfill, S., Sebanz, N.: Psychological research on joint action. In: Psychology of Learning and Motivation, vol. 54, pp. 59–101. Elsevier (2011)

Krzyżaniak, M.: Musical robot swarms, timing, and equilibria. J. New Music Res. **50**(3), 279–297 (2021). https://doi.org/10.1080/09298215.2021.1910313

Lee, M., Freed, A., Wessel, D.: Real-time neural network processing of gestural and acoustic signals, pp. 277–280. International Computer Music Association, Montreal (1991)

Leman, M., Maes, P.-J., Nijs, L., Van Dyck, E.: What is embodied music cognition? In: Bader, R. (ed.) Springer Handbook of Systematic Musicology. SH, pp. 747–760. Springer, Heidelberg (2018). https://doi.org/10.1007/978-3-662-55004-5_34

Lewis, G.E.: Too many notes: computers, complexity and culture in voyager. Leonardo Music J. **10**(1), 33–39 (2000)

Locke, J.: An essay concerning human understanding (1959). OCLC: 371280

Lusted, H.S., Knapp, R.B.: Biomuse: musical performance generated by human bioelectric signals. J. Acoust. Soc. Am. **84**(S1), S179–S179 (1988)

Magnusson, T.: Sonic Writing: Technologies of Material, Symbolic, and Signal Inscriptions (2019). OCLC: 1023599945

Martin, C.P., Torresen, J.: RoboJam: a musical mixture density network for collaborative touchscreen interaction. In: Liapis, A., Romero Cardalda, J.J., Ekárt, A. (eds.) EvoMUSART 2018. LNCS, vol. 10783, pp. 161–176. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-77583-8_11

Maturana, H.R., Varela, F.J.: Cognitive function in general. In: Maturana, H.R., Varela, F.J. (eds.) Autopoiesis and Cognition: The Realization of the Living. Boston Studies in the Philosophy and History of Science, vol. 42, pp. 8–14. Springer, Dordrecht (1980)

McNeill, D.: Hand and Mind: What Gestures Reveal About Thought. University of Chicago Press, Chicago (1992)

Miller, L.E.: Cage, Cunningham, and collaborators: the odyssey of variations V. Musical Q. **85**(3), 545–567 (2001)

Minsky, M.: Music, mind, and meaning. Comput. Music. J. **5**(3), 28–44 (1981)

Miranda, E.R. (ed.): Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-72116-9

Moss, D. (ed.): Humanistic and Transpersonal Psychology: A Historical and Biographical Sourcebook. Humanistic and Transpersonal Psychology: A Historical and Biographical Sourcebook, pp. pp. xxi, 457. Greenwood Press/Greenwood Publishing Group, Westport (1999)

Motl, K.R.: Multiphonics on the double bass: an investigation on the development and use of multiphonics on the double bass in contemporary music. Ph.D. thesis, University of California San Diego, San Diego (2013)

Nort, D.V., Oliveros, P., Braasch, J.: Electro/acoustic improvisation and deeply listening machines. J. New Music Res. **42**(4), 303–324 (2013)

Ortiz, M., Coghlan, N., Jaimovich, J., Knapp, R.B.: Biosignal-Driven Art: Beyond Biofeedback. CMMAS (2011). Accepted 2017-11-29T13:51:16Z

Paul, E.: Subsonics - Episode 4 (2009). https://vimeo.com/3799720

Reichardt, J.: Cybernetic serendipity-getting rid of preconceptions. Studio Int. **176**(905), 176–77 (1968)

Rosenboom, D.: Method for producing sounds or light flashes with alpha brain waves for artistic purposes. Leonardo **5**(2), 141–145 (1972)

Rowe, R.: Machine listening and composing with cypher. Comput. Music. J. **16**(1), 43–63 (1992)

Russell, S.J.S.J.: Artificial Intelligence: A Modern Approach, 3rd edn. Prentice Hall, Upper Saddle River (2010)

Sawyer, R.K., DeZutter, S.: Distributed creativity: how collective creations emerge from collaboration. Psychol. Aesthet. Creat. Arts **3**(2), 81–92 (2009)

Schaeffer, P.: Traite des objets musicaux: essai interdisciplines. Éditions du Seuil, Paris (1966). OCLC: 301664906

Schiavio, A.: Action, enaction, inter(en)action. Empir. Musicol. Rev. **9**(3–4), 254–262 (2015)

Schiavio, A., Jaegher, H.D.: Participatory sense-making in joint musical practice. In: The Routledge Companion to Embodied Music Interaction, pp. 1–9. Routledge (2017)

Shanken, E., Clarke, B., Henderson, L.: Cybernetics and Art: Cultural Convergence in the 1960s (2012)

Smalley, D.: Spectromorphology: explaining sound-shapes. Organ. Sound **2**(2), 107–126 (1997)

Straebel, V., Thoben, W.: Alvin Lucier's music for solo performer: experimental music beyond sonification. Organ. Sound **19**(1), 17–29 (2014)

Straussfogel, D., von Schilling, C.: Systems theory. In: Kitchin, R., Thrift, N. (eds.) International Encyclopedia of Human Geography, pp. 151–158. Elsevier, Oxford (2009)

Tanaka, A.: Musical technical issues in using interactive instrument technology with application to the BioMuse. In: Proceedings of the International Computer Music Conference (1993)

Tanaka, A.: Musical performance practice on sensor-based instruments. Trends Gest. Control Music **13**(389–405), 284 (2000)

Tanaka, A.: Sensor-based musical instruments and interactive music. In: The Oxford Handbook of Computer Music, roger t. dean edn. Oxford University Press (2011)

Tanaka, A.: Intention, effort, and restraint: the EMG in musical performance. Leonardo **48**(3), 298–299 (2015)

Tanaka, A., Donnarumma, M.: The body as musical instrument. In: The Oxford Handbook of Music and the Body (2018)

Tatar, K., Pasquier, P.: MASOM: a musical agent architecture based on self-organizing maps, affective computing, and variable markov models, p. 8. MuMe, Atlanta (2017)

Tatar, K., Pasquier, P.: Musical agents: a typology and state of the art towards musical Metacreation. J. New Music Res. **48**(1), 56–105 (2019)

Thom, B.: BoB: an interactive improvisational music companion. In: Proceedings of the Fourth International Conference on Autonomous Agents - AGENTS 2000, pp. 309–316. ACM Press, Barcelona (2000)

Uliam, H., de Azevedo, F.M., Ota Takahashi, L.S., Moraes, E., Negrao Filho, R.D.F., Alves, N.: The relationship between electromyography and muscle force. In: Schwartz, M. (ed.) EMG Methods for Evaluating Muscle and Nerve Function. InTech (2012)

Varela, F.J., Thompson, E., Rosch, E.: The Embodied Mind: Cognitive Science and Human Experience. MIT Press, Cambridge (1991)

Watakabe, M., Mita, K., Akataki, K., Itoh, Y.: Mechanical behaviour of condenser microphone in mechanomyography. Med. Biol. Eng. Comput. **39**(2), 195–201 (2001)

Whitelaw, M.: Metacreation: Art and Artificial Life. MIT Press, Cambridge (2004)

Wiener, N.: Cybernetics; or control and communication in the animal and the machine. In: Cybernetics; or Control and Communication in the Animal and the Machine, pp. 1–194. Wiley, Oxford (1948)

Zicarelli, D.: M and Jam factory. Comput. Music. J. **11**(4), 13–29 (1987)

Author Index

A. R. Jensenius (Ed.): SSD 2022, CRSM 12, p. 343, 2024.
https://doi.org/10.1007/978-3-031-57892-2