Composing and Decomposing Electroacoustic Sonifications:

Towards a Functional-Aesthetic Sonification Design Framework

A Dissertation

Presented to

The Academic Faculty

by

Takahiko Tsuchiya

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy in the

School of Music

Georgia Institute of Technology

May 2021

Composing and Decomposing Electroacoustic Sonifications:

Towards a Functional-Aesthetic Sonification Design Framework

Approved by:

Dr. Jason Freeman, Advisor
School of Music
Georgia Institute of Technology

Dr. Claire Author
School of Music
Georgia Institute of Technology

Dr. Grace Leslie
School of Music
Georgia Institute of Technology

Dr. Hiroko Terasawa
Faculty of Library, Information and Media Science
University of Tsukuba

Dr. Carrie Bruce
School of Interactive Computing
Georgia Institute of Technology

Date Approved: February 1, 2021

ACKNOWLEDGEMENTS

This thesis was only made possible by the tremendous help from my advisors, mentors, colleagues and friends. Dr. Jason Freeman has supported me and this research for so many years with genuine care. His guidance ensured that both scientific vigor and creative / aesthetic values unique in music technology are present in this research. Dr. Grace Leslie took generous time to understand my research goals and complexity, providing many practical advices and resources for the experimental design. Dr. Claire Author shared with me examples on conducting an online listening test as well as meticulous feedback on the manuscript. Dr. Carrie Bruce was also extremely helpful with constructive suggestions and insights on qualitative user studies. Perhaps more importantly, she first introduced me to a sonification project that ultimately grew into the present thesis. Dr. Hiroko Terasawa's research in timbre and sonification was a direct inspiration for my thesis. She also provided me positive views on academic career and encouragement when I most needed. Special thanks also go to Bill Finzer, Sherry Hsi, and Tim Erickson at the Concord Consortium, who offered honest and thoughtful feedback on my research from data-science perspectives. My colleagues, particularly Dr. Chih-Wei Wu, offered me support in many challenging situations without hesitation. Lastly, my mother Hisayo Tsuchiya as well as my longtime partner Zhongran Wang aided me throughout this prolonged endeavor with full trust. Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

General

F/A: Functional-Aesthetic

RQ: Research Question

Frameworks

SMSon: Spectromorphing Sonification

DTM: Data-to-Music

SPE: Spectral Parameter Encoding

MDM: Musical Data Moves

Spectromorphing Sonification

Spks: Spectral Peaks

Senv: Spectral Envelope

Sefx: Spectral Effect

Aenv: Amplitude Envelope

CD: Spks:Constant / Senv:Data

DC: Spks:Data / Senv:Constant

AD: Spks:Auxiliary / Senv:Data

DA: Spks:Data / Senv:Auxiliary

Listening Test

DF: Description-First

EF: Estimation-First

Gold-MSI: The Goldsmiths Musical Sophistication Index

Design Test

T/E: Trial-and-Error

P/E: Plan-and-Execute

DF: Data-First

SF: Sound-First

PO: Pitch-Oriented

TO: Timbre-Oriented

SUMMARY

The field of sonification invites musicians and scientists for creating novel auditory interfaces. However, the opportunities for incorporating musical design ideas into general functional sonifications have been limited because of the transparency and communication issues with musical aesthetics. This research proposes a new design framework that facilitates the use of musical ideas as well as a transparent representation or conveyance of data, verified with two human subjects tests. An online listening test analyzes the effect of the structural elements of sound as well as a guided analytical listening to the perceptibility of data. A design test examines the range of variety the framework affords and how the design process is affected by functional and aesthetic design goals. The results indicate that the framework elements, such as the synthetic models and mapping destinations affect the perceptibility of data, with some contradictions between the designer's general strategies and the listener's responses. The analytical listening nor the listener's musical background show little statistical trends, but instead imply complex relationships of types of interpretations and the structural understanding. There are also several contrasting types in the design and listening processes which indicate different levels of structural transparency as well as the applicability of a wider variety of designs.

CHAPTER 1. INTRODUCTION

1.1 Sonification, an Interdisciplinary Field

Sonification is a research of mapping non-auditory entities to sounds. This general idea, however, entails multiple definitions with multiple areas of applications. For example, it is employed as a non-visual interface to digital data for scientific analysis[1] [70] or for assisting the visually impaired [13]. It is also used to enhance our motor and sensory experiences as part of a human-machine interactive system, such as vehicle operations [51] and sensor-augmented exercises [34]. Furthermore, it is a popular form of experimentation for musical compositions [72] and artistic displays of technical data for public outreach [80].

The field is also interdisciplinary, inviting theorists and practitioners with a wide variety of backgrounds. Often, the 'practitioner' of sonification entails both the designer and the listener [87]. The designer (who might also be a theorist themselves) includes sound artists, audio-software developers, acousticians, auditory psychologists, etc. who would implement new instances of sonification with a certain methodology. The listener, or the end-user, may include data analysts, machine/interface operators (e.g., vehicle drivers), and the general audience who might be newly introduced to this relatively unconventional medium. To make sonifications more approachable and useful for the end-user, the field presents a great opportunity for theorists and practitioners to collaboratively explore new designs.

[1] Sonifications with such an application are alternatively called auditory displays, as opposed to graphical displays.

1.2 Limitations in Collaboration and Communication

Despite the prospect of cross-domain collaboration and outreach, the general field of sonification appears to suffer from several pragmatic issues. First is the long-standing incompatibilities between scientific and artistic approaches. That is, many novel and expressive sound-organization techniques cannot be easily applied to 'practical' sonifications without negatively affecting the comprehensibility of the entities being represented. Hermann articulates the case with auditory displays in the following way [39]:

> *Think of scientific visualization vs. art: what is the difference between a painting and a modern visualization? Both are certainly organized colors on a surface, both may have aesthetic qualities, yet they operate on a completely different level: the painting is viewed for different layers of interpretation than the visualization. The visualization is expected to have a precise connection to the underlying data, else it would be useless for the process of interpreting the data. In viewing the painting, however, the focus is set more on whether the observer is being touched by it or what interpretation the painter wants to inspire than what can be learnt about the underlying data. Analogies between sonification and music are close-by.*

The second general limitation is the difficulty of conveying auditory concepts or intuitions to the listener, in a way that they can understand the structural elements of a new type of sonification as the designer intends. If the designer is to introduce any unique musical ideas in sonification, such as the use of micro rhythms or clustered chords, communicating their functional behaviors can be nontrivial, especially for non-expert listeners. As a result, the use of musical techniques in a functional sonification is often limited in both the range and variety of sound organization.

These two limitations – the functionalist vs. artist dilemma and the difficulty of conveying new auditory ideas even among practitioners – inspire the author to delve into the ambiguous low-level components of sound and find a way to present them in a functional and transparent form. In response to Hermann's "painting" analogy of a musical sonification, the author asks: is it possible to create an 'accurate' painting with well-defined colors and tools, allowing the listener (viewer) to dissect the layers or strokes of auditory paint into individual color elements with a close observation?

1.3 The Approach and Goals of the Research

Throughout the thesis, the author employs an expression “functional-aesthetic (F/A)” to denote dualistic goals: A F/A sonification aims to be both intelligible, accurately representing and conveying the underlying data, and designable, accommodating a wide range of musical aesthetics. To achieve these, the thesis proposes a new sonification design framework, spectromorphing sonification (SMSon). The theoretical approaches of SMSon are informed by the works in auditory perception, signal processing, and music theories, discussed in detail in chapters 2, 3, and 6. In addition, empirical and pragmatic aspects of F/A are explored through the author's previous work in musical sonifications and the development of general-purpose frameworks (chapters 4 and 5).

The main scope of this research is set in the ‘electroacoustic’ paradigm, where theories of novel sound organization are actively investigated. Here, the author explores sound design approaches for short-durational expressions, aiming to establish rather primitive variants of a F/A sonification focused on the intricacy of timbre. Short timbral expressions, lasting only several seconds, can potentially encode a large amount of both physical and abstract information, as investigated by electroacoustic music theorists [75][99][92]. Sections 2.1 and 3.2 further elaborate on this choice of research direction.

The thesis aims to contribute to the field of sonification with the identification of new design factors that are tested from multiple angles. For example, it theorizes and seeks to capture how the end listener and designer engage aesthetically complex sonifications in a ‘sound first’ manner, where they are guided to spontaneously explore the various characteristics of timbre rather than following a rule for encoding / decoding data. The user experiments are designed particularly to observe the unseen relationships between quantitative and qualitative understanding of sonification. With the factors or engagement patterns being identified, the development of future design frameworks may benefit from additional evaluative metrics and also create design-communication strategies with commonly used sound characterizations. The benefit of findings may also extend to the design and presentation of contemporary electroacoustic music, where connecting the compositional intentions and receptions appears to be particularly difficult [57].

SMSon is intended for two target user groups. The first and primary target is the music composer who has experience in music technology or electroacoustic music but not necessarily in sonification. SMSon invites composers with various musical backgrounds to apply their aesthetic choices while guiding them to create discernible timbal sonifications. The second target is the non-expert listener with limited exposure to sonification or electroacoustic music. As the novelty of aesthetic expressions also implies the difficulty of learning / training, the use of the framework is examined for how the design and sound-structural information can be communicated.

Finally, this thesis presents three deliverables. SMSon, a methodological model, provides concepts and guidelines for creating custom timbral sonifications. Sonar.js, an audio programming library, facilitates the implementation of SMSon in the web environment. Finally, as a byproduct of the designer experiment (see chapter 8), it presents a simple web-based sonification-design environment utilizing both SMSon and Sonar.

1.4 Chapters Overview

Chapter 2 provides a general review of related work. Chapter 3 presents a roadmap of the research, expanding some of the general concepts from the related work. Chapters 4 and 5 review the author's preliminary work that informs the design of the new framework. The present research consists of the development of methodology and technology (chapter 6), an online perceptual listening test (chapter 7), and a laboratory design study (chapter 8). The thesis concludes with a summary and application of the findings (chapter 9).

CHAPTER 2. BACKGROUND

Investigating the generally divergent factors of aesthetics and functionality in sonification, the research structure can easily become complicated. To set a concrete and feasible scope, the following compares related work organized in two problem areas: the handling of complex aesthetic sounds (without data) and the structuring of design frameworks (with data). The concepts introduced here inform the design of the spectromorphing sonification (SMSon) framework as well as the user studies, as elaborated in chapter 3.

2.1 Properties of Musically Complex Sounds

Previously, sections 1.2 and 1.3 have indicated that the use of novel aesthetic designs in sonification may require the objective assessment of sound properties prior to the application of non-auditory data. The following examines the general issues and theories of aesthetic sounds, particularly electroacoustic timbral arrangements, from signal-processing, perceptual, and musicological perspectives.

2.1.1 Analysis and Synthesis of Complex Audio Signals

The sonification framework proposed in this research utilizes the timbral properties of short (e.g., 0.5-5 seconds) sounds, evolving continuously over the span of a single or several musical notes. Contemporary electroacoustic music largely explores this subsymbolic domain, where the microstructure of a sound is manipulated to create dynamic spectra, (in)harmonicity, transients, and micro rhythms [92][74]. Challenging the static and physics-bound treatment of musical notes in symbolic music, Schaeffer, Roads, and many others postulate the aesthetic potentials of equivalent “sound objects” [75][99]. The manipulation of sound objects could, as Roads argues, scale to a complete timbre-based musical piece [68][47]. The scope of the present study is limited to composing individual sound objects. However, the design process with SMSon is generally inspired by a larger-scale compositional goal, creating a variety of sonic expressions, as closely observed in the designer study in chapter 8.

In contrast with Schaeffer's original ambition to create aesthetic sounds free of physical limitations [82], new sound objects are often sourced from or modeled after existing complex auditory materials, such as human voices and musical instruments. With granular synthesis, for example, a recorded sound is sliced and spliced into a different sound expression [75][100], sometimes with a perceptual or acoustic constraint [85]. Another perhaps most well-accepted approach is the source-filter model, where a speech-like complex signal is analyzed [105] and resynthesized with, e.g., subtractive synthesis or spectral filtering [88]. While not basing on existing sounds, SMSon expands on both synthetic approaches, employing the spectral source-filter model to generate granular slices, which are rearranged by physical (data) constraints.

General theories of synthetic (generative) microsonic structures are still scarce today [57][44]. In contrast, the analysis of instrumental sounds has a long history by (psycho-)acoustic researchers as well as electroacoustic composers [73][62]. For example, the "spectralist" composers such as Murail and Harvey have focused on the compositional potential of existing instrumental timbres (rather than melodies) or environmental sounds, and documented their creative processes with the analytic / transformative techniques being used [67][37]. Such spectral transformations are also explored in SMSon, adding a layer of aesthetic creativity.

Compared to symbolic organizations, the possibilities of timbral organizations are still barely explored, presenting both an opportunity for novel experimentations as well as the difficulty of understanding and evaluating. Timbre-oriented musical sound organizations prompt the direct measurement of acoustic information in contrast to symbolic music. As various musicologists observe [57][74], a symbolic organization of sound (e.g., reusing an instrumental voice as notes, harmonic voices, etc.) leads to nonlinear and hierarchical qualities with unique traits such as time / harmonic synchronization, meanwhile diminishing our attention to the microevolution of timbre [75]. The subsymbolic timbral domain does not entail such a hierarchical structure, while generally regarded as continuous and multidimensional [21].

2.1.2 Measurement and Description of Timbre

In utilizing aesthetic timbres to represent or convey quantitative 'data,' it is imperative to consider what aspects of complex sounds are / could be quantified in relation to the listener's understanding. This measurement problem is also a primary concern in the fields of psychoacoustics and audio content analysis (ACA). This section discusses a few out of numerous implications of the measurement of timbre, namely parameterization and the mapping of auditory and perceptual attributes. The artifacts of both inquiries are often equivocally called timbral "descriptors" as well as features, parameters, and attributes. They are used interchangeably in various contexts as their problem areas tend to partially overlap [90].

As a common analytical approach, both psychoacoustics and ACA employ time-frequency transforms (e.g., Fourier transform) extensively for extracting a wide range of information from continuous acoustic signals [8]. A frequency-domain representation facilitates spectral, harmonic, and perceptual parameterizations [86]. The MPEG-7 standard defines a set of such analysis features (i.e., parameters) that are commonly used as building components for higher-level information retrieval [71]. As general approaches, applying simple descriptive statistics on the modulus of audio magnitude spectra (e.g., mean and variance; spectral features), modeling the instantaneous frequency distribution with parametric statistics (e.g., the normal distribution), as well as employing more perceptually-inspired feature analysis (e.g., Mel-frequency cepstrum coefficients or MFCC, tonality, and “fullness”) [71] reveal many possible angles for measuring and parameterizing an expressive sound. The present thesis examines some of the most common as well as less ordinary parameters of the sound [106][35] for the purposes of both better measurement and synthetic flexibility (see chapter 6).

Several of the timbral parameters defined in ACA (e.g., spectral centroid) are often heuristically correlated to human perception (e.g., brightness) [62]. However, as Siedenburg, et al. point out, it is not the general intention behind such computational heuristics to identify perceptual correlates, but rather to automatically classify timbres [90]. Psychoacoustic studies in search of individual perceptual attributes have employed other contrasting approaches, including multi-dimensional scaling (MDS) [31] and perceptual structure analysis [14]. MDS, a (dis)similarity-based dimensionality-reduction technique, particularly has laid out a foundation for identifying salient parameters in multidimensional musical timbres. A major challenge, however, remains with all these perceptual analyses – that they do not “explain” how the parameters are understood by the listener, which in return impedes the design of general listening tests with the lack of metrics or verbal definitions of the sound. Bech and Zacharov elaborate on the complexity of mapping the audio attributes (i.e., parameters) to response attributes (i.e., perceived impressions) [3]. The development of lexicon for the latter (response attributes) alone entails systematic efforts of eliciting verbal descriptors, reviewing, listener training, refining, and defining scales / rating [116]. Concerning the present thesis, the parameterization of electroacoustic sound designs lag far behind in this development of response attributes for ordinary sounds (e.g., speech) with a severely limited number of verbal descriptors available [117][11]. To remedy the shortage of descriptors in the listening test, the author combines the aesthetic theories of timbre (see the following section) as well as the vocabularies developed for general audio reproduction quality tests (see section 7.2.6).

Though not a common practice for sonification or composition, there are creative attempts connecting the measured parameters of sound directly to the generation of expressive sounds. Jehan has, for example, explored the resynthesis of complex tones from only a few well-defined perceptual attributes [50], as well as the feature-based composition of musical mosaic pieces [49]. Collins has also explored the use of ACA algorithms within a live coding composition, creating an adaptive feedback loop between the audio output and input [15]. As more general frameworks, the analysis/synthesis systems creatively utilize compressed descriptions or features of a complex sound. For example, the classic spectral-modeling synthesis facilitates the storage and manipulation of high-fidelity instrumental sounds with low-dimensional sinusoidal and noise representations [88]. More recently, corpus-based concatenative synthesis laid out the possibility of creating or resynthesizing complex sounds according to the feature space of a huge collection of sounds (grains) [85]. CataRT [86], major implementation of this concatenative synthesis, greatly informs the design and implementation of SMSon.

2.1.3 Analytical Listening of Aesthetic Sounds

Despite the general development of descriptors for existing musical instruments in psychoacoustics and ACA, novel electroacoustic sounds remain highly elusive as they continue to evolve and mutate [117]. Composers in search of new aesthetics inevitably develop new strategies for listening to capture such elusive and dynamic characteristics of creative sounds. As Schaeffer describes, "it is the listening itself that becomes the origin of the phenomenon to be studied [82]." Schaeffer introduced the concepts of acousmatics and reduced listening, where by "reducing" the scope of attention from the normal associations of a sound with physical / visual phenomena to purely auditory sensations, one is able to discover new aesthetic relationships within or among unfamiliar sounds. Such mode of listening is typically associated with 'tape music' from the mid 20th century, which affords the listener repeated and nonlinearly-ordered aural analysis [93]. The listening test in the present study adopts and examines this type of critical listening (chapter 7).

This critical listening practice has since evolved into various aesthetic / musicological theories [99][83][107]. Spectromorphology, which the author adopts the name for the proposed framework, was established mostly by a single electroacoustic composer Denis Smalley [92]. He describes extensive details and interrelationships of spectral typology and temporal morphology that emerge in the process of critical listening. Smalley's spectromorphology is unique in being analytical rather than compositional, with an approach of embodying abstract timbral phenomena with extrinsic human gestures. For developing SMSon, the author employs the term 'spectromorphing' in a more generic way, integrating multiple theories of temporal-spectral analysis. In SMSon, this term entails (1) a perception-oriented (rather than process-focused) synthetic sound organization, and (2) various types of spectrum (spectro-) evolving dynamically (morphing) over time following certain functional and aesthetic principles (see chapter 3). Its largest difference from the original theory is being a generative framework, as opposed to an analytical framework for general electroacoustic compositions. Spectromorphology and its influence on SMSon are described in detail in chapter 6.

2.2 Structural Elements of Design Frameworks

The investigation into aesthetic possibilities of timbre may reward us with a broader range of sonification design that has not been fully utilized [2]. However, this design flexibility is also largely affected by additional goals of sonification, particularly the major premise of accurately representing and/or communicating data to, ideally, the general listener. This section reviews several aspects of design challenges such as defining functional and aesthetic goals, generalizing a methodology, and the systematic evaluation of design frameworks.

2.2.1 Functional and Aesthetic Dimensions

The first challenge entails the often-diverging and conflicting relationships between functional and aesthetic design goals. One reason may be the general sparseness of aesthetic research in sonification. Except for a handful of in-depth discussions [2][107][33], concepts defining or attributed to aesthetics are loosely defined[2] compared to the fields of visualization, human-computer interaction, and contemporary music theories. The present research, therefore, is compelled to reference these related domains to theorize the possible relationships between, as well as what it entails to pursue both, functionalities and aesthetics in a timbral sonification.

It should be noted that the individual discussions of functionality and aesthetics may be situated anywhere across the fuzzy boundary of system-level (e.g., methodologies and user engagement) and material-level (i.e., the properties of musical languages or aesthetic timbres) concerns, as observed in the following.

First, there are several discussions / manifestoes from functional and aesthetic perspectives in the sonification community. Their arguments are relatively isolated from each other (either as orthogonal or incompatible views), except in terms of what is "optimal." Hermann, an advocate for scientific sonification, proposes four functional criteria for sonification systems [39]: (1) mapping objective properties of data to sound (2) systematic transformation, (3) reproducibility, and (4) exchangeability of data. The proposed SMSon framework adheres to all of them either entirely (3 and 4) or fully but with added aesthetic complexity (1 and 2). That is, e.g., there is a

[2] For the auditory properties in timbral sonifications, the innate fuzziness of definitions may be greatly impacted by the presence of sharply defined symbolic sounds, as observed in intervals-vs.-morphologies discussion by Roads [74].

potential interference of musical intentions in data representation in (1) with the exploitation of multidimensionality of timbre. As highlighted by Vickers and Hogg, "one criticism levelled against using musical aesthetics in sonifications is that the musical grammars add another language level to the interface which would get in the way of the underlying data the music would be another language to learn [107]." Many functionalists in sonification as well as visualization, therefore, opt for a minimalist approach [104][59], removing elements that are not directly related to data.

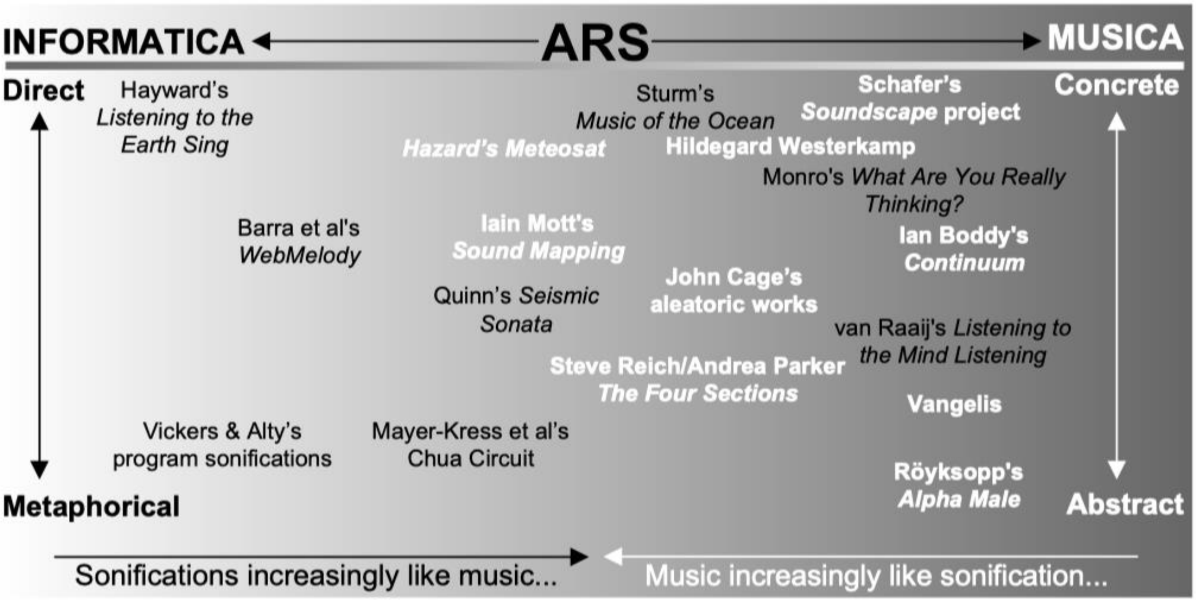

INFORMATICA
ARS
MUSICA
Direct
Metaphorical
Concrete
Abstract
Hayward's Listening to the Earth Sing
Sturm's Music of the Ocean
Schafer's Soundscape project
Hazard's Meteosat
Hildegard Westerkamp
Monro's What Are You Really Thinking?
Barra et al's WebMelody
Iain Mott's Sound Mapping
Ian Boddy's Continuum
Quinn's Seismic Sonata
John Cage's aleatoric works
van Raaij's Listening to the Mind Listening
Steve Reich/Andrea Parker The Four Sections
Vangelis
Vickers & Alty's program sonifications
Mayer-Kress et al's Chua Circuit
Röyksopp's Alpha Male
Sonifications increasingly like music...
Music increasingly like sonification...

Figure 2.1 Ars Informatica-Ars Musica Continua organized by Vickers and Hogg [107]. Notice that the boundary between musical works (in white text) and sonification (in black) is not clear, as both can take direct and indirect forms.

From an aesthetic proponents' perspective, such a directness of representation becomes a subject of exploration. A major difference in their assumption from the functionalist is that there may always be multiple optimal ways to solve a design problem with sounds [2][74], sometimes with a novel sound organization. Such examples of sonification are often complex, requiring both the designer and the listener to actively explore and discover aesthetically "optimal" forms with critical listening. Vickers and Hogg map various types of aesthetic auditory work with regard to how data or information are handled (Figure 2.1). Here, the directness of mapping (which does not imply that indirect is suboptimal), can be found in both sonification and music in a parallel relationship (the vertical axes in Figure 2.1) [107]. They employ the term "indexicality" to capture these correlated (in)directness in both domains. In the present thesis, however, the author opts for independent control of data representation and aesthetic sound structures. While exploring various types of sounds, SMSon treats data without any form of interpretation or metaphoric representation, as discussed in section 3.2.

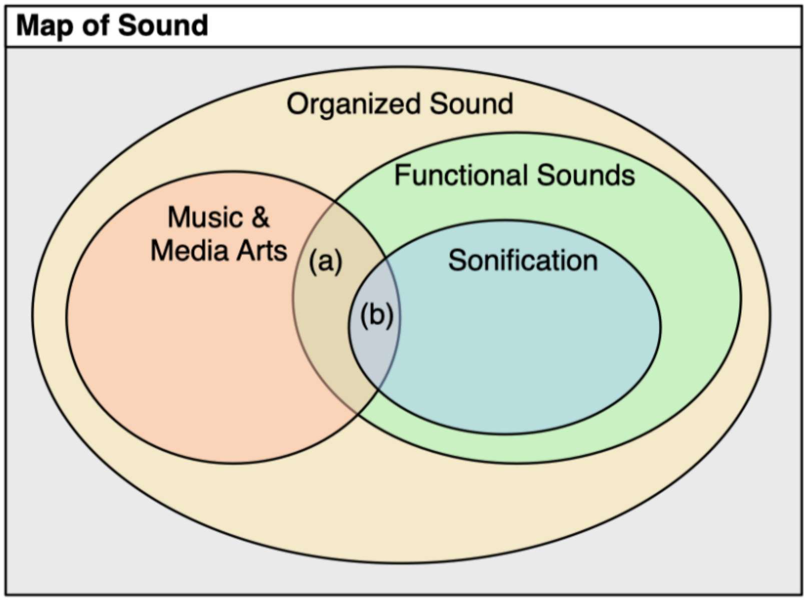


Figure 2.2 A systemic map of organized sound, as depicted by Hermann [39].

At the material (sound) level, the functionalist (or formalist [74]) tends to distinguish musical and functional sounds with minimal overlap in between, virtually separating the types of sound usable for music and sonification [39] (see Figure 2.2). Hermann suggests that functional-musical sounds (the intersection) are such as musical instruments used as a social / cultural cue, or styles of music associated with a certain social / emotional context (e.g., evoking shopping mood). Aesthetic and explorative designers, on the other hand, may view the roles of sound (or organizational approaches) in much more diverse and continuous spectra. Exploring the methodologies of electroacoustic compositions, Roads discusses the concept of “aesthetic oppositions” [74]. These aesthetic dimensions[3] may not necessarily imply a direct connection to the functionalists’ typologies of sound, but instead provide insight for structurally controllable properties of sounds. The following selects several of the nine dimensions that may supplement the previous functional arguments.

First, in the formalism vs. intuitionism observation, Roads overlays some of the extreme design processes of sonification and abstract / arbitrary compositions where the rendered sound becomes an afterthought. This entirely systematic approach may perhaps work with symbolic music or sonification with a very predictable outcome of the sound. However, electroacoustic sound organizations typically interwind the design of form / system with active listening and adjustment of sound, searching for acoustically, perceptually, and psychologically optimal designs. In this regard, aesthetics is part of creating a functional structure.

The dimension of coherence/unity vs. invention/diversity also concerns the design of an electroacoustic work. A rigorous formal consistency based entirely on data or an algorithm, as Roads argues, may result in a dull or incomprehensible aural experience. Inventiveness in the sound-organization process, on the other hand, does not necessarily break the musical coherence, as the change and dynamics are ingrained in many time scales / dimensions of musical aesthetics. A novel invention also signifies a “non-obvious extrapolation of the prior art,” which expands the

[3] Roads explores various dichotomies such as formalism vs. intuitionism, coherence/unity vs. invention/diversity, spontaneity vs. reflection, intervals vs. morphologies, smoothness vs. roughness, attraction vs. repulsion in time, parameter variation vs. strategy variation, simplicity vs. complexity in synthesis, and sensation vs. communication.

choice of sound for a more optimal experience of the material (e.g., compositional intent or represented data). SMSon embraces such a potential for changes, aiming to facilitate multiple perspectives for understanding data (see section 3.5.1).

The way to diversify the design choices may take the form of parameter variation vs. strategy variation. For the former, Roads employs the analogy of exploring in a fixed parameter space that an audio synthesizer provides. While this common practice provides stability and coherence in aesthetic exploration, eventually the choice of sound expressions may be exhausted. It then requires the change of synthetic strategy by altering the underlying algorithm or source sound. This brings an entirely new set / space of parameters to explore. To ensure aesthetic continuity, Roads suggests an approach of stepping up or down to control a different level of structure (time scales). The present study observes such an incremental change of algorithms in design practices (see chapter 8), facilitated by the SMSon framework.

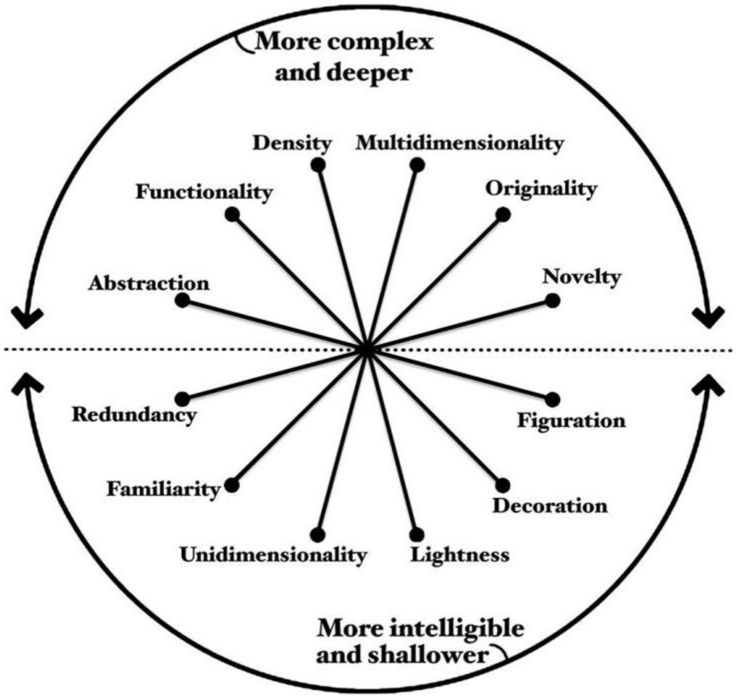


Figure 2.3 The Visualization Wheel introduced by Cairo [10].

Shifting the focus back to the system-level functionalities and aesthetics, in the realm of visualization and infographics, Cairo identifies six axes of functional-aesthetic dimensions[4] [10] (Figure 2.3). With a stark contrast to Hermann's functional principles, these dimensions are defined and discussed primarily with the user's receptions in mind[5]. The following introduces, again, several of the dimensions that are particularly relevant for creating functional-and-aesthetic visualizations / sonifications.

First, one of the continua is called "functionality-decoration" with interestingly very different implications from the previous functional-minimalists' assertions. Here, functionality is correlated to complexity, while decoration is attributed to a shallower design. Functional design (visual) elements, Cairo argues, may be largely aesthetic but still carry information related to data, implying that some design elements, at least in visualization (e.g., colors and shapes), can take an aesthetic and complex form while assuming a functional role in the presentation. On the other hand, purely decorative (non-functional) aesthetics may not necessarily contribute to added complexity, if clearly communicated to the user.

The notion of multidimensionality-unidimensionality also primarily takes the user responses into consideration. Not only with the innate multidimensionality of the medium (e.g., visual shapes and colors, or timbre in sonification), a multidimensional representation of data is achieved with the viewer's (or listener's) awareness that different forms of representation enabling multiple view angles to the data, and also their willingness to explore multiple depths of information within a single design entity. An effective visualization balances the weights in such

[4] Cairo cautions that these mappings are based on his empirical observations and not intended for formal evaluation of designs. This thesis, therefore, only references some of the concepts independent from the paired relationships and develops more appropriate definitions based on electroacoustic concepts.

[5] Note that Hermann and colleagues device a specific type of user engagement, physical user interaction with the sonification system, that complements their functional principles [87]. The user engagement in Cairo's discussion encompasses more general and simple viewing and understanding of the system / data.

multidimensionality that all the visual elements do not grab the viewer's attention simultaneously, but instead await their spontaneous and careful inspection.

Closely related to multidimensionality, Cairo's observation of novelty-redundancy provides additional meanings to the novelty (changes) of design. Here, a novel expression not only expands our choice of material (i.e., visual / sound) for optimal representations but also increases the ability to present more information in a single rendering of the design. A functional-aesthetic representation of data "can explain many different things once (novelty) or it can explain the same things several times, by different means (redundancy)." Therefore, Cairo argues, that it is important to explore and find a balance between novelty and redundancy, which contrasts with minimalist takes of functional designs.

2.2.2 General Design Frameworks

Although the design of a sonification is unrestricted from introducing unique sound and mapping structures according to specific data sources or applications, theorists aim to standardize several design methodologies. Perhaps the most commonly used technique, parameter-mapping sonification (PMSon) [32], maps data to synthetic model parameters with routine analyses and transformations (e.g., pitch scaling). Such a system is rather trivial to build within various audio-programming environments available. However, assessing the design outcome of PMSon is generally considered challenging [6]. The foremost obstacle is that the structure of PMSon designs may significantly vary according to the type and structure of a data set (e.g., numerical, categorical, hierarchical, time series, etc.). Such a structural difference of design hinders systematic assessments and also demands the user to (re)learn the mapping scheme. Nonetheless, parameter mapping may be the most straightforward method for structuring as well as communicating designs.

Researchers have also attempted to create more reusable frameworks. Possibly as the most simplistic approach, the audification technique maps each value of single-dimensional time-series data directly to a sample of a digital waveform [1] with optional preprocessing such as scaling, resampling (time mapping), and noise removal. Alexander et al. discuss the efficiency of this methodology with the human auditory perception for consuming the data at the audio sampling rate. They argue the optimal perception of thousands of data points per second as the data create a complex spectrum, oscillatory patterns, or transient shapes. Grond and Hermann point out, however, that these auditory events lose their distinctiveness as the underlying data get complex, eventually becoming an auditory noise [33]. Also, audifications do not ensure the retrieval of individual data points from the rendered sound with neither perception nor acoustic measurement, hindering the assessment of representation.

Another category of systematic approaches in sonification is model-based sonification (MBS) [41]. Unlike PMSon, in which the sound structure / generation is affected directly by data, MBS employs data to (re)configure a general-purpose virtual-acoustic model. Such a model is data agnostic, that it may accept various shapes and dimensionalities of numeric data, placed in a virtual space, without the need for explicitly mapping the parameters. As the sound-manipulation mechanism is decoupled from data sources / types, MBS allows flexible alteration of acoustic designs, providing the opportunity for aesthetic design explorations. Its major restriction may be the fundamental requirement of user interactions with data points to trigger a sound, distancing itself from composed and/or passively consumed musical sonifications.

2.2.3 Evaluation of Design Frameworks

The previous section alluded to the difficulty of systematically evaluating a sonification framework. This challenge is, again, shared with modern visualization systems with novel features.

Considering both sonification and visualization literature, this section briefly reviews the existing approaches in two broad categories: the user performance of tasks with a design framework or system artifacts (e.g., rendered sound), and unique values of the system that yield, e.g., high-level and semantic information gain. One of the 'functional' goals of the present thesis, a transparent representation of data, mostly falls into the first category. However, unique utilities of sound, system learnability, and aesthetic potentials (e.g., the usability of sonification for musical compositions) may be examined in the latter perspective.

To test and benchmark the performance capability of a sonification or visualization, many suggest the classic controlled task-solving experiment with well-defined user-performance metrics. In the field of sonification, multiple researchers have expressed concern for the general lack of statistical performance evaluation for the systems created [18]. The Sonification Evaluation eXchange (SonEX) proposes the standard metrics, including the accuracy, precision, error rate, and reaction speed, and a reusable testing environment [18]. Visualization studies also formalize the accuracy and efficiency tests, with the accuracy metrics including precision, error rate, the average number of incorrect answers, and the number of correct documents retrieved, while efficiency is the average time of completion and the performance time. For comparing multiple systems or improving a single system, these metrics are to be used as variables, along with user's known cognitive abilities as other variables, and analyzed statistically with measures such as F-test, t-test, correlation coefficients, and p levels [12]. Recently, however, these metrics have been criticized for over-generalization, comparing different systems that are designed for different purposes [96].

Another approach to the performance tests focused on the user's perception or cognition, rather than the system, is the measurement and mapping of complex parameter spaces. The stimuli used may be dependent or independent of a sonification system, although the global parameter space should often be known. The above-mentioned MDS technique is often employed for aligning stimuli and the general listener's perceptual responses. Its major advantage is the fact that it does not require any prior knowledge or training for the user and is solely based on intelligibility or the perceived distance between two stimuli. It is also a popular approach when the vocabulary for perceptual description is not well established [3]. While there may possibly be use in mapping the (descriptive) parametric space to a perceptual space, this approach has three disadvantages for the present research: it assumes that the parameter space is much higher in dimension than the perceptual space, implying that the decoded "parameters" are always a combination of multiple encoding parameters (i.e., not a one-to-one relationship that is utilized in the proposed framework); it is difficult to linearly map continuous input parameter changes to continuous output parameter changes; it requires pair-wise testing of every possible stimuli combination, though such tests may be distributed to multiple listeners assuming their mental model of the "timbral space" are similar to each other's [62]. Other comparative analyses of complex objects include the Bradley–Terry model, which computes an ordinal relationship with a pair-wise subjective judgment [56]. As a statistical technique, this model does not require a single test subject to answer every possible combination.

Beyond measuring the effectiveness of retrieving values from sound, or querying other types of information (e.g., sample by recurring patterns), there may be a need to assess the utility of the system and compare it with others. This is a complex issue even with more popular visualization systems, often with a large number of low-level questions to be examined [96]. Stasko and Wall et al. proposed a low-cost approach to developing assessment heuristics and quantifying the "value" of uniquely designed visualization systems [96][109]. The value embroils time, insights (e.g., semantics) and essences (e.g., core structural sense) discovered, and confidence (e.g., the context and trust about the data). These key qualities are further subdivided and rated by visualization experts. Vogt has proposed and experimented with a similar expert-driven assessment of multiple sonification designs, where additional unique qualities such as potential, amenity, and intuitiveness are quantified and rated by sonification designers [108].

The present research is inspired by both quantitative task-based approaches as well as heuristics-based evaluation of unique values. It aims to establish a symmetric experience for the end-user and the designer through the use of common sound-design and aural-analysis concepts. The following chapter 3 discusses the core concepts employed in the experimental design.

CHAPTER 3. RESEARCH OVERVIEW

3.1 Core Problems and Research Statement

This thesis investigates the general incompatibility between aesthetic and functional goals of sonification from a design-methodology perspective. In order to address the limited range of musical ideas usable in functional sonifications, as well as the difficulty of quantifying and describing musical sound structures, the author proposes a new design framework. This framework, spectromorphing sonification (SMSon), entails synthetic techniques, the organization of temporal structures for data and non-data (i.e., aesthetic treatments), and evaluative heuristics based on electroacoustic theories. The development of SMSon is also guided by pragmatic needs: (1) To implement the technical requirements of the user experiments (see chapters 7 and 8); and (2) to lay a foundation for creative applications of the framework beyond this thesis (see chapter 9).

3.2 Scope of the Research

This thesis mainly investigates complex timbral structures for encoding data, while leaving out the implications of real-life data or data-driven optimal designs (which are instead explored in the preliminary-work chapters 4 and 5). The study generally treats 'data' as short (e.g., less than 10 data points), one-dimensional, and arbitrary numeric sequences without any structural (e.g., relational attributes, a self-correlation, or a skewed distribution) or contextual information. SMSon aims to be data-agnostic but with the capability of accurately representing arbitrary numeric quantities, adhering to the functional principles discussed by Hermann (see section 2.2.1). Such a bare-minimum utility is a baseline for building further complex and idiosyncratic representations of data, yet still is generally impeded in musical sonifications.

Another constraint of the research is the non-interactivity of the listening-test tasks, concerning the listener’s typical engagement to auditory media. The use of physical interactions that alter the synthesis or data input, while argued to be highly beneficial and/or essential in many sonification systems [40], raises at least two concerns: (1) The majority of listening experience for electroacoustic works is not interactive but passive (e.g., with a recorded medium or at a concert). This also applies to the consumption of many sonifications with which an interactive user interface is not available to every listener. (2) Not many aspects of organized sound cannot be explored with real-time interactions. For instance, there can be aesthetic sound properties that change much faster, or slower, than a human gestural speed. Integrating gestural interactivity may prohibit the use of musical designs at certain time scales carefully arranged by the designer. Roads, in the context of electroacoustic designs, articulates the limitation of real-time systems [74]:

> *In my work, synthesis is the starting point for sound design, which is itself only the beginning of composition. I inevitably edit and alter raw sound material with a myriad of transformations. This editing involves trial-and-error testing and refinement. Because of this, my strategy would be impossible to bundle into a real-time synthesis algorithm.*

While the present study leaves out various other critical aspects of a F/A sonification, such as the application of short-durational design to a full-scale musical composition / sonification, as future work, the author hopes to identify some of the unexamined low-level relationships among sound structure, design intentions, and listener receptions for the next step of exploration.

3.3 Research Questions

Throughout the development of the framework, a listening test, and a design study, this thesis explores the following research questions (RQs)[6]:

1. How would the new F/A design framework facilitate the general listener's structural understanding of complex and variable electroacoustic sonifications?
2. How would designers with unique musical backgrounds be able to integrate their personal aesthetics into the design of a sonification, while maintaining the functional values?

The first question (RQ 1) mainly concerns the structural and quantitative understanding of sonification by the general listener. Their aesthetic responses are also carefully examined, as a qualitative interpretation may affect the quantification processes. The author broadly hypothesizes that the listener's understanding of sonification, such as the identification and measurement accuracy of the data in a timbre, improves or degrades with (1) types of sound-organizational elements and (2) types of interpretation of the sonic designs by the listener. This question is further elaborated and discussed in detail in the listening test (chapter 7).

The second question (RQ 2), in contrast, focuses on evaluating the creative potentials for the designer. It explores how different designers / composers would be able to plan, structure, and implement an electroacoustic sonification using SMSon for both aesthetic and functional goals. The author expects that SMSon enables the designer to independently adjust the intelligibility of data and aesthetic characteristics as they explore multiple options of spectromorphological

[6] Note that both RQs are intentionally phrased broadly to capture the qualitative intricacies and diversity of design / listening processes. To examine with statistical and qualitative approaches, the RQs will be subdivided into more specific sub-questions in the respective chapters 7 and 8.

behaviors, such as spectra, time scales, and motional behaviors. This question is further developed in the designer test (chapter 8).

3.4 Concepts for Aesthetic Design Exploration

To develop SMSon, the subsequent chapters explore auditory and design concepts that are somewhat either esoteric, ubiquitous, and/or not strictly defined just as the generic term "timbre [21]." This section briefly outlines their implications in the context of framework development and experimental design.

3.4.1 Time Scales and Timbral Dimensions

Regarding the temporal 'morphing' aspect of SMSon, the design methodologies, as well as the analysis of user studies, utilize the concept of time scales. The recognition of different time scales is critical for organizing electroacoustic music, as discussed in great detail by Roads [76]. He explains that a mere change in time duration implies a significant and often a discontinuous change of one type of perceptual auditory sensation (e.g., rhythm) to another (e.g., pitch). Focusing on timbral intricacies, this study explores sample, micro, and sound-object levels of time scale. The sample-level time scale captures the most fine-grained sound organizations such as a continuous amplitude envelope or the location of each spectral peak. The micro time scale presents the "grains" of a sound, or just-noticeable characteristics of a timbre. SMSon organizes a collection of such unique timbral profiles in the form of spectral models. Lastly, the sound-object time scale usually corresponds to a single musical note with a sense of beginning and termination. The user studies utilize sound objects as a general unit of sonification, addressed as either ‘stimuli’ or ‘snippets.’ Each sound object encodes a small sequence of one-dimensional data as well as multiple types of aesthetic timbral modulations. They are generally short and monophonic, but potentially musically complex and unpredictable sonic expressions. The implications of time scale are further explored in the discussions of the time-resolution experiment (chapter 4) and the new design framework (chapter 6).

As a parallel concept to time scales, the thesis also studies the multiplicity of timbral dimensions in relation to data encoding, perceptual decoding, and sound organization. Timbral dimensions are highly elusive, as section 2.1.2 describes the absence of standard definitions especially for electroacoustic sounds. Given this uncertainty, the present research first defines process-oriented 'synthetic' dimensions, representing spectral parameters separated by temporal motions in different time scales (discussed in detail in chapter 6.1.3). Roads suggests, "as Berry, Narmour (1990, 1992), and many others have pointed out, real music plays out through the interplay of multiple parameters operating simultaneously and sometimes independently [75]." With such an organizational scheme, the author observes how the listener identifies a complex stimulus with multiple characteristics.

3.4.2 Types of Analytical Listening

Related to the rather abstract 'timbral dimensions,' the general lack of shared vocabulary for describing synthetic expressions [91] can be problematic in the experimental design. For example, instructing the tasks to the first-time listener and collecting or comparing the verbal responses can be complicated. With different levels of sound literacy, some listeners may be able to quickly identify subtle aspects of timbre than others. For the possible disparity in verbal interpretation and communication, the author considers that there needs to be an active effort by all the listeners in discovering and connecting the unfamiliar characteristics of sound to personal and/or familiar textual descriptions.

The listening experiment aims to guide the listener toward the microstructure of sound with two contrasting types of focused ‘analytical listening’ (see section 2.1.3): interpretive and estimative. An interpretive listening task encourages the listener to explore the intricacy of stimuli for their nuances and motional characteristics, aided by a large list of non-technical textual descriptors. With estimative listening, on the other hand, the listener's focus is on approximating the 'degrees' of timbral motions created by data. Both types, however, involve a careful analysis of the internal structure of sound, identifying and tracing the 'data' timbral dimension among other concurrent motions.

3.4.3 Variability as an Experimental Device

The user studies put a unique emphasis on the design variety of snippets (stimuli). As argued by Roads and Cairo (see section 2.2.1), the variability or novelty of the design is integral in aesthetic explorations. Using novel sound expressions / organizations is also arguably a key element in organized sound, as Grisey observes, "this game between predictability and unpredictability, expectation and surprises is what makes time living and musical (quoted by Roads in [74])."

In this research, not only the designer study but also the listening test incorporates novelty of stimuli as a key experimental device, where there is always some level of unpredictability and freshness with stimuli. This may contrast to typically static or uniform structures in a functional sonification. To illustrate, one could consider the (lack of) aesthetic values of a 'data soundscape,' in which the types of sounding object, mappings, musical rules (e.g., tonal scales), or the compositional structure (e.g., sections) are fixed. While there may be a strong argument for the learnability of such a sonification, the more the listener listens, their musical interest may fade away rather quickly compared to a typical musical composition that incorporates structural changes in different parts of sound organization.

The thesis considers two categories of experimental factors in the human subjects tests: framework and design attributes. Framework attributes encompass the common structural elements in the sound synthesis of SMSon, such as the methodologies regarding time scales and timbral dimensions, spectral models, and data mappings. These internal elements have either deterministic behaviors or a finite set of choices.

Design attributes, in contrast, entail variable, personalized, and likely external factors (to the framework) embedded in the resulting sound design. It may be related to unique design intentions (e.g., a decision to create aggressive and intense timbral motions) but may chiefly be observed as how the framework attributes are utilized or combined. Such configurations of framework attributes are then perceived and interpreted by the listener, but likely result in different and nonuniform types of understanding (e.g., as a musical tone, an environmental effect, or an incomprehensible noise). In qualitative analyses, variable design attributes may anchor the elusive nature of design intentions and receptions. As such relationships are not predefined, the user studies aim to discover some of the common patterns from raw and diverse user responses towards uniquely designed sounds[7] [8].

[7] While design intentions may be verbally communicated from the designer to the listener, not every element can be efficiently or effectively communicated. The user studies examine the issues of both verbal and nonverbal communication of designs.

[8] It should also be noted that framework and design attributes do *not* correspond to functional and aesthetic elements of a sonification. Framework attributes are intended to support both functional and aesthetic goals, while design attributes uniquely configure the framework attributes to create a bespoke design.

3.5 The Guiding Principles and Evaluative Heuristics

Based on the above auditory and experimental concepts, this section proposes several F/A principles and evaluative heuristics for SMSon, aiming to set a directionality in otherwise infinitely open-ended design possibilities.

3.5.1 Functional-Aesthetic Principles of SMSon

For F/A sonifications, what exactly qualifies as functional or aesthetic may depend on contexts and applications (see chapter 2.2.1). As a low-level and data-agnostic sonification framework, SMSon sets four F/A principles: (1) the measurability of numeric data, (2) multiple and separable timbral dimensions, (3) the temporal dynamics of timbre, and (4) the variability of design.

First, the measurability of data is a functional principle for ensuring a direct and accurate representation of numeric fractional values. This can be considerably difficult in symbolic music with nonlinear or discretized structures. With timbral organizations, the author aims to facilitate continuous and adjustable modulations in the perceptual range and linearity (e.g., with a logarithmic scaling in frequency) for accurate retrieval of data. The measurability of different sonifications may be compared and examined in two approaches: with physical (acoustic or computational) and perceptual measurements. The present study focuses more on the latter while the development of different designs (model) for the study is informed by the computational measurement of spectral analysis and synthesis (see chapter 6.1.5).

The multidimensionality of timbre is, as already discussed, largely empirical but also an integral element in organizing aesthetic sounds. A F/A sonification may be composed of single or multiple dimensions of data and aesthetic treatments, and ideally decomposed into the original separate elements. The separability of timbral dimensions may be, therefore, a prerequisite for the data measurability. The composition and decomposition of dimensions are explored in both synthetic organization (chapter 6) and the perceptual listening test (chapter 7).

The temporal dynamics of a timbre are a critical factor for characterizing unique and novel sounds, providing a complex interface to the underlying synthetic parameter / data. This may contrast to a musical note with a fixed spectral pattern (which the author explores in chapter 5.2). Temporal dynamics may take multiple forms, including the changes at sample and structural levels. A sample-level change signifies that a short-time (e.g., 0.1 seconds) sound structure is not locked in by fixed oscillatory components, so that the instantaneous characteristics can evolve flexibly in both time (amplitude) and frequency domains[9]. Such continuous dynamics may enable complex and expressive timbral designs, but also introduce a challenge in physical and perceptual measurability. Structural dynamics, on the other hand, may signify a fundamental change of acoustic characteristics. As Roads observes, the voice of a musical instrument is often "static" in the acoustic structure, with each voice bound to a note duration [75]. With SMSon, the author aims to facilitate electroacoustic expressions with both sample- and structural-level changes, with multiple ways of evolving and morphing unbound to a musical note.

Lastly, closely related to the structural changes over time, the variability of design facilitates a wider range of musical expressions, while also inviting designers with different musical preferences. Roads describes “the economy of selection,” which affords “choosing one or a few aesthetically optimal or salient choices from a vast desert of unremarkable possibilities. [74]” Alternative representations of data may lead to a better listener engagement and/or different perspectives to data. The variable design also facilitates novel sound expressions.

[9] Roads may instead call this a micro-level change [76].

3.5.2 Three Problem Scopes

Variable design attributes are possibly largely qualitative and may not present a regular pattern to analyze. In order to capture the design attributes with listener's receptions, or for the designer to review the intended 'effect' of their designs, it is beneficial to have higher-level heuristics to situate the variable elements. The study sets three view angles to examine: data intelligibility, sound complexity, and musical appeals. The formation of these heuristics is broadly informed by the disciplines that SMSon uses as bases, which are (functional) sonification, audio analysis and synthesis, and musical compositions, respectively.

Data intelligibility is mainly of interest to functional sonifications. As alluded to in sections 2.2.1 and 2.2.3, intelligibility may entail various different metrics such as indexicality (directness or concreteness) [107], intuitiveness, and gestalt clarity [108]. With multidimensional sonifications composed with SMSon, it mainly signifies the ease of decomposing and estimating the physical quantities encoded in timbral dimensions.

Sound complexity may be primarily attributed to signal processing. The author envisions it as a controllable and a perceived information depth[10] of sound. In general, the term ‘complexity’ has various definitions such as the amount of randomness (cf. information entropy [89]) and the amount of information being encoded (e.g., semantic associations or logical depths [4]). In the context of creating / listening to electroacoustic sonifications, it signifies the intricacy of temporal and multidimensional structures carefully arranged by the designer.

A musical appeal is somewhat complementary to sound complexity – A low-complexity sonification (e.g., a single-dimensional melodic sonification) may perhaps lack the capacity to be musically appealing than a more uniquely-crafted timbral expression. With the listener, this

[10] The auditory information may include not only data but also aesthetic treatments.

heuristic aims to capture their emotional and/or aesthetic responses when engaging a complex and potentially unfamiliar sound. The analysis of listener responses seeks to find the types of subjective musical valuation (e.g., a snippet might evoke an imaginary soundscape) and how they influence the listener's objective aural analysis. For the designer, on the other hand, musical appeals may indicate how unique, versatile, or inspiring a designed sonification is for composing a functional-aesthetic musical piece (as a future direction of this study).

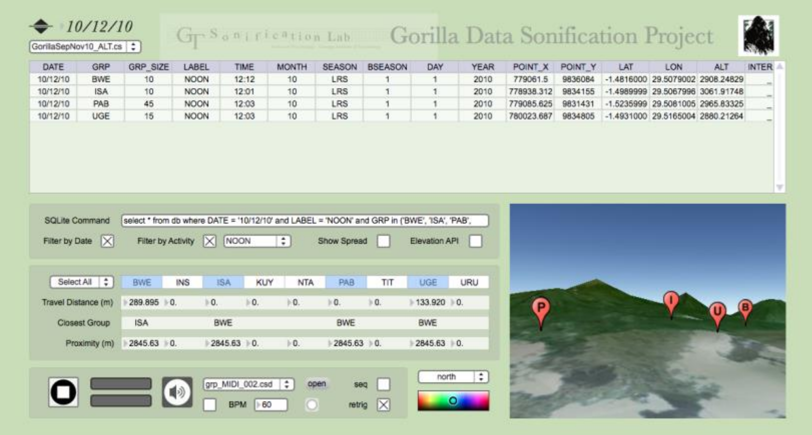
The present definitions of the principles and problem scopes are largely inspired by the author's prior works, exploring various pragmatic issues of F/A designs with real-life data. The following chapters 4 and 5 recapitulate the previous works from the lenses of these F/A problem definitions.

CHAPTER 4. PRELIMINARY WORK 1: MUSICAL SONIFICATIONS

This chapter portrays the author's previous musical sonification works. In contrast to the present framework study, these projects employ real-life data, exploring functional-aesthetic (F/A) design ideas in multiple directions. Along with the previous frameworks (see chapter 5), they contextualize the core F/A principles of spectromorphing sonification (SMSon; see 3.5.1).

All projects, except for the melodic-sonification listening test (section 4.4), were interdisciplinary collaboration extending for around 0.5-1 year. They share a similar goal of making raw or complex data to be more accessible and understandable for the general audience by using musical representations.

4.1 Gorilla-Ecology Musical Sonification



DATE	GRP	GRP_SIZE	LABEL	TIME	MONTH	SEASON	BSEASON	DAY	YEAR	POINT_X	POINT_Y	LAT	LON	ALT	INTER
10/12/10	BWE	10	NOON	12:12	10	LRS	1	1	2010	779061.5	9836084	-1.4816000	29.5079002	2908.24829	_
10/12/10	ISA	10	NOON	12:01	10	LRS	1	1	2010	778938.312	9834155	-1.4989999	29.5067996	3061.91748	_
10/12/10	PAB	45	NOON	12:03	10	LRS	1	1	2010	779085.625	9831431	-1.5235999	29.5081005	2965.83325	_
10/12/10	UGE	15	NOON	12:03	10	LRS	1	1	2010	780023.687	9834805	-1.4931000	29.5165004	2880.21264	_

SQLite Command: select * from db where DATE = '10/12/10' and LABEL = 'NOON' and GRP in ('BWE', 'ISA', 'PAB',

Filter by Date ☒ Filter by Activity ☒ NOON Show Spread □ Elevation API □

Select All	BWE	INS	ISA	KUY	NTA	PAB	TIT	UGE	URU
Travel Distance (m)	289.895	0.	0.	0.	0.	0.	0.	133.920	0.
Closest Group	ISA		BWE			BWE		BWE	
Proximity (m)	2845.63	0.	2845.63	0.	0.	2845.63	0.	2845.63	0.

grp_MIDI_002.csd open seq □
□ BPM 60 ○ retrig ☒
north

Figure 4.1 The Gorilla-Ecology Sonification software.

The first musical experiment with data was the gorilla-ecology sonification (2013), developed in collaboration with Georgia Tech (GT) Sonification Laboratory[11], GT Center for

[11] http://sonify.psych.gatech.edu/

Geographic Information Systems, and the Zoo Atlanta[12]. The project featured time-series data of wild gorilla activities near the boundary of Uganda and Congo. The daily observation data include gorilla-group coordinates, sizes, and qualitative remarks. The project aimed to explore these unanalyzed records through sound and (consequential) musical effects that would ideally be able to 'tell a story' about the gorilla-group interactions (e.g., musical patterns emerging from the group sizes and inter-group proximity). The team examined varieties of mappings that would allow a simultaneous playback and perceptual tracking of gorilla-group activities up to nine groups, while also considering the scientific validity on a psychological basis. The author developed an interactive sonification system with a graphical and programmatic (SQLite) data-exploration user interface (UI; see Figure 4.1) and a modular sound-design pipeline (with Csound). This separation of data handling and real-time audio synthesis led to (re)programmable sonification designs. The system enabled any data-viewing configuration (e.g., input value range and dimensionality set dynamically by a particular SQL query) can be seamlessly mapped to audio without breaking.

4.1.1 Sonification Models

Among more than 10 types of sonification models created in the environment, this review introduces four of them. The most simplistic pitched-pulse model mapped a data dimension to the time onset interval. A time-stretched decaying amplitude envelope was applied to a pitch-modulated oscillator with a static timbre. While the perceptual measurability for each voice may be considered high [29][6], this method had a limited potential of creating a multi-stream (polyphonic) sonification for variety as well as discernibility.

[12] https://zooatlanta.org/project/dian-fossey-gorilla-fund/

The second, modal-resonance synthesis, was a classic yet effective way to produce a static and characteristic voice. This method used a set of predefined frequency peaks[13] to create narrow bandpass filters, applied to a white noise. In this project, it was combined with a pulsating amplitude envelope to create a somewhat realistic voice (e.g., resembling a wine glass, woodblock, etc.) assigned to multiple gorilla groups. With modal filters, using a pitched oscillator (a wide-band oscillator such as sawtooth wave) was less effective for achieving a well discernible timbre, losing a common salient parameter for data mapping.

Several musical examples focused on aggregate representations, such as the spread of selected gorilla groups. The sound-beacon sonification provided a general sense of location by anchoring a continuous sound source, such as a looping wind noise, at a point of interest in the map (e.g., corners of the map or the hills). As a gorilla group traveled close to the beacon, one would be able to hear more of these static environmental sounds. While it was not feasible to map the group identification to the variation of the activated sound, the more group nearing the beacon contributed to the more intensity of these resulting sounds. Aggregate mapping methods, in general, were observed to have the best flexibility for controlling the time scale of musical events as they were not bound to individual data points. Sound beacons were a many-to-one mapping, while other methods would utilize a many-to-many configuration.

Lastly, the polyphonic-melody piece realized the most complex but potentially discernible 9-part polyphony in a micro-rhythmic manner. Using a symbolic sound organization, it created a three-note melodic pattern based on the ascii-code representation of each gorilla group's name. The melody, with a General-MIDI SoundFont instrument arbitrarily assigned to each group for perceptual voice separation, was pitch-shifted (with the proximity to any other group) and time-

[13] https://csound.com/docs/manual/MiscModalFreq.html

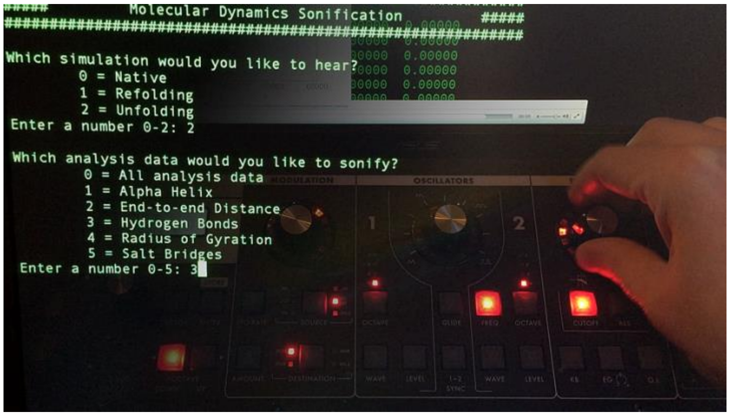
compressed or expanded (with the travel distance per day). The sound-beacon technique was also used simultaneously to present the general (unidentified) group locations on the map. The musical piece, slowly playing back the data for the three-month duration, was captured as the final demo of this sonification system as a potential museum installation at the Zoo Atlanta[14].

Despite being the first sonification project for the author, given the creative freedom, the team actively explored the unique potential of variable musical sound organizations for a single data source. Different musical models signified new perspectives to the same data. However, the environment lacked the ability to explain the mappings without the user spontaneously interacting and isolating each gorilla group from the others.

The variable models also indicated that a particular musical structure only worked in an isolated musical style from others. For example, the use of a continuous-pitch pulsating instrument, being in somewhat of timbral and subsymbolic time scale, was not directly compatible with the polyphonic-melody model, which compromised some continuity / perceptual resolution of data to retain the distinctions of the voices. A designer / composer might develop a very different musical style than others depending on which basic technique to employ. A question remained unresolved that how different musical models or their components may be analyzed and compared with others, and potentially be integrated into a single musical structure.

[14] https://vimeo.com/113269881

4.2 Protein Music



```
#####         Molecular Dynamics Sonification                   #####
######################################################################

Which simulation would you like to hear?
        0 = Native
        1 = Refolding
        2 = Unfolding
 Enter a number 0-2: 2

 Which analysis data would you like to sonify?
        0 = All analysis data
        1 = Alpha Helix
        2 = End-to-end Distance
        3 = Hydrogen Bonds
        4 = Radius of Gyration
        5 = Salt Bridges
 Enter a number 0-5: 3
```

Figure 4.2 A version of Protein Music installation. Simulated protein data are streamed to a monophonic virtual-analog synthesizer.

Continuing on the idea of 'music to tell an emerging story about data,' the author’s next project was the molecular-dynamics sonification (2014), a public-outreach project co-developed with GT school of biology and GT research institute (GTRI) configurable computing laboratory (Config Lab). The general goal of the project was to convey the dynamic behavior of protein molecules in a novel and intuitive way to the non-expert audience in chemistry. The team employed simulated data of a protein folding and unfolding, a time-evolving three-dimensional structure, from which the Config Lab extracted five high-level attributes that characterize the dynamic and complex states. To aid the first-time audience in learning the likely unfamiliar interface of sonification, the initial approach was the simplest sonification mapping each data attribute to the most salient attributes of a sound (Figure 4.2), such as pitch [22]. However, with the challenge of representing the time-evolving characteristics of a geometric structure, the team also sought the inherent "musicality" or perceptible proportions of the high-level attributes.

4.2.1 Types of Sonification

The team developed four versions of sonification, showcased at a science festival booth as well as an art gallery. Two of them had the author’s significant design contributions.

The first "voice-mixing" piece, programmed with Python and Csound, featured an interactive UI with a multi-slider MIDI controller. Each slider was mapped to a musical voice (e.g., instrument and synthetic sound) representing an analyzed attribute of data. The voices could be freely mixed or turned on / off by the user to focus on fewer voices. Each voice was generated by the variable-speed and -directional playback of the data sequence, with "one-to-many" mapping of attributes to synthesis parameters that utilize various statistical summarizations in multiple time-window lengths (this could be also considered as a "many-to-many" mapping [79][45]).

The second sonification piece, Protein Symphony, was a non-interactive audio-visual composition developed in Csound and Max/MSP. As a theatrical piece projected on a large screen in a dark room, it aimed to tell a story about the dynamic features of protein unfolding with more of the author's musical interpretations. The screen showed a high-definition visual rendering of the surface and skeleton structures of the protein, the bar graphs of the five attributes used, and short texts explaining the data context and sound mappings. The piece gradually introduced new voices and transitioned to a new sound texture in the middle presenting different musical perspectives synchronized to an alternate visual structural representation.

4.2.2 Listener’s Understanding of Sonification System

In contrast to the gorilla sonification, which realized the on-the-fly change of music through data queries with user interactions, the protein voice-mixing sonification focused on the user interaction with the sound interface. It was intended to help the listener learn the surface-level mapping relationships between the sound behaviors and a data dimension, while it did not detail the mapping processes such as which data points contributed to which musical notes.

The Protein Symphony, as a fixed media piece, opted for textual descriptions on top of the visualization of the data to help the listener make sense (e.g., how to interpret 'high' or 'low' values when they are not mapped to the pitch or loudness).

For both protein and gorilla sonifications, as the syntheses and mappings were mostly hard-coded, the range of musical expression was fixed during the usage / listening. As a musical sonification, perhaps the most "hard-coded" elements were the compositional decisions used for, e.g., creating the many-to-many mappings and musically sounding structures. The complex and partially hidden process of the composition may hinder the understanding of sonification even for other sonification designers. How to make the process of composition more transparent, accessible, simpler, and flexible was a common question asked in the following projects.

4.3 Streaming Civic Data Sonifications

Continuing on the collaboration with the GTRI Config Lab, the next project explored F/A sonifications with a large amount of streaming data.

Similar to the molecular sonification, the civic data sonification (2014-15) set off as a public outreach demonstrating alternative ways of handling and representing inaccessible data. Combining the web technologies, live data monitoring, and automated data integrations and sense-making algorithms, the team explored the idea of how to handle a rapid and large stream of data such as live Twitter feeds and environmental sensor arrays installed across the busy town center of Decatur, Georgia. This was also the author’s first web-based sonification experiment, serving as a testbed for building a reconfigurable sonification framework / library for the web.

Along with the sonification engine data-to-music (DTM; see chapter 5.1), the author developed two types of audio-visual system: (1) data-streaming browser dashboards and (2) a data-driven performance piece featuring live musicians, live coding, and real-time score generation.

4.3.1 Web Dashboards

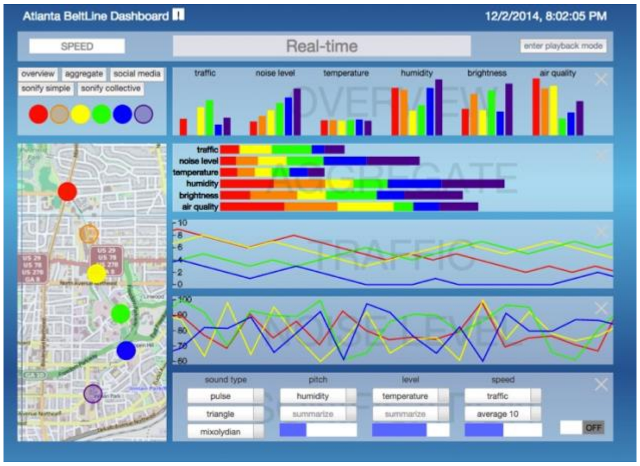


Figure 4.3 The web sonification / visualization dashboard for real-time civic sensor-data streams.

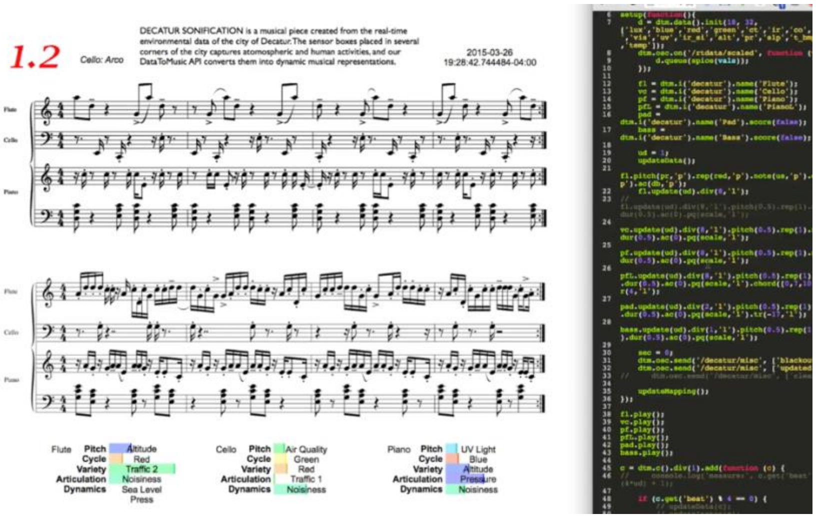
The web dashboards were intended to bring an interactive audio-visual experience to the end-user. In contrast to the Gorilla and Molecular sonifications, which were bound to a specific software environment, the browser dashboards rendered the audio entirely on the user's side. This raised experimental opportunities for user-specific exploration of data, a more personalized listening experience, and potentially a closer design-feedback loop involving the user in the mapping / sound-design process. These were further examined in the following online listening test.

The dashboard sonifications were also distinct from Gorilla / Protein projects as they focused on facilitating the discovery of patterns or anomalies from dense unstructured or unknown (live) data. The Twitter sonification featured the tweets (messages) related to the natural disaster events of Atlanta snowstorm (2014) and Hurricane Sandy (2012). The tweets had been collected and processed by the GTRI high-performance computing laboratory adding sentiment values, estimated locations, topic relevance, etc. Within the dashboard, the tweets were played back at normal or higher speed, displaying a flood of information at the peaks of the event. For the user to make sense of such high-volume timed events, the dashboard provided two main features: (1) A facility for the on-the-fly categorization of tweets by user-specified keywords, and (2) The user-controlled reconfiguration of the synthesis and mapping. For each category of tweets, the user would be able to set a sound type that could be distinguished from others. For instance, they may create a category with the term “water,” and map a transient sound such as a snare-drum sample. The sound could be further configured with continuous modulations such as band-pass filtering, mapped to the sentiment or location values. While this “one-to-one” datum-and-sound mapping did not provide much of a complex musical structure, it arguably demonstrated a type of interactivity for establishing a creative but also understandable sonification by involving the user in the design process.

The civic-data dashboard (Figure 4.3) similarly featured a customizable sonification. It was audio-visual monitoring and navigation of the environmental data from seven sensor boxes, custom-built and installed around a town square by the Config Lab. This dashboard facilitated the loading of multiple visualization or sonification widgets interfacing the constant stream of data. While the sonification widgets allowed for a little beyond one-to-one mappings with variable-length aggregation, this dashboard was not particularly successful in providing creative or sense-making opportunities.

Typically, a dashboard (e.g., of a vehicle) would feature multiple infographics to provide the user immediate access to various information. With Twitter and Decatur sonification, besides the concurrent use of visualizations, the project did not explore the possibility of assembling a multi-modal interface, though it was possible to stack multiple voices of similar type sonification. The potentials of multimodal sonification, creating a complex soundscape, was explored in a later project with the CODAP plugin development (see chapter 5.3).

4.3.2 Live-Score Piece



DECATUR SONIFICATION is a musical piece created from the real-time environmental data of the city of Decatur. The sensor boxes placed in several corners of the city captures atomospheric and human activities, and our DataToMusic API converts them into dynamic musical representations.

```
setup(function(){
    d = dtm.data().init(10, 32,
['lux','blue','red','green','ct','ir','co',
','vis','uv','ir_sl','alt','pr','slp','t_bm
,'temp']);
    dtm.osc.on('/rtdata/scaled', function (
        d.queue(splice(vals));
    });

    fl = dtm.i('decatur').name('Flute');
    vc = dtm.i('decatur').name('Cello');
    pf = dtm.i('decatur').name('Piano');
    pfL = dtm.i('decatur').name('PianoL');
    pad =
dtm.i('decatur').name('Pad').score(false);
    bass =
dtm.i('decatur').name('Bass').score(false);

    ud = 1;
    updateData();

fl.pitch(pr,'p').rep(red,'p').note(us,'p').
p').ac(db,'p');
    fl.update(ud).div(8,'1');
//
fl.update(ud).div(8,'1').pitch(0.5).rep(1).
dur(0.5).ac(0).pq(scale,'1');

vc.update(ud).div(8,'1').pitch(0.5).rep(1).
dur(0.5).ac(0).pq(scale,'1');

pf.update(ud).div(8,'1').pitch(0.5).rep(1).
dur(0.5).ac(0).pq(scale,'1');

pfL.update(ud).div(8,'1').pitch(0.5).rep(1)
.dur(0.5).ac(0).pq(scale,'1').chord([0,7,10
r(4,'1');

pad.update(ud).div(2,'1').pitch(0.5).rep(1)
.dur(0.5).ac(0).pq(scale,'1').tr(-17,'1');

bass.update(ud).div(1,'1').pitch(0.5).rep(1
).dur(0.5).ac(0).pq(scale,'1');

    sec = 0;
    dtm.osc.send('/decatur/misc', ['blackou
    dtm.osc.send('/decatur/misc', ['updated
//      dtm.osc.send('/decatur/misc', ['clea

    updateMapping();
});

fl.play();
vc.play();
pf.play();
pfL.play();
pad.play();
bass.play();

c = dtm.c().div(1).add(function (c) {
//      console.log('measure:', c.get('beat'
(4*ud) + 1));

    if (c.get('beat') % 4 == 0) {
        // updateData(c);
```

Figure 4.4 Live score generation with civic sensor data.

The live performance piece, called the Decatur Sonification, explored how human musicians and system reconfiguration could render a stream of real-time sensor data into a piece of structured polyphonic music. Together with the live dashboard example, which used the same streaming data from sensor boxes, the team showcased the piece to the public audience in the Atlanta Science Festival 2015.

In this live piece, three professional musicians from Sonic Generator[15], a pianist, cellist, and flutist, participated. They would sight-read a musical score generated in real time (Figure 4.4).

[15] http://www.sonicgenerator.gatech.edu/

The generated score, a four-bar snippet in traditional western notation, would start simple with long and isolated notes as part of the compositional control of data mapping, but later evolve into somewhat complex melodies using sixteenth notes. To ease the sight-reading, as well as to aid the audience make some sense of the mapping of the rapid data sequence, the musical style of the sonification incorporated several layers of repeating patterns. First, the sensor stream, updated every second, was aggregated and buffered for the largest repetition cycle of 15-20 seconds. Each cycle, consisting of four measures, was repeated twice, while the next cycle was already rendered and displayed in the area below. The melodies were generated with five parameters: the phrase length (this was also called the "cycle" as the phrases would repeat), variety (note irregularity), pitch modulation and range, the introduction of note articulations, and dynamics (introduction of rests). The instrument parts having different phrase lengths would create a poly-rhythmic pattern resembling the minimalist style of Philip Glass.

The live performance used two versions of data: (1) prerecorded and sped-up sensor data that contained a variety of interesting motions and arching dynamics over time, providing the sense of beginning and end; and (2) the slow real-time streaming from the live sensor boxes. As a conductor / director, the author joined the performance by live coding the overall structure of the piece. On the author's laptop, a browser page running the DTM live-coding editor would receive the sensor data from the central hub, process them with mappings and melody generations, and forward them to Max/MSP running the Guido notation library [43] for rendering the score. The audience would see the score, data visualization, and live coding controlling the mappings. The structural control of the piece entailed two approaches: (1) The melodic variety and intensity were reduced when more steady data channels, e.g., the temperature, were mapped and increased when more active data, e.g., microphone and light sensors, were used. (2) The available notes (pitches) for representing data were arbitrarily reduced / increased according to the current section of the piece. The more pitches available, the more complex harmonies appeared.

4.3.3 Synthesis / Analysis Window Lengths and Time Variance

Both of the protein sonification pieces combined various time-windowed aggregates mapped to musical phrases of various lengths (i.e., many-to-many). The Decatur Sonification, on the other hand, utilized a fixed time window for summarizing the inputs and mapping them simultaneously to high-level parameters of melody generation, such as “cycles” and “variety.” The four-bar melody cycle arguably made the visual and sonic comparisons easier. The matching, or dynamics, of the time window for sound generation and analytical listening has become a major topic of exploration in later studies, including this thesis research.

The use of aggregate mappings provided flexibility to create musical phrases, cadence, and form over time. However, it was also evident that some level of time-evolving characteristics of the data (e.g., self-correlating structure) was lost in the process of summarization. A question lingered that if it would be feasible to organize a time-evolving musical sound that conveys the information about data (values or otherwise).

4.4 Time Scales of Melodic Sonification: A Listening Test

With musically complex sonifications, it was generally difficult to observe the relationships among user interactions, their understandings of the system, and the retrieved information (data). To address this challenge, the author conducted an online listening test (2018) with a very simplistic but interactive sonification model [102].

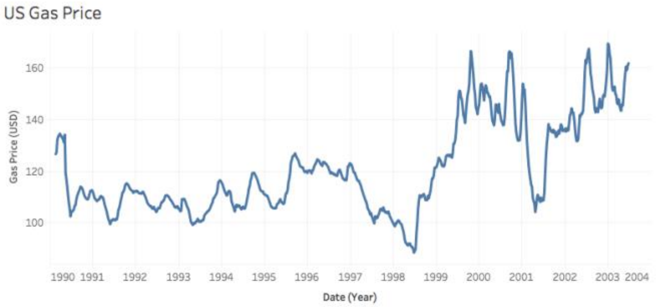


Figure 4.5 The US Gas Price data used in the melodic sonification.

The listening test employed arguably the most primitive sonification of all, a pitch-modulated monophonic melody. Correspondingly, the data used in the test[16] had a single-dimensional structure (Figure 4.5). Limiting the aesthetic and structural complexities, the study focused on examining the process of the listener's aural analysis: When trying to make sense of a monophonic 'melody' representing certain data, how does the listener recognize the patterns of the underlying data, such as recurring shapes, directionalities, or the local / global contours? Also, how can such analytical listening be observed and quantified?

4.4.1 Background: Multiplicity of Musical Time

To design an experiment about aural sense-making, the authors primarily referenced musical theories of time, including the musical time scales. The theories of time scale may be categorized as either structural, focused on the organization of sound, or durational, focused on human listening.

The structural view of time scales, as already introduced in section 3.4.1, considers musical time to be hierarchical, even for a simple monophonic melody. Roads describes nine levels of time scale that contribute to the formation of a musical piece, sections, rhythms, pitch, timbre, etc. for

[16] https://vincentarelbundock.github.io/Rdatasets/doc/DAAG/carprice.html

both perception and generation [76]. Here, different time scales imply different resolutions or levels of detail. The sensation of tone (pitch), for instance, happens at the “sound object” (i.e., note) and “microsonic” levels when there are enough repetitions of a microsonic pattern, while timbral effects are noticed when there are “micro” and “sample” level irregularities. The melodic patterns are recognized in the “meso” time scale, one level above the sound objects. The listening test inquires the following questions: How would these structural time scales or resolutions play a role when analyzing the data within a melodic sonification, such as value-by-value fluctuations, local and global peaks, and gradual shifts of the central point? How does the listener identify these scales in a new melody?

The durational view of time scales, in contrast, argues that our experience of musical time can be multiple (instead of hierarchical) while listening to and analyzing a melody. Jonathan Kramer expresses such a multiplicity of experienced time as the “polyphony of viewpoints [54].” He describes that, when we listen to, familiarize, and compare different parts of a melody, our mind does not simply follow the absolute time with a linear progression. Instead, we gradually learn the structure of the melody in terms of the duration of auditory patterns and their various proportions (cf. time scales). As we listen to the piece / melody, Kramer observes, we acquire new information about the proportions. As such, the focal length of time may be constantly adjusted with two cognitive processes present: one following the durations in passing (from a still sounding past moment until the current moment), and the other experiencing and comparing the remembered durations in retrospect. While we may continuously learn the time structure of the melody as it progresses, he also argues the importance of repeated listening for a thorough analysis [55]. The listening experiment heavily incorporates these ideas into the task design as well as the analytical approach.

4.4.2 Interactive Listening Environment

The listening test featured a minimalistic user interface consisting of a playback control, a time-resolution control slider, and a pitch-resolution control slider. The data were played back at the rate of 39.7 data points per second[17] for the original time / pitch resolution. The amplitude of the melody stays constant. In the experiment, the listener did not see the visualization of the data, but only heard the sonified result.

The more unique and interactive features of the experiment, the time / pitch resolution controls, would modify the melody being played in real time by altering the underlying data. They essentially emulated the varying time scales or 'focal ranges' of analytical listening, explicitly controlled by the user. The pitch-resolution control applied a uniform but variable-step-size scaler quantization onto the log-scaled (i.e., MIDI-note) frequency. As the user lowers the pitch resolution, for example, the output MIDI range is rounded by the factor up to 9.0. The time-resolution control, on the other hand, would down-sample and then interpolate the data. With a lower time resolution, the model emulated a linearly interpolating characteristic between data points (as in a line graph) rather than a step interpolation (as in a bar graph) and rendered the pitch-modulated sound accordingly. When the time and pitch abstractions are combined in some tasks, the pitch quantization was applied after the time abstraction (interpolation) process.

4.4.3 Listener Tasks

The main purpose of the experiment was to observe the subject's exploration over time as they perform an analytical task on a melodic sonification. However, designing an analytical listening task required a significant care. For instance, it would be impossible to ask them to

[17] 40Hz is used typically for speech analysis and modeling as the minimum threshold for the perception of a tone [75].

identify a specific pattern in the melody (as ground truth) without giving away the information about the data and biasing their exploratory behaviors. Instead, the tasks instructed the user to look for a “balancing” point or a perceptual boundary between two high-level categories of data attributes that were most likely in different time spans.

The study categorized the target attributes into the time spans of “details,” “local,” and “global” levels. The “details” level would include perceived attributes such as rapid fluctuations or continuities of individual data points. The “local” time span may include small peaks and periodic patterns, as well as the current central tendencies. The “global” attributes may be such as the overall contour of the data, the total value range, and general directions of the data sequence. The analytical tasks were grouped into two main stages: the first three tasks asked the user to find a boundary between the “details” and “local” characteristics, and the second three asked to identify the balancing point between “local” and “global” levels. Each task required different combinations of the time / pitch resolution controls. Finally, an extra task asked them to find the most musically balanced resolutions without the consideration for data analysis.

While the authors hoped to find general correlations between their “final” parameter values indicating the optimal perception of certain data attributes, such results would be also dependent on the data source and, ultimately, the user’s personal preference. Instead, greater interest was in finding patterns in their exploratory behaviors on a newly encountered sonification, measured over several different durational time scales. In other words, this study questioned if the progress of time correlates the way they like to hear the melody in certain time / pitch resolutions for a particular analytical goal.

The study hypothesized that, taking Kramer’s argument of the listener gaining new information about melodic proportions through the progress of time and repetitive listening, there may be converging behaviors or common directionalities for both time- and pitch-resolution adjustments over time. There was also a question of if the total duration spent for a task would have a positive correlation to their final values. As for the use of the time / pitch resolution controls that emulate the structural analysis of listening, Roads' hierarchical time-scale theory suggested to us that creating an appropriate pitch resolution may be more important to capture the phrase-level and longer time-scale structures of the melody, such as the global contour, than for local details. On the other hand, the time-resolution control may have stronger relationships especially to the perception of local-level data attributes, such as small peaks and rapid fluctuations.

4.4.4 Summary of the Results

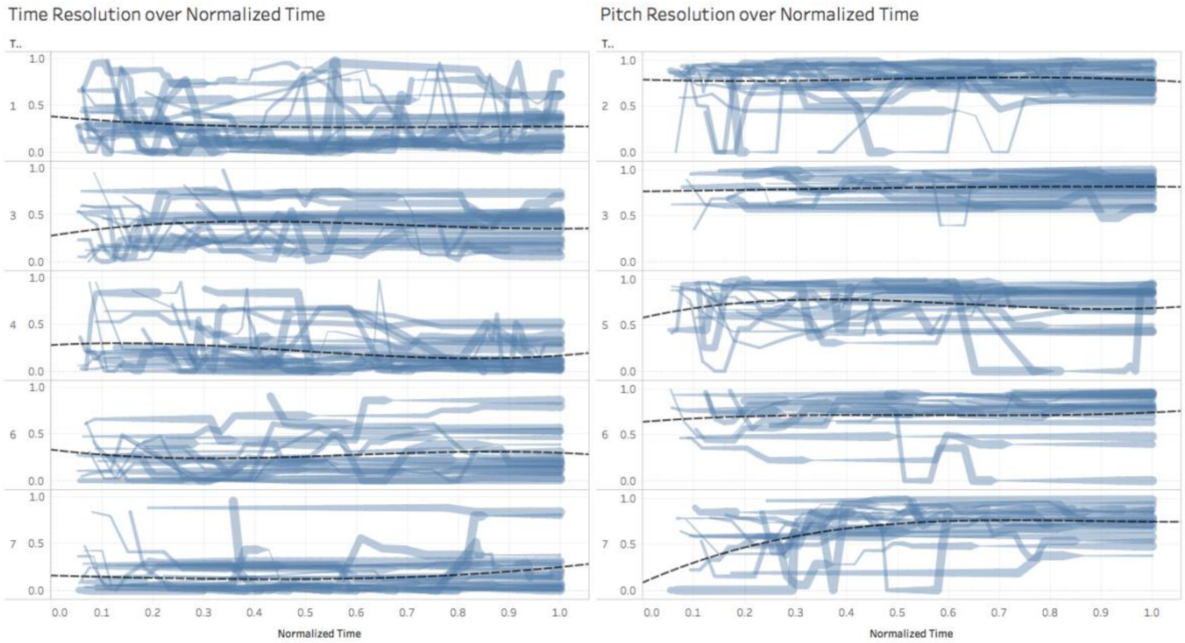


Figure 4.6 The non-aggregated time / pitch resolution controls over normalized time. The dotted lines are the 3rd-order linear regression.

The study collected data from 20 subjects, which were analyzed with the first and third-order polynomial regression to map the variable relationships. The variables, such as the time- and pitch-resolution values (indicating the exploratory behaviors and user analyses results), various time units (e.g., normalized task time and data phase), and task types were compared in terms of the resolution-values alone, with the time progress, and the relationships between analytical and musical balances.

In general, the regression models with the time-resolution or pitch-resolution as outputs with any input variable (e.g., time) showed very low R^2 values, commonly below 0.1 (10%), suggesting that they could not be used for a precise prediction of the “exploratory” behaviors. The authors speculated the factors to be either the generally random nature of user exploration, strong personal biases of perceived structures in the melody, or potentially the ineffectiveness of the task instructions. On the other hand, some models had a p-value below 0.01, indicating the presence of common directionalities of two variables when compared to the case where the input coefficient is set to zero (i.e., a null hypothesis). The model directionalities were, however, generally very moderate as seen in Figures 4.6.

The results indicated several contrasting patterns in how the listener explores and matches the resolution controls to the structures of the melody (data). At the task level, which disregards the time progression, there were strong correlations between the musical preferences and all pitch configurations, whereas the time resolutions varied for both the perception of data and musical balances.

Over the timeline of data exploration, on the one hand, the time resolutions had much larger effects in all tasks until arriving at the final configurations. This defied the hypothesis that time / pitch resolutions would affect separately on local and global structures. On the other hand, in repeated listening analyses over the data phase, the pitch resolution had more significance in the exploration over time. These suggest the durational time scales for listening (i.e., separate tasks, normalized time, and phase) may utilize different combinations of the resolution controls, but not as simple as the time to the low-level and pitch to the high-level relationships.

4.4.5 Informing the Present Research Design

The findings of this listening test showed nonlinear relationships between different time scales and analytical goals, and how one might design a musicality with just pitch and rhythmic components. Aiming to identify more linear relationships, the present listening test (chapter 7) focuses on more subtle timbral components and gestural / non-gestural time scales, untangling the all-encompassing sensation of symbolic pitches. In the development of frameworks (chapters 5 and 6), the author attempts to expand the use of hierarchical time scales for mapping data and aesthetic components separately, examining how they contribute to the design flexibility, measurability, and the user's understanding of a complex sound structure. The analytical listening experience will be observed without the use of interactive sound modification, but with how the listener describes their aural perceptual experience with concentrated listening.

4.5 General Reflection on the Musical Sonifications

This chapter described the development of musical sonification projects and the discovery of unique F/A values such as design variability and time-scale organizations for both sound and data. On the other hand, all the musical sonifications with malleable designs posed a fundamental challenge in systematically assessing or improving the functionality and aesthetics, as well as defining their relationships. The quantitative listening test with a melodic sonification mainly only indicated nonlinear relationships among dynamic time scales, masked by unpredictable exploratory patterns. The present user tests, therefore, take the exploratory behaviors into account for the experimental design / assessment.

The F/A elements, however, were effective in the listener engagement as well as gathering informal feedback. All three of the musical-sonification projects incorporated real-time, interactive, and variable sonic designs. At the showcasing venues, the curious audience explored the musical varieties to interact with the underlying data, learn the sonic interface, and occasionally inquired how the data were prepared, analyzed, and transformed to musical expressions.

Between the projects, the mode of interaction has shifted from the UI for data query and exploration, sound alteration, to the reconfiguration of the mapping. These raised a question of where or whom the agency of design should belong in order to create a F/A sonification. In some system designs, interactivity was simply not available. The thesis further explores this problem.

Almost all of the sonifications used the symbolic-music paradigm, with clearly defined notes, melodic phrases, and harmonies, but with less of timbral complexities. The data points were mapped to individual notes or even to fewer note events as an aggregate (the “many-to-one” mapping [83]). While some sonification models, such as the modal-resonance synthesis, explored the timbral variety, the relationships of how individual sound forms a bigger and structured experience over time were left for future inquiries.

The styles of music were also largely constrained by data contexts and available technologies (e.g., synthesis engines), which led the author to develop more general and data-agnostic frameworks introduced in the next chapter.

CHAPTER 5. PRELIMINARY WORK 2: DESIGN FRAMEWORKS

This chapter reviews the previous design frameworks and toolsets that the author has (co-)developed. As general-purpose framework research, they explore methodologies, implementation, and functional-aesthetic (F/A) concepts. Some projects are focused more on technical aspects (e.g., data-to-music) while the others explore theoretical implications more.

As observed in the previous projects (chapter 4), timbral or musical sonifications may face functional challenges such as the difficulty of mapping various types of data to time-based signals and effectively communicating the mapping and sound properties to non-expert listeners. The author explores these issues using live reconfiguration (see section 5.1), quantity-driven sound organization (see 5.2), and the compositional use of data analysis (see 5.3). Each framework further engenders unique functional and aesthetic values (not limited to the ones discussed in chapters 2 and 3). However, similar to the musical sonification projects, these frameworks highlight the difficulty of (1) simultaneously achieving multiple F/A values and (2) systematically assessing them. Such challenges are carried over to the present framework research (see chapter 6).

5.1 Data-to-Music

After designing the gorilla and molecular sonifications, the author and collaborators at GTRI Config Lab identified two pragmatic issues in real-time interactive sonifications: (1) The use of native audio programs such as Max/MSP and Csound hindered the availability for the end-user; (2) Many aspects of data handling, mapping, and sonic designs were hardcoded for a particular data structure and musical style, making the system not reusable when a new set of data were brought in. To facilitate more reusable designs with end-user availability, the author developed the data-to-music (DTM) application programming interface (API), a JavaScript library for web browsers[18]. As the name implies, this API aims to adapt and transform data with any shape (i.e., dimensionality and cardinality), structure (e.g., patterns within values), and quality (e.g., cleaned, analyzed, or not), to an aesthetic auditory representation.

5.1.1 Developmental References and Motivations

A major inspiration for developing DTM came from data-driven documents (D3) [7], a low-level visualization library for creating custom and experimental visualization. D3 provides links between foreign data structures, particularly tree-structured HTML / SVG and flat relational data. DTM, similarly, aims to connect two separate domains, offline relational data and web audio [110], a graph-based real-time audio technology for web browsers. It provides several modes of reading data and mapping to various aspects of sound from the rhythm of notes to waveform shapes. In contrast to the long-existing audio programming environments such as Csound and Max, creating a multi-time-scale musical expression in real time with Web Audio can be challenging with the limited number of unit generators and fewer ways of event scheduling or updating. DTM overcomes this limitation by mixing offline and online (i.e., real-time audio) rendering in the process of nested schedulers [103].

DTM incorporates a system model similar to the classic digital audio synthesis, creating a complex chain of audio transformations by combining and connecting unit generators [61]. At a lower (sample) level, the audio and control signals are handled as a stream of buffers, where each buffer is a time frame of one or more audio channels to be processed iteratively. DTM utilizes both the modular (mapping) and streaming (iteration) operations to connect offline data to real-time

[18] Browser-based technologies are commonly available for the end user as opposed to native applications. A web-based library is also able to seamlessly integrate to other domains of technology such as databases and visualization tools.

audio outputs. The API also features chained transformations, where the result of an operation in one function is used directly in the next.

5.1.2 Functional-Aesthetic Values in DTM

The core F/A values of DTM include the reconfigurability and type-agnostic mapping of data to musical expressions. Particularly, the author examined how multiple types of non-musical data could modulate a single musical structure without losing musical complexity, similar to the expressive virtual acoustics with data-agnostic model-based sonification [38].

DTM emphasizes the "non-musical" origins of input data – Music technologists may argue that most, if not all, generative and interactive musical systems today may be regarded as "data-driven" or even a sonification. Such a perspective may be attributed to the abstract nature of mapping, as digital technologies allow us to flexibly rewire the control data of digital musical instruments unlike traditional acoustic instruments [111]. Also, many musical and even sonification applications opt to use human-controlled data sources, such as wearable sensors, with complex and nonlinear transfer functions to synthesize audio [27]. However, such gestural data are often generated with some level of musical intent and goals, sometimes with an immediate feedback loop to adjust the gestures [16], and without the interest of preserving the original state of the input data. Purely functional sonifications, on the other hand, focus on the representation of non-musical information by avoiding distortions with an added layer of aesthetic mappings [39].

While retaining the transparency of sonified data should not be ignored, DTM mainly focuses on novel ways of handling data that lack inherent musical qualities (e.g., periodic patterns in a gestural time span) or intents. As a design environment, it aims to facilitate connecting any data structure[19] to any aspects of musical expression. A musical sonification model created in such a manner, on the other hand, should have an adaptive quality capable of handling various input data structures.

5.1.3 Live Coding as a Design Methodology

DTM facilitates live coding for reconfiguring sonification models. Live coding in audio and musical contexts adds unique constraints for handling time-sensitive events [64]. They include "just-in-time" evaluation of events [77] and audio or musical state management. In a design exploration with DTM, when transitioning between sections or mappings of data, ensuring acoustic continuity without resetting the synthesis process is beneficial. Such transitions themselves may become part of a sonification piece (see sections 4.3 and 5.3). Live-codable systems also facilitate designing a deeper user interaction, providing end-users some level of a design agency. The Twitter sonification exemplified this well, where the user could customize instruments assigned to arbitrary keywords filtering a large number of Tweets. This facility was carried on to the newly proposed framework (see chapter 6).

In addition, live reconfigurations facilitate the exploration of adaptive designs for various non-musical data mapped to a musical structure. By "adaptive," DTM aims for non-disruptive handling of incoming (stream of) data of any shape and format [26], rather than a parameter estimation algorithm [52]. This challenge is well illustrated when handling unseen data, such as a real-time data stream from sensors (see section 4.3). A musical model may need to account for data with different ranges and value distribution, value dropouts or missing values, and the data

[19] A "data structure" here signifies not an abstract organizational format but the underlying characteristics such as predictable development over time, periodicity, and non-uniform distribution. These may be considered as the opposite of randomness.

rate in relation to the time-evolving musical structure. The following implementation examples illustrate how DTM handles and transforms data to overcome these challenges.

5.1.4 Symbolic Structure Modeling

```
var articulations = ['rest','staccato','half','tenuto','slur']
var durations = [0,0.1,0.5,0.9,1]
var initLevel = [0,1,0,0,0]

model.articulation = function (datum) {
  var n = datum.range(0,5).round()
  var init = initLevel[n]
  var dur = durations[n]
  var atk = atkSlope[n]

  // Note division may be determined with another data modulation.
  var ampEnv = dtm.data(init,1-init).line(div.recip().mult(dur)).etc()
  model.amp(ampEnv)
}
instr.articulations(datum).pitchRange(...)
```

Code 5.1 A snippet of a symbolically modeled instrument with the articulation parameter.

For the styles of music, DTM initially focused on modeling symbolic expressions [101]. The first project with DTM, the civic sonification, incorporated simple but contrasting types of voice organizations such as harmonized pads (the sensor dashboard), percussions (Twitter dashboard), and metered melodies (Decatur sonification). These musical styles typically required one-to-many or nonlinear mapping schemes, connecting flat continuous data to a discrete or hierarchical structure. The general approach was to (1) utilize multiple layers (time frames) of data aggregation to control the rate of changes in sound, and (2) modeling the sound generators at a phrase level, with a high-level parameter / construct that describes the modulation of symbolic elements, such as phrase dynamics and note articulation. Code 5.1 illustrates a simplified process of symbolic modeling, utilizing quantization and multiple dimensions of a note event.

Table 5.1 The primary list analysis and transform functions in DTM.

Category	Examples	Descriptions
Interpolation	line, step, cubic	Fits to or stretches by a target size factor.
Scaling	range, expcurve, limit	Controls the range and shape.
Modulation	amp, freq, phase	Applies synthesis-inspired modulations to data.
Distribution	dist, CDF, entropy	Analyzes and manipulates the value distributions
Conversions	miditofreq, beatstointervals	Converts between musical (or non-musical) units.

For facilitating the modeling of high-level musical expressions, DTM also provided a large collection of data transformation tools inspired by such as Elsea's LObjects[20] and Spiegel's Music Mouse [95]. They are essentially a collection of generic or unique list operations, applicable to both data and musical sequences (see Table 5.1).

Categorical and ordinal data typically need to be quantified to numeric measurements before applying other transformations. While lacking elaborate multi-attribute analytical functionalities used in later frameworks (see sections 5.3 and 6.2), DTM offers the histogram and distribution-related measurements, also utilized in the structural analysis of the following framework (see SPE). Non-time-series data, on the other hand, often need to be ordered or mapped to time by a sorting technique, including the frequency in the distribution. Then, the ordered attribute is used as a time-based query call, as described in the next section.

[20] http://peterelsea.com/lobjects.html

5.1.5 Clock and Phase

```
// Load relational data (table).
var data = await dtm.csv('sample.csv')

// Column 1 values normalized to a MIDI pitch range.
var pitch = data('attr1').scale(60,100)

// Column 2 values rescaled exponentially to 1, 1/2, 1/4, etc.
var rhythm = data('attr2').scale(1,8).round().powof(2).reciprocal()

// Pre-generate 5 voices.
var voices = dtm.range(5).scale(0,1).each(v => dtm.music().pan(v).play())

// At every onset interval defined by `rhythm`, modulate the pitch (note) of
all the voices.
dtm.music().interval(rhythm).every(e => {
  var query = dtm.range(voices.len).add(e.index)
  pitch(query).each((v,i) => voices(i).note(v))
}).start()
```

Code 5.2 A nested clock and query structure.

A large part of technical effort went into creating a real-time clock (a cyclic event scheduler) that would act as an interface between offline data and real-time audio events. Offline and real-time events are organized into two data structures: dtm.data and dtm.music. The latter serves either as a synthesizer or a clock with no sound. It may be nested in multiple ways to realize a complex time structure. For example, a dtm.music can act as a data scanner to incrementally read a section of a dtm.data, while its rhythms were modulated by another dtm.data (Code 5.2).

```
voice = dtm.music().play()

dtm.music().phase(p => {
  voice.freq(data.phase(p).range(500,1000)
}).scan().for(10)
```

Code 5.3 Interpolating a value with instantaneous phase values of the clock.

dtm.music as a clock creates callbacks at multiple time events such as voice onsets, offsets, and the phase. The phase signifies the time position during a voice is active, linearly moving from the value of 0 to 1. Correspondingly, dtm.data also features an array-item accessor named "phase," which linearly interpolates values, or vectors for matrices, at a fractional index. Combining these phase callback and accessor enables the basic form of continuous data scanning (Code 5.3).

```
var data = dtm.data('PeriodicTable.csv')
var atomicWeight = data.col('AtomicWeight').range(0,1)
data.col('MeltingPoint').phase(atomicWeight)
```

Code 5.4 A data attribute used to query (estimate) another attribute’s values.

While the assumption of linearity between discrete data points may be naive, this simplistic approach enables creative and complex ways of scanning and exploring data. For example, one might use an audio-rate sawtooth wave to create a type of audification or use another data dimension as a data reader (i.e., read indices; Code 5.4).

5.1.6 Limitations: Understandable Musical Sonification

As an informal assessment, the musical and interactive potentials of DTM were presented in public performances including the Civic sonification and a real-time remote collaboration piece showcased in Audio Mostly 2017[21]. The reconfigurable design techniques using a live coding practice were also demonstrated in workshops at ICAD 2017[22] and Moogfest 2017[23].

Regarding the present thesis objectives such as realizing a measurable and variable sound structure, DTM has several technical shortcomings. First, with the chained transformations from a data object to a musical event, the resulting sound representation is often far detached from the acoustic or perceptual measurability of data because of scaling, quantization, and the one-to-many

[21] http://annaxambo.me/music/group-performances/transmusicking-i-audiomostly-2017/
[22] http://icad.org/icad2017/program-2/workshops.html
[23] https://moogfest2017.sched.com/event/AGbC

relationships between a data point and multiple nested time scales. In addition, the intermediate data structures during transformation, e.g., state parameters and time scales, are not generalized across different models, hindering the use of varying musical structures that may be compared to each other or summarized in the same manner. Second, DTM heavily utilizes the classic unit-generator paradigm for the scheduling and modification of sound in real time with the parameters abstracting away the structural details. This also results in more difficulties in comparing or interchanging different sound-organization techniques, especially when accommodating different data structures. The following frameworks depart from this module-based approach and provide more access to non-parametric aspects of sound design.

5.2 Spectral Parameter Encoding

The spectral parameter encoding (SPE) is a design framework for creating acoustically measurable sonifications with flexibly designable aesthetic characteristics.

5.2.1 Functional-Aesthetic Values in SPE

One of the ultimate goals of sonification may be to convey encoded information as accurately as possible. Musical sonifications are, however, often susceptible to the loss or distortion of information [113]. The loss may occur at various stages in sound creation or reception, including the auditory perception, the listener's familiarity with the sound structure, and the mapping and transformation processes of data to sound. SPE focuses on the last, aiming to preserve the data quantities in musical expressions using acoustic measurements. As it tries to minimize the loss of information, it also examines the ways of increasing the dynamic range of musical expressions through structural analyses and mapping of data. This framework was the first step towards a composable and decomposable timbral sonification.

SPE is implemented with DTM in combination with Max/MSP. DTM handles the symbolic organizations of additive or spectral synthesis to a limited capacity of the single-threaded processing in web browsers, while Max analyzes the spectral features, using the Zsa audio descriptor library [60], to verify the encoded data. The timbral models using spectral filtering are also handled in Max with the difficulty in DTM employing STFT with overlapped frames.

5.2.2 Methodologies

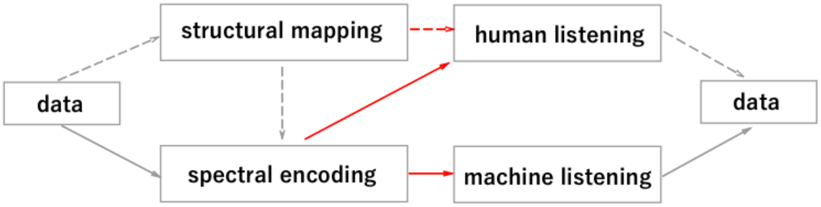


Figure 5.1 The overview of SPE. The red lines indicate acoustic signals, while the black lines are offline data. The dotted connections are optional.

The workflow of SPE consists of five general steps: data input, structural analyses of the data, spectral encoding with a selection of musical mapping, audio output, and data recovery. The input data are limited to, or handled as, a one-dimensional array of numeric values with other contextual information being detached – only focused on the physical quantities. As shown in Figure 5.1, some paths are optional and the whole process could consist of only the data input (no analysis), spectral encoding, and the acoustic recovery of data. The structural analysis and mapping provide additional musical organization to the spectral encoding process. Both the mapping and encoding stages employ the technique of dividing the signal into two distinct components, the structured signal and random residual components, with somewhat different implications for data-analysis and sound composition or analysis stages.

5.2.3 Structural Analysis and Mapping

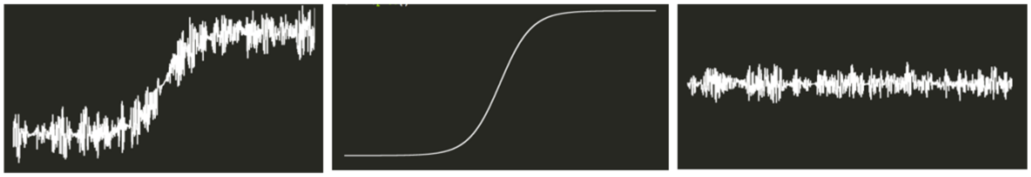


Figure 5.2 An example decomposition of a one-dimensional structure. Left: input; Center: estimated structure; Right: residual signal.

The input data for functional sonifications may often lack an innate "musicality" such as periodicity or gestural slow / rapid changes. To elaborate, such a structure or characteristic "signal" may sometimes exist but are hidden under non-structure, not readily mappable to common musical structures such as the amplitude envelope of a note (see Figure 5.2). In other cases, a signal / structure of data may be visible but not in a gestural time scale, such as the slow rise and fall of the temperature in daily weather records, and requires an arbitrary amount of time compression to be audible.

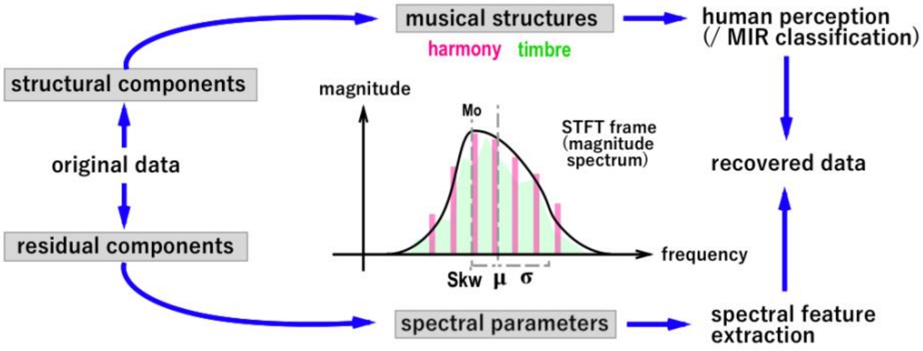


Figure 5.3 A simplified view of structural mapping in SPE.

SPE aims to engage these types of non-musical but potentially structured data, analyze and repurpose the inherent structured and non-structured components of data for musical expressiveness (Figure 5.3). SPE employs the additive-error model, a common data modeling technique used in compression, audio and vocal synthesis, and statistical signal processing. The analysis process separates the input signal (data) into rough structural and residual components, such that

$$x(t) = f(t) + \mathcal{E}, \tag{1}$$

where x is the input signal, f is the structural component, t is the time index, and $\mathcal{E}$ is the non-structured component. In a data-compression or audio-synthesis context, the aim of a structural decomposition would usually be the reduction of complex data into more concise parametric representations. The compact structural part may be further coded or transformed, while the residual part is retained at every data point but in a narrower and more stationary dynamic range than the original form. For the purpose of mapping to musical structures, instead of size reduction, the decomposition of data allows us a more flexible mapping of non-musical data to appropriate musical structures without losing the information as a whole.

To illustrate the structural mapping process in a very simple sonification problem, consider a single-dimensional time series with slowly increasing central values with somewhat stationary noise deviating around them (Figure 5.2), which may be expressed as

$$x(t) = \frac{1}{1 + e^{-t}} + \varepsilon. \tag{2}$$

Mapping the original signal x to, for example, the volume of an oscillator has the potential of impeding the musical balance or perceptibility as the slow increase of the volume may be hard to hear in the beginning, and it does not take advantage of the dynamic range of human hearing at any given moment. Similarly, if x is mapped to the frequency of an oscillator with a fixed amplitude, although it may represent the data faithfully, it may also produce a sense of "unstable" pitch slowly evolving that does not reside well in more complex sonifications where multiple sound dimensions are presented. The extraction of a larger non-stationary envelope allows repurposing of such non-musical data sources by, for instance, mapping the residue to a full range of amplitude to hear the detailed fluctuations while the slow-moving central values could be assigned to the frequency to produce more stable and "organized" pitch-sweeping gesture.

With unordered data, the focus of analysis typically shifts to, for example, clustering, cross-correlation, or observing the distribution. SPE utilizes the shapes of distribution for a musical organization. For example, if a given distribution does not fit to a Gaussian function, SPE may instead parameterize it into line segments with a fewer number of breakpoints. Such an envelope can then generate a percussive or metallic timbre with sinusoidal oscillators with the frequency randomly sampled from the parameterized distribution, as described in the next section, while using the residual signal for amplitude modulation.

Another approach for utilizing the value distribution rather imposes an existing musical structure, such as a musical scale that introduces the minimum amount of distortion, to transform the data non-linearly while retaining the residual values for additional parameter encoding. For example, the author modeled the algorithm for selecting the best musical scale by computing the signal-to-noise ratio after quantizing the data points with the common musical scales in all transpositions. This example may provide improved musical perceptibility but does not assure a spectrum-based data recovery. For that, the quantization residual signal may be mapped to, for example, a parameter of the magnitude spectrum.

5.2.4 Timbral and Harmonic Mappings

SPE considers two groups of sound organization: timbral (subsymbolic) and harmonic (symbolic). The timbral organizations employ a creative application of the classic source-filter model in the frequency domain. It shapes a dense and uniformly distributed source spectrum (e.g., a white noise or flat spectral peaks with linear / random phase) with a spectral envelope. The resulting distribution is played back in various ways over time such as crossfading between multiple distributions or random sampling. As the thesis work expands this timbral approach in much greater detail (see chapter 6), this chapter mainly introduces the harmonic-mapping techniques.

The harmonic organization, a bottom-up generative approach, structures time-domain sinusoidal voices under certain statistical constraints. The main idea is to add or redistribute templates of sinusoidal components, such as harmonic voices and chords while satisfying simple statistical moments of a magnitude-frequency spectrum. The statistics include the spectral mean (centroid) and spread (standard deviation). The statistical parameters are determined by scaling the input data (numeric) of one or two dimensions.

For instance, a spectral centroid value set by a data point can be "expanded" into a single note with N harmonic partials with a fixed unit amplitude, such that

$$\begin{gathered} n = 1, 2, \ldots, N;\ N \in \mathbb{Z}, \\ v_n = \frac{nN\mu}{N!}, \end{gathered} \tag{3}$$

where v is the vector of frequencies for sinusoidal additive synthesis, μ is the spectral centroid in Hz, and N is the number of the harmonics. We can also generate any pitch by adjusting the amplitude and number of overtones and undertones accordingly. The following example computes a single tone conforming to a given spectral centroid and spectral spread values. For an odd number of natural harmonics with an identical gain for non-central oscillators,

$$for\ N \in \{1, 3, 5, \ldots\}, \tag{4}$$

$$g = \frac{\sigma^2}{\sum(v_n - \mu)^2 + \sigma^2(1 - N),}$$

where a is the gain coefficients with a symmetric form $\{\dots, g, g, 1, g, g, \dots\}$ and σ is the square root of the spectral spread (i.e., standard deviation). Similarly, we can construct an arbitrary harmony with sinusoidal oscillators centered around a given spectral centroid, such that

$$v_n = \sum_{n=0}^{N-1} \frac{\mu C_n}{\bar{C}}, \tag{5}$$

where C is a vector of normalized frequency coefficients in Hz for creating a chord 2 and C is the average frequency of them. Combining both additive synthesis and harmony, it is also feasible to generate any chord on any root note from given spectral parameters. The flexibility in generating the pitch or the chord quality can be utilized to encode additional data dimensions for human perception. These harmonic techniques are relatively robust for retrieving the spectral parameters, provided that the data mapped to the spectral centroid is scaled properly so that the harmonic or chord-voice frequencies exist within the spread range.

5.2.5 Limitations and Future Work

With many mapping options for the data structure and residue with spectral, time-domain, and sometimes symbolic parameters, SPE lacks a uniform evaluation scheme with computational or computer + human listeners. Simple distortion metrics, such as mean-square error, between the input data and the spectral feature employed in a sound organization provide relative performance comparisons among mapping combinations. In informal evaluations, however, the author observed a somewhat complex and unpredictable degradation of measurement with the choice of algorithms, frequency range, and additional audio processing such as reverberation. The present thesis organizes these complex variables to perform a more formal evaluation with computational and human listeners.

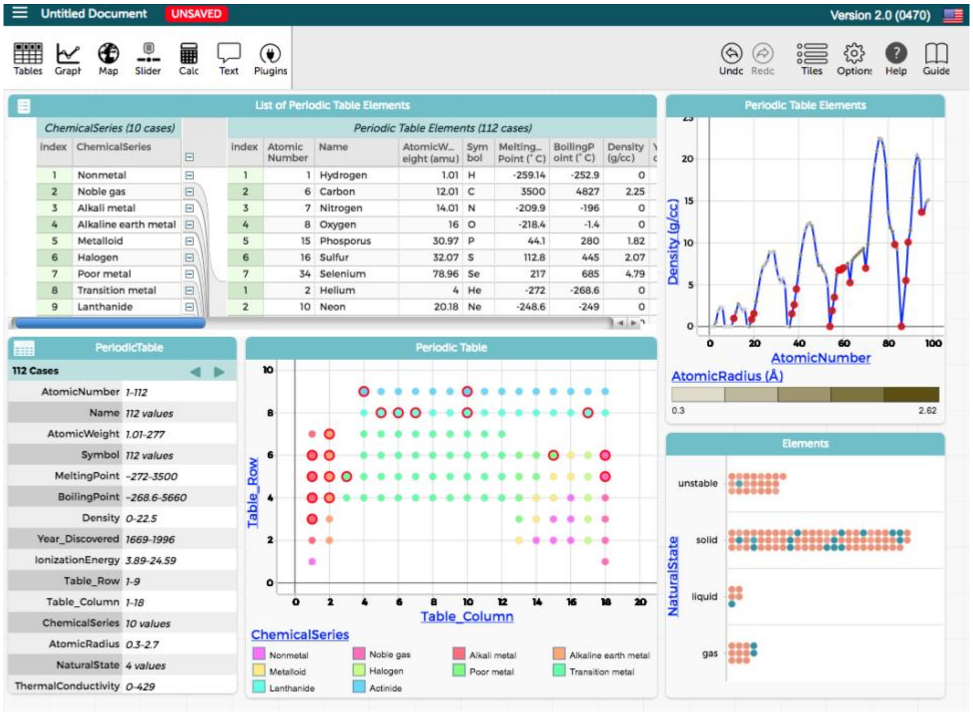
With the priority of retaining the acoustic measurability of spectral features in a stable manner, SPE is limited in creating time-evolving expressions. Continuously modulating a spectral or time-domain parameter would often alter the spectral content in an unpredictable way. To mitigate this, a sound object representing a data point would have a static magnitude distribution and either linear or random phase effects.

5.3 CODAP and Musical Data Moves

SPE, while focused on the physical measurability as well as musical design opportunities, also left a major concern unaddressed: Would a computationally measurable sonification also be perceptually understandable? If so, would the listener be able to know how much of the sound material came from the designer's aesthetics, and how much came from the data? The author explored these questions during an NSF-funded internship[24] at the Concord Consortium[25], a non-profit research group specialized in K-12 and college-level education of general and data science.

[24] This industry internship program was supported by the National Science Foundation under Grant No. 1417835.

[25] https://concord.org/



ChemicalSeries (10 cases)

Index	ChemicalSeries
1	Nonmetal
2	Noble gas
3	Alkali metal
4	Alkaline earth metal
5	Metalloid
6	Halogen
7	Poor metal
8	Transition metal
9	Lanthanide

Periodic Table Elements (112 cases)

Index	Atomic Number	Name	AtomicW... eight (amu)	Sym bol	Melting... Point (°C)	BoilingP oint (°C)	Density (g/cc)
1	1	Hydrogen	1.01	H	-259.14	-252.9	0
2	6	Carbon	12.01	C	3500	4827	2.25
3	7	Nitrogen	14.01	N	-209.9	-196	0
4	8	Oxygen	16	O	-218.4	-1.4	0
5	15	Phosporus	30.97	P	44.1	280	1.82
6	16	Sulfur	32.07	S	112.8	445	2.07
7	34	Selenium	78.96	Se	217	685	4.79
1	2	Helium	4	He	-272	-268.6	0
2	10	Neon	20.18	Ne	-248.6	-249	0

112 Cases

AtomicNumber	*1-112*
Name	*112 values*
AtomicWeight	*1.01-277*
Symbol	*112 values*
MeltingPoint	*-272-3500*
BoilingPoint	*-268.6-5660*
Density	*0-22.5*
Year_Discovered	*1669-1996*
IonizationEnergy	*3.89-24.59*
Table_Row	*1-9*
Table_Column	*1-18*
ChemicalSeries	*10 values*
AtomicRadius	*0.3-2.7*
NaturalState	*4 values*
ThermalConductivity	*0-429*

Figure 5.4 An interactive exploratory analysis of data in CODAP.

The author developed a set of sonification tools for the Common Online Data Analysis Platform (CODAP) [28]. CODAP is a web application that facilitates interactive data inquiry for beginners through a novel graphical interface. In CODAP, the user engages various data utilizing a hierarchical spreadsheet, individual and aggregate data views, and graph widgets, requiring almost no programming (Figure 5.4).

5.3.1 Functional-Aesthetic Design Principles

Unlike SPE, the major goal here was to help the user of sonification (in CODAP, both the designer and listener) with not just perceptual but conceptual understanding of data with sound as well as the sound properties themselves. With multiple approaches, including designing an echolocation-themed data-science game and an exploratory sonification of the Lake Tahoe data[26], the authors sought possible intersections between data-science practices and musical sonifications.

As a prototype F/A framework, the authors proposed a model for 'composing' a musical sonification with data. Such composition is mainly to tell a story about data rather than exploring novel aesthetic possibilities. The piece should express how the analyst reached the sonic organization of the information through experiments and explorations. This framework postulates two principles: (1) When organizing the sounds, the designer should draw most of the sonic materials from either the data themselves or the analytical processes, but not from purely aesthetic decisions (e.g., applying a musical-scale quantization) or arbitrary interpretations. (2) To 'compose' a musical sonification piece, the designer should incorporate not only the final analysis result and mapping of the data but also the gradual progress of the data exploration.

The framework was named musical data moves (MDM) after an existing conceptual framework for data science education. In contrast to the previous sonification frameworks by the author, the sound-design process in MDM is largely data-dependent, where the analytical / compositional decisions are largely dictated by the data context.

[26] Exploratory analysis of Lake Tahoe data with sonification: https://vimeo.com/289564413

5.3.2 Data Moves

Table 5.2 The list of core and additional data moves.

Name	Type	Goals
Filtering	Core	Scoping and exploration
Grouping	Core	Compare subgroups
Summarizing	Core	Aggregate group features
Calculating	Core	Add new information from attributes
Merging	Core	Add information from other data sets
Making Hierarchy	Core	Provide alternative view
Temporary Attribute	Additional	Data preparation
Sorting	Additional	Data preparation
Stacking	Additional	Data preparation
Sampling, etc.	Additional	Simulation-based inference

This section briefly illustrates a common procedure of data exploration in CODAP. To make the examples concrete, the following discussions use a data set of the chemical elements periodic table, which consists of various numeric and categorical attributes for us to explore and analyze (Appendix A). Suppose the analyst aims to uncover correlations between various attribute pairs of the periodic table. The default flat spreadsheet, while perhaps intuitive to chemistry experts, may not reveal obvious patterns or provide analytical perspectives to many. A non-expert analyst needs to strategize how to find a more useful data representation for further analytical tasks. This type of challenge is even more pronounced with datasets in real life that are huge, unstructured, and messy (e.g., with missing values, inconsistent labeling, or added noise in measurements). Before employing any statistical or machine-learning techniques, analysts would use more fundamental techniques for engaging and navigating through "unprepared" data. One would reorganize data by filtering, grouping, summarizing, generating new conceptual attributes, and integrating multiple data contexts with the aid of responsive graph visualizations. Erickson et al. identify these data-navigation or reorganization techniques as the data moves [25]. For the direct application in sonification, the following discussion introduces two out of the six core data moves (see Table 5.2): filtering and grouping.

The filtering moves either highlight or hide data points to focus our attention on reduced and simplified data. In CODAP, filtering can be done by selecting data points in the spreadsheet or graph, either manually or by utilizing group structures. A selection visually highlights the table and graph items, while also triggering some of the sonification sounds. The user can further filter the view by temporarily hiding the highlighted or non-highlighted items and apply more transformations.

The grouping moves bind multiple data points by common values. CODAP facilitates grouping by visually dragging an attribute (column) in the spreadsheet to a parent / child layer. Grouping creates a hierarchical structure by labeling each subgroup with a common numerical / categorical value (or even multiple joined attributes). With the periodic table, for example, we may group the elements by the table row (orbitals) or column (outer electrons) numbers, the natural states (gas, solid, etc.), the metallic properties, or any combination of them.

5.3.3 Musical Data Moves

Similar to the high-level goals of visualization [66][53], a sonification may be used to understand data (exploration) and to communicate the understanding (explanation) through sound. The musical data moves (MDM) extend the concept of data moves for both exploratory and explanatory sonification. In a data exploration, MDM utilizes the analyst’s interactions and data transformations (i.e., data moves) as the elements of sound generation. This process combines various sound generators with data moves to create sound snippets, representing moment-to-moment perspectives to data. Then, in the explanation phase, MDM reorganizes the data moves in order to sequence the sound snippets, building a time-varying and multidimensional data soundscape. In short, MDM facilitates the generation, recording, and reorganization of the sound materials associated with data moves. The specific examples are, however, better explained in the context of an actual data analysis (see sections 5.3.5 and 5.3.6).

5.3.4 Sonification Plugins

Table 5.3 The list of CODAP sonification plugins.

Name	Type	Description
Simple Spectrum	Generator	Maps scatter plot to the pitch and magnitude of sinusoidal partials.
Micro Rhythm	Generator	Maps scatter plot to pitch and time offset of short grains.
Robot Voices	Generator	Generates timbral signature with inverse MFCC for each group.
Noise Color	Generator	Unpitched version of Robot Voices.
Glissando	Generator	Continuous transition of spectral partials upon remapping.
Sound Beacons	Generator	Plays and mixes sound-clip loops placed on a visual map
Csound Playground	Generator	Live-codable plugin prototyping system with Csound WASM and DTM.
Case Sequencer	Controller	Automatically selects cases or groups in order.
Data Moves	Controller	Records and plays back data-move operations.
Plugin Transport	Controller	Synchronizes multiple generators.
Plugin Mixer	Misc.	Records the output of the generators.

The author developed a series of sonification plugins (see Table 5.3) utilizing DTM and Csound WebAssembly module [115]. The sonification plugins implement the concepts of MDM where the sound reacts interactively to data moves such as filtering, grouping, and (re)mapping. The plugins attempt to capture both the resulting sounds and the source actions in a way that they can be reorganized. The plugins are categorized into sound generators and controllers.

The generators mostly facilitate sound organizations manipulating short-time sound spectra. For example, the Simple Spectrum plugin takes a two-dimensional graph as a static magnitude-frequency spectrum, where each data point represents a sinusoidal component. The

Micro Rhythm plugin adds time to Simple Spectrum as a mappable dimension, decomposing the static timbre of Spectrum into a time-evolving melody or texture similar to the granular synthesis [75].

The controller plugins are typically meta instruments for creating polyphonic and/or time-varying organizations as well as capturing user interactions and data moves. For example, the Transport plugin synchronizes the playback of multiple instances of monophonic generators. The Data-Moves Sequencer facilitates recording and replaying various data moves for musical reorganizations. The recorded data moves are somewhat more specific than the general definition: selecting, moving attribute positions (including hierarchical grouping), sorting, formula editing, and changing the parameter mappings with the graph and sonification plugins.

5.3.5 Phase 1: Exploring Data and Collecting Sonic Materials

Using the aforementioned periodic-table data, this section illustrates an example of MDM for the first exploratory phase. To simplify the descriptions, it focuses on only several plugins and data attributes among many others. The analysis begins with the Simple Spectrum plugin with only the atomic density attribute mapped to the pitch and with the constant loudness[27].

For an analyst, the primary goal here is to recognize, memorize, and understand various sonic patterns manifested by the data. Can the listener (analyst), for example, hear and understand the global shape of the density distribution? Simple Spectrum creates a static but complex sound texture representing the entire data context by selecting every data point in the table. This may sound loud, inharmonic, and noisy. As the sound texture is quite complex, the analyst may instead select individual points to hear where each element / density values are positioned in the spectrum,

[27] MDM demonstration video 1: https://vimeo.com/392067108

and how they contribute to the collective texture. Selecting several points exhibits the intervallic relationships between the data points, where clustered values create phase-canceling 'beating' effects. Rather than manually selecting data points, however, the analyst may also want to compare the density distributions between different categories of elements. Grouping by the natural states reveals sparser and more perceptible sounds for gas, liquid, and unstable elements, which the listener can compare back and forth to identify the ranges and other statistics. With these real-time filtering-based interactions and timbral experiments, they gradually learn the relationships between sounds and data.

From the composer's standpoint, the primary interests may be in creating and filling the sonic palette with various representations of the data, while adhering to the rule of not adding materials unrelated to the data or distorting the original data for purely musical benefits. For example, the composer (= analyst) might find that the previous grouping of the chemical elements by the natural state left the sound of "solid" atoms still too chaotic and complex. They may further decompose it by, e.g., subgrouping with the metallic properties, which yields a slightly less-complex texture (for solid-metal) and two sparse and harmonic sounds (for solid-metalloid and solid-non-metal). Each sound color in the palette, varying in intensity, harmonicity, and the amount of contrast with others, represents a unique configuration of data and the history of data moves taken to reach there.

Filtering and grouping generally reduce the complexity of sound and data. However, the composer may also be interested in creating a perceptually complex soundscape of data, layering multiple sound textures in a polyphonic manner. The CODAP sonification facilitates such organizations with the time synchronization with the Transport plugin, as well as with the category-based spectra generation with the Robot Voices plugin[28].

5.3.6 Phase 2: Rearranging Moves for Storytelling

After the exploratory analyses, the analyst / composer's sonic canvas may be filled with timbral snippets representing various perspectives of data with sound structures from simple to complex. Now, the main focus of MDM shifts to the end listener. How would a new listener be able to learn about the data with sound, just as the analyst / composer did in the exploration?

A musical sonification about data may benefit from being self-explanatory, with the likely absence of a prior model for the listener to understand the meanings of sound. This, however, signifies a fundamental issue in auditory interfaces; an abstract (synthetic) sound has difficulties conveying "designative" meanings [65]. Such meanings may include the data-to-sound mappings, units, and referential axes, which would be highly difficult to convey without an accompanying visualization or a verbal explanation. At the same time, a spoken description might not be efficient for explaining a musically complex sound organization (e.g., suppose describing the sound of dynamically regrouped and filtered data over time).

While completely avoiding the use of visuals or words is not practical [113], with musical storytelling, the author focuses on the aspect of the "embodied," or internal, meanings of sound [65]: the unique structural relationships among various sonic patterns in a specific sonification / data. This relational sense-making differs from having external structural references such as musical scales or semantic associations (used, e.g., for 'earcon' sonification [36]) for understanding data. Murail, a leading spectralist composer, argued that the absence of absolute

[28] MDM demonstration video 2: https://vimeo.com/392067234

reference points with timbral (electroacoustic) music would prompt us to focus on listening to the changes and relationships among sound elements [67]. In the preceding data exploration, the sonification started with a plain and simple mapping between data and sound, but gradually evolved towards multiple and distinct auditory patterns. In the beginning, it might be feasible for the composer to verbally and/or visually explain the designative meanings behind the sounds. However, this may not scale with diverging and evolving sound snippets that represent multidimensional or transformed data.

To create a coherent and possibly self-explanatory sonification, MDM reorganizes the data moves employed in the exploratory phase. In CODAP, the Data-Moves Sequencer captures the history of a data exploration (e.g., filtering, mapping, formula editing, etc.) and enables an on-demand playback of points from the history. Recording the history of data moves allows the analyst / composer to rearrange and connect the storytelling components into a methodical but time-evolving musical piece, through which the end listener may experience the process of exploration. In addition, as a fixed sonification piece may not afford the listener learning by physical interactions, re-experiencing the composer's explorations may be an alternative for that.

Effective visual storytelling with data often follows a logical flow of establishing a context, guiding the listener's attention to various aspects of the data, and highlighting the main message or findings [53]. Similarly, a musical sonification may guide the listener's learning and exploration by utilizing various electroacoustic compositional concepts such as repetitions, variations, controlled intensities, and continuous transitions [74][92]. The following illustrates each of these sound (re)organization approaches[29].

[29] MDM demonstration video 3: https://vimeo.com/392068275

Repetition is a simple but effective technique for introducing and communicating the notable patterns in sonification. For the first-time listener, it typically requires a considerable effort to familiarize themselves with, memorize, and be able to compare the details of multiple sonic patterns. Deliberate repetition of a few alternating patterns may ease this problem, while also introducing a forward flow of musical expectations with repetitions and variations [46].

Variations in music typically occur as part of a repetition. With MDM, variations can be created by the arrangement of parameter mapping pivoted by a fixed parameter. A Micro Rhythm may, for instance, use a single time attribute (e.g., density) with alternating pitch attributes (e.g., melting and boiling temperatures) that creates a slight variation of the pitches with an identical rhythm. This enables the comparison of three or more dimensions of data in a coherent musical sequence.

The controlled intensity aids the listener's understanding of how a single timbre is formed over time, while also creating a small-scale narrative. Even a single evolving sound object can tell a story of exploration – as Roads observes the role of individual sound in composition, that "to serve a structural function, an element either increases intensity, decreases intensity, or stays the same. Processes such as increasing intensity (heightening tension) and decreasing intensity (diminishing tension) are narrative functions [74]." In MDM, the control of intensity works similarly to the interactive filtering but instead is intentionally composed based on the data-moves history. For example, a spectrum representing the data points may be gradually built up with random sampling (instead of manual selections) or reduced into summarized values (e.g., bins), creating a continuous change of timbral intensity.

The transitions from one state in the data-move history to another can be discrete for the purpose of comparisons but may also benefit from continuous expressions of spectral musical transformations. For example, the composer may want to track how the raw data points of densities converge to the summarized values (e.g., the mean) of subgroups. Such continuation can be realized with interpolation between the states A and B, as all the data points move on the same plane of measurement. The composer may also be interested in a diverging expression, such as from a summary to the raw values, as spectral convergence and divergence are both important techniques in timbre-based compositions [92].

5.3.7 Towards a New Framework Design

In summary, MDM models a design process of a F/A sonification incorporating exploratory data-science and timbral sound organization techniques. The examples only touched upon a fraction of possible sound organizations with dynamic spectra [75]. This aesthetic possibility is, however, uniquely confined by data and interactivity. MDM maintains that, to provide sense-making values for data science activities, a musical sonification should deliberately display how the sounds have been designed and evolved. It also relies mainly on the analyst / designer's interactive explorations in order to form a time-varying sound structure, rather than the compositional intent of the designer. The resulting piece, while being non-interactive when listening, embodies the interactive processes of exploratory analysis and sound design / mapping taken by the designer.

The present study contrasts with MDM in terms of where aesthetic values are sourced from. MDM, as well as SPE, are concerned with how we could repurpose the unique structure of data, and an exploration specific to the data context, to the compositional strategies of a musical sonification. The present spectromorphing sonification (SMSon), on the other hand, allows the incorporation of an arbitrary control and mixture of purely musical vs. data-driven physical values to colorize a measurable and identifiable sound.

SMSon also disregards the compositional rules of MDM for data storytelling (section 5.3.1). These rules assume the inherent musicality of the data context or the CODAP plugins, and prohibit the use of external aesthetics such as the innovative ways spectralist composers exploit an existing audio recording.

In terms of the range of expressions, the CODAP / MDM plugin designs are somewhat limited compared to SPE and the present synthetic model (Sonar; see chapter 6.2). While CODAP's spectral synthesis plugins share the same basic principles of SPE with the acoustic features closely being tied to the descriptive methods in audio analysis, their main goal is in establishing basic and easily explainable sound expressions for non-musician listeners.

The focus of the present study may be comparable to the development of a single timbral-sonification plugin for CODAP with a vastly increased range and complexity of expression, whereas MDM is more about how to employ and combine multiple of them in a larger time scale.

CHAPTER 6. NEW DESIGN FRAMEWORK: SMSON AND SONAR

This chapter provides a closer look into the new sonification design framework, spectromorphing sonification (SMSon), along with a programming library, Sonar.

6.1 Spectromorphing Sonification

SMSon models the organization of musically complex and bespoke timbral expressions that can also convey external non-musical quantities (i.e., data). It is inspired primarily by Smalley's spectromorphology [92], an analytical theory for electroacoustic compositions, as well as time scales (TS) and timbral dimensions (TD) as the core organizational elements of spectral audio synthesis.

6.1.1 The Integration of Functional-Aesthetic Principles

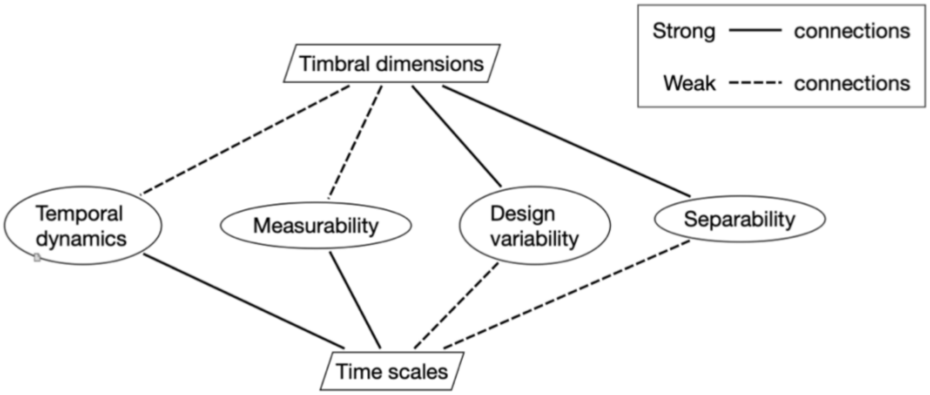


Figure 6.1 The strong and weak relationships among the F/A values, timbral dimensions, and time scales as implemented in SMSon.

Chapter 3.5.1 proposed four functional-aesthetic (F/A) principles: the measurability of data, separable multidimensional timbre, temporal dynamics of spectra, and the variability of design. In order to apply all the four F/A values in a sonification design, SMSon utilizes the properties of synthetic timbral dimensions and how they interact with different time scales. These structural elements and the F/A principles are connected in many-to-many relationships, some with stronger and some with weaker implications (see Figure 6.1).

First, the organization of time scales mainly concerns the measurability and temporal dynamics of a sound. The difficulty of establishing both a continuously evolving timbre and computational measurability was a major limitation in SPE (see chapter 5.2). The time scales are defined to organize the synthesis structure as well as independent motions of sound parameters for perceptual and computational analyses.

A separable and variable design of timbre is approached by defining four synthetic dimensions, each with a corresponding category of generative models. These dimensions are spectral peaks, envelope, effect, and amplitude envelope. Models for each dimension may be combined and modulated simultaneously and continuously over time by the designer. While these dimensions are utilized for individually controlling synthesis and approximating sound complexities, the user studies examine if they also directly or indirectly align with perceptual experiences for the listener and designer.

6.1.2 Structural Attribute: Time Scales

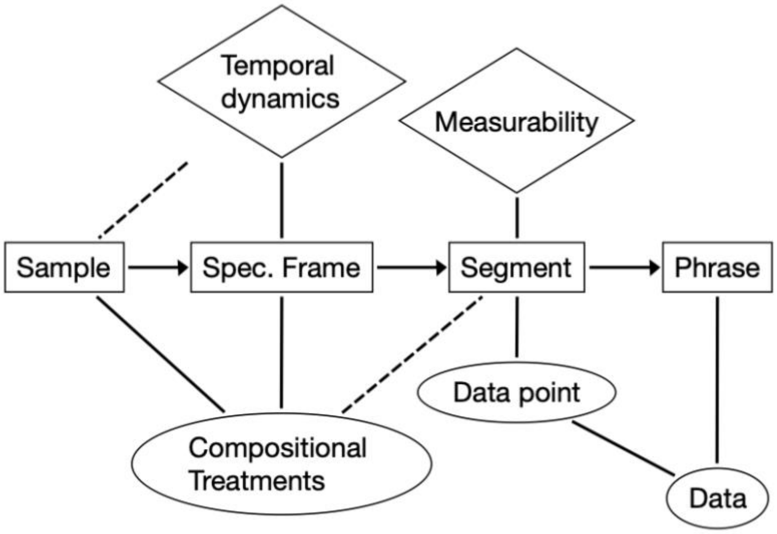


Figure 6.2 An overview of the time-scales organization. Solid lines indicate strong dependencies, while dotted lines imply possible / weak influences. The arrows indicate the increase of structural time scales (represented with rectangular boxes). The oval items denote variable factors. The diamonds are the F/A values.

Recall the ‘structural’ and ‘durational’ time scales explored in the melodic-sonification user study (see chapter 4.4.1). Structural TS are musical interpretations of absolute time spans inspired by the extensive discussions by Roads [76], which define hierarchical time measurements contributing to the sensation of pitch, rhythm, a micro grain of timbre, etc. On the other hand, durational TS [55][54] are more closely related to the listener’s memory and cognition, connecting the musical information (e.g., the form, melodic patterns and evolutions, timbral dynamics, etc.) of the current moment, from the beginning of the piece / sound, and from the past iterations of listening. While the listening test (see chapter 7) is examined with regard to both durational and structural TS, the design process below focuses mostly on structural TS.

SMSon formalizes the organization and parameterization of time-evolving sounds using several distinct levels (Figure 6.2). The most important may be at an intermediate TS – the duration associated with a data point / quantity. Given a sequence of one-dimensional numeric values (regardless of them being a time series or not), each data point is assigned to a segment of a sound object. The duration of the segment, determined by the designer, may correspond to a typical musical note or a short sound object [76] spanning, e.g., anywhere from 0.1 to 1 second. The segment duration is then quantized to the nearest integer multiple of the short-time analysis hop size with the sampling rate at 44100 Hz, the block size of 4096 samples, and the hop size of 1024 samples.

The data value within a segment is static, setting the timbre generation algorithm to a certain fixed state. We may call this the “primary” modulation by data. Within the sound segment, however, additional timbral embellishments may take place at a faster rate of every 1024 samples (~ 0.023 seconds). This compositional treatment is to animate and characterize the sound object and may be considered an “auxiliary” modulation. An auxiliary timbral modulation may occur once or repeat multiple times in each segment, to the extent while it retains the perceptual and computational measurability of the spectral features. The modulations typically control a parameter of a spectral envelope, convolved with the spectral peaks of every block.

Table 6.1 Examples of auxiliary aesthetic modulation patterns.

Name	Type	Description
Constant	Lossy	Sets a static modulation value.
Linear ramp and triangle wave	Zero-mean, uniform	Sweeps through the parameter range N times in upward or downward direction
Exponential ramp	Skewed	Provides a sharper attack, slower rise, etc.
Sine wave	Zero-mean, skewed	Typical musical LFO but skewed towards the margins of the parameter.
Modulated sin	Zero-mean, less skewed	The modulation range is gravitated toward the center.
Square wave	Zero-mean, lossy	Alternates between two values N times.
Random steps	Zero-mean, uniform	Randomly sample N parameter values from a uniform distribution.
Random slopes	Zero-mean, uniform	Randomly samples and interpolates between steps.

An auxiliary (non-data) modulation is constrained to particular shapes / patterns with either zero-mean or a stationary distribution across every segment (see Table 6.1). These constraints are introduced to balance or potentially cancel out the effect of aux modulation over the period of a datum / sound segment, preserving the measurability of the primary (data) modulation. For example, if we use an aux modulation that completely filters out half of the spectrum, either the human or computer listener may lose the measurable changes introduced by data. With a zero-mean modulation, typically an odd or even oscillatory function (e.g., a sawtooth function) satisfies to retain the center position, despite potentially affecting the overall spectral distribution. A stationary distribution may be satisfied by using a shape based on exhaustive sampling, either ordered or random.

Further down the time scales, the frequency-domain representation and manipulation of the sound allow us to compose sample-level cyclic patterns. This is done with the placement and modulation of spectral peaks (the magnitude of individual frequency bins) and the phase linearity over half the block size (Nyquist frequency). The framework proposes several templates / models to generate simple and complex patterns in this domain (see the last section of this chapter).

Above the segment-level time scale resides a phrase or sequence of segments, comprising a longer sound object with a consistent and reused structure[30]. The simplest approach concatenates the same type of segments with block overlaps. The user experiments employ this methodology with seven data segments. A resulting sound object may have a duration anywhere between one to ten seconds.

A more elaborate organization, however, incorporates the concept of "sound-unit" structuring[31] [5], utilizing multiple types of synthetic algorithms in a sequence. A sound object or snippet may consist of three or four spectromorphing templates (models) concatenated or morphed from one to another. With the communication of quantity / data in sonification, this combination process may require a tight synchronization of the synthesis parameters. For example, to concatenate an attack transient with a noise content to a harmonic sustained segment, these sound units need to have the same spectral parameter (e.g., spread) for computational measurability.

While the thesis focuses on the study of timbre at the sound-object level, the ultimate interest of SMSon is to facilitate the composition of a complete electroacoustic piece employing multiple sound objects. The arrangement of sound objects extends horizontally in time scale, as well as vertically in polyphony. The user listening study examines perceived relationships among individual segments, the sequence of sound objects, and beyond (e.g., cultural associations). The designer study, on the other hand, compares multiple sound objects in a nonlinear order, with polyphonic implications observed in some responses.

[30] The "consistency" of sound may be attributed to perceptual sound integration [63].

[31] This term is introduced by Blackburn in a compositional context of spectromorphology, whereas Smalley uses "gesture unit" in analyses to denote sound events associated to a musical-instrument performance.

6.1.3 Structural Attribute: Timbral Dimensions

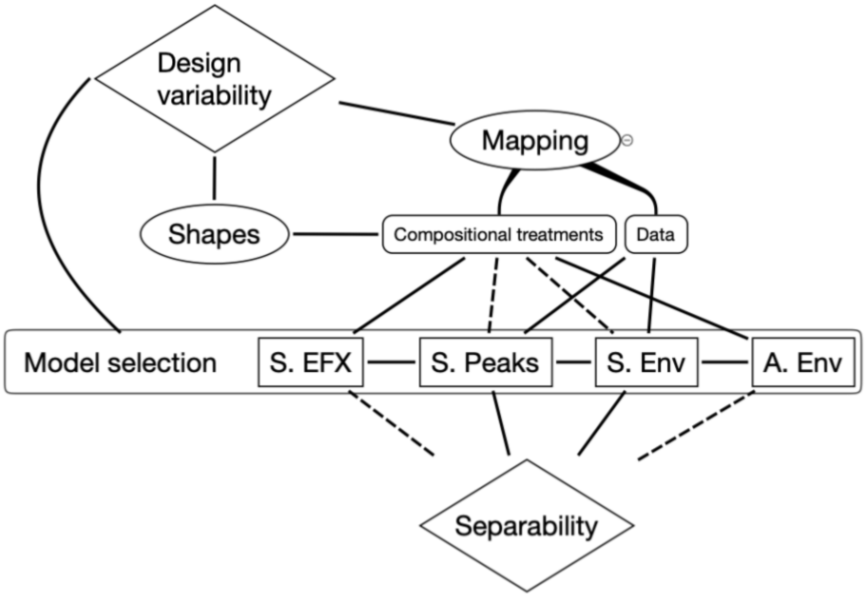


Figure 6.3 An overview of the timbral-dimension organization in SMSon. See Figure 6.2 for the definition of graphical notations.

With the lack of standard definitions for timbral definitions (see 3.4.1), SMSon proposes several synthetic dimensions for exploring the implications for perceptual timbral dimensions (Figure 6.3). The following theoretical perspective of timbral dimensions draws the concepts from signal processing and electroacoustic literature, focusing on the generative elements of time-evolving sound expressions.

Leveraging the general approaches for creating a computationally measurable sonification in SPE, SMSon adopts the classic source-filter model as the baseline of audio synthesis. SPE was limited to creating static or at most stationary time-varying expressions using a single magnitude spectrum. SMSon removes this constraint by providing continuously morphing generative models for synthesis and analysis. The models for organizing timbral dimensions are grouped into spectral peaks, spectral envelope, spectral effects, and amplitude envelope. As the names suggest, they are not categorized by perceptual dimensions in a similar way to pitch and loudness [62], but rather based on common methodologies in time-frequency analysis and synthesis [81][94]. In what ways these dimensions are mapped and modulated by the designer, in spectromorphological assessments, may be the main connections to the listener's experiences.

A bare-bone process of a single-frame spectral composition may be expressed as following,

$$y = STFT\{f([s.pks]) * [s.env]\} \cdot [a.env], \tag{1}$$

where f denotes a spectral-effect algorithm, $*$ and $\cdot$ each indicates an element-wise multiplication of vectors in the frequency and time domains, respectively. To animate the timbre, independent modulations and time progressions are applied to each component over multiple time frames.

The spectral-peaks models generate flat-magnitude spectra with various methods. The generated shapes may be broadly categorized into impulse trains and noises. The impulse configurations vary greatly for creating mixed harmonic and inharmonic flat spectra.

Spectral peaks also concern the phase information where the peak magnitude is non-zero. Typically, for any peaks algorithm, the starting phase (angle) of each bin is randomized to avoid the cumulative 0-phase magnitude peak in the beginning. With spectral peaks with a relatively static magnitude, for example, each phase value may be linearly incremented according to the current position of the sliding STFT window and bin number. This increment may however be modulated to create a frequency modulation, allowing a micro tuning of each sinusoidal component.

The spectral-envelope dimension filters the spectral peaks with convolution, i.e., the element-wise multiplication of the magnitude spectra. The envelope defines the general smooth shape of a magnitude spectrum. There are a number of ways to generate and animate such an envelope. The most generic approach uses parametric distribution models such as the log-normal function, while defining the way to approximate the parameter used for modulations, such as the spectral centroid. More creative or specialized approaches for audio processing may further transform the parametric models, extract spectral envelopes from existing audio sources, or use perceptual models such as Mel-frequency cepstrum coefficients (MFCC) to approximate an envelope [97].

Spectral effects modify the contents of spectral peaks, both magnitudes and phases, while retaining the spectral envelope mostly unmodified. This modulation is optional in the design process, and often more difficult to achieve perceptual linearity or salient changes to the timbre. While this decreases the opportunities for representing data with the primary modulation, it may be suitable for creating subtle "textural" expressions [92] with auxiliary compositional treatments. Electroacoustic composers have explored to create a large list of transforms potentially usable with the above constraints [24][114].

The amplitude envelope / modulation is also an optional high-level aspect of a sound object for adding a layer of temporal dynamics. The slow-moving envelope attenuates the overall magnitudes of short-time spectra, while avoiding altering the structure of the spectral peaks and envelope of each frame. With a time-evolving spectral content, the amplitude / loudness is not ideal for mapping quantitative information for either perceptual or computational decoding. However, it may contribute greatly to aesthetically characterizing the sound object[32]. While the spectral templates are designed and modulated at the frame level, amplitude modulations are organized mainly at the data-segment time scale. The modulation is applied either to each time-domain sample or to each spectral frame as a summary (single-value) modulation.

[32] For the user listening test, a stimulus with no amplitude modulation was commonly perceived as fatigue-inducing in a pilot study. The actual stimuli for the tests, therefore, employ simple amplitude modulations to reduce such annoyance.

6.1.4 Design Attributes: Spectromorphology and Aesthetic Practices

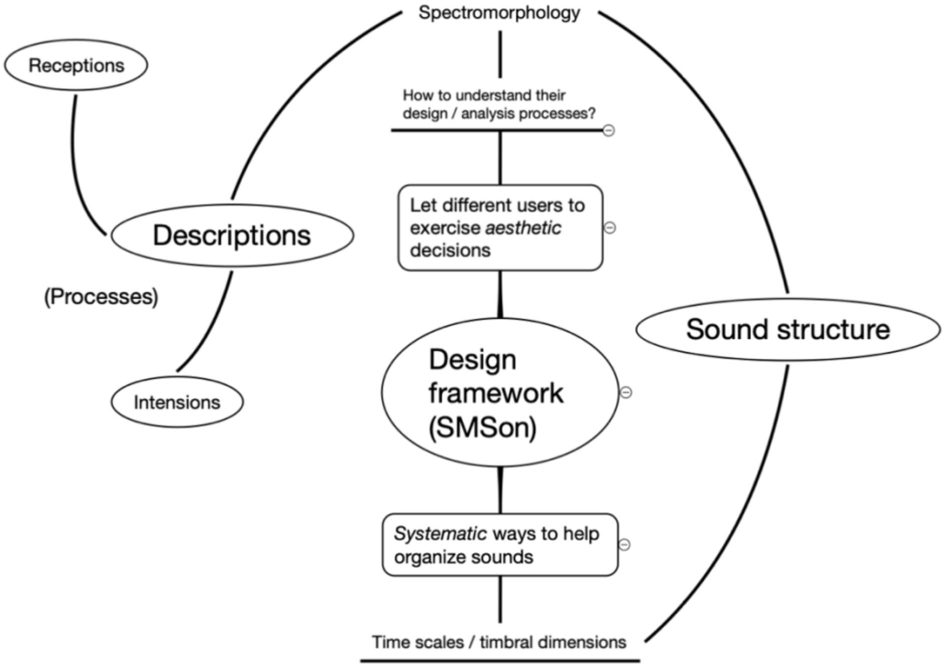

Spectromorphology
How to understand their design / analysis processes?
Receptions
Descriptions
(Processes)
Intensions
Let different users to exercise *aesthetic* decisions
Design framework (SMSon)
Sound structure
Systematic ways to help organize sounds
Time scales / timbral dimensions

Figure 6.4 The structural (bottom half) vs. aesthetic (top half) components of SMSon.

Chapter 3.4.2 has alluded to the difficulty of aligning the perspectives between the designer and the listener for an electroacoustic sonification. SMSon uses the theory of spectromorphology to potentially link (1) different personal descriptive approaches to a musically complex sound and (2) the sonic components of a generated sound back to the structural elements in SMSon (i.e., time scales and timbral dimensions; Figure 6.4).

Table 6.2 Summary descriptions of spectromorphology.

Organization	Description
Source bonding	Intrinsic or extrinsic reasoning of sound components.
Temporal phases	Archetypes of sound-object segments.
Gesture and Texture	Spectral motions over time in human-gesture or environmental (slow) time scale.
(Non-)Hierarchy	Multiple time scales of gestures and textures.
Structural functions	Analytical archetypes of sound event in variable time scales.
Behavioral references	The types of motion behavior over spectrum and time.
Types of spectrum	Pitched, harmonic, or noise characteristics in a time snapshot.

Spectromorphology is a body of analytical principles for general electroacoustic music developed by the musicologist and composer Denis Smalley [92]. From various analytical perspectives (see Table 6.2), Smalley defines a terminology about sound structures shaped over time, which are recognized and identified mainly with focused and repeated listening[33]. The following discussion applies several of his "sound-forming" principles[34] to the organizational and aesthetic processes in/with SMSon.

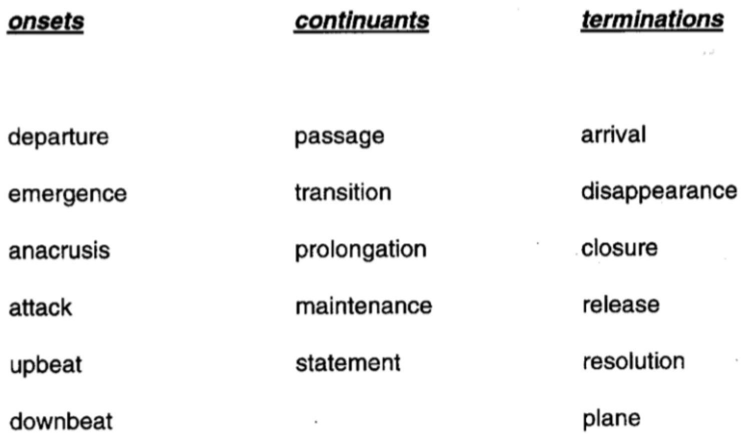
First, stepping back to the technical process of sound organization, creating a sonification snippet with SMSon roughly takes three steps: (1) Modeling the TD algorithms and TS modulation shapes. (2) Deciding the mappings, shapes, and models with aesthetic preferences. (3) Rendering the sound with the TD/TS pipeline.

The present study focuses on (2) and (3), as (1) is prefixed by the author (see section 6.3). The TD/TS organization in (3) shares some similarities to a generic view of vertical (spectrum) and horizontal (time progression) timbral layout [17]. Smalley, in addition, discusses the typology of motion behaviors, with various categories of spectral motions connecting one moment of time

[33] Though Smalley is responsible with introducing the term and the extensive research over decades, many point out the overlap of similar studies in the field of acousmatic music since Schaeffer and INA/GRM [57][99].

[34] The original forming principles are not necessarily concerned about composing a sound, but more of a kinetic approach to making sense of a complex sound.

to another. Smalley distinguishes spectral motions (or growths) between "gestural," which is bound with the sense of physical and forward motion, and "textural," which lacks a motion in human-gestural time scale but is rather extremely slow or still.



onsets	***continuants***	***terminations***
departure	passage	arrival
emergence	transition	disappearance
anacrusis	prolongation	closure
attack	maintenance	release
upbeat	statement	resolution
downbeat		plane

Figure 6.5 The structural functions for common temporal phases [92].

A gestural spectral motion can be explained from numerous perspectives. Firstly, there are several archetypes of motions specific to the location within a sound object. Smalley identifies such temporal phases to be an onset (attack), continuant (body), and termination (decay) (see Figure 6.5). These segments interconnect with each other, sometimes discrete and sometimes morphing, with an expected or unexpected (deceptive) psychological effect. Such temporal-phase structures largely contribute to establishing unique and memorable sound expressions. For example, as an extended design process with SMSon, one may employ several gesture-unit archetypes corresponding to the onset and the rest (continuant and decay). While the continuant may be composed with the basic routine of model-shape-mapping selections, the onset segment may have its own set of models and shapes mapped, with either a constant (dummy) modulation in place of data or an aggregate value of all the data segments.

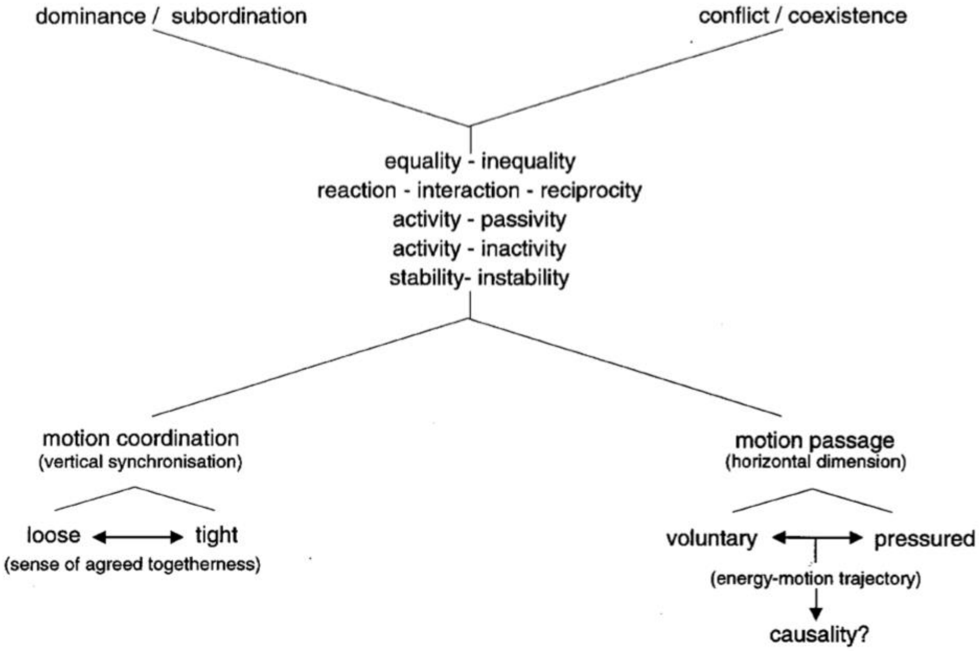


Figure 6.6 The vertical and horizontal motion behaviors (bottom) [92].

Secondly, realizing a gestural motion applies not only to the selection of temporal modulation shapes (2) but also to the design / selection of the models (1). Smalley describes the coordination of vertical (spectrum) and horizontal (temporal) motions as the “behavior” of a spectral motion (Figure 6.6). For example, a gesture may start “loose” and arrive “tight,” or gain a forward motion over time toward a “stable” spectrum. While not easily generalizable, the design of a model may be based on such loose-tight continuum so that the shape modulations are more predictable. The ‘quasi-harmonic’ model for the spectral peaks (see section 6.3.1), which is employed in the subsequent user studies, may possibly be described to have such behaviors by the user.

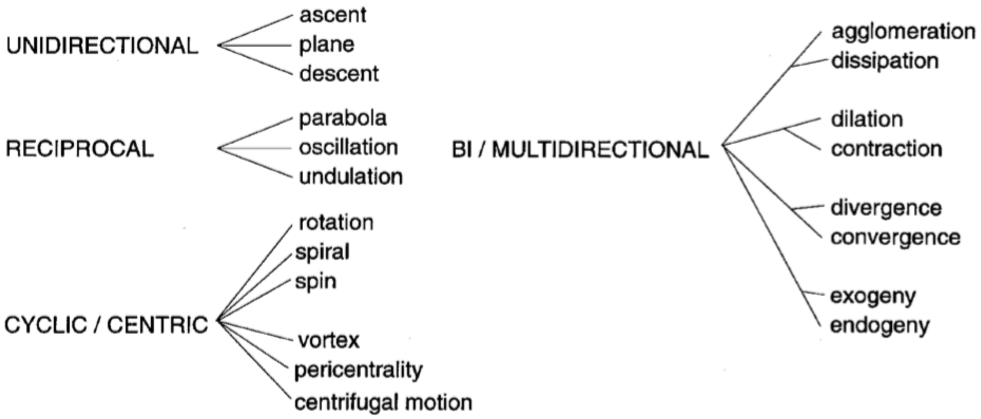


Figure 6.7 The types of motion growth (bottom) [92].

There is a presumed separability of timbral dimensions in SMSon, where multiple TDs modulated by different shapes or data can be decomposed / isolated to a certain degree. For example, one may be able to hear independent motions between a set of harmonic spectral peaks with its fundamental frequency modulated by a fast-moving sine wave, and a phase-distortion spectral effect with a linearly increasing randomization. Smalley addresses the existence of multiple spectral motions, with varying time scales, as the structural levels with non-hierarchical time-scale layers. In an extreme case, two layers of motions may be perceived as one with a gestural and another with a textural speed. However, in most cases with SMSon, multiple gestural motions may coexist in similar durational time scales. Smalley utilizes the structural functions (see Figure 6.5 for temporal phases) scaled to various time spans to discover or identify multiple motions. The shape-based motions in SMSon are, however, mostly cyclic locked into the same data-segment duration, making it more challenging to isolate as a temporal event. Instead, the listener may perceive lower level "directional tendencies" that describe the motional behavior of each dimension (Figure 6.7).

6.1.5 Composing and Decomposing Sonifications

To recapitulate the various structural and design elements of SMSon, it may be worth considering the role of SMSon in composing (designing) and decomposing (listening to) a sonification, the thematic elements of the research questions (see chapter 3.3). How would the framework aid the user in exercising aesthetic decisions during designing or listening / sense-making processes?

Both the composition and decomposition of an electroacoustic work are highly individualized and intricate processes [117][91], likely sharing little regularity among the users besides the acoustic stimuli and embedded data. The assessment of the framework necessitates ways to compare unique approaches taken by designers or subjective responses by listeners and find possible connections between these two groups of users. The role of SMSon in composition / decomposition processes is, therefore, structured in two ways: (1) Using the TS and TD organizations to simplify or formalize the user activities; and (2) Employing and linking descriptive materials, including textual stimuli and spectromorphological concepts to the attributes of TS/TD.

In the composition phase, SMSon facilitates the designer creating sound objects representing a small segment of data while incorporating variable aesthetic decisions. It encourages aesthetic explorations with quick and iterative sound generation. This is realized by providing basic synthesis building blocks consisting of several fixed and variable controls. The fixed parts are the sequence of data values as well as predefined generative models and modulation shapes. Design variabilities are introduced with (1) the selection of timbral-dimension models (algorithms), (2) the selection of shape and rate of parameter modulations, and (3) mapping data and shapes to the dimension models.

As an example, a designer may decide to create a metallic percussive pattern where the resonance is controlled by data. Among many possible approaches, one may first select a static inharmonic material (a bell-like spectrum) filtered by a bandpass spectral envelope with the spread parameter mapped to the data. The use of a decaying amplitude envelope may contribute the most to the percussive quality. Then, the designer may experiment with animating the inharmonic peaks with a smooth sine curve or random slopes, and swap the model to an algorithm featuring inharmonic peaks and noise.

For observing the listener's engagement to various and musically complex stimuli, the selection of stimuli needs some level of organization. As the combinations of model-shape-mapping patterns can be enormous, the observation cannot be comprehensive for creating a timbre space [112]. Instead, limited types of stimuli are broken down by the timbral-dimension layers and are organized in terms of the degree of "complexity." What is considered as the factor of sound complexity may differ greatly among designers and listeners. For the listening test, however, the author takes some of the simplest models and shapes, including the 'constant' shape, and applies them incrementally to simulate an increasing level of sound complexity. The listener's reactions are then compared to the increased level of complexity as well as the combinations of mapping and shapes rather than the stimulus type.

Following SPE, SMSon also upholds the measurability of physical quantities, both computational and perceptual, mapped to various spectral features. Although a computationally measurable audio feature does not simply imply perceptual measurability [48][90], it is useful as one of few quantitative benchmarks for organizing and analyzing various complex sounds. Quantitative metrics provide a possible layer / connection between what aural components a designer considers measurable and how different listeners perceive a composed sound to be measurable (which may vary drastically).

SMSon may specify a decoder that estimates the values of encoded data according to the mappings and algorithms being employed. Each of the spectral peaks, envelope, and effect algorithm is categorized by parameter types such as the centroid and tonality. The details are discussed in the implementation section below with Sonar.js. However, different decoders may share a general structure as follows:

1. If an amplitude modulation was applied to the sound segment, demodulate it by dividing with an estimated amplitude envelope. The amplitude envelope, if cyclical or bounded by silences, is also used to estimate the sound duration and the time window for the spectral analysis.
2. If the data is mapped to the spectral envelope, approximate the feature / parameter values of the envelope according to the model type information.
3. If the data is mapped to the spectral peaks, first estimate the spectral envelope (time aggregate) and deconvolve to obtain a flat spectrum. Then, extract the spectral feature according to the model type.

6.2 Sonar

SMSon as well as the user-study environments are implemented with a new sonification toolset. Sonar, a shorthand for ‘sonic array,’ is a programming library for web browsers that succeeds DTM (see chapter 5.1). It addresses the limitations of DTM in audio-processing performance and design generalizability, with the focus of sound organization in subsymbolic time scales. The new programming interface and design features facilitate the composition of time-evolving timbres with data. Sonar is also capable of implementing the SPE designs, granular and concatenative synthesis [85], and mapping spectromorphological characteristics [5][42] to multiple time scales of a sound.

6.2.1 Motivations: Limitations of DTM

While the core F/A principles could be satisfied with DTM or other general audio environments, end-user operations, both for the designer and listener, require some level of streamlining to provide an optimal (high-fidelity and highly flexible) experience. DTM has faced several performance bottlenecks with its generic interfaces for dtm.data and dtm.music: (1) The difficulty of optimizing the real-time performance, mainly because of a considerable overhead for adapting to different data types / inputs / situations; (2) Inability to extend the interface for specialized tasks, such as time-evolving audio synthesis with overlap-and-add rendering. Sonar employs functional programming principles (e.g., immutability and statelessness) and type interoperations to circumvent these bottlenecks.

Sonar provides high structural generalizability over DTM by minimizing the implicit stateful behaviors. In DTM, transformations of data are recorded in the dtm.data object itself in order to assess the loss of information (e.g., the error between before and after a transformation of data) or to retain the history of data transformations. In Sonar, in contrast, data and synthesis are always handled in explicit ways. Immutability facilitates the referencing of data at any point of time. Such multiple reference points for a single data object also allows complex “many-to-many” mappings.

6.2.2 Modularization

Table 6.3 A selection of SMSon modules

Name	Web Audio	Extends	Description
Seq	No	Array	Unit-less array with various accessors
Mono	No	Seq	Translates between various musical units
Poly	No	Mono	Manipulates child Monos simultaneously
Tempo	No	Poly	Overlap-and-add (OLA) pipeline
FFT	No	Poly	Analysis and synthesis facilities
Stream	Yes	None	Real-time OLA pipeline
Clock	Yes	None	Reads / maps data in real time
Waveform	Yes	Clock	Plays back Mono / Poly as waveform
Wavetable	Yes	Clock	Plays a waveform repeatedly
Fetch	No	None	Loads data sets and audio sources
Table	No	Poly	Relational data representation

Unlike DTM, Sonar organizes offline and time-related data structures into many specific types (Table 6.3). These are explicitly defined structures as opposed to transitional states in dtm.data, easing a systematic comparison of one component of sonification to another. All of them are, however, also interoperable using the type-casting methods as they derive from the native JavaScript Array class. The Seq, Mono, and Poly classes are the basis of all other sub-types, providing the essential accessor and transformation methods.

Table 6.4 The core accessor methods of Sonar

Name	Description
phase	Linearly interpolated estimation of a value or Vector.
pframe	Samples a section of data with variable size using phase.
at	Samples values at indices.
put	Replaces values at indices.
with	Applies type-interoperation and transformation to all values.
do	Applies type-interoperation and transformation at indices.
peek	Exposes the copy of current values in a callback function.
poke	Exposes the reference to current values in a callback.

```
let k = 1024
let phase = 0.2
let width = 0.1
data.pframe(phase,width,k) // -> [p0,p1,…,p1023]
```

Code 6.1 A frame of a data sequence, linearly sampled and interpolated at k points between the phases [0.15,0.25]. This facilitates a sliding window just by updating the phase value.

Extending the effective design elements of DTM, most classes and methods incorporate the relative phase metrics, a linear positional value ranging between 0 and 1. Phase may be used to query a single or a range of values, interpolate (linearly or otherwise) between multiple values or vectors, and synchronize data queries to audio playback or offline rendering. The code example below, the phase-frame (pframe) accessor, is utilized frequently for mapping a snippet of data to the stream of an audio signal (Table 6.4, Code 6.1).

```
let k = 4096
let wav = snr.tempo(state => {
  let p = state.phase
  return spectra.phase(data.phase(p)).cast(snr.fft)
    .increment(state.cycle).synthesize()
}).block(k).hop(k/4).duration(3).render()
```

Code 6.2 The construction of a time-evolving expression with Tempo and phase. The phase value (p) increments linearly, which may be further remapped through data.phase(p) to create a nonlinear playback of spectra over time.

The Tempo module provides a pipeline for the discrete short-time Fourier transform (STFT)[35] and overlap-and-add (OLA) methods. Tempo is a basic building block for generating time-evolving sound snippets. In an offline rendering, it queries and maps data to synthesis using

[35] The fast Fourier transform implementation used in Sonar is KissFFT by Borgerding (https://github.com/mborgerding/kissfft) transpiled to a WebAssembly module.

the block index, total-clip phase, and group delay (used for calculating or modifying the STFT phase-angle increment), with analysis windowing and compensation (“deframing” [20]) for the amplitude modulation introduced by the windowing function (Code 6.2).

```
let k = 4096
snr.stream(state => {
  let p = state.phase
  return spectra.phase(data.phase(p)).cast(snr.fft)
    .increment(state.cycle).synthesize()
}).block(k).hop(k/4).duration(3).play()
```

Code 6.3 An example of real-time streaming temporal expression. Notice that same synthesis designs from offline snr.tempo can be employed (see Code 6.2).

Similar to dtm.music, some modules operate in real time utilizing the Web Audio API. Stream, for example, mirrors the Tempo API but performs buffer-based instantaneous OLA, allowing an immediate and indefinite duration of real-time synthesis (Code 6.3). It facilitates real-time interactions by the user. The downside, however, includes the difficulty of deframing, normalizing the output level, and reusing the rendered output for, e.g., further hierarchical transformations.

6.2.3 Type Interoperation

DTM attempts to bridge the gap between the behaviors of offline data operations and the real-time streaming / scheduling of samples in the Web Audio API. In contrast to DTM, Sonar is generally decoupled from Web Audio while optimizing the speed of offline calculations by selecting the best iteration techniques supported by the browsers.

Another overhead in DTM is in instantiating or cloning the dtm.data object. The high number of parameters, including the previous array contents before transformations, inline function definitions, and varying types and dimensions of arrays, consume a considerable amount of time for memory allocation and initialization. By modularizing the data class to more minimal and specialized classes, Sonar minimizes the memory-allocation overheads.

```
const k = 4096 // Short-time A/S block size
const a = await snr.fetch(‘sample.wav’) // A multi-channel audio sample.
const b = a[0].chunk(k,k/2) // Slice into blocks of size=k with the sliding
factor of k/2.
const c = b.sortby(v => v.apply(snr.stats,’RMS’)) // Quickly cast and reorder
by a feature.

// A one-second overlap-and-add operation with the data / waveform blocks
queried by the clip phase.
snr.tempo(s => { // s (state) provides several different time-scale values
for the moment.
  const d = data.phase(s.phase) // Interpolated value from a 1-D data.
  return c.phasestep(d).peek(v => v.multiply(d/v.apply(snr.stats,’RMS’)))
}).block(k).render()
```

Code 6.4 Sequencing of a windowed audio sample, ordered by RMS mapped to data. This facilitates sample-based synthetic techniques such as granular and concatenative synthesis.

Modularization, however, introduces an extra challenge in chaining or composing functional units, sometimes interoperating between different types. Sonar employs type casting for changing the behavior of the data being handled. This includes switching between common list operations and audio analysis. For instance, a very simple analysis-synthesis approach may take an existing audio sample, slice them into windowed blocks, reorder the blocks with root-mean-square (RMS), a statistic of the overall volume of the block, and use them in a sonification with queried blocks with the volumes micro-adjusted to the data value (Code 6.4).

Such temporary conversions facilitate the analysis of the output of each transformation such as the spectral-peaks modulation, spectral-envelope modulation, and block or sound-object-level amplitude modulation.

6.2.4 Towards Subsymbolic, Continuous, and Non-Parametric Sound Organizations

```
let pitches = dtm.data(0,3,5,7,10).add(60)
data.each((v,i) => {
  dtm.music().note(pitches(i)).after(v).repeat().every(1)
})
```

```
let k = 4096
let pitches = snr.mono(0,3,5,7,10).add(60).from(‘notes’).to(‘bin’,k)

data.map(v => snr.poly.harmonics(k, pitches.phasestep(v))
  .cast(snr.fft).synthesize())
```

Code 6.5 Generating similar pitched melodies in DTM (above; with unit generator) and Sonar (below; with spectral synthesis). This allows creating symbolic expressions in the frequency domain in Sonar.

DTM mainly facilitates symbolic expressions, such as the use of pitched notes and harmonies in equal-tempered tuning. The structural organization is often discrete (e.g., see Code 6.5 for comparison with sonar.js) or parametric (a high-level parameter controlling multiple aspects) that the audio results are often nonlinear and difficult to measure / quantify. With the Mono module, Sonar aims to resolve such nonlinear relationships with bidirectional conversions between symbolic pitches and harmonics (magnitude spectra), as well as rhythmic offsets and magnitude or magnitude/phase spectra. Conversions to intermediate spectral representations facilitate unique morphing from one state to another (e.g., different chords or rhythmic onsets), which is used as the basis for various continuous expressions introduced in the following section.

6.3 Synthesis / Analysis Models

Combining the concepts for SMSon and Sonar, this section illustrates the generative and transformative algorithms for creating F/A sonification snippets. These algorithms were designed by the author as the ‘fixed’ components for the user studies (see chapters 7 and 8), with simpleness and computational efficiency as the priority.

The audio waveforms are rendered with the offline Tempo module. In many cases, the processing block size K is set to either 4096 or 8192 samples with a sampling rate of 44100 Hz to retain a reasonable frequency resolution of the analysis bins in the lower frequency range. A ‘reasonable’ resolution is determined with the fundamental pitches to be used in a musical scale. With the equal-tempered tuning, if we were to use the pitches among {C3,...,B3} (or MIDI note numbers between the interval of [48,59]), K=4096 would round the note D#3 to a frequency bin with a quantization error bigger than 50 cents. Therefore, for models with a pitched / symbolic structure, it is preferable to use the notes above D#3 or the block size of K=8192[36]. The hop ratio uses the ¼ of the block size. The Hamming window function, which does not taper down to zero, is employed for deframing (i.e., demodulation by division) the rendered waveform.

6.3.1 Spectral Peaks

Spectral-peaks algorithms serve as a baseline or source material to be further processed by other algorithms. It provides room for a wide range of creativity with a constraint that non-zero magnitude peaks must always be uniformly distributed. While retaining the global distribution, local peaks may have uneven magnitudes or intervals to create an inharmonic structure. The peaks as a whole may be attenuated by the amplitude envelope to a non-zero value.

[36] Pre-analysis zero-padding is also a commonly used technique to increase the frequency resolution. The models in this study do not employ this technique with the primary interest in timbral structures rather than pitches.

6.3.1.1 Tonal – Noise

```
// A static spectrum with the mixture of two sets of comb-shape peaks.
let tonal = snr.mono.mask(
  snr.mono.comb(k/2,30),
  snr.mono.comb(k/2,40)
)
let genNoise = v => snr.m.random(k/2,0,v)

snr.tempo(s => {
  let m = mod.phasestep(s) // The value may vary between [0,1].
  let noise = genNoise(m)
  return tonal.multiply(1-m)
    .add(noise).zeropad()
    .cast(snr.fft).increment(s.cycle)
    .take(1, p => p.add(noise)) // Randomize linearly-incremented phase.
    .synthesize()
}).block(k).hop(k/4).render()
```

Code 6.6 A simplified view of the spectral-peak generation and rendering. With larger modulation, the noise component becomes louder and also adds randomization to the spectral phase values.

This generative algorithm interpolates between two states. The “tonal” state consists of static magnitude peaks with linearly incrementing phases. Each peak resides in a single frequency bin with surrounding zero magnitudes, akin to the Dirac-delta function. The N peaks are distributed evenly across the spectrum between the first (sr/nfft) and the Nyquist frequencies, with fractional indices quantized to the nearest bin. The tonal peaks are not necessarily harmonic (i.e., not the integer multiples of the lowest peak). Optionally, several sets of peaks with varying N (different intervals) and a uniform magnitude can be overlaid (bit masked), creating a more complex and inharmonic pattern while retaining the uniform distribution.

The noise component is created with a combination of the uniformly distributed random magnitude and phase spectra. The phase values for each sliding window are thus interpolated between random and linearly incrementing values.

6.3.1.2 Noise – Buzz

```
// Create a k/2-by-res matrix interpolating from random to constant spectra.
let mag = snr.poly(
    snr.mono.random(k/2,0,1),
    snr.mono.const(k/2,1)
).linear(res)
```

Code 6.7 Magnitude spectra gradually interpolated from noise to buzz. Note that the phase delay is modulated with a similar matrix that masks a random-generator output.

Similar to the first algorithm, but the tonal component is replaced with densely placed peaks with each peak not always being separated by zeros. This results in an intense amplitude modulation between neighboring bins, creating a classic sawtooth timbre (without a dumping factor for higher-frequency partials).

6.3.1.3 Inharmonic Mixture

```
let k = 4096
let nPeaks1 = [31,41,47]
let nPeaks2 = [29,37,43]
let mags1 = [0.7,0.4,1]
let mags2 = [0.6,1,0.5]

// Poly provides generators with multiple spectra as outputs. Mix() method
masks and overlays the spectra into a single Vector.
snr.poly(
  snr.poly.comb(k/2,nPeaks1,mags1).mix(),
  snr.poly.comb(k/2,nPeaks2,mags2).mix()
).linear(res)
```

Code 6.8 Two sets of a complex mixture of harmonics, linearly interpolated to create a gradual transition.

This algorithm crossfades between two mixtures of equally spaced magnitude peaks. The example below creates two sets of inharmonics, each mixing three comb shapes with varying magnitudes (still providing a mostly flat distribution). As the partials are not integer multiples of the lowest harmonic, they result in a metallic and complex pitch mixture.

6.3.1.4 Quasi-Harmonic

```
let k = 4096
let numPeaks = 30
let jitter = [0,1]

snr.poly.harmonics(numPeaks,jitter)
  .linear(res)
  .map(v => v.comb(k/2).zeropad())
```

Code 6.9 Linearly interpolated magnitude spectra that transitions from a harmonic (jitter=0) to inharmonic (jitter=1) spectrum.

This model gradually transitions between purely harmonic peaks (with integer-multiple frequency bins) of N partials and N peaks with randomized partials except for the lowest harmonic. The perceived dynamics tend to be more subtle than the Inharmonic mixture, with the constant fundamental pitch dominating the moving partials.

6.3.1.5 Musical Scales

```
let k = 4096
let notes = snr.mono(0,2,4,5,7,9,11).octaves(2).add(60)
snr.poly.note(k/2,notes)
```

Code 6.10 A simplified method of converting symbolic pitches (a major scale) to magnitude spectra. Each spectrum (representing a pitch) may be further interpolated to morph into other spectra.

This is largely a “symbolic” model rather than timbral, providing a strong impression of traditional Western music in equal temperament. The example below produces 24 peak states that are not interpolated but discrete, each corresponding to a Major scale note between C4 and B5 (MIDI note numbers [60,83]). The notes / partials are ‘mostly’ uniformly distributed, akin to a sawtooth wave with no dumping factor for upper harmonics. It should be noted that, as the fundamental pitch rises, the total number of harmonics decreases (with 83 peaks at C4 and only 22 peaks at B5) and the distribution may also be skewed, potentially hindering the computational measurability.

6.3.1.6 Chord Scales and Harmony

```
let k = 4096
snr.mono(0,2,4,5,7,9,11).octaves(2).add(60).miditobin(k)
  .peek(v => {
    let idx = snr.seq.range(v.len()-3).map(x => snr.seq(0,2,4).add(x))
    return idx.map(x => v.at(x).map(y =>
      snr.mono.harmonics(hk,y)).cast(snr.poly).mix())
  }).cast(snr.poly)
```

Code 6.11 The musical scale data is windowed and expanded to create a chord structure.

Similar to the single-pitched musical scale algorithm, but this model groups the notes into tertian triads (i.e., (C4,E4,G4), (D4,F4,A4), etc.) creating a diatonic parallel-chord expression.

6.3.2 Spectral Envelope

A spectral envelope takes several parameters to create characteristic contours and resonant structures.

6.3.2.1 Centroid (Log-Normal Function)

```
let k = 4096
let means = snr.seq(-3,1).linear(30)
let variance = 2
let envs = snr.poly.lognormal(k/2,means,variance)
```

Code 6.12 A simple parametric model for generating log-normal envelopes of size k/2 by means.length (30).

This algorithm employs a common parametric-distribution model to generate a smooth magnitude envelope. The spectral centroid is regarded as perhaps the most salient timbral attribute in psychoacoustic and audio content-analysis studies [58]. Normal or log-normal functions can emulate and modulate the spectral centroid through the function centroid or the peak location of the magnitude distribution[37].

[37] It is, however, worth noting that the majority of musical audio signals do not have a single-peak envelope structure similar to a normal-distribution function. This simple model, therefore, provides a relatively limited range of musical expressions.

6.3.2.2 Spread (Log-Normal)

```
let k = 4096
let mean = -1.5
let variances = snr.seq(0.25,30).reverse().linear(30)
let envs = snr.poly.lognormal(k/2,mean,variances)
```

Code 6.13 Similar to centroid, but multiple values of variance create spectra from narrow to wide distributions.

Using the same log-normal function from the centroid model, the spread model emulates the spectral spread of a timbre from the standard deviation of the magnitude distribution.

6.3.2.3 Resonances

```
// The resonant-envelope collection.
Let means = snr.seq(-3,1).linear(30)
let src = snr.poly.lognormal(k/2,means,30)

let sections = snr.poly.pframe(20,0.5,snr.v.phasor(res))
  .map(v => src.phase(v).mix())
```

Code 6.14 Windowing and mixing a subsection of a matrix (the last line) to create unique combinations of an envelope with peaks.

This algorithm is based on a log-normal distribution consisting of 30 resonant tapered peaks. Each peak itself is generated with the same log-normal function with a varying centroid and small variance factor. The parameter modulation for this algorithm filters these peaks with a sliding rectangular window, with fractional indices at the margin linearly interpolating the peak-function shapes.

6.3.2.4 Block-Windowed Envelopes (“Stria”)

```
let interval = k/4
snr.seq.phasor(res).map(v => {
  return snr.seq.lognormal(k/2,-1)
    .multiply(snr.seq(1,0).step(interval/2).phaseshift(-v*.5))
}).cast(snr.poly)
```

Code 6.15 A more elaborate spectral envelope with a 4-cycle square wave starting at various phases.

The “stria” model filters a log-normal envelope with sliding low-frequency square wave or ‘stripes.’ There are three variations of stria with noticeably different timbral characteristics. They are determined by the modulating square wave with the interval of either K/4, K/8, or K/16, where K is the FFT size of 4096 samples.

6.3.3 Spectral Effects

A spectral effect modifies either magnitude or phase of already-generated spectral peaks. It is usually more appropriate for an optional embellishment rather than the primary modulation. Even though some algorithms may be used for mapping data to the amount of modification[38], the effects tend to be either perceptually subtle or complex and nonlinear. Another drawback is the relative difficulty of parameter decoding with the requirement of reference or buffered spectra.

[38] For example, noise-tonal spectral peaks configured with a (lossy) 'constant' shape may be, instead, modulated with a spectral effect.

6.3.3.1 Peak Jitter

The peak-jitter algorithm randomly shifts individual magnitude peaks to neighboring bins with a controlled range. This range is a segment of the magnitude spectrum, increased exponentially as {1,2,4,...,K/2}. The range-parameter modulation is, instead, scaled logarithmically.

6.3.3.2 Peak Fluctuation

This algorithm applies randomization to each magnitude peak where the value is non-zero. The overall amount of fluctuation may be controlled continuously between [0,1], with 0 not altering the original magnitudes.

6.3.3.3 Phase Distortion

Similar to peak fluctuation, the phase distortion adds random modulation to each bin phase. The effect is only audible with tonal peak algorithms with linearly incremented phases.

6.3.3.4 Peak Sampling

This effect is perhaps the most lossy and destructive, transforming a static magnitude spectrum into a series of individual peaks that are played back randomly.

6.3.3.5 Mirrored Peaks

This algorithm gradually attenuates the original spectral peaks and mix in a reversed shape.

6.3.4 Amplitude Envelope

Unlike the spectral algorithms, the amplitude envelope modulation does not use morphing states or multiple shapes to interpolate between. Rather, a single modulation shape is directly applied as the envelope to the time-domain samples, over either each data-point segment or the entire duration. The reason for this simplicity is to (1) not to draw too much of the listener's attention from the spectral timbral aspects to the rhythmic complexity and (2) to afford the envelope demodulation to estimate the original flat-amplitude signal.

```
Snr.tempo(s => {
  // STFT OLA block
  return spectra.phase(s).synthesize()
}).block(k).hop(k/4).render().peek(v => {
  // Time-domain multiplication
  return v.multiply(aenv.linear(v.length))
})
```

```
snr.tempo(s => {
  // Time-domain multiplication within OLA.
  // This smears out the transients.
  Return spectra.phase(s).synthesize()
    .multiply(aenv.pframe(s.phase,0.1,k))
}).block(k).hop(k/4).render()
```

Code 6.16 Above: Applying an amplitude envelope in the time domain. Below: Attenuating each STFT block with segmented values of the envelope at each phase.

The amplitude modulation is typically applied to the rendered result of STFT synthesis. While it is also possible to apply segmented amplitude envelopes directly to the time-domain output of each STFT block, the overlapped and tapered blocks tend to lose transient details (see Code 6.16).

The listening test employs three simple and repetitive shapes: (1) an exponentially decaying “sharp” attack envelope, (2) a slow-attack “swelling” ramp, and (3) a pulsating square-wave modulation (with non-zero amplitude). To demodulate, the signal is denormalized and divided by its analytic signal, yielded by the Hilbert transform.

This chapter, in summary, provided a theoretical framework and implementation for creating bespoke and expressive timbral sonifications, with examples of spectral algorithms (models). The following chapters 7 and 8 discuss how these categories of algorithms are employed along with various temporal modulation shapes (mappings) to (1) prepare simple and complex stimuli for the listening test and (2) facilitate the rapid creation of snippets in the design test.

CHAPTER 7. ONLINE LISTENING TEST

To evaluate the proposed design framework, spectromorphing sonification (SMSon), the author conducted two sets of human subjects study. The goal of these studies is to observe how various elements of design (e.g., the organizational concepts, aesthetic choices, and synthetic algorithms) relate to the effective and intuitive understanding of a sonification. The first study, an online listening test, focuses on the listener's structural understanding of several SMSon examples. It explores their aural analysis behaviors toward musically complex stimuli from various angles using quantitative and qualitative methods. There are many methodological challenges in assessing a functional-aesthetic (F/A) timbral sonification. Two issues, concentrated on the aesthetic side, are finding (1) an approach to instructing or training the listener in how to engage with a musically complex sound organization and (2) how to balance the training / practice with the retention of the effect of creativity / musical novelty.

Assessing the listener's effective engagement to electroacoustic works is largely an open issue in contemporary musicology. Whether it is a purely aesthetic musical piece or functional-aesthetic sonification, there is often a wide gap of "understanding" an organized sound between the designer (composer) themselves and the listener. Discussing the (lack of) accessibility of electroacoustic music, Landy points out the general absence of "investigations into aesthetic responses" (e.g., listening experiences) with the sound itself, and instead, most of communication efforts being rather focused on revealing the construction process of a piece [57].

A listening experience is, however, difficult to moderate or navigate. Complex and artificial sounds can be quite foreign to non-musician listeners who may lack previous experience in formally engaging (carefully listening and contemplating) them. How they approach such an unfamiliar situation may vastly vary from one to another, perhaps dependent on their musical background. In this regard, it is also difficult to employ a generalized training scheme for the listener that sets them to a similar level of musical / timbral literacy. For a timbre-oriented study, options for interactivity [87] are also limited, as a continuous time-varying sound structure may not always be organized with real-time interactions and sound synthesis. This limitation also carries to the public presentation of sonification where the interactivity may not be available to all the audience.

Issue (2) concerns the dynamic (non-static) nature of the design framework. A listener training generally sets an expectation toward the test stimuli yet to be presented. This introduces a delicate dilemma to the experience of aesthetic work. Novelty, or variability, is an aspect of musical aesthetics where the designer incorporates personal design choices and innovative ideas as opposed to reusing a fixed musical structure. Such changeable design demands the listener to learn the design / system on the fly or learn and recall as different expressions are presented in succession. However, extensive or repeated training may work negatively as the process of listening becomes more systematic, diminishing the chance of observing how the listener encounters and discovers information from a novel sound expression.

This listening test is itself a creative exploration to address the above challenges. In order to observe the sense-making process of the listener with a timbral sonification, the author finds that the experiment needs to guide their listening, navigating their attentions toward the details of the timbre but without directly explaining or revealing the actual content of the sound stimuli by words. As such, this listening study applies the classic electroacoustic concept of “reduced listening,” a passive (non-interactive) but focused and repeated listening typically with a recorded sound [82]. A reduced, interpretive, and qualitative listening phase may precede a transcribing phase for quantifying the data in sonification, serving as a substitute for a formal training process.

This may be viewed as a brief training stage that does not involve written instructions or interactions but, instead, nudges the listener to closely explore the sound and make sense of its internal structures.

7.1 Research Questions

Earlier, chapter 3 has presented the following research question: With SMSon, how would the general listener be able to recognize and quantify the multidimensional timbral structure of a F/A electroacoustic sonification? By observing the listener's engagement with a musically complex sonification with elements of reduced listening, this listening test seeks to address three subdivided questions.

A. What structural elements of SMSon contribute to or hamper the listener's intuitive understanding of a complex timbre and estimation of values?
B. How does the presence or absence of qualitative (interpretive) listening affect the quantitative (estimative) listening process?
C. What are the possible connections between the design elements and the listener's understanding, in terms of intended / perceived complexity, intelligibility, and musicality?

The first question aims to examine the relationships between synthetic models / mappings (as discussed in chapter 6) and the listener's general responses. The models entail several types of generative algorithms for creating dynamic spectral peaks, spectral envelope, and amplitude envelope. The mappings denote the mapping and combinations of data and auxiliary modulations for each timbral dimension. The author hypothesizes that changes to the spectral models may affect the accuracy of data values being estimated. On the other hand, model-agnostic mapping schemes may have an effect on the speed of aural decomposition.

The second question is mostly concerned with the listener's perceptual decomposition of a novel and complex timbre. This is observed through two different listening tasks: a qualitative interpretation of timbral characteristics and the quantitative estimation of data encoded in the timbre. The listening test presents various configurations of sonification that feature multiple timbral dimensions with independent temporal modulations. To retrieve quantities from such a complex stimulus, the listener must be able to identify the temporal motion of data, which can be either distinct or subtle, among other simultaneous motions. The author sets a presupposition that a perceptual decomposition generally entails segregation of timbral dimensions followed by a retrieval of the encoded data values. That is, the presence of a focused interpretive listening phase should have a positive effect on the estimation of the encoded data. The positive effect may be observed as the increased accuracy, efficiency, and/or confidence. Furthermore, there may be potential relationships between how the listener interprets (e.g., aesthetic valuations, the perceived complexity of sound, etc.) and the estimation performance.

Lastly, the third question explores the possible connections between the designer's intentions and the listener's reception. More specifically, it aims to find the connections between how the designer manipulates the characteristics of a F/A sonification with intended qualities such as data intelligibility, musical appeals, and sound complexity, and how the listener evaluates these qualities. As this experiment only includes limited types of stimuli, the analysis focuses on how to observe the listener's responses through these lenses. The author speculates that spectromorphological typologies, such as sound references and gestural associations, may provide a way to describe the organizational and perceived qualities of timbral sonification.

To address these questions (shortened as RQ 1-A, 1-B, and 1-C), the user study employs statistical and qualitative methods and analyzes multiple aspects of the subject's listening experience. The following discussions are structured into four parts: (1) The experimental design and analytical approaches, (2) the overview of collected data and preprocessing, (3) statistical analyses, and (4) qualitative analyses and discussions.

7.2 Experimental Design

The listening test investigates how the listener approaches to comprehend electroacoustic sonifications with a musically complex sound design. The listener analysis of F/A sonification is a complex problem. To examine from various possible angles, this user study presents two contrasting types of listening tasks, one focused on the quantitative estimation of encoded data and the other on qualitative interpretations of a complex timbre. A qualitative "focused" listening is an elusive concept with little standardization in musical analysis [11]. The study employs non-auditory materials, namely the textual descriptor of timbres, to guide the listener toward a qualitative listening experience.

This listening test also aims to establish a unique listening experience for the subject with the sense of creativity preserved in various musical stimuli. Simultaneously, it is critical to make the aural analytical tasks accessible and intuitive for general non-expert listeners, who are the main target of this listening test. On the other hand, this experiment does not aim to find out what synthetic types generally yield better measurability or decomposability, nor rank them in order for, e.g., a design recommendation.

7.2.1 Preliminary Tests and Design Iterations

With the multiple goals of a F/A sonification, the design of a listening test can easily become complicated for both the participant and analyst. The experimental design has gone through several iterations with preliminary tests recruiting six music-technology graduate students and a post-graduate. The primary areas of iterative adjustment were the stimuli and interface, including the musical complexity of the timbre, the selection of descriptors, and balancing the length and details of the instructions. The measurements are also analyzed for estimating the statistical power.

In earlier prototypes, all the auditory stimuli consisted of combinations of very simple spectral models (e.g., a filtered sawtooth) and auxiliary shapes for the purpose of intentionally setting the complexity of sound in an additive manner. That is, the number of simultaneous modulations could be 0, 1, or 2, presuming the increase of the perceived complexity in a similar manner. This approach, however, required the statistical design to present the same synthetic models with different permutations at least eight times (without repetitions), resulting in listening fatigues after 20-30 minutes. Some signals with constant amplitude (no modulations) caused noticeable discomfort and eventual loss of attention for the analytical tasks. This led the author to (1) focus on the variety of synthetic models rather than the variation of a single model and (2) introduce a certain amount of musical complexity at minimum with generally more constant motions of the spectrum, together with amplitude modulations always present. These changes have improved the general quality of response and engagement while compromising the opportunity to observe the effect of an ‘additive’ complexity.

In addition, adjusting the language of the sound descriptors and the task instructions took a large effort, as this turns out to be a significant barrier in successfully recruiting non-musician listeners with acceptable work quality.

7.2.2 Online Study on Amazon Mechanical Turk

This listening test leverages the web-based workflow of sonar.js (see chapter 6.2) together with Amazon Mechanical Turk (MTurk)[39]. MTurk is a platform for hosting surveys and data-collection projects, or so-called human intelligent tasks (HITs), accommodating requesters and anonymous participants for a paid contract work. The actual listening experiment, however, takes place in a custom-built audio-visual environment. The MTurk survey page, which is used solely for recruiting, compensating, and occasionally contacting the participants, links to the listening-test web application hosted at Georgia Tech[40]. The web application consists of a client-side static page with dynamically generated audio-visual[41] materials and a PHP server application with MySQL database recording listener responses.

MTurk invites listeners from across the world with varying auditory and musical knowledge. On the one hand, including non-musician participants is essential in this study for assessing the practical value of sonification for not only audio experts. It is also beneficial for creating a statistical study based on the general population. On the other hand, it poses a nontrivial challenge of making the experiment accessible to all knowledge levels, as well as qualifying the listener for English literacy, task comprehension, etc. Compared to in-person studies, anonymous studies have a noticeably higher risk of including ‘unqualified’ users. For a listening test with

[39] https://www.mturk.com/

[40] https://networkmusic.gatech.edu/son_test_1_p4/

[41] The visual UI employs d3.js and NexusUI libraries.

quantitative and qualitative inputs, unqualified jobs generally result in lower performance for quantitative tasks and random or rushed responses in qualitative tasks, as described in section 7.3.1.

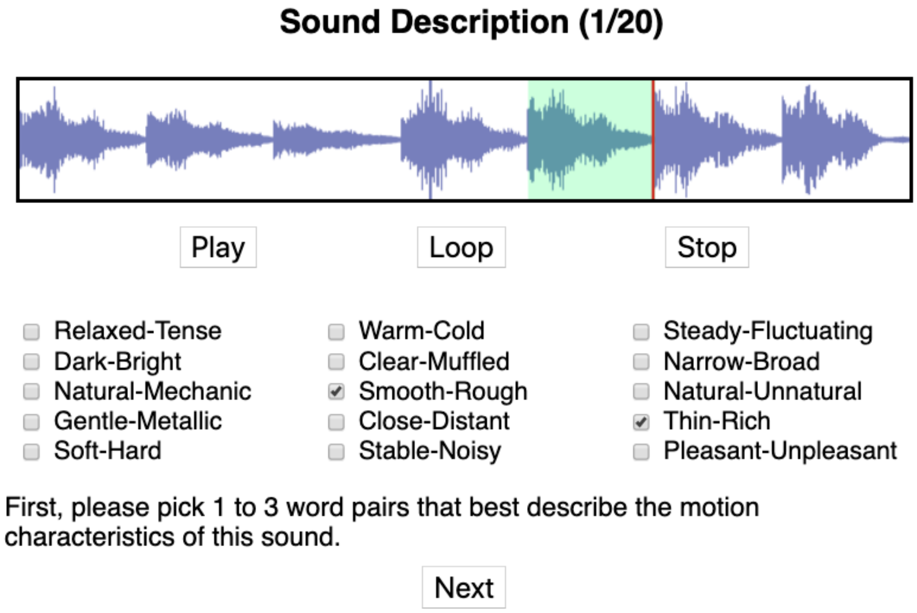
Furthermore, an online public listening test cannot control the listening environment of the participant, such as the choice of speakers / headphones and the ambient noise. It is likely that the audio stimuli are played with limited bandwidth and dynamics. The stimuli, therefore, need to retain certain qualities: (1) they should be content-based (cf. semantic information encoded in a speech signal), which withstands added noise or distortion to some extent, rather than being fidelity-based (cf. audio reproduction and spatial imaging), and (2) they should have highly contrasting characteristics within and between different stimuli types.

7.2.3 User Interface and Procedures

This listening test is relatively extensive in its length that a single session lasts for 52 minutes on average (N=116). Each session consists of three sections: an introduction and practice stage, the main listening test, and a survey.

The introduction provides the consent form and the general descriptions of the test, emphasizing the minimum requirements for participation such as English literacy and browser compatibility for sound-based tests. The practice stage introduces the interactive user interface for analytical listening tasks, with a simple auditory stimulus and an example response to it. After being familiarized with the interface and format, the subject proceeds to the main listening test.

The main listening test consists of 20 audio stimuli / trials. These trials are mostly identical in the format, each consisting of three tasks: sound ‘description,’ data ‘estimation,’ and sound-data ‘matching’ tasks. All of these tasks use the same auditory stimulus, generated according to randomized value sequence, synthesis models and mappings (see chapter 6 for details). The description and estimation tasks are presented in a semi-random (shuffled) order as an independent variable, always followed by the matching task.


Sound Description (1/20)

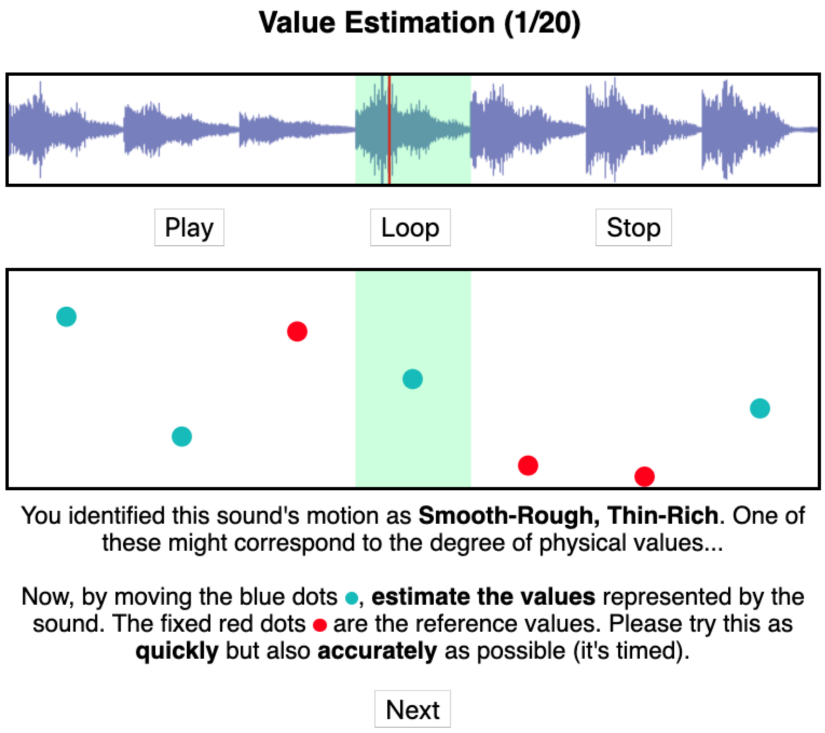
□ Relaxed-Tense
□ Dark-Bright
□ Natural-Mechanic
□ Gentle-Metallic
□ Soft-Hard
□ Warm-Cold
□ Clear-Muffled
☑ Smooth-Rough
□ Close-Distant
□ Stable-Noisy
□ Steady-Fluctuating
□ Narrow-Broad
□ Natural-Unnatural
☑ Thin-Rich
□ Pleasant-Unpleasant

First, please pick 1 to 3 word pairs that best describe the motion characteristics of this sound.

Figure 7.1 The user interface for the sound description task.

The description task aims to prompt the subject to listen to a stimulus in a focused, explorative, and interpretive manner. The task page (Figure 7.1) presents an interface with the stimulus' waveform and playback controls. It instructs the subject to select one to three adjective pairs, out of 15 choices, that best describe the temporal characteristics of the sound. The main expectation for the subject is to explore and carefully examine various aspects of the timbre with repeated listening[42]. The descriptors (word pairs) are expected to serve multiple purposes affecting the listener responses, as discussed in section 7.2.6.

[42] In the description task, the listener is not urged to select the words quickly (although is prohibited from pausing) but can spend arbitrary time needed to fully explore the aspects of sound.



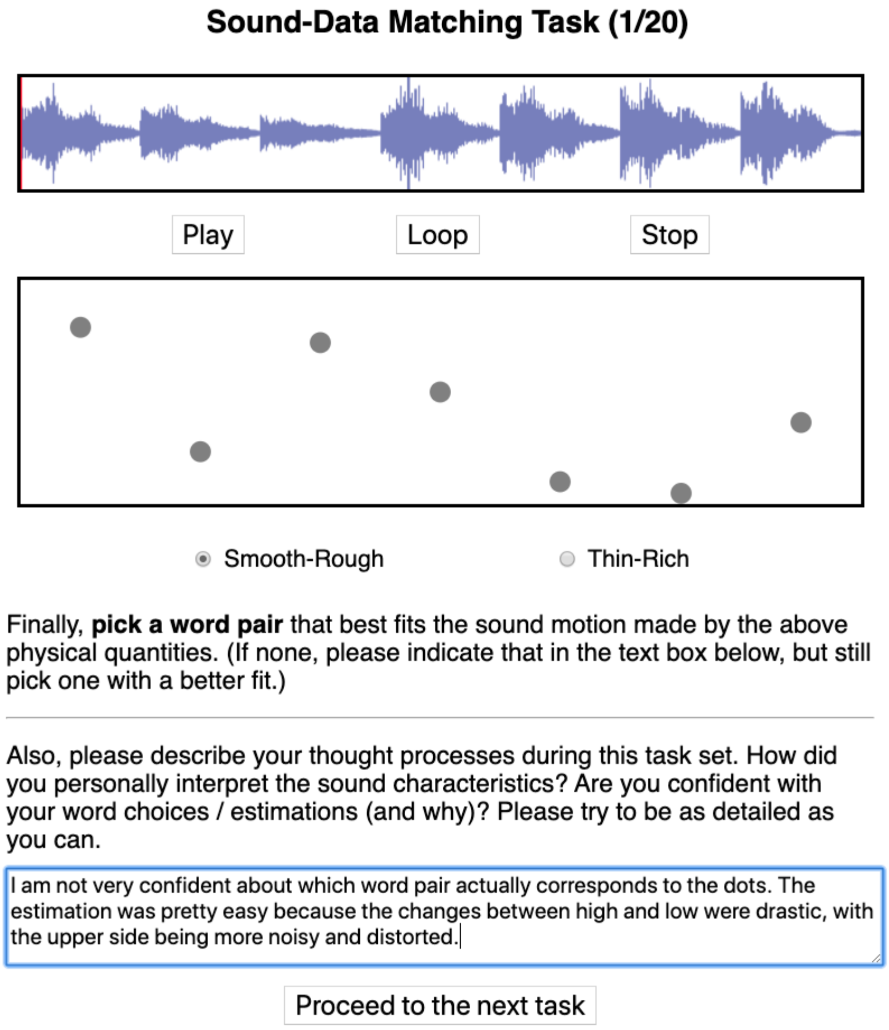
Value Estimation (1/20)

You identified this sound's motion as **Smooth-Rough, Thin-Rich**. One of these might correspond to the degree of physical values...

Now, by moving the blue dots ●, **estimate the values** represented by the sound. The fixed red dots ● are the reference values. Please try this as **quickly** but also **accurately** as possible (it's timed).

Figure 7.2 The user interface for the value estimation task.

The estimation task aims to gauge a ‘quantitative’ listening of sonification. It presents the playback interface as well as a set of vertically movable ‘dots’ (or sliders), each corresponding to a segment of the stimulus being displayed (Figure 7.2). The listener is prompted to place the dots to where they perceive the underlying values to be as quickly as they can. Out of seven dots, three are colored in red as referential points, placed at the true position of the data values and are immovable. The referential segments, with their timbral content and positional cue, aid the listener in orienting the vertical direction, as timbral changes may lack inherent high-low or up-down associations. Also, they imply the range and scaling factor, with the maximum, median, and minimum of the seven values being preselected as the references. The experiment only discloses to the subject that the referential points indicate the true values, but not the functions regarding directionality and scaling to minimize the risk of visual-focused estimation attempts ignoring the sound information[43]. The other movable points, colored in light blue, are set randomly for each trial.



● Smooth-Rough ○ Thin-Rich

Finally, **pick a word pair** that best fits the sound motion made by the above physical quantities. (If none, please indicate that in the text box below, but still pick one with a better fit.)

Also, please describe your thought processes during this task set. How did you personally interpret the sound characteristics? Are you confident with your word choices / estimations (and why)? Please try to be as detailed as you can.

I am not very confident about which word pair actually corresponds to the dots. The estimation was pretty easy because the changes between high and low were drastic, with the upper side being more noisy and distorted.

Figure 7.3 The user interface for the sound-data matching task.

The matching task asks the listener to reflect on the previous two tasks, displaying their word-pair choices and the estimated values (Figure 7.3). It first prompts them to select a word pair closest to the temporal sound behavior created by the data as estimated (hence the “matching” of

[43] The presence of visual reference points may arguably affect the response behavior for value estimation tasks, as such references are typically not available in real-life applications of sonification. The author observes whether if visual-centric approaches affect the auditory analysis in the matching task responses.

the description and estimation). If the number of the previously selected descriptors is one, it highlights the item as a tentative match. This is followed by a textual response to the general experience with the stimulus / trial, asking the confidence of the word-pair selection and estimation as well as the thought processes while undertaking the tasks. The nature of the question and expected answers are rather inexact, prompting spontaneous selection of the topic by the user. This is to mitigate the challenge for non-musician users who may not be accustomed to describing specific types of listening activities in detail.

After all the 20 listening trials are completed, the experiment concludes with a two-part survey. The first part consists of four (+ one optional) questions for textural feedback directly related to the research questions and experimental procedure. The second part inquires the listener's musical background, using the Goldsmith Musical Sophistication Index (Gold-MSI) [98]. This information is used to construct high-level categories of users as additional explanatory factors and qualitative analyses. See Appendices B.1 and B.2 for the list of questions.

7.2.4 Measurements and Dependent Variables

This listening test presumes that the decomposition of electroacoustic sonification entails the segregation of timbral dimensions (separability) and the retrieval of encoded quantities (measurability). While being explored in later qualitative analyses, the perceptual or cognitive segregation of a complex (time-varying) timbre is nontrivial to quantify and measure, especially when non-expert listeners are concerned. Unlike the segregation of voices from a mixture in auditory streaming researches [19][84], little has been established as a theoretical framework for separating timbral dimensions beyond dissimilarities-based unlabeled timbral spaces [23][31]. The quantitative part of this study, instead, focuses on the retrieval of values, which may follow a varying degree of successful isolation of the target dimension.

Table 7.1 The collected listener responses and the primary usages in analyses. "Quantitative" implies that the measurement is treated as a dependent variable either as it is or after a quantification process. "Qualitative" and "mixed" imply the use of responses in case analyses. "Filtering" responses are used to disqualify certain users from the analysis.

Task	Listener Response	Usage
Estimation	Error	Quantitative
Estimation	Duration	Quantitative
Estimation	Playback count	Quantitative & filtering
Estimation	Invalid inputs	Filtering
Description	Word-pair selection	Mixed-method
Description	Duration	Quantitative
Description	Playback count	Quantitative & filtering
Description	Invalid inputs	Filtering
Matching	Word-pair selection	Mixed-method
Matching	Textural feedback	Qualitative & filtering
Survey	Study feedback	Mixed-method
Survey	Musical background	Quantitative

The study collects various quantitative and textual responses from the listener (Table 7.1). In data analyses, these measurements are further standardized (normalized), combined to calculate higher-level attributes, or thematically analyzed.

Error, duration, etc. of estimation are general benchmarks also seen in the task-oriented study of visualization systems. Also called time-and-error benchmarking, they are sometimes criticized for not capturing unique values of data interfaces such as providing insights and analytical confidence [96]. This listening study, in contrast, attempts to analyze how the characteristics and structure of sound are learned and utilized for data extraction.

For statistical analyses, the raw time-and-error measurements are processed into the ‘accuracy’ and ‘efficiency’ metrics informed by the collected data. The details are discussed in the following Data Collection and Processing section.

7.2.5 Independent Variables and Randomized Elements

Table 7.2 The models and mapping sources of the audio stimuli. The leading numeric value indicates a model with fixed spectral / amplitude algorithms. (X) denotes the mapping source, with C: constant (no modulation), D: data, and A: auxiliary modulation.

Type	Spectral Peaks	S. Envelope	Amp Env.	Repeat
1-CD	Quasi Harmonics (C)	Centroid (D)	Swell	2
1-AD	Quasi Harmonics (A)	Centroid (D)	Swell	2
1-DC	Quasi Harmonics (D)	Centroid (C)	Swell	2
1-DA	Quasi Harmonics (D)	Centroid (A)	Swell	2
2-CD	Noise-Tonal (C)	Centroid (D)	Decay	2
2-AD	Noise-Tonal (A)	Centroid (D)	Decay	2
2-DC	Noise-Tonal (D)	Centroid (C)	Decay	2
2-DA	Noise-Tonal (D)	Centroid (A)	Decay	2
3-DA	Chord (D)	Centroid (A)	Pulse	1
3-AD	Chord (A)	Centroid (D)	Pulse	1
4-DA	Noise Color (D)	Resonance (A)	Decay	1
4-AD	Noise Color (A)	Resonance (D)	Decay	1

For statistical analyses, the experiment specifies two primary independent variables that are both sequenced in semi-random (shuffled and sampled) orders: the types of sound organization (described in detail below) and listening tasks. The order of listening-task types may either be description-first (DF) or estimation-first (EF). Both of these orders are presented randomly but once for each stimulus type (Table 7.2), resulting the same stimulus type repeated twice (except for the stimuli types 3 and 4).

The audio stimuli (i.e., trials) are scrambled to mitigate perceptible relationships among them, with the main focus of attaining unpredictable / novel listening experiences. Each stimulus also embeds certain "data" values in a preselected timbral dimension (mapping). The "data," a numeric sequence of seven values, are random-based but considered a controlled nuisance variable [3] rather than an independent variable. With the main interest of decomposing a complex sound structure rather than making semantic or structural sense of data, the "data" only signify individual quantities with no intended structure or pattern. In the listening test, the values are randomly generated for each trial / stimulus, with the possible maximum interval of [0,1] and the minimum delta (`max(seq)-min(seq)`) of 0.5. The following code is used to generate such sequences:

```
// A random delta between 0.5 and 1. E.g., 0.7
let d = snr.v.random(1,0.5,1)

// A random 7-value sequence of the interval [0,d].
let data = snr.v.random(7).rescale(0,d)

// Shift the sequence by a small amount. E.g., 0.1
data.add(snr.v.random(1,0,1-d))

// -> E.g., [0.44,0.07,0.47,0.65,0.77,0.59,0.67]
data.print()
```

Code 7.1 The generation of a range-controlled random sequence.

Each listener session contains a total of 20 audio stimuli generated with randomized "data" quantities. From the synthetic models listed in the previous chapter, 16 stimuli (types 1 and 2) uses more musically simple combinations mainly for quantitative analyses, and the other four stimuli (types 3 and 4) employs more complex patterns for observing qualitative responses as well as for adding contrasting and/or unexpected patterns (cf. nuisance variables [3]). While the stimuli share the limited few synthetic models to generate from, the author expects them to have some level of novel and distinct impressions for the listener with different mapping combinations and when presented in a non-repetitive order.

The sound organization may be explained in terms of the synthetic model (algorithm), mapping, or their combinations. Table 7.2 summarizes the models and mappings. The repeated stimulus types use the same models and mappings, but with different data values. They are also presented in mirrored orders of tasks (i.e., either the estimation-first or description-first). In the analyses, the mapping schemes (e.g., CD, AD, etc.) are further examined in terms of mapping targets and auxiliary motions (see section 7.4 for details). All the auxiliary modulations are a sinusoidal LFO of three cycles for each data point, modulating the parameter over the maximum range as specified by the algorithm. See chapter 6 for the details of each algorithm.

The stimuli with different model-mapping combinations are presented in a shuffled order with several deterministic patterns:

- The trials 5, 10, 15, and 20 should be musically complex stimuli (i.e., the types 3 and 4).
- The same model-mapping with a different order (i.e., description- or estimation-first) should not appear simultaneously in the first half (i.e., trials 1-10) or the second half (11-20). This is to minimize the short-term memorization of the same stimuli type.
- The same model should appear only twice in each quarter of trials (e.g., 1-5).

7.2.6 Guided Interpretive Listening and Sound Descriptors

The description and matching tasks aim to facilitate a qualitative exploration of the stimuli and collect subjective interpretations in the forms of word pair selections and freeform writing.

Table 7.3 The timbral descriptors presented in the listening test. The types are assigned for the data analysis purpose by the author.

Word Pair	Type
Natural-Mechanic	Gestural
Steady-Fluctuating	Gestural
Relaxed-Tense	Gestural
Natural-Unnatural	Gestural
Stable-Noisy	Textural
Smooth-Rough	Textural
Thin-Rich	Textural
Soft-Hard	Textural
Clear-Muffled	Spatial
Close-Distant	Spatial
Narrow-Broad	Spatial
Gentle-Metallic	Cross-Sensory
Dark-Bright	Cross-Sensory
Warm-Cold	Cross-Sensory
Pleasant-Unpleasant	Cross-Sensory

Table 7.4 The lexical rules for mapping the descriptors to different types.

Type	Appended phrases
Gestural	Motion
Textural	Pattern or texture
Spatial	Placement or presence
Cross-sensory	Sensation or feeling

In description tasks, the interface presents 15 word-pairs from which the listener selects best-fit descriptions of timbral characteristics (Table 7.3). As the task instruction explains, these adjective pairs signify a continuum of timbral qualities. Such continuum should not only capture a static quality of sound, but also dynamic characteristics or multiple timbral states as perceived. The listener is prompted not only to identify the timbral motions created by the mapped data (i.e., the quantities to be measured), but also potentially multiple timbral qualities and motions (up to three selections) present in the complex sound.

This textual device brings a common pitfall that is often overlooked. A word pair, e.g., Soft-Hard, does not have a fixed directionality, as timbral motions or data quantities may go in multiple directions within a stimulus. While this fact is emphasized in the instructions, some listeners still appear to misunderstand that a word pair has a predetermined directionality (thus making them unfit to describe certain timbral motions).

The descriptors are composed of non-technical English words considering non-expert subjects with varying literacy in music and sound. They are mostly based on the work in perceptual audio-reproduction quality tests, where descriptors in plain language appear to be most well developed [3] (see chapter 2.1.2).

Considering the elusiveness of textual expressions for the general listener without a significant amount of training, the use of descriptors in this study entails two purposes: (1) For the listener, they serve an "anchoring" or attention-guiding elements for qualitative sound analysis, while also easing the challenge of describing unfamiliar auditory phenomena for some listeners; (2) For the analyst, they serve as a classificatory response (of the listener) to analyze. In addition to these roles in the listening test, at a higher level of F/A design, the use of textual descriptors also concerns the communication of design between the composer and listener (see chapters 8 and 9).

To elaborate further on (1), the variety of words may provide a starting point for exploratory analysis and more personalized textual descriptions. They are used for a cognitive anchoring effect [30], guiding the listener's attention to various aspects (cf. the timbral dimensions) of an electroacoustic sound by 'qualifying' their subtle / complex characteristics[44]. Such characteristics might otherwise be left unnoticed or treated as latent qualities in their analytical listening. On the other hand, anchoring descriptors also have a risk of introducing cognitive

[44] Some listeners instead preferred to use "quantifying" to explain this effect according to the post-experiment survey.

mismatch, especially when the listener already has a preconceived idea about how certain types of sound should be best described. Therefore, they are utilized as secondary and hypothetical devices in the analyses.

For (2), the descriptors as a response variable are utilized as a metric for timbral multidimensionality. First, the selected word pairs are converted to the number of selections, which is simply attained from the one, two, or three selections made in the description task. The number of descriptors may relate to the perceived complexity of the sound. This is later interpreted as a perceived plurality of timbral characteristics or an ambiguity of sound / descriptors.

For the purpose of high-level analyses, the descriptors are also categorized into four general groups by lexical and/or spectromorphological (SM) rules (see Table 7.4). The lexical groupings are proposed by the author by appending common phrases, such that, e.g., "a steady-fluctuating motion" may make sense but "a steady-fluctuating texture, presence, or feeling" may do less so. In addition, the gestural, textual, and spatial categories encompass technical implications for the spectromorphological analysis of sound [92].

A gestural sound is typically characterized by temporal changes of the spectrum in a human-gesture time scale. This is akin to the physical control of musical instruments, although Smalley defines several distance levels ("surrogates") within and beyond the gesture-to-instrument relationships. For example, a "relaxed-tense" quality of sound may be explained by the listener that the sound is generated directly by different human-gesture or physical / mechanical motions.

A textural sound, in contrast, may allude to slower (or in some cases much faster) timbral motions found in non-instrumental expressions. For example, "stable-noisy" may imply that a spectral pattern sustains beyond the sound snippet being heard in the trial. The listener may potentially interpret such sounds with reference to environmental sounds that constantly and automatically produce a type of sound (e.g., a wind).

A spatial description may not be fully applicable to the present ‘content-based’ listening test with monaural signals. However, a sensation of spatiality may attribute to the quality of acoustic realization of a source sound. For that, a source sound may be identified as a separate physical entity that automatically emits a characteristic sound while moving through the space in relation to the listener.

A cross-sensory description encapsulates other miscellaneous types that may lack motional or textural associations. They may induce some level of emotional response or other sensations that are difficult to translate to the previous motion characteristics.

It should be acknowledged that the grouping thresholds between these categories are defined relative to each other within the 15 word-pairs and with some forced disregarding of inherent ambiguity / overlapping. It is likely that the word pairs are actually interpreted completely differently by each listener. For example, a "natural-unnatural" motion may perhaps be understood as an emotional response (feeling) instead by the listener. In such a case, the assignment to an SM category is prioritized. Qualitative study also attempts to capture the potential mismatch of the theoretical grouping and their responses.

7.2.7 Statistical Design

The quantitative analyses examine the estimation performances from four different perspectives: sound organization, listening-task order, listener's musical background, and interpretive response categories. For each perspective, an analysis observes the statistical difference of standardized 'accuracy' and 'efficiency' scores within subject (see the section Measurement Processing).

For observing the task-order effect, the within-subject records are grouped into two main factors: the description-first (DF) and the estimation-first (EF) trials. While both task orders use the same auditory stimulus, DF asks the listener to qualitatively interpret the characteristics of the sound not limited to the data-modulated aspect. EF, on the other hand, challenges the listener to immediately engage in a quantitative estimation of the main (data) modulation without qualitative insights.

For the analysis of DF/EF ordering effect, only the 'low-complexity' stimuli (types 1 and 2) are considered, as 'high-complexity' (types 3 and 4) stimuli are only presented in DF order without repetition. When focused on the effect of sound organization and disregarding the potential ordering effect, however, the analysis includes high-complexity stimuli as well.

7.2.8 Regression Analyses

Using the accuracy and efficiency measures as the primary response variables, the analysis employs the paired (dependent samples) t-test for univariate analysis as well as the two-way analysis of variance (ANOVA) methods for various combinations of two explanatory factors. Both approaches observe the statistical difference of group means with the significance (p-value) threshold of 5%. These are computed respectively using R language's t.test (with the paired parameter enabled[45]) and car:Anova functions. The results are validated in terms of the normality of residuals (visually with a Q-Q plot) and homogeneity of variance with Levene’s test (car:laveneTest). The analysis also measures the standardized effect size between individual factors using pair-wise Tukey test (stats:TukeyHSD) and Cohen's d, in which the score above 0.2

[45] The independent t-test is not applicable as the DF/EF conditions are applied within subjects.

generally indicates a small effect (i.e., not obvious but not ignorable), above 0.5 a medium effect (i.e., a noticeable effect), and so on. This is calculated with

$$d = \frac{|\mu_1 - \mu_2|}{(\sigma_1 + \sigma_2)/2}. \tag{1}$$

The present study, however, uses a sample-wise standardized t-score, resulting in close-to-1 mean-square residuals of the ANOVA model in many cases. This may effectively set the mean differences in the pair-wise comparisons to be close to Cohen's d.

For various configurations of fractional factorial analyses (combining several independent variables), first, the balanced-group two-way ANOVA (type II) method is used. This method takes the organization of stimuli (i.e., synthetic models and mappings) and task orders (DF/EF) as the independent variables, and estimation performance (accuracy and efficiency) as the response variables. The null hypotheses are that (1) the means of estimation performance for different stimuli types or listening orders are the same, and (2) there is no interaction between stimuli types and listening orders.

Other explanatory factors may be based on the listener's attributes. For example, the author examines if musical experiences and/or musical knowledge, based on the Gold-MSI survey, make a difference in the response patterns. The null hypothesis is that the means of estimation performance for different musical backgrounds are equal. A factorization based on participant attributes (e.g., musical background and listening typologies) produces unbalanced groups. Such analyses may utilize the type III unbalanced-group ANOVA method. Moreover, the musical background divides the subjects into separate groups, necessitating the use of a mix-method ANOVA with within-subject and between-subject factors.

7.2.9 Meta Subgrouping Analysis

The previous ANOVA methods compare the means of subject groups, who repeat the trials with different conditions. In contrast, this meta subgrouping analysis divides the responses into subgroups of trials and compare the repeated trials within subjects for each subgroup.

This analysis observes the listener's estimation accuracy and efficiency when they perceive a stimulus to have a certain timbral quality. It, therefore, uses a dependent variable as the grouping factor. As this grouping may create unbalanced pairs of trials (i.e., a discrepancy of interpretation between the same stimulus type) within subject, the analysis also incorporates a filter to only include matching results between repeated trials (i.e., the listener selects the same descriptor type for a particular stimulus type).

The subgroups, or subcategories, are established in two ways: by the number of descriptors selected and by the 'primary' descriptor type selected for a particular stimulus. The primary descriptor signifies the word-pair selection corresponding to the data-modulated timbral dimension. It is selected by the listener in the data-words matching task, which displays the previous selections from the description phase and the data visualization from the estimation phase. As previously discussed, the 15 word pairs are mapped to four higher-level categories corresponding to spectromorphological typologies.

7.2.10 Qualitative Analyses

In addition to the general statistics, the author examines the qualitative responses of the listener in order to theorize the connections between design elements and listening experiences and to interpret the statistical results in further detail. The qualitative data are based on 24 (or 25) short written responses per listener, acquired in the post-experimental survey (see Appendix B.1) and the 20 'matching' tasks. The survey questions generally prompt the listener direct feedback on the experimental design and experience. Some of the responses, however, also reveal how they interpret and utilize potentially unfamiliar devices presented in the experiment. Such devices include the word pairs, the complex and variable audio stimuli, as well as their potential musicality. The matching task, asking their confidence and general thought processes while analyzing each stimulus, captures unique perspectives for analyzing and interpreting the sounds.

The qualitative analysis starts with thematic coding. The coding aims to find recurring patterns (themes) in personal approaches to describing and quantifying the timbral characteristics. The codes are first generated / extracted from texts in a ground-up manner, then categorized into the organizational elements or concepts of the SMSon design framework where appropriate. Such conceptual groups may include spectromorphology, time scales, and timbral dimensions. Within spectromorphology, the analysis borrows the concepts of intrinsic and extrinsic references, typologies of spectra, and gestural / textural distinction of time scales. For the quantification process, the author primarily seeks patterns in (1) the types of temporal references and (2) the types of analytical devices. For the sound descriptions and interpretations, the coding process particularly looks for (1) intrinsic / extrinsic references, (2) how the listener elaborates their selection of word pairs, and (3) indications or the use of multiple descriptors. The survey questions also employ the same approach of analysis.

Besides the collective thematic analysis, the study also selects several listeners with representative and contrasting behaviors. They may provide more coherent illustrations of the overall experience, aesthetic responses, and quantitative performance.

7.3 Data Collection and Processing

Through MTurk, 116 participants have completed and submitted the HIT. This amounts to 2340 trials (estimation + description + matching), of which, 1872 (2 x 936) trials with type 1 and 2 stimuli are configured for a pair-wise repeated analysis for the DF/EF ordering.

Table 7.5 The summary statistics of unfiltered listener responses.

Attribute	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Est. Score	0.0000	0.3744	0.5334	0.5365	0.7117	0.9747
Est. Playback	1.000	1.000	2.000	3.226	4.000	34.000
Est. Duration	3.114	10.918	17.380	27.160	29.044	3572.449
Desc. Num. Selection	1.000	1.000	2.000	2.173	3.000	3.000
Desc. Playback	1.0	1.0	1.0	2.7	3.0	37.0
Desc. Duration	3.658	10.680	16.536	27.237	28.522	1112.291
Mch. Feedback Words	1.00	5.00	10.00	18.32	28.00	97.00

The summary statistics of quantitative data (Table 7.5) show that there are signs of erratic inputs and outliers. For example, both the duration of estimation and description tasks include the records of the listener spending 20 minutes to 1 hour, which is likely a pause during a task that was prohibited. Also, some listeners spent only 3 to 4 seconds in listening tasks, which is less than the duration of a stimulus and indicates a rushed / random response. These signs of low-quality work necessitate a significant cleaning of data as detailed next.

7.3.1 Filtering Unqualified Subjects

A unique challenge in MTurk studies [78], as already discussed, is the relatively high number of unqualified participants, often blindly signing up for the HITs. Despite geographical restrictions to English-speaking regions and multiple warnings about required English literacy and potential rejections, out of N participants who attempted the HIT, roughly 50 exhibited limited

English literacy and/or insufficient understanding of the instructions. 25 obvious offenders, based on their use of scripts or duplications to fill nonsensical responses to the textual inputs, were rejected immediately after the submission of HIT.

Table 7.6 The range and quantiles of the filtering attributes. Cases between Min-Lower and Upper-Max are considered as outliers.

Attribute	Min.	Lower	Q1	Median	Q3	Upper	Max.
Invalid inputs for estimation	0	0	0	1	5	12	38
Invalid inputs for description	0	0	1	3	5	9	30
Total playback count	40	40	46.5	74	156	308	549
Average Mch. feedback words	1	1	5.03	10.78	28.35	57.20	72.95

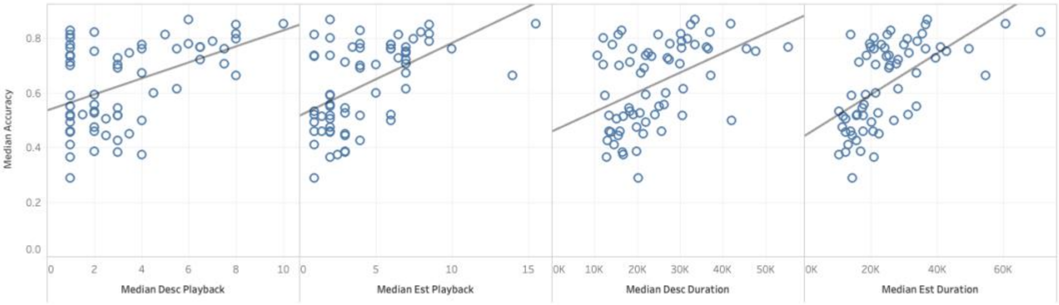


Figure 7.4 The correlation between the number of playbacks and the median estimation accuracy per listener.

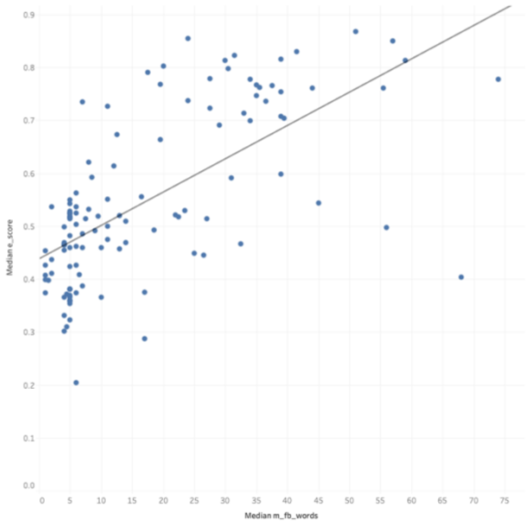


Figure 7.5 The correlation between the average word count for the sound descriptions and the median estimation accuracy per listener.

The collected data have a large likelihood of being erratic and noisy, demanding a careful cleaning. It is, however, important to acknowledge the risk of data fabrication in the filtering process. For example, Figures 7.4-5 indicate that there are apparent correlations between the amount of invalid or rushed inputs and the performance of the estimation task, suggesting that removing unqualified work may likely skew the response distributions. It is important to clearly and carefully define what constitutes work with minimum effort and/or random / erratic inputs and apply the rule systematically.

To filter out unqualified works, many MTurk studies employ quality-validation subtasks, commonly called attention checks, within their experiments. The present listening task includes a rather strict validation of user inputs in every task. The estimation task, for example, prohibits the subject from proceeding to the next task without either listening to the stimulus or moving the slider UI, and displays a warning message as they are detected. Similarly, the description task invalidates the listener inputs trying to mindlessly proceed without playing the stimulus or selecting any word-pair item. The accumulated invalid inputs are counted toward immediate rejections or pre-analysis filtering of participants.

Other indications of rushed work include the observed physical "effort" in listener responses that are within validation thresholds. For example, the total number of stimuli playback can be, at minimum, 40 times if the listener plays each stimulus only once in the estimation and description tasks. It is, however, highly unlikely to sufficiently answer every task with a single playback (especially considering novel / unexpected stimuli). The filtering process consists of eliminating the subjects with the following statistics:

1. The word length of the average matching-task feedback is below 5.03 (Q1).
2. The total number of invalid inputs for either task exceeds the upper whiskers.
3. The total number of playbacks is below 46.5 (Q1).

Table 7.7 The summary statistics of filtered listener responses.

Attribute	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Est. Score	0.000	0.4511	0.6277	0.6054	0.7904	0.9747
Est. Playback	1.000	2.000	3.000	4.546	6.000	34.000
Est. Duration	5.381	15.074	23.088	29.158	34.857	266.677
Desc. Num. Selection	1.000	1.000	2.000	2.038	3.000	3.000
Desc. Playback	1.000	1.000	2.000	3.697	5.000	37.000
Desc. Duration	4.243	14.268	21.728	32.651	34.572	1112.291
Mch. Feedback Words	3.00	11.00	22.00	25.03	35.00	97.00

As a result, the analysis includes a total of 69 listeners (out of 116), 1380 trial responses (out of 2340), and 1104 responses for the paired-analysis tasks (out of 1872), with the range of responses shown in Table 7.7.

7.3.2 Processing the Measurements

The primary metrics of perceptual measurability are the accuracy in estimation tasks and the efficiency in either estimation or combined (estimation + description) tasks.

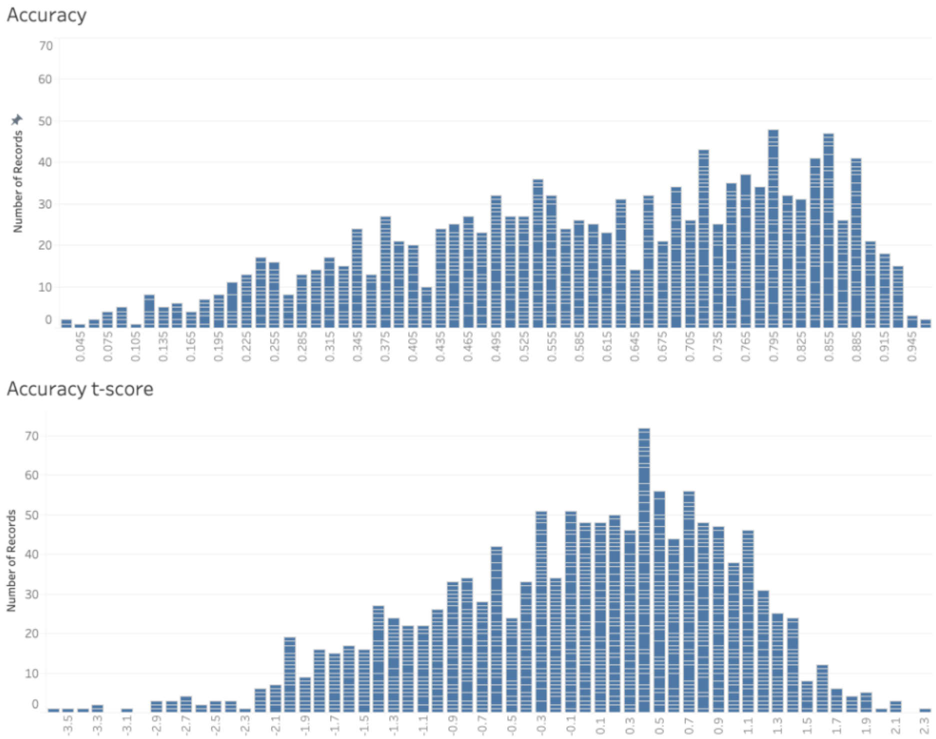


Figure 7.6 The histograms of the raw (above) and standardized (below) accuracy scores for all listeners. The per-listener standardization attains reasonable normality for the entire distribution.

The accuracy is attained simply from the closeness between the estimated values and the actual data. Both values have the maximum interval of [0,1], and the error is approximated using the Euclidean distance. The summarized accuracy score is:

$$acc = 1 - \sum_{i=1}^{N-M} \frac{\sqrt{(x_i - \widetilde{x_i})^2}}{N - M}, \qquad (2)$$

where $(x_i - \widetilde{x_i})$ is the error between the actual and estimated values, N is the total number of values, and M is the number of referential segments.

Since the raw accuracy scores across all subjects are distributed unevenly (Figure 7.6) with highly varying means and variances, they are standardized using the t-score[46] treating each listener as a sample:

$$t_{acc} = \frac{s_{acc} - \overline{s_{acc}}}{std(s_{acc})}, \qquad (2)$$

where s denotes a sample (listener) and t_{acc} is the t-score of each accuracy value.

The accuracy t-score (or standard score) is interpreted as the performance (ratio) on a task relative to the other tasks / stimuli by the same subject and dismisses the differences of average performance across subjects.

Another performance metric is the "efficiency" of timbral decomposition, indicated by the duration and the number of playbacks required for the task. The heuristic is that the more difficult the decomposition is for the complexity or the lack of familiarity with an auditory stimulus, the longer and/or more playbacks it takes. Following the accuracy, the duration and playback count used in the analysis are standardized within each listener using the t-score.

Compared to the accuracy, however, the efficiency is harder to quantify for two reasons: (1) The possible discrepancy between measured and actual listening durations and (2) the training (familiarization) effect by the task order.

[46] A sample (i.e., within subject) version of the z-score.

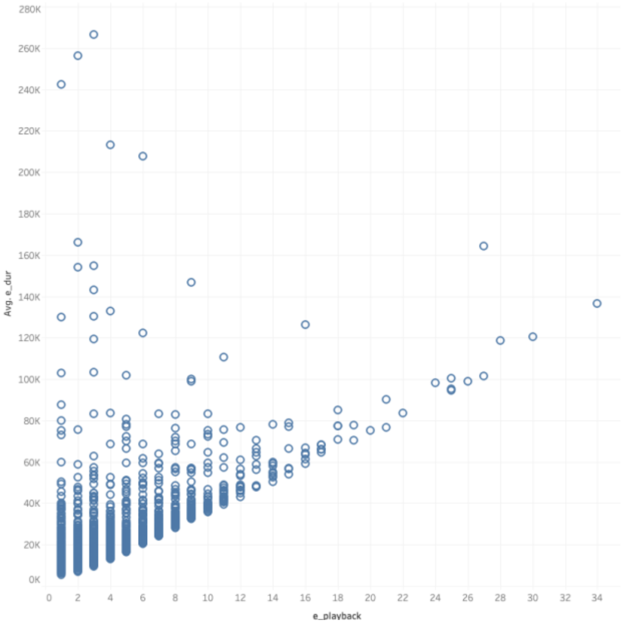


Figure 7.7 The playback count and duration for the estimation tasks. There are some cases where one metric overweighs the other.

First, the efficiency may not be attributed to only either the duration or playback count. For example, Figure 7.7 shows that there are long-durational responses with only one playback, in which the actual duration of listening to the stimulus may be much smaller (sometimes with the listener ignoring the instruction of not taking a break during the task). On the other hand, it is also possible that some listeners rely more on their short-term memory of auditory sensation, only listening to several times and then interacting with the UI while in silence. In this study, as it is not trivial to generalize these behaviors, the normalized t-scores of the duration and playback count are simply added together with equal weight, such that

$$t_{eff} = -\frac{t_{dur} + t_{pb}}{2}, \tag{3}$$

where a positive t_{eff} means a higher efficiency than a negative value.

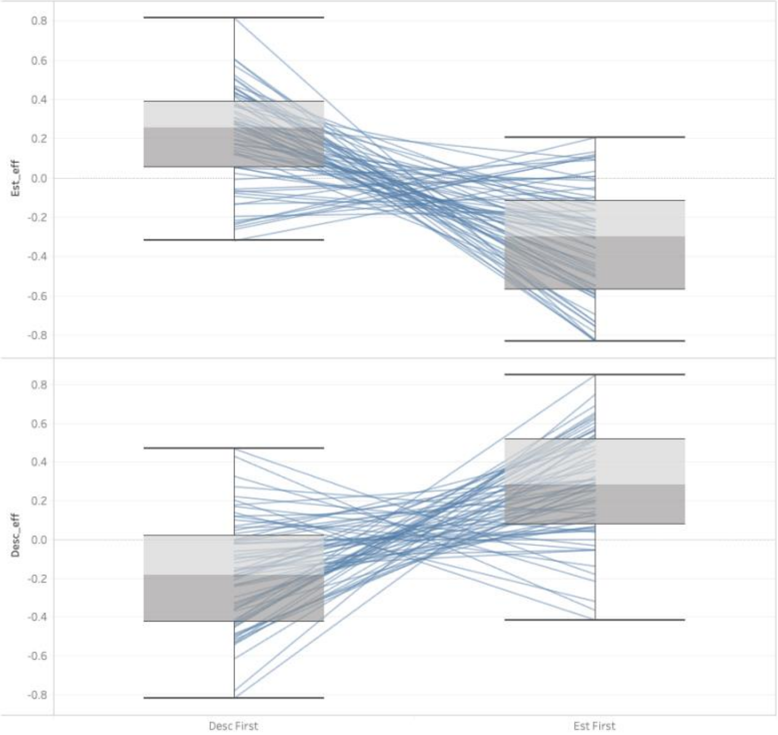


Figure 7.8 The correlations between the DF/EF (left and right) order and efficiency scores. The efficiency for each task is naturally higher when preceded by another task type, where the listener likely already familiarizes themselves with the stimulus.

Also, it should be highlighted that the speed of the estimation phase alone is ill-suited for explaining the effect of the listening task order (DF/EF) on the efficiency of a perceptual decomposition. That is, as the listener encounters a novel timbral expression, DF introduces a familiarization / training effect, and the estimation speed is almost always higher than EF (see Figure 7.8). In comparison, in EF, the average duration decreases in the description phase by a smaller amount, suggesting a smaller contribution in the familiarization with a stimulus by the quantitative listening for the qualitative listening. To partially rectify this issue, the author employs a broader / higher-level efficiency measure combining the measurements of the estimation and description phases, with

$$t_{eff:comb} = \frac{t_{eff:est} + t_{eff:desc}}{2}. \tag{4}$$

The combined efficiency is interpreted as the total quantitative + qualitative analysis efficiency, agnostic to the DF/EF order. This metric is only used when comparing the efficiencies with regard to DF/EF effects.

7.4 Statistical Results

7.4.1 The Overview of the Responses

Table 7.8 The abbreviations used in the statistical analyses. They represent four different mappings for each model type (which is either 1 or 2 for low-complexity types).

	Spectral Peaks	Spectral Envelope
DC	Data	Constant
CD	Constant	Data
DA	Data	Aux. motion
AD	Aux. motion	Data

This section describes some of the statistical outcomes that are particularly significant. Insignificant results are included in Appendix B.3. Many of the graphs refer to the structural elements of SMSon, i.e., models and mappings, denoted as shown in Table 7.8. Note that the mappings (DA, AD, etc.) recur for different model types, while model types (1-4) alter the algorithms assigned to spectral peaks and envelope dimensions. Also, the listening task settings are denoted as estimation-first (EF) and description-first (DF).

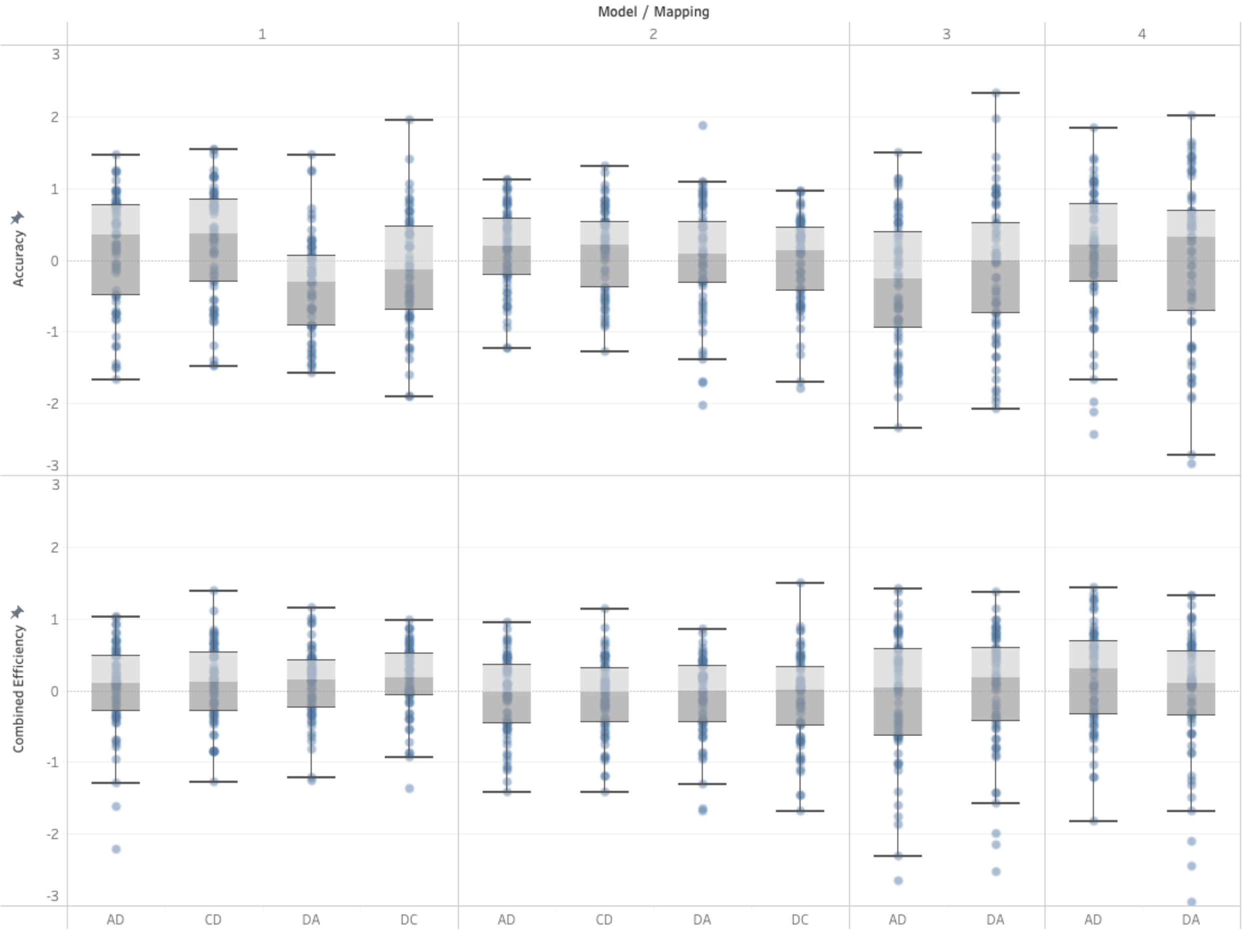


Figure 7.9 Individual accuracy (top) and Desc./Est.-combined-efficiency (bottom) scores grouped by the models (four boxes) and mappings (columns). Note that most of the following analyses only use low-complexity (types 1 and 2) models as they are tested with the full combinations of factors.

The listener's estimation-task performance was analyzed in terms of sound structure, DF/EF order, the listener's musical background, and their choices of descriptors for the estimated stimuli. The results generally indicate that the models / mappings have significant effects on the listener's estimative accuracies with varying strengths, but less so on the efficiencies (Figure 7.9). In addition to the structural elements, meta-analyses employing the listener's interpretive response types also indicate some influence on the estimative performance. On the other hand, there is little to no significance in DF/EF orders nor the listener's musical background (Appendix B.3), contradicting the hypothesis for RQ 1-B.

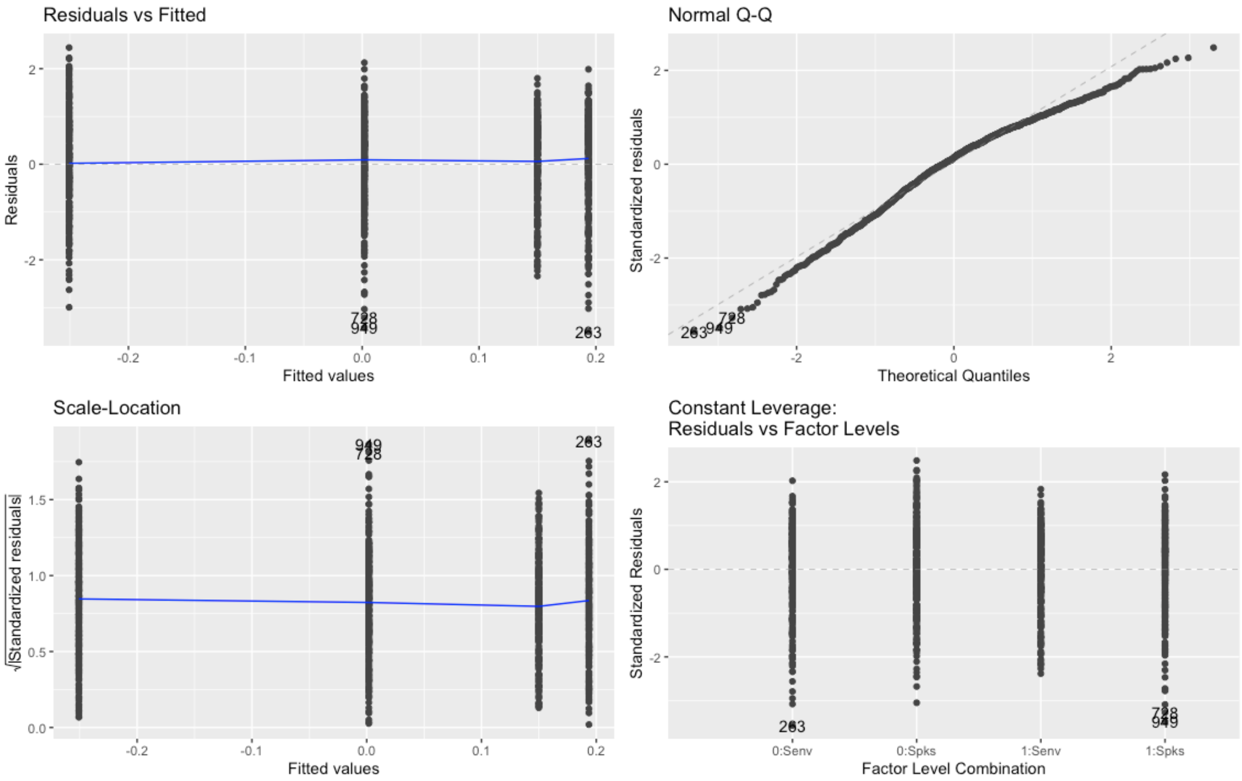


Figure 7.10 The linear relationship (top left), normality of residuals (top right), homogeneity of variances (bottom left), and outlier influences (bottom right) all indicate the statistical assumptions with this ANOVA regression model (model x mapping) is generally valid.

Before examining the results in detail, it should also be noted that the validity of statistical assumptions using ANOVA models is generally sufficient for all the significant results. Figure 7.10, for example, shows the validation plots for the model and mapping subgroupings, generally satisfying the regression assumptions. Different subgroup analyses appear to have similar levels of validity and will only be addressed if a violation occurs with, e.g., Levene's test for normality.

7.4.2 The Effects of Sound-Organization Factors

Table 7.9 ANOVA type II tests with accuracy t-score as the response variable. The stimuli only include low-complexity models, with DF/EF trials treated as repetitions.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Significance
Model	1	2.8	2.838	2.931	0.0872	.
Mapping	3	23.9	7.953	8.213	2.1e-05	***
Model:Mapping	3	8.1	2.697	2.786	0.0397	*
Residuals	1032	999.3	0.968			

RQ 1-A speculated that there may be a general change in the estimation accuracy between different spectral models, while alternating the mapping sources and targets within a model may not affect the accuracy as much. On the contrary, the results indicate the model to have less than significant effect, at least between these low-complexity stimuli (Table 7.9). The mappings seem to have a significant effect on accuracy, with a sign of interactions with the models. This suggests that, while the models are not the main factor for general accuracy differences, they may influence the general estimation performance with different mappings.

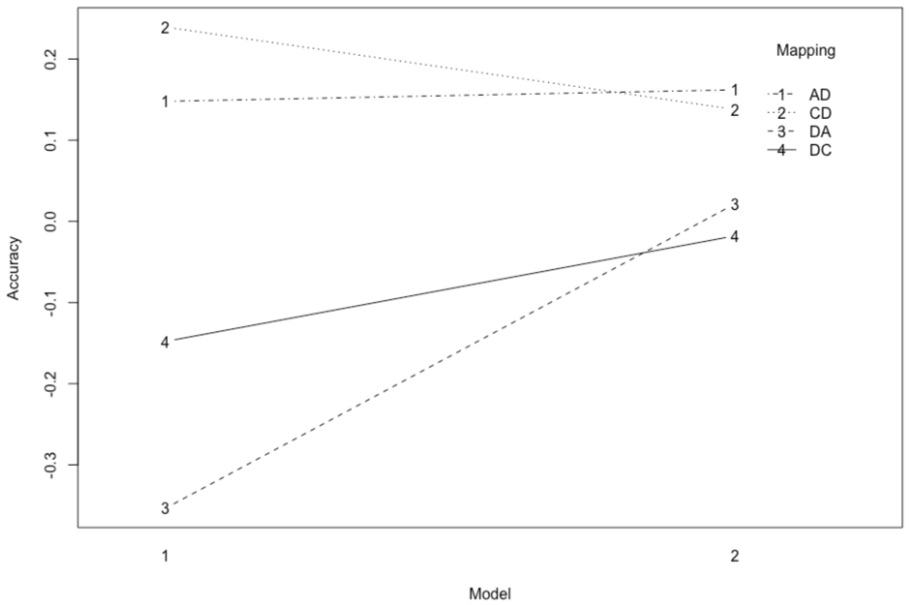


Figure 7.11 The interaction of group means of accuracy for model (columns) and mapping (lines) combinations.

The interaction plot (Figure 7.11) shows that there are relatively huge gaps between different mappings for model 1, especially between the groups of AD-CD (lines 1 and 2) and DA-DC (3 and 4). These combinations signify the target location of data (D) being either the spectral peaks (left; Dx) or spectral envelope (right; xD). The gaps between these groups diminish for model 2, although not in a uniform direction. Interestingly, some interactions occur between the models and separate mappings AD and CD (1 and 2) as well as DA and DC (3 and 4), although the distance between these mapping pairs is considerably smaller. The interactions may relate to the fact that a mapping either contains an auxiliary motion of timbre or not. These mapping types warrant a closer look at their individual effects.

Table 7.10 Type II ANOVA tests with accuracy as the response variable, breaking down the mapping factors. It shows the presence of auxiliary motion does not generally affect the estimation performance and does not interact with the location of data in timbre.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Significance
DataAt	1	22.8	22.815	23.396	1.52e-06	***
WithAux	1	0.9	0.886	0.908	0.341	
DataAt:WithAux	1	0.2	0.159	0.163	0.687	
Residuals	1036	1010.3	0.975			

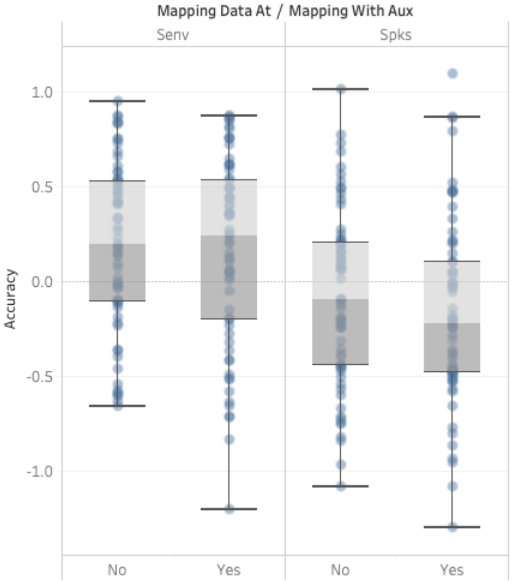


Figure 7.12 The distributions of estimation accuracy grouped by the mappings of data dimension (left and right halves) and the presence of auxiliary motions (columns). The changes are more significant between the data dimension mappings.

Combining the estimation responses for both models, this analysis regroups the mapping schemes by the data location and the presence / absence of an auxiliary motion. The ANOVA results (Table 7.10) show that only the location of data has a general influence on the accuracy.

The presence of auxiliary motion appears to have a slight negative effect on some individuals (Figure 7.12), but this effect may be rather determined by (interacts with) the change of models.

Table 7.11 Tukey pair-wise comparisons of the mapping values. The pairs that exchange the position of data (D) between left (spectral peaks) and right (spectral envelope) appear to have a significant difference.

	Mean Diff	Lower	Upper	P Adjusted
CD-AD	0.03365870	-0.1892066	0.25652397	0.9800615
DA-AD	-0.32093747	-0.5438027	-0.09807220	0.0012689
DC-AD	-0.23785502	-0.4607203	-0.01498975	0.0311380
DA-CD	-0.35459616	-0.5774614	-0.13173089	0.0002664
DC-CD	-0.27151372	-0.4943790	-0.04864845	0.0095554
DC-DA	0.08308245	-0.1397828	0.30594772	0.7726035

Pair-wise analyses on the mapping values (Table 7.11) also reaffirms the insignificance of the presence of auxiliary motions. The mean differences between target positions of data indicate a small to below-medium effect size with the range of [0.23,0.35]. These values roughly correspond to standardized Cohen's d measurement, as the residual mean-square of one-way ANOVA (0.95) is close to the unit.

Table 7.12 ANOVA type II tests with Desc./Est. combined efficiency as the response variable.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Significance
Model	1	9.2	9.202	11.749	0.000633	***
Mapping	3	0.1	0.037	0.048	0.986159	
Model:Mapping	3	0.5	0.173	0.221	0.881583	
Residuals	1032	808.3	0.783			

While the change of models had insignificant effects on the estimation accuracy within the low-complexity models, additional analyses indicated their potential roles in estimation. For example, the combined efficiency metric (i.e., adding the description and estimation efficiencies) appears to be largely affected by different models and no other factors including mappings, DF/EF orders, musicianship, etc. (Table 7.12). Pair-wise test reports that the models have a mean difference of -0.188 (from model 1 to 2). With the residual mean square of 0.779 in one-way ANOVA, this yields a small effect size of d = 0.213. This may imply that there is a slight gain of easiness to describe the sound and/or estimate values with model 2.

7.4.3 Meta-Analysis Using the Choice of Descriptors

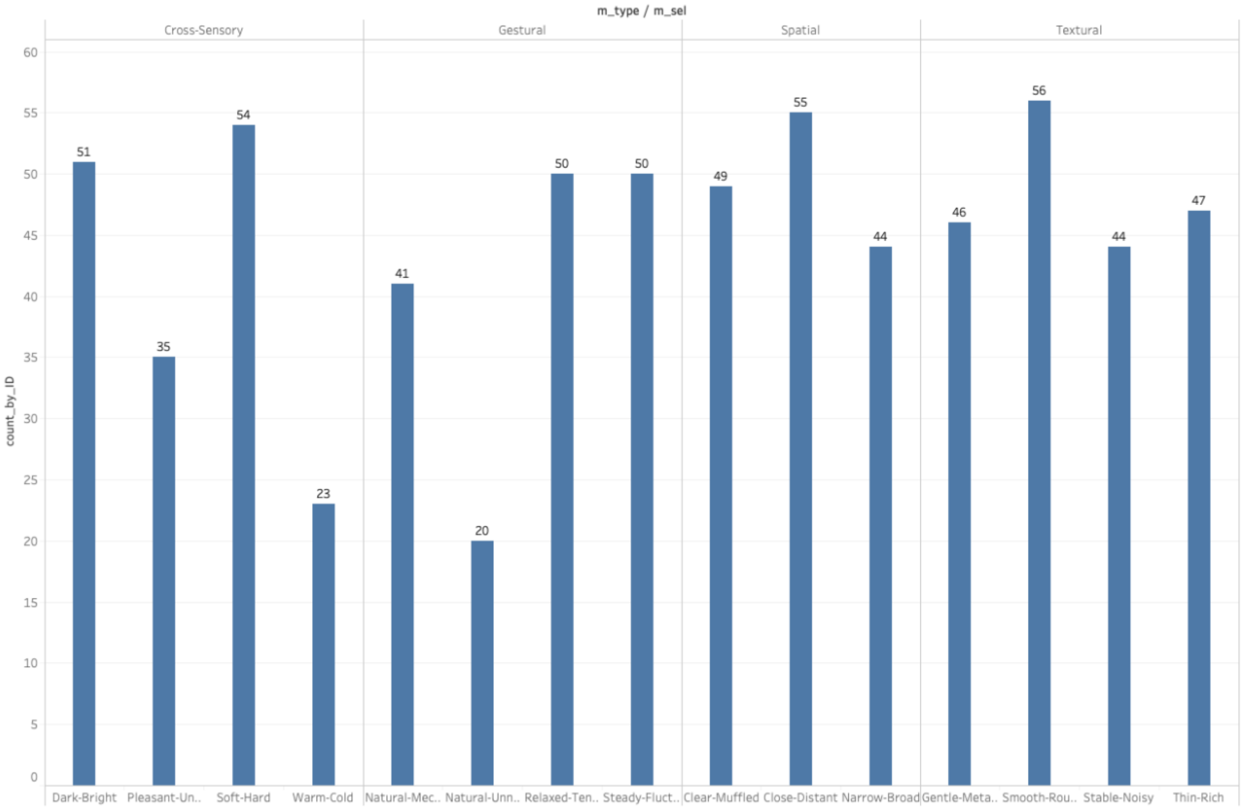
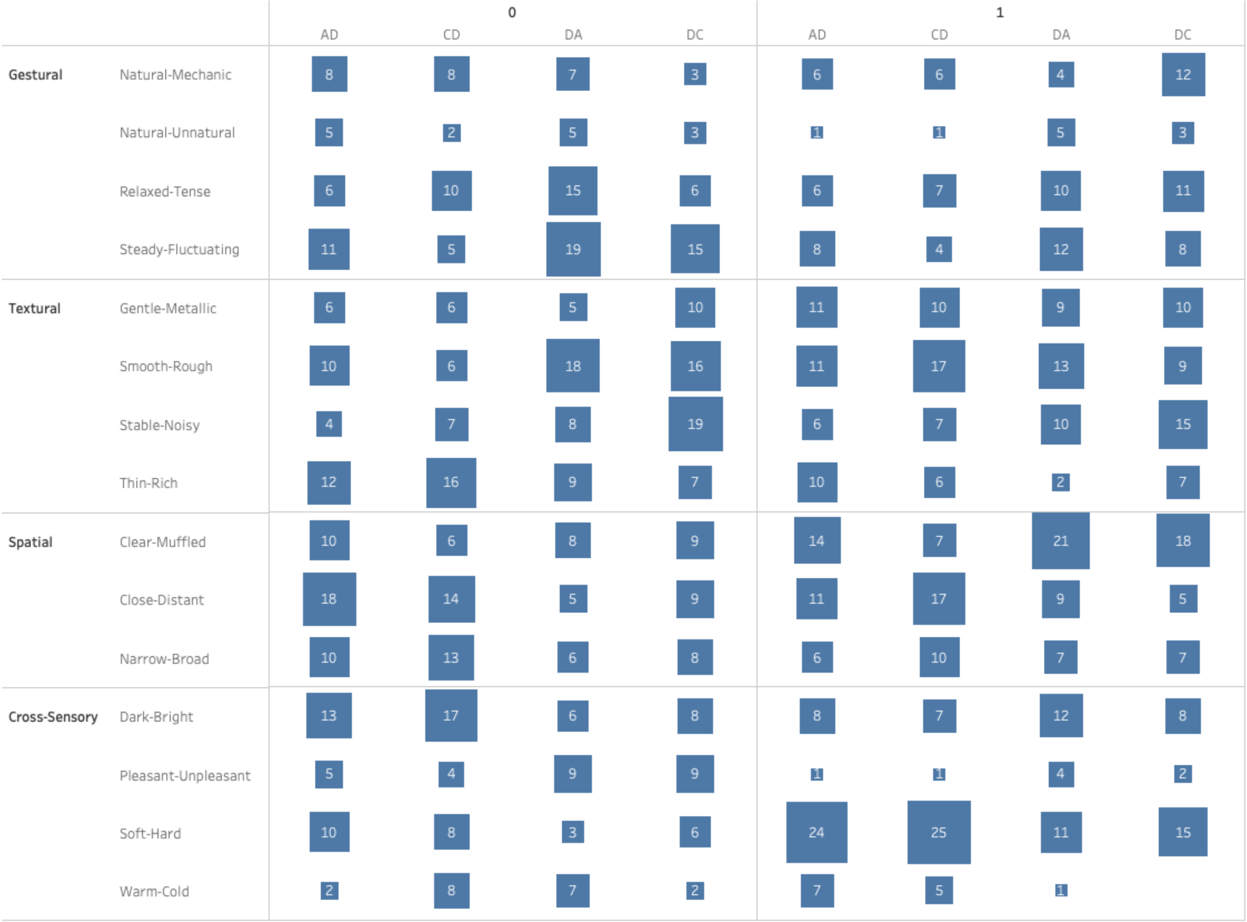


Figure 7.13 The histogram of matching-task descriptor selections after filtering. Individual descriptors (columns) are grouped by the spectromorphological types indicated above the graph.



		0				1			
		AD	CD	DA	DC	AD	CD	DA	DC
Gestural	Natural-Mechanic	8	8	7	3	6	6	4	12
	Natural-Unnatural	5	2	5	3	1	1	5	3
	Relaxed-Tense	6	10	15	6	6	7	10	11
	Steady-Fluctuating	11	5	19	15	8	4	12	8
Textural	Gentle-Metallic	6	6	5	10	11	10	9	10
	Smooth-Rough	10	6	18	16	11	17	13	9
	Stable-Noisy	4	7	8	19	6	7	10	15
	Thin-Rich	12	16	9	7	10	6	2	7
Spatial	Clear-Muffled	10	6	8	9	14	7	21	18
	Close-Distant	18	14	5	9	11	17	9	5
	Narrow-Broad	10	13	6	8	6	10	7	7
Cross-Sensory	Dark-Bright	13	17	6	8	8	7	12	8
	Pleasant-Unpleasant	5	4	9	9	1	1	4	2
	Soft-Hard	10	8	3	6	24	25	11	15
	Warm-Cold	2	8	7	2	7	5	1	

Figure 7.14 The number of matching word-pair selections for each stimulus type (models and mappings).

Contradicting the general assumptions of RQ 1-B, the mere presence of an interpretive listening before an estimation (i.e., DF or description-first) did not have a uniform effect on both types of estimation scores (Appendix B.3). However, as speculated in the context of RQ 1-B, there seems to be a sign of general effects in estimation performances when the listener has a particular interpretation of a sound.

First, it should be acknowledged that subgrouping estimation responses by subjective responses, both corresponding to the same stimulus, requires caution. As Figures 7.13 and 7.14 show, the selection of descriptors for the data dimension, as part of the matching task, appears to be relatively evenly distributed for the low-complexity stimuli. This nonuniformity of selection also applies to the high-level categories of the descriptors (i.e., gestural, textural, spatial, and cross-sensory) based on spectromorphology (see section 7.2.6). This suggests the general detachment between estimative and descriptive behaviors.

Table 7.13 ANOVA type III (unbalanced) tests with estimation accuracy as the response variable.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Significance
(Intercept)	1	1.27	1.27	1.2901	0.256293	
Model	3	0.09	0.03	0.0878	0.767004	
Mch. Desc. Type	3	12.35	4.11	4.1750	0.005974	**
Model:Desc.	3	3.39	1.13	1.1468	0.329128	
Residuals	1032	1017.35	0.992			

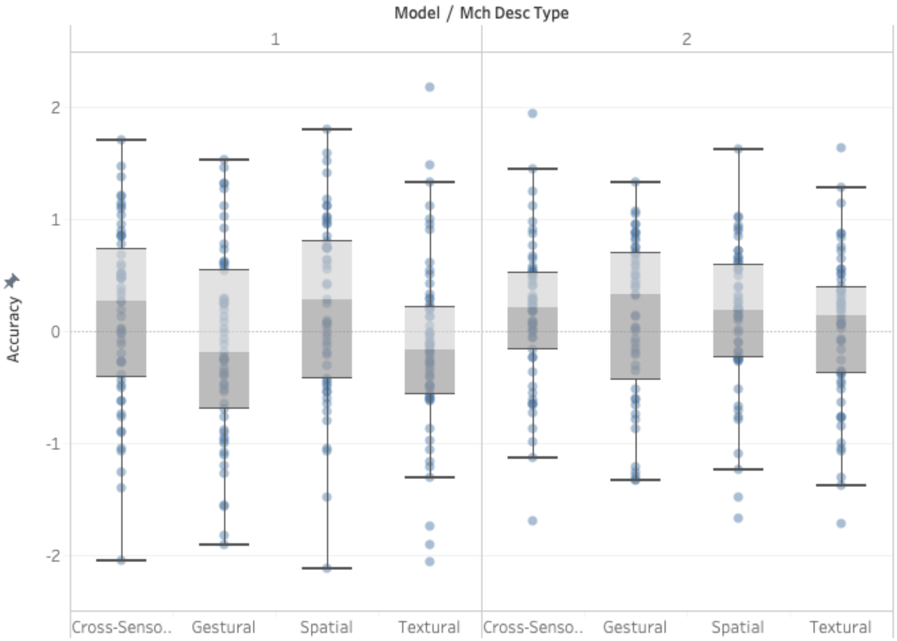


Figure 7.15 Estimation accuracy grouped by the data-dimension descriptor types (columns) and low-complexity stimuli model (left and right).

One-way (by the descriptor types) and two-way unbalanced ANOVA (by models and descriptors; Table 7.13) reveal that the subjectively selected descriptor type overweighs the model differences, affecting the accuracy of estimation[47]. This might perhaps imply that associating the characteristics of a stimulus to a certain descriptor type improves the quantitative estimation. However, there appear to be no clear relationships between the descriptor types and models (Table 7.12 and Figure 7.15) besides the differences are more pronounced for model 1, just as the wider variance with the mapping values.

Table 7.14 Tukey pair-wise comparisons of estimation accuracy between the matching-task descriptor types.

	Mean Diff	Lower	Upper	P Adjusted
Gestural - Cross-Sensory	-0.21772125	-0.42370055	-0.011741957	0.0335298
Spatial - Cross-Sensory	0.03517243	-0.17198349	0.242328358	0.9721148
Textural - Cross-Sensory	-0.15853502	-0.35533862	0.038268590	0.1627930
Spatial - Gestural	0.25289369	0.04623466	0.459552709	0.0091140
Textural - Gestural	0.05918624	-0.13709426	0.255466735	0.8654830
Textural - Spatial	-0.19370745	-0.39122236	0.003807464	0.0569256

Combining the models, pair-wise comparisons (Table 7.14) show two types of descriptors (Spatial and Cross-Sensory) outperforming the other two (Gestural and Textural) by a small effect size of around d = 0.2. The differences within both groups are insignificant.

7.4.4 Findings and Discussion

The statistical analyses explored the first two sub-questions of RQ 1: the listener's estimation performance in relation to (A) the elements of sound organization and (B) the subjective

[47] When subgrouping the accuracies by descriptor types and mappings, however, the mapping overweighs the descriptors while both are being statistically insignificant.

interpretation of the stimulus. For RQ 1-A, the result generally indicated that the structural elements, particularly the location in timbral dimensions where the data are mapped (spectral peaks vs. spectral envelope), had a direct effect on the estimation accuracy. Between the low-complexity stimuli, however, the change of models had less significance, defying the hypothesis. On the other hand, models appear to affect the general efficiency or intuitiveness in aural analyses.

Unexpectedly, the addition of auxiliary motions to the timbre had little effect on the estimation accuracy. While there may be various unaccounted factors, the author speculates several implications: (1) For these low-complexity models, the motions in different synthetic dimensions are reasonably orthogonal to each other as they are perceived by the listener. (2) There may be flexibilities in sound design for adding extra temporal motions without negatively affecting the perceptibility of data.

For RQ 1-B, the statistical results were less convincing, with the DF/EF orders appearing to have little effect on estimation. It is highly possible that the type of subjective analysis task employed was generally not optimal for perceptually deconstructing a complex sound. However, a meta-analysis between the selection of descriptors (for the data dimension) and the estimation performance revealed that some descriptor types contribute to better accuracy scores. When the listener selected either gestural or textural types for the data dimension, they appeared to perform slightly worse. Although it is not straightforward to interpret this result, the author speculates a peculiar pattern between the descriptor types and the timbral motions with data. As defined by Smalley [92], gestural and textural interpretations are more directly associated with the temporal motion of sound contents (e.g., spectra) than spatial descriptions. SMSon abides by this principle when mapping different types of motions to the synthetic dimensions. It allows auxiliary aesthetic motions to use the continuous and more ‘gestural’ time scales, while the data is typically allocated to more stationary segment-wise step motions. The author speculates that listeners who found the data dimension to have gestural or textural characteristics may have, perhaps, confused the continuous auxiliary motions and the stepwise data motions.

7.5 Qualitative Analysis of Listening Experiences

The preceding statistical analyses mainly explored the general influence of different sound organizations and types of listening task on a value estimation. Such factors were also examined with the subject's musical background and preferred ways of interpreting the stimuli, which were treated as preconditions to estimation. With the general difficulty of quantifying aesthetic responses, however, there are likely other unnoticed behavioral traits that interact with the measured responses. The following qualitative inquiries aim to uncover intricate patterns in the listener's personal reflections on their task / listening experiences. The inquiries aim to address all three subsets of the main research question (RQ 1-A, B, and C) but particularly focus on RQ 1-C: what factors might intermediate between the design intentions of an aesthetic sonification and the end-listener's understandings.

The listener responses in general offer a great variety and contrasting perspectives for thematic analysis, but also pose challenges in directly addressing the RQs. One such challenge is the diverging nature of typological analyses with highly idiosyncratic responses. This is mainly due to somewhat open-ended questions asked in the experiment for accommodating listeners with various levels of sound literacy. As a result, the patterns discovered are rather enormous, with each response almost always containing multiple patterns in an overlapping manner. The analysis attempts to frame them into fewer high-level 'themes' in an attempt to fill the spectrum between (1) design elements and intentions and (2) the listener's receptions. These themes for pattern organization are namely data intelligibility, sound complexity, and musical appeals, as previously introduced as evaluative heuristics (chapter 3). Among them, the sound complexity metric may need additional details / adjustment for the listener analysis. Here, in the context of design intentions and receptions, the term is used to gauge and compare the listener's impressions of how elaborate and/or convoluted an electroacoustic sound is. For the designer, this may instead be called an intended intricacy or aesthetic depth that may influence the sound composition for creating simple / primitive or detailed / expressive sonifications[48]. While data intelligibility and sound complexity might perhaps be regarded as opposites, the author speculates that it is possible to create an intelligible sonification with an intricate expression (see chapter 8.6 for continued discussions).

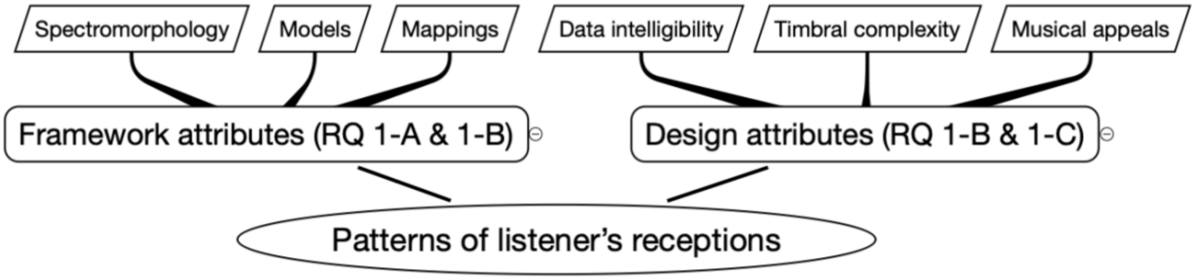

Spectromorphology
Models
Mappings
Data intelligibility
Timbral complexity
Musical appeals
Framework attributes (RQ 1-A & 1-B)
Design attributes (RQ 1-B & 1-C)
Patterns of listener's receptions

Figure 7.16 The problem scopes of the qualitative analyses, which mainly focus on the right-hand side, concerning the highly variable elements of design intentions and receptions.

As alluded to in chapter 3.5.2, the heuristics (or themes) for framing design intentions / receptions need to be distinguished from the fixed structural components proposed in SMSon (see Figure 7.16). The designer would employ the framework concepts, such as models, mappings, and spectromorphological time scale / timbral-dimension guidelines (see chapter 6.1), while also

[48] The following discussions use a more neutral word 'complexity' instead of 'intricacy,' as some listeners find the details of timbre to be rather perplexing.

incorporating their own design decisions, to organize a new sonification. As such, the variable elements related to data intelligibility, sound complexity, and musical appeals are addressed as "design attributes," while the structural sound-organization elements are called "framework attributes." The design attributes for each heuristic are discovered and/or defined along with the patterns of aural analysis.

7.5.1 Subjects Overview and Representative Listeners

Within the 69 anonymous subjects from the statistical analysis, 58 are individually examined for the qualitative analysis of their textual responses to 1160 stimuli. 11 subjects are excluded from the qualitative analysis for providing insufficient textual responses, such as mostly parroting the selected word pairs (e.g., "The sound is very smooth" with the word pair "Smooth-Rough" being selected) or answering with gibberish.

The following sections examine each listener's reflections on (1) overall experience and perceived challenges, (2) the order of interpretive and estimative listening, (3) the utilization of descriptors, (4) the interpretation processes, and (5) the estimation processes. Sections 1 through 3 include group observations that are accompanied by the number or percentage of listeners (out of 69 listeners) and direct quotes from the subjects. Sections 4 and 5, on the other hand, focus on individual responses with thematic analyses.

Some of the quotes are from 'representative' listeners. Six such listeners, labeled A through F, displayed relatively distinct (archetypal) and contrasting characteristics in their responses. This section first summarizes the characteristics of the representative listeners, and as the discussion unfolds with their responses and quantitative data, aim to illustrate high-level relationships between their unique listening approaches and task performances.

Table 7.15 General characteristics of the representative listeners A through F. Raw Score denotes the average and standard deviation of their estimation raw (non-standardized) accuracy scores. DF and EF scores are the average raw scores for the description-first and estimation-first tasks, respectively. (All accuracy scores here exclude the stimulus types 3 and 4, which are always DF.) Desc. and Est. PB are the average playback counts for the estimation and description tasks, respectively. Num. WP denotes the average number of word pairs selected in the description tasks.

ID	Musician	Raw Score	DF Score	EF Score	Desc. PB	Est. PB	Num. WP
A	No	0.695 ± 0.160	0.713	0.677	6.45	7.55	1.40
B	No	0.808 ± 0.075	0.768	0.849	8.90	7.30	2.20
C	No	0.772 ± 0.120	0.807	0.736	7.55	5.80	1.60
D	Flute	0.640 ± 0.173	0.603	0.677	9.85	15.70	2.70
E	Voice	0.861 ± 0.064	0.875	0.847	5.85	2.05	2.05
F	Drum set	0.769 ± 0.119	0.816	0.721	5.40	6.35	1.65

As Table 7.15 shows, the six representative listeners attain the raw average scores in accuracy from above average (> 0.605) to considerably high among others, and also provide informative textual responses (with the average word lengths of textual feedback > 25). The selection of these listeners is inadvertently biased towards better-performing ones because of the correlation between the quantity of textual response and estimation scores (see section 7.3). These listeners form two general subgroups of non-musicians (listeners A-C) and musicians (listeners D-F). The following summarizes the key characteristics of their analytical behaviors.

Listener A, a non-musician, provides elaborate descriptions of stimuli focused on timbral motions. They actively take advantage of interpretive tasks to explore and familiarize themselves with novel sounds. Their choices of word pairs, either only one or two selections, are generally consistent for each stimulus type, suggesting confidence in their interpretations. For many complex stimuli, they recognize multiple qualities / ambiguities. For example, for a thin-rich stimulus, they describe that its characteristics "change from dark to bright but [also] were so distinctly thin in some parts and really rich in others." For musical appeals, listener A generally does not find the stimuli in the experiment to be aesthetically pleasing, as “[they] are not the type of sounds that people associate with popular music."

Listener B, also a non-musician, achieves relatively high and consistent (low deviation) estimation accuracies among all. They particularly perform well in the estimation-first trials, while the (combined) efficiency stays nearly unchanged for both orders. They select either two or three word pairs for all stimuli, but the selections are typically inconsistent for type 1 and 2 models. For many stimuli, they acknowledge the plurality of timbral impressions, such that "all three pairs fit pretty well but I chose Relaxed-Tense because the noisier parts had a very palpable feeling of tension." When asked about the musical appeals of the stimuli, they answer "yes but not all, some were actually unpleasant to listen to. However, some sounds could easily fit into electronic music (which I listen to often)."

Listener C, a non-musician, describes the general process to be "really challenging (only in the way that you really have to concentrate to feel what the sounds are conveying to you) but fun and different." They often display a creative use of the provided word pairs, including the combinations of multiple words, elaboration with their own terminologies, and the use of qualitative descriptors to identify the segments in estimation. A stimulus, for example, "starts with a steadier tone that broadens to static tone but it fluctuates back to the steadier tone and then back to the noisier static." In terms of aesthetic values, the listener finds emotional and cross-sensory effects in the timbres, such that "some of them yes and depending on the what the music is for maybe all of them could be used in a musical piece. Many of the tones convey excitement, harshness, darkness or brightness."

Listener D, a flutist, reveals interesting relationships among musical appreciation, the quantification process, and the characterization of sounds. Compared to other listeners, this listener performs poorly in estimation tasks. Their description-first listening generally decreases the accuracy and efficiency compared to estimation-first, suggesting an added confusion or uncertainty by the qualitative listening phase. They find the general experience of this study to be challenging, as "[this] one was way outside of the box, so I don't mind that it went on rather long. I completed this one for the enjoyment of doing something truly different." This listener also does not hesitate with making negative emotional and meta-experiment remarks. For example, for a response to a trial, they show frustration as "I'm really wondering what the purpose of this study is, exactly." Sometimes even disagreeing with the visual references, they state "I can't really make these clues work, the last red dot is ill-placed, not where it makes sense to me." The listener is also often frustrated with the word pairs being misfit, that "I understand the attempt to provide useful word pairs, but the ones offered only go so far." Their general responses indicate preconceived musical knowledge actively hindering the timbral decomposition.

Listener E identifies themselves as a singer, who displays one of the highest and consistent accuracies in the estimation tasks with only two playbacks on average. They gain slightly more accuracy in description-first trials but are generally critical about the provided word pairs. Their textural analyses are detailed with structural information, such as: "a lot of the word choices seemed to fit this round, but soft-hard seems to fit the most. One sound is very harsh and loud, while the other sound is light and almost pleasant. I'm only moderately confident with the values though because there were two fixed values that seemed very different but were fixed very close to each other on the scale." For the general musicality of the stimuli, the listener finds "maybe they could be used as samples in the background of music, but I can't imagine them as a major part."

Listener F, a percussionist, also attains a relatively high accuracy in estimations, especially in description-first trials. Their interpretations are often extrinsic and metaphorical, almost resembling spectromorphological analysis with various gestural and spatial depictions. For example, "I see this as someone going through a tunnel and someone is opening and shutting the tunnel exit for him to get out. I am most confident in this answer because [I] can see the person in the tunnel seeing the door close on him several times and then the door squeaking open some of the times which resembles the dark-bright word pair." The listener finds the stimuli to be aesthetically appealing, that "yes! Some of them definitely seemed to be from a beat mixer. Especially 16 and 20 in my opinion I could see in a techno song."

7.5.2 General Responses to the Experimental Design and Process

Since the premise of this listening test, a 'timbral sonification,' may have been rather foreign to many participants, it is worth examining what elements are recognized as most challenging or confusing and where in the tasks they have lacked the confidence. This inquiry relates to the difficulty of matching the intended complexity of sound and the actual receptions (RQ 1-C) as well as identifying the potential sound-organization elements that hinder the analytical listening (RQ 1-A). The coded responses reveal that there are roughly equal numbers of listener who perceived three different challenges: the general high complexity of stimuli (25% of the total responses), the generic difficulty of the estimation task (25%), and the elusiveness of textual description of sounds (25%). The spread of challenges shows that no single element alone can explain the irregular or contradictory listener responses, such as the perceived plurality / ambiguity of timbral dimensions.

The perceived complexity of stimuli, or perhaps more of obscurity so as to distinguish from an intentionally designed intricacy, appears to affect both estimation and description tasks. The complexity / obscurity is attributed to either the synthetic models, multiple timbral motions, subtlety, familiarity and musicality. In terms of the synthetic models, the variations (novelty) appear to disorient some listener's expectations, with the challenge being, e.g., "just trying to describe my thoughts for each sound since they were so different and not at all what I was expecting to hear." Occasional insertions of musically complex models (types 3 & 4) appear to greatly affect at least three participants, recalled by one as "[the most difficult part was] when the sounds were segmented within each block. That made it more difficult to determine the pitches[49]." For individual stimuli, the motions in multiple timbral dimensions posed a challenge to at least eight listeners. For example, listener B experiences difficulties "knowing which element of the sound the estimation was tracking," recognizing the multidimensionality of the timbral motions.

The perceived complexity also manifests as a subtlety or unpredictable (unnatural) behavior intrinsic to the stimuli. Despite the designer (author) configuring the timbral modulations to have a minimum parametric deviation of 0.5, four listeners find some of the motions indistinguishable. For instance, a listener reports that "sometimes the sounds remained pretty consistent over the whole time, so I had to listen very carefully to detect what was changing." On the other hand, however, possibly related to the (un)familiarity of musically structured tones, listener C reports contrasting behaviors as "some of … more mechanical or … synthesizer tones are difficult because they seem to be all over the place like up and down." Another listener remarks, "the sounds were kind of confusing because they're unfamiliar." Listener E is more specific about the type of expression, that "the most challenging was the sections that had multiple notes in each segment."

The reported complexity / obscurity of stimuli also reveals its interesting relationships to the musical appeal (aesthetic value) of sounds. Some of the negative responses to musicality, such

[49] This listener appears to intermix the data quantities with musical pitches, as several others do.

as a conflict with their preconceived musical values, appear to hinder the analytical processes. For example, listener D reflects, "I'm a musician, so am used to identifying beats with note length and pitch. This is something altogether different, and quite challenging."

The difficulty in listening also manifests as fatigue. As discussed in the sections Preliminary Tests and Independent Variables, 16 out of the 20 stimuli presented are based on 'lower-complexity' models with a minimum amount of timbral embellishments. This, however, may have resulted in a strained listening experience with rather artificial and repetitive timbral expressions, to which N listeners reported listening fatigue and unfamiliarity. One complains "the most challenging elements were the high-pitched screech noises that sounded like [mosquitoes]. They were horrible to sit through."

Among the listeners who struggled with the estimation tasks, seven listeners indicated a general lack of confidence in the timbre-based quantity plotting in relation to other auditory or non-auditory elements. More than a few listeners, for instance, are confused by the disparity between the perceived pitch and main timbral modulations, stating "[the challenge was] trying to guess the pitches of the notes on the graph and if they were higher or lower" and "definitely the dots since I know my musical pitch isn't the most perfect." For some listeners, the referential segments sometimes become a source of confusion, as "placing the dots was sometimes super difficult. The red dots weren't always the most reliable references and were sometimes flipped." One also indicates the reliance on time-domain waveforms being displayed, writing "the most confusing was trying to determine the estimations. Even basing the blue dots off the red dots was hard because I couldn't really find a correlation even when looking at the width of the sounds in parts." Three listeners expressed frustration in their general efficiency, such that "estimating the sounds was the most challenging because I would often be too slow in recording my estimations and would have to repeatedly listen to it."

For the interpretation of stimuli, many listeners appear to feel unsettled with the gap between the aural and textual expressions. 13 listeners express frustration over the sheer difficulty of verbalizing the perceived timbral events, including the association of those to word pairs. One writes "I think the most difficult elements was my own lack of experience describing sounds and my worries that I was using the vocabulary wrong." Similarly, listener F is ultimately unsure about their selections, as there was a challenge "seeing if any of the word pairs actually matched the sound motion." 15 listeners (partially overlapping with the previous 13) indicate their discontent with the forced selection of word pairs. For example, one states, "I felt that the most challenging aspects were choosing a word pair to describe the subjective differences that weren't really fully captured by the options available," and another writes "I wish I could have [written] in my own elements for the sound." The more detailed responses to the use of word pairs are discussed in the following sections.

One listener elaborates the cognitive dissonance in quantifying / verbalizing their mental image of a timbre in an orderly manner, such that "I didn't find any of it confusing, the directions were clear, and the example was useful. I did find it somewhat challenging though, to apply these types of thought to noises, like more tactile descriptions or thinking of noises in a more concrete form, I had to, I feel like, access the part of my brain that didn't analyze too much, that would just listen and feel, and then move into explaining that feeling after that." This response indicates the potential risk of a textual sound analysis being completely retrospective, not contributing to the intuitive understanding of a complex timbre (for a better value estimation).

7.5.3 Order and Interactions Between Interpretive and Estimative Listening

This textural analysis also revisits the question of if the order of interpretive and estimative listening influences the overall perceptual decomposition processes (RQ 1-B). The study hypothesizes that focused qualitative listening should generally contribute to the accuracy and/or efficiency of the estimation tasks with a complex sonification. On the contrary, the survey shows that 39 listeners (56%) find no perceptible difference in interpretation / estimation orders. Nine listeners (13%) also prefer estimation-first, while only six (8%) prefer description-first listening for comfortably quantifying the sounds.

Among the listeners who indicate that the order does not matter, four listeners focus on the auditory stimuli themselves but not their approaches in listening, claiming the stimulus and references carried most of the information necessary for decomposition. One writes, e.g., "for me, it was really the sound itself, as well as the position of the points that I could not move." Listener F also recounts, "[the order] did not affect the difficulty in any way. I saw the entire sound motion task the same difficulty throughout." Listener C points out the redundancy of description tasks, writing "[description tasks were] not really [helpful]. I preferred to do the value-estimation first because it had me thinking of descriptions already." For both listeners F and C, however, their estimation accuracy is significantly higher with description-first, despite not perceiving a difference (see Table 7.14). Two listeners do not benefit from the description phase as they rely heavily on, or are occupied by, the visual-based analysis approach. As listener D reports that the order had no effect as they "primarily relied on the dot placement, but didn't understand the logic of some of the red locked placement." Their DF accuracy is heavily hampered compared to EF tasks.

Listener A also elaborates the lack of effect on the estimation efficiency, claiming "I didn't notice it taking me longer to come to a conclusion based on which I did first." However, they also remark that "sometimes moving the blue dots to fit the sounds I would have to listen to the sounds quite a bit until I was satisfied," indicating the lack of efficiency by the estimation task itself. Assuredly, they achieve somewhat better performance in DF trials. Similarly, listener E indicates that they benefit from a repeated listening, such that "description-first might have been easier since in the description I listened to it more times than in the value task." They also gain some accuracy in DF. In contrast, for listener B, the DF accuracy is considerably decreased from EF despite reporting that "it felt easier to do the description first[. T]he estimation was typically harder so getting more familiar with the sound before doing it seemed to help."

7.5.4 The Interpretation and Utilization of Sound Descriptors

RQ 1-B and 1-C concern how to facilitate and observe the qualitative interpretation of complex timbres by the end listener. As discussed in section 7.2.6, this study employs 15 word-pairs that are intended to have multiple implications such as the anchoring effect and the relatability for non-musicians, moderately stimulating their analytical listening. In the survey, listeners expressed either positive (23 subjects), mixed (12), or negative (20) reactions to the provided word pairs, where each may be explained in terms of reinforcing, classificatory, or eliciting values.

7.5.4.1 Reinforcing Effects

The reinforcing, or "quantifying[50]," effect of descriptors may help the listener verbalize their understanding of sound. Five listeners report positive reinforcing effects, that descriptor(s) "helps to characterize the sound track," "… were generally helpful and helped me to visualize the sounds in my mind better" and "… were helpful because it let me quantify what I was feeling." Listener F, who utilize the word pairs often in spectromorphological ways, also affirms, "yes they were! I could see more clear images in my head of how the word pairs actually fit the sound motions. None of them were misleading." On the other hand, at least five listeners reported that the provided word pairs are rather imprecise in quantifying in the sense of range and directionality. For example, two listeners mention the one-sidedness of timbral motions, such that "sometimes … there was only one word in a pair that accurately described aspects of the noises I was hearing." Similarly, for listener B, "[the descriptors] were somewhat helpful but often one word fit very well while the other word in the pair either fit not at all or only in a vague sense." Listener E finds a mismatch of the ranges, describing "a lot of the sounds weren't really opposites, which is what the word pairs were." In addition, two listeners are confused by the directionality the word pairs indicate, reflecting "they were helpful, but at the same time, I wish they were reversed. I feel like the reverse sometimes fit the sounds I heard better."

7.5.4.2 Classificatory Effects

For the usefulness of word pairs, the mention of classificatory value predominated (52%) with nine positive, three neutral, and 24 negative responses. The positive responses include, e.g., “the word pairs gave me a wide range of selections to choose from and there [were] always choices

[50] This is a different type of quantification than a value estimation process.

that matched” and “they were mostly very useful, because [without,] I wouldn't have been able to really come up with descriptive words that were more accurate.” Two listeners preferred to categorize iteratively with the provided list, explaining "they helped me group my thought together easier when comparing what I was listening to against the different word pairs," and "it was very helpful to listen to the sound on loop and keep looking at the choices. It helps to scan the choices while on loop" (listener C).

The negative and mixed responses, on the other hand, illuminate several distinct issues with the provided descriptors. 13 listeners (18%) find either partial or general disparity between the timbres and descriptors. For example, the available descriptors are too limiting to some listeners, that "most of the time they did not fit any of the sound motions, so I was forced to make selections I did not truly believe fit the sound motions," "most of the time they didn't exactly fit for what I was hearing. Like "natural-metallic" when everything sounded artificial. … [I]t might have been better if I could have chosen individual words" (listener E), and "I felt too restricted and they did not seem properly descriptive to what I was hearing." To some listeners, the stimuli themselves also contribute to the descriptors being misfit, that "the sounds were hard to describe to me and some of sounds didn't go along with any of the options," and "most of the words selections didn't match the sounds because almost all of the sounds were loud or soft and very synthetic." Listener D also points out the classificatory limitations, reflecting "honestly, these [particular] pairs were more limiting [than] helpful. Words like ascending, descending would have been good. Maybe just too subjective? Square peg round hole?"

7.5.4.3 Eliciting Effects

While the classificatory approach may be limited only up to selecting the best-fit descriptors, the word pairs sometimes inspire the listener toward further analyses of the timbral details. For several listeners, the word pairs function as an analytical device for closely investigating the sounds, with the variety of descriptors eliciting unnoticeable characteristics. For example, listener A reflects "I think [the word pairs] were helpful because they helped me examine the sounds more closely as I selected which ones were the best choices," while another utilize them with an iterative and reinforcing manner that "[a word pair] is very helpful to [analyze] the sound description, because I compared the sound with the word pair, then I have [a] clear idea to proceed." On the other hand, the eliciting effects work on some listeners in misleading or conflicting ways. Some, for example, pointed out the inexactness when interpreting the words, that "I think it was misleading because everyone has their own definitions for words so I wasn't sure I was using them correctly." Another, however, views in a reverse way, that "I did not think they were misleading because they were up to me to decide on in the first place." These views highlight the future need for a more precise definition of the role of textual materials.

7.5.5 Patterns in Subjective Interpretations of Timbre

From the experimenter's viewpoint, the word-pair selection in a description task is merely an entry point for analytical listening, where the listener records a quick impression of a newly encountered stimulus. The matching task, on the other hand, demands further details of stimulus impression and listening experience. This inquiry particularly concerns RQ 1-B – how different interpretations affect the estimation performance. For this reason, the following observations include the quantitative estimation results (the raw / standardized accuracy and efficiency scores) together with corresponding quotes.

Table 7.16 The measured responses for selected individual trials. Listener Ref. provides unique identifiers corresponding to the quotes (referenced in the text as, e.g., Table 7.16.1). Representative subjects are indicated with "(letter ID)". Stimulus denotes the model-mapping combinations (with D, C, and A respectively indicating Data, Constant, and Auxiliary modulations). "First" indicates the within-trial task order, either estimation-first (Est.) or description-first (Desc.). Word pair denotes the selection for the primary data dimension. Raw score (0-1) indicates the non-standardized accuracy of estimation. The standardized accuracy (the within-subject t-score) generally spans between -1.5 and 1.5, with a positive value indicating a better performance among tasks and 0 indicating the average. The same applies to the efficiency score, which describes the total standardized duration spent for the estimative + interpretive analyses.

Listener Ref.	Stimulus	First	Word Pair	Raw Score	Accuracy	Efficiency
1 (B)	2-AD	Est.	Close-Distant	0.8059	0.215	-1.133
2 (B)	3-DA	Desc.	Warm-Cold	0.7008	-0.625	-1.426
3 (E)	1-DC	Desc.	Dark-Bright	0.9412	1.147	0.997
4 (C)	4-AD	Desc.	Stable-Noisy	0.4936	-1.330	-0.417
5 (C)	2-DA	Est.	Steady-Fluctuating	0.7777	0.399	-1.915
6 (C)	1-DC	Desc.	Narrow-Broad	0.7955	0.507	0.388
7 (C)	3-DA	Desc.	Dark-Bright	0.5664	-0.887	0.854
8	1-DA	Desc.	Thin-Rich	0.8926	0.558	-1.069
9 (A)	1-DC	Est.	Soft-Hard	0.3819	-2.055	-0.606
10 (B)	2-DC	Est.	Stable-Noisy	0.7777	-0.835	0.190
11 (C	1-CD	Est.	Clear-Muffled	0.7698	0.351	0.557
12	4-DA	Desc.	Clear-Muffled	0.4273	-1.733	1.177
13 (A)	2-DC	Est.	Gentle-Metallic	0.5393	-1.044	-1.943
14	4-DA	Desc.	Dark-Bright	0.8555	0.820	-1.331
15	2-AD	Est.	Gentle-Metallic	0.8076	0.342	-0.597
16	1-AD	Est.	Dark-Bright	0.9313	1.123	0.618
17 (D)	1-AD	Desc.	Soft-Hard	0.8283	1.229	-1.305
18 (D)	1-AD	Est.	Smooth-Rough	0.5841	-0.271	0.660
19 (F)	1-DC	Desc.	Relaxed-Tense	0.7395	-0.156	-0.354
20 (F)	1-DC	Est.	Clear-Muffled	0.5988	-1.093	0.204

In the matching task, the listener writes a reflection about the audio stimulus and/or their analytical processes in the current trial. The question provides flexibility for the style and content of the response (considering different levels of auditory / musical literacy), but many listeners tend toward elaborating their experience in the description task. Such responses generally contain a mixture of (1) the elaboration of the word pair(s) they have selected for the timbral dimensions and (2) detailed analyses of sound in intrinsic or extrinsic manners. Many also concluded the explanations with their confidence for the task results.

7.5.5.1 Elaboration of the Descriptors

Elaborations of word pairs are quite diverse in types, but often contain either (1) the assessment of their functionalities, (2) acknowledging a timbral plurality or ambiguity (in relation to the provided word pairs), or (3) the introduction of the listener's own terminologies. Other recurring patterns, e.g., the application of words to different types of motion, are discussed in the next section in the context of estimation processes.

First, many of the responses to the functionality of word pairs expand the previously discussed classificatory values, such as the contrast or one-sidedness, fitness, and directionality. The majority of detailed responses are critical to these functionalities. Listener B, for example, struggles with the lack of contrast in a stimulus, that "[m]ost of it sounded about the same for the estimation part and it made picking a pair tough because it only fit one or the other." The corresponding estimation result (Table 7.16.1) indicates that the accuracy is above average, but the efficiency of estimation and/or interpretation is considerably hampered. The same listener is then confused with the directionality of a word-pair continuum, where "I'm not sure which end I pin to warm and which [end to] cold." The estimation following the description task appears to be severely disrupted for both accuracy and efficiency (Table 7.16.2). Listener E compromises with their selection as "I chose dark-bright because one of the sounds was higher pitched than the other. They were both very similar though. I'm not at all confident on the values since I couldn't tell very much of a difference between the tones." Contrarily to their unsureness, their approach to mapping the spectral peaks (Quasi-Harmonics) to the pitch appears to yield a very high estimation performance (Table 7.16.3).

Some listeners elaborate a word pair with motional analogies. For example, Listener C finds "it has a more focused narrow tone that almost explodes into a broad noisy tone." This turbulent sensation, however, appears to result in a highly inaccurate estimation (Table 7.16.4). The same listener also combines multiple descriptors to depict timbral states and motions, that "it starts with a steadier tone that broadens to static tone but it fluctuates back to the steadier tone and then back to the noisier static." With the less-turbulent description of motion ("fluctuates"), the estimation retains accuracy but still considerably suffers in efficiency (Table 7.16.5). Listener C also struggles with a better description of a particular timbral motion, such that "the first tone is narrower and then it opens up to a much broader almost vibrating sound. At first I thought steady-fluctuating but the vibrating isn't really fluctuating because it's more constant." Despite the unsureness of the word selection, the estimation performance further improves with the depiction of a 'predictable' motion (Table 7.16.6). These examples highlight different implications of the use of a word (not necessarily a pair) to describe a timbral state (an instantaneous quality), a motion between states, and a repetitive pattern over time.

The complexity of timbre is also recognized as a plurality or ambiguity of characteristics. A common response is the mixed use of multiple word pairs. For example, listener C is overwhelmed with the multiplicity of a stimulus' characteristics, describing "it's all over the place. From softer to more harsh and darker to brighter. [S]ome tones seem … steady while others are more mechanical." Implying frantically altering timbral quality over time, the following estimation task suffers significantly in accuracy (Table 7.16.7). One listener finds "the two sounds sounded nearly identical to me, one seeming to overlay the other at points which I guess made it richer." This response first points out the subtlety / lack of contrast between two timbral states, and then maps the inherent timbral plurality to a word pair. Although this aural analysis appears to take more time, the estimation (after the description phase) achieves a high accuracy (Table 7.16.8). With a complex multidimensional timbre, many listeners experience an ambiguity where precisely mapping a timbral quality to a word pair becomes difficult. Listener A reports, "while there were parts that sounded smooth and rough, the sounds were also soft and hard." Such ambiguity of mapping may be attributed to the inseparability of timbral characteristics from one another and/or the unfitted-ness of a word pair (as might be expected with the classificatory limitations). Interestingly, the accuracy of estimation (preceding the description task) for this stimulus also suffers significantly (Table 7.16.9).

With ambiguous or complex timbres, some keen listeners attempt to give nuance to their observation by incorporating idiosyncratic expressions. Listeners elaborate, e.g., their word-pair selection, various timbral states in the stimulus, and intricate motions connecting the timbral states. Their original descriptors typically augment or adjust the meanings of provided word pairs, while some phrases are reused for multiple stimuli / trials to possibly denote their similarities. Listener B, for example, describes the sound segments of a stimulus with a simple augmentation, that "a couple pieces are simple (stable) while the others are very busy (noisy)." This analysis, perhaps rather lacking finer details, follows a poor estimation accuracy (Table 7.16.10). Listener C, for many of their responses, prefers to mix in "piercing," "harsh," and "unpleasant" with the provided descriptors to characterize the slight nuances of timbral events over time. For example, a stimulus "definitely has a clearer slightly piercing tone in the beginning and then goes into a more muffled almost vibrating tone," while another "has a clearer more piercing tone to a slightly muffled sound that repeats in this order." Both stimuli (Table 7.16.11 and 7.15.12 respectively) are responded with adequate efficiency, though the latter with complex model and mapping attain a very low accuracy.

7.5.5.2 Extrinsic References

As an interpretation becomes more complicated for time-evolving sound expressions, many listeners opt to utilize intrinsic or extrinsic references for denoting sound characteristics. Adapting Smalley's depiction [92], an intrinsic interpretation aims to find structural relationships within an organized sound, while an extrinsic interpretation incorporates external factors (e.g., sound sources) to explain the sound. The thematic coding reveals that qualitative interpretations of timbre are largely extrinsic, with the frequent use of (1) physical or gestural references, (2) cultural references, and (3) traditional musical concepts. Intrinsic analyses of timbre, on the other hand, are generally tied more to quantitative estimations, and therefore will be discussed in the following section instead (see Patterns in Estimation Processes).

The first type of extrinsic references concerns the combination of a physical sound source and how it generates sounds or is excited by a physical action. For example, listener A draws a word 'metallic' (of 'natural-metallic') to embody the image of a sound source along with an indication of physical motion, as "there was very distinct metal sound like something hitting something metal, so I was confident it was the best choice." The preceding estimation, however, suffers in accuracy, and the low overall efficiency also indicates the struggle in their analysis (Table 7.16.13). Another listener describes a complex timbral texture in a metaphoric manner, that "to me these sounds seem bright like the sound of a powerful fluorescent tube light pulsating. I am not very confident in these choices because these sounds are something which I am not familiar with." Interestingly, this listener applies the word-pair 'dark-bright' to a physical light source in a cross-sensory manner (as opposed to the pitch association of Table 7.16.3). Although taking a long time to analyze, this qualitative listening leads to relatively high accuracy in the following estimation (Table 7.16.14).

Further extending physical and gestural references with complex stimuli, some listeners exercise their aural imagination with cultural references or depictions of a soundscape with various environmental sound sources. This approach, although may appear inefficient with the number of metaphors employed, uniquely establishes the context of smaller individual sound events. For example, a listener describes "it almost feels like I am inside a factory when I listen to this sound clip. I can hear the crunching of the metal throughout the clip and then it alternates to something else, it almost feels "sandy." The pitches were tough to determine on this one. I feel more confident in the pair than the placement of the blue dots, but this is as good as I think I can do it." Highly focused on describing the timbral nuances (although rather trying to estimate the 'pitch'), the preceding estimation scores a high accuracy with a slower analysis speed (Table 7.16.15). Another listener provides a more cultural interpretation of sound events, where "all I saw was something bad happening in a video game, and then it turned into something good, and it just keeps repeating over and over. I saw a dark-bright approach more than a warm-cold just because I can see the heavier pitches in the sound motion to be more of a stopping force or boss in a video game than to where the soft ones were something good happening in the game." The description includes various qualitative perspectives, but with the main 'dark-bright' sensation mapped to data, the preceding estimation attains a very high accuracy (Table 7.16.16).

As another type of extrinsic reference, some listeners prefer to associate the timbral characteristics to familiar musical instruments. Similar to non-musical objects, this approach seems to necessitate an elaboration with how the instrument is used or a juxtaposition with other sound-source references. Listener D, for instance, elaborates a word pair with a depiction of cello performance and environmental sound sources, as "this is actually soft hard, it makes me think of some quick notes on the cello. Also a bit like bees in a tin can. Intermittently." The following estimation yields a high accuracy, but the overall analysis appears to take a long time (Table 7.16.17). For the same stimulus type in a reversed task order (estimation first), listener D applies a similar description "bees, steam, and cello all come to mind while listening to this. Almost a horn sound on the deepest note." However, the preceding estimation accuracy is lowered considerably while gaining the analysis efficiency (Table 7.16.18). Listener F references a musical instrument with a vivid depiction of a gestural motion – "I saw someone playing [the] violin and trying to tune it and get ready for a song. I see it more as relaxed to tense more than anything because most parts of the sound motion were relaxed, but there were two that were pretty tense and seemed to not go with the rest of the sound motion." The following estimation attains a little below their average accuracy (Table 7.16.19). For the same stimulus type in the estimation-first setting, however, they struggle in both estimation and description, stating not being able to describe the characteristics (Table 7.16.20).

7.5.5.3 Interpretation-Design Attribute Relationships

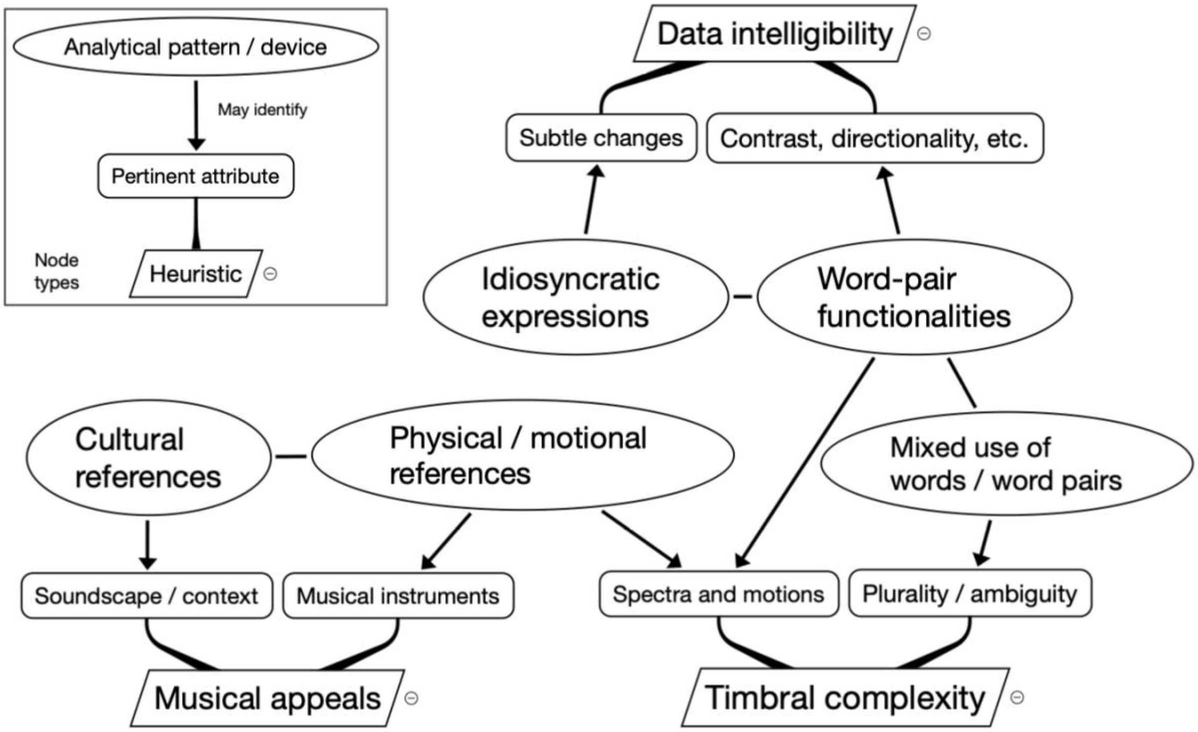


Figure 7.17 The observed relationships between the interpretive patterns and design attributes. An analytical pattern or device may identify certain attributes, but likely not all of the connected attributes simultaneously as indicated.

As shown in the graph summarizing the pattern-heuristic relationships (Figure 7.17), in general, several distinct analytical patterns interrelate with design attributes (i.e., variable characteristics of stimuli addressed by / relevant to a listener's analysis, denoted as 'pertinent attributes') as well as with other interpretive patterns. (Note that the graph only illustrates the most direct connections, while there are undoubtedly weaker / indirect connections among the interpretive patterns. These patterns, as mentioned previously, are also employed in a different combination for each interpretation.)

The observed interrelationships may provide some guidelines for how to design, communicate, and understand the perceived effects of a timbral sonification. At a higher level, some interpretive patterns, such as the use of cultural references, multiple word pairs, and elaboration with personalized expressions, appear to have a one-to-one relationship with a specific design attribute. On the other hand, physical / motional references and the elaboration of word-pair functionalities may potentially address and connect multiple heuristics, whether intended or not. With these observations, the designer may be able to anticipate unintended influence on the listener's analysis and their focused heuristics. For example, one might try to create a sonification that features multiple concurrent motions of timbral features – e.g., oscillatory and transient modulations. To communicate / explain how this sonification is structured, the designer may incorporate references to a sound source (e.g., a virtual wind instrument) and how it is excited (e.g., with a swelling blow) to draw the listener's attention to the continuous timbral changes, and use additional descriptors (e.g., smooth-rough) to signify the discrete changes of a textual quality. Such a word pair may be accompanied by suggested high-low associations to ensure some level of data intelligibility.

The observed patterns also facilitate the identification of relevant design attributes for each heuristic. For data intelligibility, this includes the overall and variable characteristics of a stimulus, each with a unique range of expressions. For this, many listeners opt to map a word or word pair (e.g., 'stable') to describe the 'averaged' characteristics of the stimulus, and then try to identify subtle motions or low / high contrasts utilizing other descriptors (e.g., 'fluctuating,' etc.).

Comparably, for sound complexity, the use of multiple word pairs typically indicates the plurality of characteristics. This is often carried out with the decoupling and recombining of a word pair in various creative ways. With timbral interpretations, however, it is often difficult to determine where exactly the perceived plurality resides – whether in different time locations (segments), in separate spectral regions, or simply as concurrent multiple qualities. In this regard, the analyses in the context of value estimation provide more concrete information with specific attributes (see the next section).

For musical or aesthetic appeals, the interpretive patterns suggest the significance of general familiarity or relatability of a stimulus. These qualities may help the listener create an easier association of timbral characteristics to existing musical instruments or imaginative soundscapes. The mentions of musical instruments largely mirror the use of extrinsic descriptors addressing sound complexity. Extrinsic references are utilized to describe complex spectra (instantaneous states) and spectral motions in a concise way. In a musical association, the reference to a physical sound source similarly defines a spectral profile (e.g., an instrument), while the gestural motions often correspond to a musical performance.

The present listening experiment provided the word pairs with a heavy focus on facilitating the analysis of variable timbral states. However, as the multiple observations signify, it may be beneficial to distinguish such descriptors into static and dynamic types informed by physical-source and human-gesture archetypes.

7.5.6 Patterns in Estimation Processes

Continuing on the observations of listener responses to individual stimulus / trial, the matching task results are also examined in relation to value-estimation strategies. The analyses here highlight various ways of how the listener understands a sound structure (RQ 1-A) and describe the data intelligibility and sound complexity (RQ 1-C). Generally, textual responses include fewer mentions of quantitative observations than qualitative descriptions of timbral characters. Yet, the line between these listening modes is rather fuzzy, as many listeners utilize the qualitative aspects of sound for depicting the process of value estimation. All the quotes in the following discussion are taken from the context of estimation processes (e.g., accompanied by a mention of estimation confidence, which may not be included in the quotes).

The experimental design presumes that an estimation process involves the listener to identify the 'primary' timbral motion created by the mapped data (among multiple simultaneous motions) and adjusting the quantity slider for data-point segments relative to each other while orienting and scaling them based on the sound / visual of the referential segments. However, hardly any listener illustrates their estimation process in such a systematic manner. Instead, they focus on one or several observed elements, or their own analytical techniques, that contribute to their confidence in estimation.

Table 7.17 The measured trial responses accompanying the selected quotes. See the previous Table 7.16 for the interpretation of the fields.

Listener Ref.	Stimulus	First	Word Pair	Raw Score	Accuracy	Efficiency
1 (C)	2-DC	Desc.	Stable-Noisy	0.7528	0.247	-1.152
2	3-CD	Desc.	Close-Distant	0.8806	1.025	-0.29
3	1-DC	Desc.	Dark-Bright	0.4884	-0.943	0.629
4	2-AD	Desc.	Smooth-Rough	0.8582	0.878	0.688
5	1-DA	Est.	Steady-Fluctuating	0.5938	-1.538	-0.912
6	1-AD	Desc.	Thin-Rich	0.6441	-0.164	0.980
7	3-DA	Desc.	Dark-Bright	0.8795	0.595	-2.004
8	1-DC	Desc.	Stable-Noisy	0.6202	-1.476	1.300
9	4-AD	Desc.	Soft-Hard	0.7028	0.226	0.592
10 (D)	2-CD	Est.	Smooth-Rough	0.7153	0.535	-1.239
11	1-DA	Est.	Narrow-Broad	0.5414	-1.688	-0.416
12	2-AD	Desc.	Clear-Muffled	0.8992	0.733	-0.062
13	1-DC	Desc.	Gentle-Metallic	0.6000	-0.918	-1.101
14	1-DC	Desc.	Steady-Fluctuating	0.5448	-0.77	0.779
15	4-DA	Desc.	Steady-Fluctuating	0.3563	-1.708	0.655
16 (F)	2-CD	Est.	Narrow-Broad	0.7908	0.186	0.65
17	3-AD	Desc.	Close-Distant	0.5560	-0.701	-1.113
18	2-CD	Est.	Relaxed-Tense	0.7956	0.501	0.107

Table 7.17 continued

Listener Ref.	Stimulus	First	Word Pair	Raw Score	Accuracy	Efficiency
19	2-DC	Desc.	Stable-Noisy	0.9073	1.157	0.221
20 (B)	3-AD	Desc.	Steady-Fluctuating	0.3354	-3.546	-0.449
21 (B)	1-DC	Est.	Pleasant-Unpleasant	0.8342	0.441	0.457
22 (B)	2-DC	Desc.	Soft-Hard	0.7911	0.096	0.114
23	1-DA	Est.	Thin-Rich	0.4002	-1.536	0.082
24 (E)	2-DC	Est.	Stable-Noisy	0.7777	-0.835	0.190
25 (E)	3-DA	Desc.	Dark-Bright	0.7495	-1.176	-0.004
26 (D)	2-DA	Desc.	Gentle-Metallic	0.3945	-1.436	0.158
27 (D)	1-CD	Desc.	Pleasant-Unpleasant	0.9032	1.689	-2.202
28 (A)	1-DA	Desc.	Steady-Fluctuating	0.6614	-0.259	1.288
29	2-DA	Est.	Clear-Muffled	0.8016	0.18	-0.012

The responses show that estimations are frequently carried out with intrinsic comparisons of the parts of sound. With comparisons, the listener typically identifies relative differences, recurring patterns, and/or anomalies. Intrinsic references may be broken down into several types of temporal / spectral analysis: (1) segmental (data point) comparisons, (2) the analysis within / across segments, and (3) cross-reference to previous stimuli or trials. Among these, the responses are predominantly about segmental observations. This is, as might be, expected with the estimated values each being mapped to a sound segment implicitly by design.

7.5.6.1 Segmental Comparisons

Segmental comparisons are elaborated with detailed analyses focused on degrees, temporal patterns, referential tones and/or visuals. Such approaches are, again, employed in various mixtures as they are often complementary. Listener C, for instance, describes the alignment of the timbral qualities of segments with referential segments, such that "the first tone is more stable and fluid and then it is noisy with a static tone. I lined up the first tone with the later red dot that shares that tone." Although the analysis takes a relatively longer time, the estimation following attains an above-average accuracy (Table 7.17.1). Similarly comparing the characteristics of segments, another listener comments, "the last 2 blue dots are similar to the first red dot. The tone starts heavy then light. The 3rd and 4th dot are almost a whispy tone." Following this depiction, they achieve considerably high accuracy in estimation but with somewhat low efficiency (Table 7.17.2).

Segmental analyses focused on assessing vertical degrees may be roughly divided into two categories: (1) comparing spectral or other timbral regions and (2) pitch-oriented estimations. For the former category, the listener often aims to establish the directionality and a rough assignment of timbral qualities to low / high regions. For example, one describes, "[t]he lower notes sound darker and more complex than the higher notes which in turn sound brighter." While acknowledging that a part of the sound is "more complex," the following estimation does not seem to align with the scaling they have established and achieves a considerably low accuracy (Table 7.17.3). The same listener, however, applies a similar approach with more confidence and details of multiple timbral qualities, that "I think smooth-rough is the best choice for this one. The higher it goes, the rougher it feels. The lower notes are a little muffled but feel smooth." This leads to estimation with high accuracy (Table 7.17.4).

Pitch-oriented estimations of timbral motions are distinct among others (partly as they contradict the user instruction discouraging this approach). Even so, there appear to be multiple implications of aural analysis within this approach, where some are a deliberate application of symbolic musical structure (i.e., notes), while others are naive or spontaneous intermixing of timbral features, perceived pitches, and the visual graph (spatial degrees). The former type is examined shortly in the context of the usage of musical knowledge (see section 7.5.6.5). The latter, a possible mixture of timbre and pitch with different levels of awareness (that some are deliberate), is observed very frequently. One naive example shows, "I interpreted it based on the pitches. To me, they sounded very robotic like." The preceding estimation falls severely short in accuracy and efficiency, perhaps indicating the total offset of the data dimension from the perceived motion of pitch (Table 7.17.5). This listener is one of a few that included the word 'pitch' in every textual response with the possible conceptual mixture with timbre and visual cues. Interestingly, however, their average estimation accuracy attains 0.7652 (in raw absolute score), a relatively high result even compared to some of the musically trained listeners. Another listener also prefers to use 'pitch' to denote a voice, such that "[the stimulus] starts off with a thin and narrow pitch then gets low and broad and finishes thin and narrow." The following estimation accuracy is a little below average (Table 7.17.6). One listener opts for a pitch analysis as they are unable to capture timbral differences, explaining "sounds like a nice tune to me this time. The only thing varying really is the perception of "pitch" as we move along the melodic line. The[re] was no pair like this, so I tried to pick the close[s]t thing." Even though struggling with the timbral analysis (low efficiency), the following estimation achieves a relatively high accuracy (Table 7.17.7). In contrast, another listener is prompted toward timbral analysis because of unchanged pitches, that "the more stable sounds, as the dots get lower, are less fuzzy and unclear. I'm fairly confident in both this word choice and my estimations because the difference is pretty clear to me, particularly because the sounds seem to have closer to the same pitch so the other subjective differences in timbre stand out more." Their confidence in timbral interpretation also shows in the high efficiency of analysis, although it does not translate to an accurate estimation (Table 7.17.8). Several listeners acknowledge that the pitch information is not the main point of interest and try to focus their attention on the timbre. For example, "one of the sounds that instead of being a high or low pitched, actually went into being softer or harder, which made rating the values different." With this observation, the following estimation achieves a moderately accurate result (Table 7.17.9).

Listener D also states, "I'm listening to the sound quality of each beat - both the 'pitch' and the roughness/smoothness of sound." This simple remark precedes an estimation with relatively high accuracy (Table 7.17.10). Lastly, there are many listeners who utilize both a rough sense of pitch and timbral descriptions to better identify / communicate the segments of interest. Such segments are commonly addressed as lower- / higher-pitched tones. For example, a listener describes, "looking at the values I have assigned in the graph, it looks more like it is going from broad to narrow, but listening to the sounds, I think I interpreted higher pitch with narrow, and lower pitch with broad." This mapping attempt, perhaps reflecting their sensory conflict, does not yield an accurate estimation (Table 7.17.11).

7.5.6.2 Recognition of Continuous Motions

While the majority recognizes the change of timbral state in a per-segment basis, some listeners describe or imply the sense of continuous motions within or throughout multiple segments. The recognition of such granular motions may be attributed to repetitions and variations. Nine listeners spontaneously use the term "vibration" to describe, most likely, the sinusoidal oscillation applied as an auxiliary modulation. For example, a listener reports "many of the sounds had no vibration in them while others did," indicating that some segments had an additional auxiliary motion more noticeable than others. This perceived intensity of motion is followed by estimation with very high accuracy (Table 7.17.12). Another listener associates the oscillation to a timbral quality, explaining "this sound was definitely gentle but with a metallic vibration to it. I felt like this one was easier to align the dots because several of the pulses were very similar." This description does not imply a temporal growth of the 'vibrating' quality, and interestingly, scores considerably poorly in their estimation (Table 7.17.13).

On the other hand, some perceive non-oscillatory motions within or extending the segmental time scale. The word pair 'steady-fluctuating' is commonly used to convey an irregularity or the novelty of events. For example, a listener applies this descriptor to the whole of a sound sequence, explaining "there is a steady tone throughout the set, but it also fluctuates a bit throughout it." This perceived "fluctuation" may be attributed to the inharmonicity of spectral peaks modulated by data, though such recognition does not appear to contribute to their estimation accuracy with a considerably low score (Table 7.17.14). Another listener, in contrast, applies the same descriptor to the segmental time scale, describing "each sound had a steady fluctuating tone within it." Again, the accompanying estimation is considerably poor in accuracy but appears to be more efficient (Table 7.17.15). In either case, the word pair is used as if both qualities coexist at any given time rather than alternating / transitioning in a gestural time scale, suggesting a form of timbral plurality.

Another example of recognizing a motion within a segment is the application of a word-pair continuum to each segment as a (fixed) motional pattern. Listener F recounts, with an interesting misunderstanding of the word-pair directionality (see section 7.2.6), "[t]his is more a hard-soft than a soft-hard. I was going to choose that one, but that wasn't one of the choices (hard-soft). I saw someone hitting two big pans together the entire time and then they just scraped one time and then went right on to doing it again." This illustration contains multiple factors – the time scale of the timbral motion is that of a human gesture with "hit" and "scrape." They also mention a physical sound source ("two big pans") creating narrow-broad variations of the projected acoustics. The estimation preceding the description scores a moderate accuracy, with high efficiency for the overall analysis (Table 7.17.16).

7.5.6.3 Cross-References Across Trials

Comparisons to previous stimuli (trials) are somewhat infrequent but are used to depict the impression of (dis)similarities or changes of characteristics. A listener recalls, "it sounded much more mechanical than the previous sounds so far. It reminds me of the kind of sounds you would hear coming the coils in a circuit board. It has a very digital sound to it given the square wavelength. Despite this I thought close-distant was the best choice because the sounds don't change much except for their volume. Quieter is more distant, louder is closer. I think this made the number estimations somewhat simple." This illustration is rich with multiple perspectives such as physical (metaphorical) sound sources, a reference to the visual waveform, and the elaboration of the word pair. However, given the stimulus type with high complexity, the following estimation does not align accurately with the actual data dimension (Table 7.17.17). Another listener details the temporal characteristics of the current and previous stimuli, as "this was somewhat similar to a previous sound in that I equated the static sounding initial part with ‘relaxed’ and the gunshot-like middle part with ‘tense.’" With the articulate account of mapping, their preceding estimation also shows a sufficient accuracy (Table 7.17.18). The same listener, again, recalls that "this was somewhat similar to a previous sound and I think the stable-noisy description also works well here. The stable portion is in the fifth segment and it gradually becomes more and more noisy. I think my estimations map well with the sound as the references indicate clear demarcations." The estimation reflects the confidence in analysis with very high accuracy (Table 7.17.19).

7.5.6.4 Multidimensional Timbral Motions

As observed in qualitative interpretations, some listeners recognize the multidimensionality in complex stimuli, which often negatively impacts the estimation performance. Listener B, for example, describes the perceptual challenge with concurrent timbral motions that "there were a couple metrics that seemed to be varied so it was difficult to determine which metric the estimation was tracking. There is a very clear distinction [between] audio and visual whether a section is internally steady or fluctuating." The last sentence particularly acknowledges a change of motion within a segment with 'steady-fluctuating.' However, this mapping of words to the data dimension leads to a significantly inaccurate estimation result (Table 7.17.20). Listener B also struggles with an estimation task with the ambiguity of timbral dimensions, recalling "I had a hard time pinning down the estimation because there seemed to be multiple qualities that varied. #2 is the pleasant side, most of the rest lean unpleasant." Their timbral description with an emotional response ('pleasant-unpleasant') mapped to data follows the estimation task, which interestingly attains relatively high accuracy and efficiency (Table 7.17.21).

7.5.6.5 The Use of Musical Concepts

Some but fewer listeners also employ extrinsic associations in their estimation strategy. Compared to qualitative assessment, however, estimations typically involve more of the concepts found in traditional symbolic music rather than environmental or cultural references. There are two common approaches observed, where the listener frames their analyses into (1) the mixing of voices (instruments) and (2) a melodic motion with a discrete 'note' structure.

The "mixing" approach maps multiple characteristics (i.e., word pairs or self-introduced terms such as instruments) in the form of polyphonic voices (cf. auditory streaming [9]). The listener B reflects, "[t]his one is a pretty clear [sound characteristics] A and B though they mix. A is soft, B is hard, C is both sounds together." Their following estimation yields an average accuracy and efficiency (Table 7.17.22). Another listener has a similar observation, that "the two sounds played were mixed together seemingly. A third or higher pitch added in the third to fo[u]rth session added to the complexity of rating and finding a word pair." This listener describes an addition of individual harmonic 'pitch' forming a 'thin-rich' sensation. However, the preceding estimation appears to be misguided with this sensation, resulting in a very low accuracy (Table 7.17.23).

Similar to the references of musical instruments, listeners with prior formal musical training tend to employ symbolic musical concepts, such as notes, melodies with pitch and rhythm, and performance aesthetics with instruments (or voice). This approach generally appears to hamper the quantitative timbral analysis, abstracting away some of the details. It also tends to limit the use of word pairs either to a single mapping or with no mentions in descriptions. Listener E, for example, maps a word pair to the perceived musicality and sound types, such that "[o]ne tone sounded like an actual musical tone, which I considered stable. The other was more like white noise and noisy. I'm also fairly confident in the values since most of them were down on the noisy end." This account follows an estimation with relatively low accuracy, indicating the disparity between the perceived musicality and physical modulations of timbre (Table 7.17.24). Listener E also relates a word pair to the perceived pitches in a compromising way, describing "the tones were all the same, only higher and lower in pitch. So, the lower sounds were dark, and the higher ones were bright." This observation, again, results in estimation with lower accuracy, perhaps as a result of failing to notice the timbral differences (Table 7.17.25). Listener D is also troubled by the elusiveness of timbral characteristics that "it's hard to describe these sorts of sounds, I almost want to hum a little tune to hold it in my memory. Sort of snare to chime on the last two notes." The following estimation is largely hindered in accuracy, perhaps reflecting that the data dimension modulates the tonal-ness of the spectrum rather than a harmonic pitch (Table 7.17.26). With the same listener, another indication of symbolic interpretation is that "this one is interesting. I think these will be easier to remember as 'notes' rather than word pairs." They appear to spend a considerable amount of time analyzing the sound (i.e., low efficiency), but the following estimation scores high as they discard verbal associations altogether but focus on familiarizing themselves with the sound itself at note level (Table 7.17.27).

7.5.6.6 Use of Qualitative Analyses for Estimation

Finally, there are also responses that highlight direct relationships between qualitative and quantitative understanding of a sonification. Many listeners actively utilize the qualitative characteristics as the basis of estimation, often consciously mapping multiple qualities to temporal events. Listener A, for instance, reflects that "the sounds were mostly clearly steady throughout but there was one where it fluctuated which is the reason I am confident in both the word choice and estimation." However, whether by misidentifying the data dimension or lacking the precision with this interpretation, the following estimation attains a relatively low accuracy (Table 7.17.28). Another listener also recalls that "I had a difficult time with the dot section and also some difficulty distinguishing the sounds to make word pairs. #2 and 4 seem relatively clear especially compared to the muffled effect of the rest." The preceding estimation, despite lacking confidence, reaches an above-average accuracy, indicating the perceived change of timbral qualities aligning well to the data modulation (Table 7.17.29).

7.5.6.7 Estimation-Design Attribute Relationships

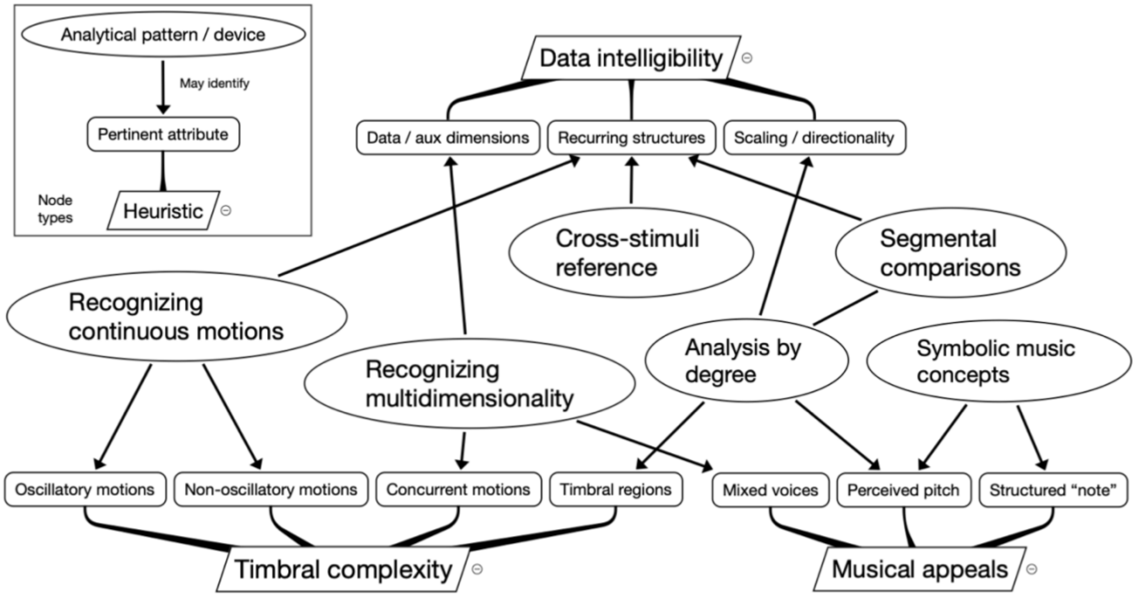


Figure 7.18 The observed relationships between the estimation patterns and design attributes. Note that an analytical pattern may possibly identify some of the attributes but it is not guaranteed.

In general, as Figure 7.18 shows, many of the observed estimation patterns seem to directly relate to multiple design attributes, while not heavily depending on other patterns[51]. The most peculiar finding is perhaps the closer connections between the attributes of musical appeals and data intelligibility, contrasting with the interpretive patterns with relative isolation between these two heuristics (Figure 7.17). To be more specific, for estimation tasks, applying certain musical vocabulary or knowledge seem to affect the perception of intelligibility / complexity attributes. The prime example is the reliance on the sense of pitches when comparing different segments, even for a stimulus without a consistent harmonic structure across segments. The observed cases

[51] Again, there are likely weak or indirect relationships not depicted in the graph. It should also be noted that most of the reflections on estimation are facilitated by the use of interpretive descriptors, thus suggesting the presence of multiple layers of connections.

indicate biased attention toward frequency-based features over other timbral nuances such as inharmonicity.

The patterns of estimative analysis also highlight intricate but also quantifiable aspects of design attributes, especially for the timbral-complexity heuristic. Instead of aggregate impressions of plurality and ambiguity in qualitative interpretations, for example, timbral dimensions are often recognized in terms of concurrent motions. They are sometimes noticed as an 'inconsistency' of structure, as one characteristic persists across segments but another changes between segments. While the recognition of such a complex structure does not always yield an accurate estimation of encoded values, it generally appears to contribute to the efficiency, possibly signifying an improved perceptual decomposition.

Between sound complexity and data intelligibility, continuous timbral motions recognized as oscillatory (rather than transient or gradual) appear to yield a better estimation accuracy more often. This may correspond to the recognition of a recurring structure, creating noticeable differences between multiple segments.

Timbral multidimensionality is perceived in at least two different ways: as a concurrent motion or a mixture of instrumental voices. The latter often appears to yield a better estimation accuracy, unlike with the former, and raises an interesting problem – For some listeners, multiple motional qualities of timbre may have a varying degree of presence over time, whether at the sound-organization or at the perceptual level. Thus, assuming an equal and persistent presence of all the timbral dimensions may not be sufficient when designing data and auxiliary modulation mappings.

Overall, the acknowledgment of continuous motions, multidimensionality, and segmental differences of timbre are observed as distinct estimative strategies, often employed independently from each other. However, as they all address unique aspects of the design heuristics, some level of combination may facilitate an in-depth understanding of timbral structures. The application of musical-structure concepts, such as note- or voice-based analysis, also may contribute positively to estimation when not overshadowed by a pitch-oriented analysis.

7.5.7 Discussion: Addressing the Research Questions

For many listeners, the verbal analysis of timbral sonifications proved to be an unintuitive process with the difficulty of pinpointing intangible / dynamic elements of sound. The listener employs various descriptive approaches to capture such fluid characteristics.

In terms of accurate and efficient value estimation, most interpretive patterns appeared to lead to both positive and negative outcomes (within the 20 trials by each subject). This inconsistency may come from multiple reasons that (1) a verbal reflection may not fully capture what actually contributes toward a sufficient analysis; (2) the goals of interpretation and estimation do not always align with each other; and (3) a reflection highlights more of personal challenges / points of interest rather than their definite solutions. With these caveats (and possibly more) in mind, the observed response patterns bring new perspectives on the research questions.

RQ 1-A examines the elements of sound organization in relation to the listener's quantitative understanding of sound. This question concerns more of the fixed structural attributes of SMSon than the design variables. In this regard, the listener's analytical patterns show some similarities to how SMSon organizes a variable sound structure in terms of models and mappings. Multiple responses show a creative use of word pairs and external references depicting conceptual sound sources and physical or timbral motions. A sound source, such as a musical instrument, would represent a fixed set of timbral states, possibly corresponding to the synthetic spectral models. Motional references, such as instrumental performance, may be associated with mappings, where different data sources or auxiliary shapes drive the spectral models over time. These potential associations could be extended to the interpretation of the statistical result of the estimation tasks. That is, the mapping of different motional descriptors to a timbral dimension may affect the estimation accuracy while the interpretations of a spectral model may affect the efficiency. The verification of this hypothesis is left for future work.

There also seem to be opposing directions in how the listener depicts the timbral states and temporal motions. That is, some listeners are more focused on 'connecting' isolated timbral states with generic verbs (e.g., go, move, etc.), while some others aim to map unique descriptors (e.g., relaxed, hard, etc.) directly to the motional behaviors. For example, listener A tends to describe the stimuli as "the sounds go from A to B, then moves between C and D" and so on. Listener F, on the other hand, would describe "I see a person interacting with an object to create a certain type of sound, first using motion A but suddenly shifting to motion B." Many listeners use mixed approaches that are between these two extremes. Corresponding to the model-mapping view above, anticipating the listener's biased attention on timbral states or temporal motions may be important in the sound-organization process.

With regard to RQ 1-B, the prior statistical results as well as the listener survey mostly disproved any direct or linear influence of a description task toward an estimation task. It appears that a guided "reduced" listening focused on the internal structures of a sound, in the form of a quick selection of descriptive word pairs, is not always sufficient for both experienced and untrained listeners to quantify the relevant timbral features. However, the detailed textual reflections indicate that some types of interpretation (often regardless of the description-estimation order, as they may occur simultaneously for some listeners) follow the outcome or confidence in the estimative analysis.

For the cases where the order does matter, some listeners take advantage of the anchoring nature of descriptors to examine multiple aspects of a complex timbre, managing to analyze and isolate inconspicuous characteristics. Such exploratory behaviors are observed with listeners A, C, and F, who also generally score considerably higher in description-first analyses (see Table 7.14).

The use of multiple word pairs in textual reflections entails several distinct patterns, including the acknowledgment of plural timbral qualities and states / motions in various time scales (e.g., within or across segments). However, with the description tasks alone, the number of selected word pairs leads to some unexpected results. Particularly, among the representative listeners, the ones who always select two or three word-pairs (the listeners B and D) perform worse in description-first settings, perhaps reflecting more of a perceived ambiguity of timbral qualities rather than a plurality.

Lastly, RQ 1-C concerns the gap between design intentions and listener's receptions, assessing more of subjective and variable factors not observed in RQ 1-A (though both questions concern some level of functionality and personal aesthetics). The preceding sections (7.5.5 and 7.5.6) presented some of the notable relationships between the analytical patterns and pertinent design attributes. In general, many of the analytical patterns (e.g., the application of multiple word pairs) can impact the perception of multiple design attributes simultaneously, creating complex dependencies among the three design heuristics.

Data intelligibility, for instance, is likely often affected by musical or aesthetic appeals of a sonification. This is frequently observed in terms of how the context of interpretation is established – some with a preconceived listening habit for symbolic music, and some with a tendency to associate everyday sound sources (e.g., flying insects) or cultural sounds (e.g., science-fiction materials) to abstract timbres. In a discussion of functional values of electroacoustic sonification, Vickers and Hogg argue the indexicality, or deictic use, to be a key value for mapping data / meanings to a sound object [107]. This assumption requires a proper establishment of a listening context, which could be approached with the types of textual descriptors used for the experiment / communication or juxtaposing the variable designs of stimuli to facilitate more cross references.

For this listening experiment, sound complexity was partially quantified with the synthetic method being employed (i.e., the spectral models and mappings). However, the indication of perceived complexity somewhat lacked a discernible pattern in relation to the sound organization elements, such that many found the stimuli with no auxiliary motions (e.g., mapping type Data-Constant with a static spectral filter) to be complex. There are general indications that simply switching the modulation mappings of the same model (e.g., Data-Aux vs. Aux-Data) providing a completely different impression of timbre for many listeners. This phenomenon could be further investigated in order to create novel but also partially predictable sound organizations.

The perceived musical appeals appear to have diverse and sometimes contradicting effects on quantitative listening. On the one hand, unfamiliar or non-relatable sound expressions, which some perceived as unpredictable or unnatural timbral motions or described as "noise," often affected the intelligibility negatively. On the other hand, when the listener fit the perceived timbral quality to a version of a musical instrument that they were familiar with, such as the violin, their listening experience often either (1) gained the efficiency of decomposing individual timbral motions with the physical (e.g., gestural) associations or (2) misguided the estimation with an illusion of harmonically structured pitches.

Responses to aesthetic appeals also appear in the form of identifying uniqueness or dissimilarity between and within stimuli. That is, while the designer (in this experiment, the author himself) has intended to create multiple and unique expressions by manipulating the mappings and models, some listeners find it difficult to qualitatively differentiate the characteristics. Likewise, some keen listeners notice a subtle difference among the timbral characteristics of segments within a stimulus and map different descriptors to denote (sometimes elaborated with emotional or musical likings such as 'this tone is stable and musical' by listener E).

7.6 Listener’s Understanding of Electroacoustic Sonifications

The listening test presented multiple variations of SMSon-based designs to the listener and observed how they qualify as well as quantify the sounds primarily through careful listening. The statistical responses indicated that some structural elements, namely the mapping location of data, may have a general relationship to the accuracy of data estimation. The presence of auxiliary timbral motions appeared to have no general influence on data estimation. However, the interpretations of timbral motions widely varied among individual listeners. While the interpretive patterns were very complex and difficult to draw a conclusion, the listener’s subjective understandings appeared to be influenced not only by the structure of SMSon, but also by how they interpret the overall impression of sound, their individual musical orientation, as well as uniquely constrained areas of aural exploration.

This experiment, therefore, investigated how the general listener would engage with unfamiliar and complex electroacoustic sonifications (RQ 1). The next chapter similarly examines how the first-time user explores various intricate elements of musical timbre in relation to data, but primarily from the designer’s perspective.

CHAPTER 8. LABORATORY DESIGN TEST

The second human subjects study examines the creative merits of spectromorphing sonification (SMSon). In contrast to the previous listening test, which recruited untrained and non-expert subjects (see chapter 7), this design experiment invites domain experts in musical composition and/or music technology, though not necessarily in sonification[52]. The author aims to verify that SMSon is capable of facilitating musical and functional outputs for multiple types of designers, so as to validate the core aesthetic principle of the 'variability of design' of SMSon, discussed in chapter 3.

Such a goal with a deliberately broader audience inevitably creates challenges in the experimental design, as did the listening test involving untrained listeners (chapter 7). For example, since each designer may have a different musical sophistication and preferred methods of sound design, comparing their processes of learning, designing, and evaluating sonifications may not be straightforward. The experimental environment needs to adapt the SMSon framework to provide a flexible but also somewhat restricted and uniform control scheme. Another challenge is guiding the designer's perspective toward non-expert listeners.

Revisiting the challenge of connecting design intentions and receptions with electroacoustic sonifications, the author aims to observe (1) how design intentions relate to the physical steps of sound design and (2) how unique designs are evaluated in anticipation of listener's receptions.

This study, therefore, features a uniquely designed environment with two goals. (1) It facilitates the creation of short sonification 'snippets,' or sound objects, with variable timbral

[52] One of the goals of SMSon is to invite a broader range of artists, even without a previous experience with sonifications, to creating a 'functional' sonification.

characteristics. Such a snippet closely resembles the novelty-oriented audio stimuli employed in the listening test in terms of the 'data' and spectral sound organization. The present experiment, however, may yield significantly more variations of snippets. With SMSon and Sonar (see chapter 6), this designer environment facilitates a quick and wide-ranging design of sonification without programming or signal-processing knowledge. (2) It facilitates the assessment of the designer's responses to the methodology, the underlying sonification framework, and the challenge of evaluation in relation to unknown listeners. However, a large portion of responses may also concern the particular elements of the experimental design. The analysis aims to clarify how such 'meta' response may or may not relate to the core questions.

It should also be noted that this design study took place subsequent to the listening test. The listening test only employs several primitive types of auditory stimuli to limit the number and complexity of parameters, not covering the wide variety of snippets created by designers. Although both experiments examine the factors of design attributes, the responses are not directly comparable with regard to specific designs.

8.1 Research Questions

Chapter 3 raised a research question (RQ 2) concerning the creative and individualized elements of the design of a timbral sonification. It asks that, by using the proposed framework, how the designer would be able to integrate personal aesthetics into the design, while maintaining the integrity of sonified data. Similar to RQ 1 explored in the listening test (see chapter 7), this broad question is subdivided into more specific forms.

A. How does the designer utilize the structural elements of SMSon to control the design and integrate personal aesthetics?

B. How does the designer analyze and manipulate the functional and aesthetic aspects of their designs in anticipation of the general non-expert listener?

The first sub-question (RQ 2-A) concerns the framework attributes as well as the unique factors of the designer-study environment. Focusing on more of the subjective design "intention" aspects, the author hopes to verify that the environment / methodology supports a wide range of musical expressions, while observing how the elements are differently interpreted and utilized.

The second sub-question (RQ 2-B) concerns the retainment of the perceptual integrity of data, exploring how the designer would be able to assess their designs from a perspective closer to an untrained listener. From the previous listening test, the study reintroduces several heuristics to the designer for exploring the elements of listener-oriented assessment. These are data intelligibility, sound complexity, and musical appeals, which are to be experimented with and defined more concretely by each of the designers in this study. The author aims to observe how such perspectives of evaluation interact with each other. For example, some designers may (anti-)correlate data intelligibility and sound complexity when creating certain types of musical expression. The author hypothesizes that the designer may be able to independently control the aesthetics of sonification while retaining the intelligibility if the interpretations of complexity align well with their aesthetic preferences (musical appeals).

8.2 Experimental Design

Each subject participates in a single one-hour design session and interview, conducted in person with the author in a studio environment with a laptop, stereo speakers, and headphones. The lab session is documented in the form of a voice / audio recording, seven to eight custom-designed sonification snippets (each with an audio file and configuration data), and evaluations by the designer themselves. From these records, the author attempts to identify some common design patterns, but more importantly, the traits of unique / diverging aesthetic decisions and how they interact with the framework elements.

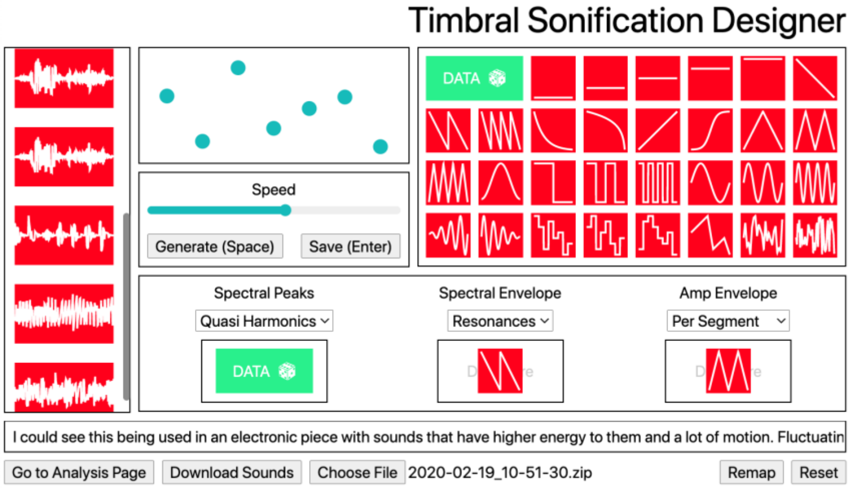
8.2.1 Software Environment and Synthetic / Analytic Concepts

This experiment utilizes an interactive audio-visual web application. The underlying technology and concepts in this environment resemble those of the listening-test web application (see chapter 7.2), including Sonar for audio synthesis, d3 [53] for drawing waveforms and modulations, and NexusUI[54] for visualizing data sequences. Unlike the listening test, however, this environment does not track the user actions such as playback count and estimation time / error, and instead stores information directly related to sound design and assessment.

[53] https://d3js.org/

[54] https://nexus-js.github.io/ui/

8.2.2 Synthesis Page and Generative Models



Spectral Peaks: Quasi Harmonics

Spectral Envelope: Resonances

Amp Envelope: Per Segment

I could see this being used in an electronic piece with sounds that have higher energy to them and a lot of motion. Fluctuatin

2020-02-19_10-51-30.zip

Figure 8.1 The user interface of the sound synthesis page.

The designer interface consists of two separate pages: The synthesis page and the analysis page. The synthesis page (Figure 8.1) features five separate components: (1) the data-sequence view, (2) the modulation-sources pool, (3) the models/mappings section, (4) the sound renderer and player, and (5) the snippet browser.

The data-sequence view (1), at the top-center, resembles the value-estimation UI used in the listening test, presenting a sequence of seven numeric values. The values, however, are set manually by the designer or randomized with a keypress. The values are also hidden by default until a keypress to reveal, encouraging a sound-focused, rather than a visual-driven, design of sonification.

The modulation pool (2), at the top-right, provides a wide variety of geometric shapes as well as the data-sequence object. They can be dragged and dropped into the timbral-dimension view and mapped to each synthetic model, creating a temporal modulation for each data-sequence segment. The modulation shapes consist of several types such as constant values, oscillatory or structured shapes, and random sequence with different time resolutions. The data sequence, when mapped as a modulation source, applies a step (constant) modulation to the synthesis parameter.

Table 8.1 The list of spectral-synthesis models (generative algorithms) featured in the designer environment. This first list includes the models for generating spectral peaks, a matrix of flat magnitude distributions. The type denotes either timbre-oriented (Subsymbolic) or pitch-oriented (Symbolic), with the latter providing more discrete and unevenly quantized structures, although maintaining flat spectra.

Model Name	Type	Description
Noise	Subsymbolic	A white noise with no modulatable structure
Inharmonic Mix	Subsymbolic	Morphs between two irregularly spaced peaks
Quasi Harmonics	Subsymbolic	Morphs between a regularly spaced and irregularly-spaced peaks
Noise-Tonal	Subsymbolic	Cross-fades between a white noise and a pitched sound
Noise-Buzz	Subsymbolic	Morphs between a white noise and densely-packed harmonics
Major Scale	Symbolic	Cross-fades between two octaves of harmonic pitches quantized to the major scale
Major Chords	Symbolic	Shifts through tertian diatonic chords in the major scale
Minor Scale	Symbolic	Cross-fades between the minor-scale pitches
Minor Chords	Symbolic	Shifts through tertian minor-scale chords
Parallel Chords	Symbolic	Shifts through tertian chords in a 'sustained' chord scale

Table 8.2 The models for generating spectral envelopes, a matrix of filter shapes.

Model Name	Description
Centroid	Filters the peaks with a log-normal curve with the centroid (center frequency) being modulated
Spread	A log-normal envelope with the standard-deviation being modulated
Mid-Side	Cross-fades between a trapezoid and its vertical inverse
Resonances	Morphs between comb-like peaked envelopes that expands / contracts over the frequency
Stria 1-3	A log-normal envelope that are truncated by equally spaced rectangles with different widths

Table 8.3 The methods for altering the amplitude envelope. The mappable shapes are multiplied in a unipolar manner.

Model Name	Description
Per Segment	Applies the modulation shape to each data-sequence segment
Entire Duration	Applies the modulation shape across the data-sequence segments

Table 8.4 The models for generating spectral effects, which modify the spectral peaks (either their magnitude or phase) in varying degrees.

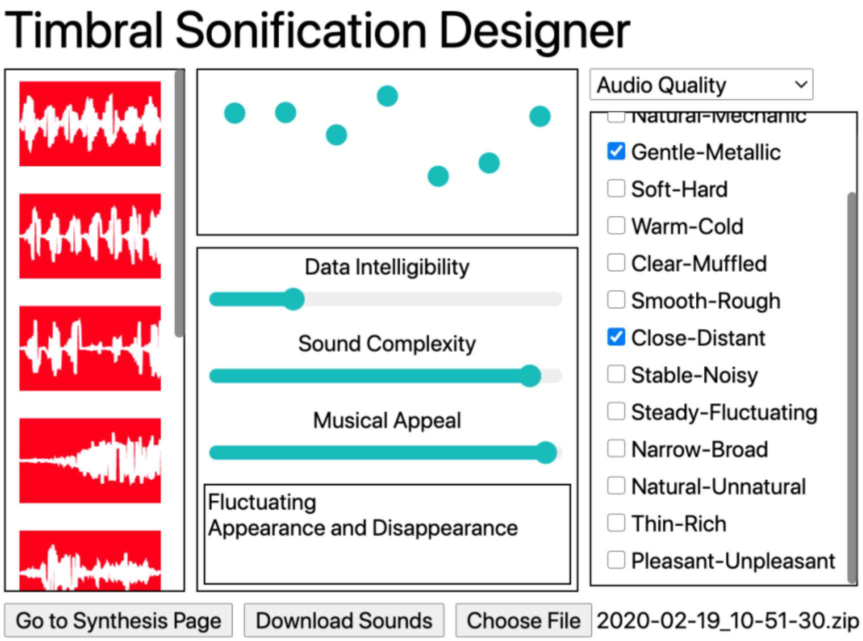
Model Name	Description
Jitter	Randomly shifts the spectral peaks over frequency, with the amount determined by the modulation size
Peak Fluctuation	Randomly alters the magnitude of the spectral peaks, with the amount set by the modulation
Phase Distortion	Randomly shifts the phase value of spectral peaks from a linear increment
Peak Scan	Samples the spectral peaks in a random order, according to the current modulation index
Odd-Even	Filters and cross-fades between odd-only and even-only spectral peaks
Mirror	Creates a mirror of the spectral peaks and cross-fades between

The models-mappings section (3), at the bottom half, presents four similar interfaces representing synthetic dimensions, to which the modulation shapes are mappable with a drag-and-drop operation. Each synthetic dimension also features a drop-down selector for synthetic models (see Tables 8.1~8.4). There are three main synthetic dimensions: the spectral peaks (denoted as Spks), the spectral envelope (Senv), and the amplitude envelope (Aenv). In addition, the designer can reveal an optional dimension, the spectral effects (Sefx), with a key press.

The sound renderer (4), at the center-left, initiates the synthesis and playback of a sonification according to the current data sequence, mappings, and model selections. It also allows for specifying the entire duration of the snippet between 0.5 to 5 seconds. Generating a sound temporarily registers the current sound in the snippet browser, which can be properly saved once the designer finishes the adjustments.

The snippet browser (5), at the left, displays all the sound objects created and explicitly saved by the designer in the current session. Selecting a snippet plays back the created sound, with the option for reproducing the mappings with the 'Remap' button. The designer can audition the current and previous snippets to compare and can also play them back simultaneously to create a polyphonic expression. The designer also can add notes to each snippet while reviewing, as displayed in the bottom text box.

8.2.3 Analysis Page and Evaluation Heuristics



□ Natural-Mechanic
☑ Gentle-Metallic
□ Soft-Hard
□ Warm-Cold
□ Clear-Muffled
□ Smooth-Rough
☑ Close-Distant
□ Stable-Noisy
□ Steady-Fluctuating
□ Narrow-Broad
□ Natural-Unnatural
□ Thin-Rich
□ Pleasant-Unpleasant

Fluctuating
Appearance and Disappearance

2020-02-19_10-51-30.zip

Figure 8.2 The user interface of the sound analysis page.

The analysis page provides the designer an interface for reviewing the snippets they have created (Figure 8.2). The review process for each snippet consists of (1) a rough estimation of the data encoded, (2) a subjective rating of three design intention-reception heuristics, and (3) verbal descriptions using notes or selecting various types of descriptors. As the number and variety of snippets grow, the designer is encouraged to adjust the reviews, browsing different snippets multiple times.

The analysis page presents the data-sequence view (1) identical to the synthesis page, which the designer can hide or reveal with a keypress. On this page, however, the values displayed indicate the quantities already mapped to a timbral dimension(s) that cannot be changed. The visual content is hidden by default, prompting the designer to quickly and roughly estimate the values before revealing the actual values.

The designer is also to utilize three continuous sliders (2) for the subjective rating of sonification snippets: data intelligibility, sound complexity, and musical appeal. These metrics follow the theoretical groupings of perceived design attributes introduced in chapter 7. In the present study, these concepts are employed as listener-oriented heuristics. All three heuristics are briefly introduced to and discussed with the participants concerning their possible definitions and implications. The author presented the concepts in the following manner.

'Data intelligibility' may signify how intuitively and accurately the listener would be able to estimate the encoded quantities. The listener, in this case, entails any untrained end-user of sonification without the detailed knowledge of design. Rather than providing an immediate assessment, the designer is encouraged to revisit the collection of sound they have created, show / hide the visualization of the underlying data, and adjust the ratings relative to other sounds. 'Sound complexity' is introduced as a high-level impression of the timbral structure that the designer finds to have a temporal or spectral intricacy, and to some extent, hinders the data measurement. This experiment uses the term 'sound' instead of 'timbral' complexity to accommodate ideas and discussions not limited to timbre, such as symbolic melodies and references to a cultural context.

Lastly, 'musical appeal' denotes how potentially useful or inspiring a sound snippet is for the compositional styles the designer is familiar with. These heuristics of variable design are discussed not only in the analysis phase but also as part of the design exercises while on the synthesis page.

The textual descriptors (3) in the analysis page include four categories: the timbral-quality word pairs, acoustic instruments, symbolic-music terminologies, and spectromorphological terms. The first timbral-quality word pairs are identical to the ones employed in the listening test, catering to non-expert listeners (see chapter 7). The instrument names and symbolic terms, collected mainly from Oxford Dictionary of Music[55], include the ones commonly used by composers of western traditional music.

The present design study, however, mostly analyzes spoken responses over the selection of the textual descriptors. While some designers utilize textual records (i.e., descriptor selections and free-form inputs) in their design assessment (see Appendix C), they do not appear to form structures / relationships to the design elements comparable to each other.

8.2.4 Design Tasks and Questions

The one-hour lab session consists of pre- and post-interviews, exploratory learning of the system, and two phases of design challenges.

1. Musical-background questions
2. An introduction to the design environment
3. Practice and exploratory sound design
4. Designing and assessing musically appealing snippets

[55] https://www.oxfordmusiconline.com/page/The-Oxford-Dictionary-of-Music

5. Designing and assessing intelligible snippets
6. General reflections and formal questions

The first set of interview questions aims to establish the general musical orientations of the designer. The inquiries include the instruments they play, the experience and types of compositions, and the styles of music they are familiar with and have the knowledge to describe the organizational elements.

The designer is then introduced to the design environment for the first time, going over the concepts of timbral sonification, the data object, auxiliary modulation shapes, spectral synthesis and generative models, mappings, and the playback and navigation of sound snippets. The designer then spends 10 minutes exploring the synthesis page, experimenting with combinations of models and mappings, and learning the aural characteristics of each spectral model in a spontaneous order. While the time allocated for learning is considerably short, the designer is inquired what elements of the system are intuitive or nonintuitive when experimenting or planning a new timbral expression.

Then, as the first of two design exercises (step 4), the author prompts the designer to create several sonifications that are musically interesting, inspiring, and/or potentially usable for their accustomed styles of composition. As they experiment for 20 minutes, the author asks them to articulate their understanding of synthetic configurations and how they would verbally describe the sound. The designer is then guided to the analysis page, where they are asked to comparatively assess the snippets with the perceptual heuristics and textual descriptions.

As the second design exercise (step 5), the designer is asked to focus on data intelligibility for the next 20 minutes. Particularly, they are requested to create low, mid, and high complexity snippets, while maintaining high intelligibility of data for all the variations. The author restates that, ideally, such intelligibility should accurately and intuitively convey the quantitative characteristics to the first-time untrained listener. Simultaneously, the designer is also encouraged to consider the musical appeals of the snippets, although not as the primary criteria.

Lastly, the author inquires their general experience and thought processes with the following questions:

- Do the design methodology and environment provide unique values to sound and/or sonification designs? In what ways?
- How do they value the musical expressiveness / limitations of the snippets or the method?
- How easy / difficult is it to predict the outcome of a design configuration, and why?
- What elements of the design environment help decide the aesthetic approaches for sound design?
- What constitutes / contributes to the complexity of a sound organization?
- How would they anticipate the reactions / utilization by the general untrained listener?

The first four questions aim to capture the factors related to RQ 2-A (the connection between organizational elements and aesthetics or utility), while the last two questions may address RQ 2-B for identifying the aspects of design communication.

8.3 Analysis of Design Session Results

The following qualitative analyses organize and enlarge the designer's responses from three angles: (1) The general design procedures, (2) designs focused on musicality, and (3) designs focused on complexity and intelligibility. For 2 and 3, the author examines several of the created sonification snippets in detail as 'case studies,' illustrating the unique design problems and solutions developed over time. The physical configurations of a design are reconstructed from the recorded conversations and the submitted snippet configurations (see Appendix C). The submitted designs are the waveform and configuration data for each snippet stored in the environment. The configuration data include the data-sequence values, model selections, the mapping of data / shapes, the snippet duration, subjectively evaluated design attributes, descriptors, and notes. In the following observations, selections of both models and mappings combined are expressed as 'value selections,' signifying the predefined symbols in the environment and not to be confused with the numeric data-sequence values.

8.3.1 Quantifying the Variety of Design Configurations

The present study is largely qualitative. However, as an indicator of creative explorations by each designer, the analyses employ the information entropy [89] of models and mappings in various partitions. The author interprets entropy as a comparative metric for the variety of designs among designers. The entropy is defined as

$$H(X) = -\sum \frac{P(x)}{\log P(x)}, \tag{1}$$

where $P(x)$ is the occurrence of each unique value (symbol) within the distribution of selected model or mapping values X. A higher entropy generally signifies a flat distribution of selected values, whereas a lower entropy indicates repeated selections of the same value. The null value (i.e., unused models or unassigned mapping slots) is treated as a symbol for both models and mappings.

8.3.2 Subjects Overview

The experiment recruits five designers, who are current and former graduate students in Music Technology. The following observations refer to them as designers A through E. First, this section summarizes their musical background and general traits in the design workflow, as observed throughout the design exercises and in the submitted designs.

Table 8.5 A summary of the participating designers. Variety denotes the utilization of different models and mappings, grouped into three levels. Model Entropy and Mapping Entropy below are calculated across all the timbral dimensions of the submitted designs (see Appendix C). Workflow, Priority, and Orientation categorize the general design traits into two or three levels. These designations, however, are not final and interchange in different situations, as observed in the following discussions.

ID	Musical Style	Instrument	Variety	Model Entropy	Mapping Entropy	Workflow	Priority	Orientation
A	Contemporary	Trombone	Exhaustive	3.90563	3.64031	Trial & Error	Sound	Timbre
B	Modern Jazz	Saxophone	Exhaustive	3.79790	3.04952	Trial & Error	Sound	Timbre
C	Classical	Piano	Neutral	3.48166	3.30595	Neutral	Data	Pitch
D	Pops	Guitar	Minimal	3.07059	2.72455	Plan & Execute	Data	Neutral
E	EDM	Trombone	Minimal	3.41379	2.67651	Plan & Execute	Sound	Timbre

Designer A is a trombone and saxophone player, who has received extensive training in composition and music theories and has composed pieces for chamber ensembles, choral groups, and electronic music. They have also previously worked on sonification projects focused on the musical potentials of data. In the present design experiment, along with designer B, this designer generally takes a highly exploratory approach, utilizing most of the available synthetic dimensions (by mapping data or motional shapes) in 98% of the snippets created, and using unique (non-repeated) models or mapping sources for 16% of the total available parameters. They exhibit the highest entropy / variety values in overall designs (Table 8.5).

Designer B is a saxophonist with degrees in jazz and music technology. They have composed and performed in a wide range of styles including classical and electroacoustic soundscapes. In the present study, they demonstrate highly exploratory, trial-and-error approaches to create a variety of timbres, with 93% of the parameters modulated and 47% of unique selections utilized for the overall designs. This designer also does not hesitate to experiment with extreme configurations for aesthetic effects, such as with minimum and maximum durations of a snippet.

Designer C is a pianist who has received formal training in western traditional and contemporary classical music. They enjoy writing scores for small- (e.g., sonata) and large- (e.g., symphony) scale ensembles, using the traditional pen-and-paper method. As a music technologist, this subject has also worked on sonification projects for science education, using data that requires domain knowledge to properly represent. Their design explorations in the present study indicate an overall moderate variety, modulating 90% of the available synthetic dimensions and trying unique models and mappings for the 25% of total parameters. However, along with designer D, they exhibit peculiar musical preferences toward pitch-based organizations, as observed below.

Designer D is a guitarist and musicologist who is most familiar with popular music, ranging from 1940's jazz to contemporary rock and pops. Besides composing in such styles, they have created several soundscape pieces for radio podcasts using commercial synthesizers. This designer also has considerable interests in sonification, while professionally being accustomed to visual representations and analytics of domain-specific data. For the snippets created, designer D takes a relatively minimalistic approach modulating only 60% of the synthetic dimensions for all the snippets, while selecting unique models or mappings for 19% of the total slots. Their entropy / variety values also confirm somewhat skewed selections in the overall design.

Designer E is a former trombone and piano player with experience in producing original compositions using commercial synthesizers. They favor electronic dance music, which is heavily reflected in their design approaches and thought processes being shared. While indicating a strong focus on timbral organizations, they also choose to take a systematic, plan-and-execute approach, generally creating concise design configurations with 60% of the parameters modulated and 19% of unique selections of models and mappings. Interestingly, along with designer B, they explore significantly more combinations of models over the mappings, as the entropy / variety metrics suggest.

8.4 Exploration and Formation of Design Procedures

As a preparatory creative experiment, the designer browses through and familiarizes themselves with the sound-design resources (i.e., models, etc.). They are to learn the system behavior primarily by carefully listening to the result of different configurations, but also by verbally describing and confirming the timbral behaviors with the author. Being tasked to simultaneously learn, discover, and manipulate the elements of SMSon, the designer provides immediate interpretations or confusions about the design methodology / environment. Such reflections may indirectly address the RQ 2-A (the relationships between the organizational elements of SMSon and aesthetics), but also provide insights in how they see the system from a relatively detached perspective akin to a first-time listener.

First, from an aggregate review of the designer's verbal responses as well as the submitted snippets, the author proposes several perspectives on the design procedure. They are discussed in terms of workflows, organizational priorities, and musical orientations.

8.4.1 Design Workflows

As briefly discussed in section 8.3.2, general design workflows entail a noticeable dichotomy: trial-and-error and plan-and-execute. They are signified by multiple factors, including the exhaustive vs. minimal use of the available configurations, the consistency vs. contrast between snippets, and verbal reflections. These workflows also suggest unique relationships to the designer's aesthetic preferences as further examined in the following sections.

A trial-and-error (T/E) workflow tends to involve rapid experimentation of mapping multiple different shapes, alternating between models, and frequent auditions. As indicated by the works of designers A, B, and C (see Appendix C), T/E approaches typically lead to an exhaustive use of synthetic dimensions and mapping sources with fewer unused (null) dimensions[56]. When asked about the reasons and implications of using more elements in general, designer A confirms, "I'm a trier so that's why I was trying so much with the different dimensions. But it probably would be a good idea to focus, if I had tried just one or two, because that didn't cross my mind. I just wanted to try everything. ... Some of the first sounds I made were just really, a lot [of sound elements]. But if I had just changed one or two it wouldn't have been so drastic sounding."

A plan-and-execute (P/E) workflow, in contrast, typically involves the manipulation of fewer and more focused sets of models and/or mappings. With this approach, typically, the designer identifies an effective configuration for satisfying musicality and/or intelligibility, and

[56] Designer C is, however, more towards neutral as they alternate the model types (i.e., symbolic and subsymbolic) infrequently.

experiments around this core structure with less deviation. An 'effective configuration' may take two separate directions: (1) value selections and (2) synthetic procedures. The selection of fewer particular models and/or mappings are well observed with designers C and D. Designer C, for instance, largely prefer to employ pitch-based models for Spks, articulating each 'note' with oscillatory / structured Aenv modulations in the segmental time scale. Designer D mostly only utilizes two synthetic dimensions, Spks and Senv, with Senv:Resonances and two types of Spks models (pitched scales and Spks:Noise-Buzz) as the dominant combinations. A P/E approach focused on synthetic procedures, instead of reusing the same value combinations, defines priorities or effectiveness of the synthetic dimensions, while exploring different models and mappings within such dimensions. For example, designer E explains their general approach that "definitely putting a pattern on or data on spectral peaks is the clearest way I think to put your data in the sound. And then the second clearest way was to pick centroid on spectral envelope." In another conversation, designer E reiterates, "spectral peaks is kind of the most important one, I think. Spectral effects seem to be one of the least important because you can turn it off completely."

8.4.2 Orders and Priorities of Handling Data

Another dualistic design perspective is the data-focused and sound-focused organizations of a sonification. Such a contrast in priorities can manifest, for example, as (1) the order of incorporating data and non-data (auxiliary shapes) into a sound organization, as well as (2) the way of creating a musical structure with or without data at its core.

With data-first (DF) organizations, there is often a clear separation between the primary timbral motion with data and secondary timbral ornamentations. The snippets created by designer C (e.g., the second, third, and sixth snippets in Appendix C) may best exemplify DF, where they frequently map the data to symbolic Spks models, which they argue to be most salient. Then, they map either the copies of data or non-oscillatory motions to other dimensions but avoids more distinct oscillatory patterns suppressing the timbral dynamics. Designers B and C, and to a limited extent A, appear to take the DF approach, verbally indicating that data being sonified should inform the musical characteristics of the sonification.

With a sound-first (SF) organization, the designer would establish a complex sound structure while switching, shuffling, mapping or discarding the data object. This approach is perhaps most evident with designer E, who explores the variety of musical expressions often without mapping the data object[57]. The SF approach highlights that the contents or the mapping destination of the data object is interchangeable, and the data have less effect on the (already-established) musicality of the sonification. Interestingly, designer C with primarily the DF approach also discusses the interchangeability of mappings, that "I think just modulating [the amp envelope and spectra effects] is [enabling] these two (points to Spks and Senv) to change [the mapping of data] … If I had a parallel like changing instruments … each instrument has a different timber and these [synthetic dimensions] are all … creating different timbres for an instrument … no matter [where] I put data [at]. It's kind of a different medium to go to [depending] on what I'm trying to do – The amp envelope is [for] more subtle effects … The spectral effects do the complexity effect."

The difference of priorities is generally more distinct during the early phase of a snippet design and becomes obscure as the design evolves. Sometimes, it can be unclear even in the early phase of a design, as some designers prefer to reuse a previously made snippet as the starting point of a new design.

[57] Designer E does experiment with the data object in design processes, but does not hesitate to prioritize non-data mappings, as described in the case study in the next section.

8.4.3 Symbolic and Subsymbolic Orientations

In terms of musical appeals, perhaps as expected, designers with symbolic (i.e., pitches and notes) and subsymbolic (i.e., timbral) orientations use contrasting design strategies. These tendencies emerge as how frequently the designer combines a harmonic pitch-based model (e.g., Spks:Major Scale) and the data object, which creates a distinct melody-like expression. In the present experiment, the designer groups with pitch vs. timbral orientations appear to largely overlap with the aforementioned data-first vs. sound-first advocates, respectively.

Designers with pitch orientations (PO), notably designer C, often aim to create a data-dependent melodic shape. This correlates with the popular concept that the pitch is the most intelligible feature for creating a sonification (and perhaps overused by musical sonifications overshadowing other perceptual features [69]). Designer D explains, for the first musicality-focused snippet they created, "I would say data intelligibility is high because I was making this into a melodic arc." When requested to create a 'high-intelligibility / low-complexity' snippet (task 2-A), all but designer D resorted to a symbolic organization[58].

Designers with timbral orientations (TO), such as designers A, B, and E, actively explore non-pitched timbral motions using both data and auxiliary shapes, with more emphasis on the non-Spks synthetic dimensions. Designer B, for example, creates the third musicality-focused snippet focusing on the non-pitched aspects by randomizing the symbolic Spks:Major Chords but assigning data to all the other dimensions. Such an exploration typically involves a T/E workflow. Designer A relates the timbral exploration to the unpredictability of a design process, recalling "when I started off with something where I could clearly hear the pitches, that wasn't too bad[59].

[58] Designer D appears to deliberately stay away from a symbolic model in this task, after exploring other timbral dimensions.

[59] This "not too bad" is a response to a question – if the general design outcome was predictable and/or controllable.

But I wanted to make something that was more timbrally interesting, and that's when I was really exploring and didn't always know exactly what I was gonna do, but I was trying a lot of stuff out." For the outcome of the exploration, designer A recalls "there wasn't any disappointment but there was a lot of element of surprise. A lot of times, I feel like, wow, I did not expect this to happen."

The perspectives of the design workflow, data organization, and musical orientation appear to interrelate in a complex way, sometimes the same designer displaying both of the opposite traits, depending on the design problem at hand. The next two sections map the actual aesthetic and functional design decisions to these perspectives to further solidify their traits.

8.5 Musicality-Focused Designs and Assessments

This first design exercise asks the subject to create three or more musically appealing and distinct snippets while allowing any levels of data intelligibility and sound complexity. The author aims to observe (1) the range and variety of designs that SMSon facilitates, both per and across individual designers, and (2) what the designer identifies as the elements of musicality and/or musical potentials. These are analyzed along with the designer's subjective responses on the analysis page, where they verbally review their designs.

8.5.1 Design Configurations, Variety, and Subjective Ratings

First, the configurations of the musicality-focused snippets (see the Trial 1 snippets in Appendix C) as well as the variety metrics reveal several unique patterns in design focus and exploration.

In terms of the individual configurations, trial-and-error (T/E) designers (A, B, and to some extent C) show different focuses of exploration in data mapping and selections of shapes. Designer A (T/E, sound-first, and timbre-oriented) explores (1) different Senv responses all modulated by data and (2) primarily oscillatory motions mapped to non-Senv dimensions. They seem to provide contrasting ratings with 'low-intelligibility + high-complexity and musicality' and vice versa. Designer B (T/E, sound-first, timbre-oriented) appears to (1) flexibly explore the data mapping targets and (2) animate all the synthetic dimensions with unpredictable non-oscillatory patterns or data. Their subjective ratings are mixed and do not highlight noticeable patterns. Designer C (data-first, pitch-oriented) experiments with (1) flexible mappings of data and (2) more oscillatory motions, with largely symbolic Spks models. They also provide generally mixed ratings with a clear dislike in musical appeals for a subsymbolic-based combination (snippet #4).

The plan-and-execute (P/E) designers (D and E) show more focused experiments with fewer model and mapping combinations. They avoid the use of random motions, in contrast to the trial-and-error designers. Designer D (P/E, data-first) indicates a strong preference of data mapped to symbolic Spks models (snippets #1 and #3), which they rate with high-intelligibility and high-musicality indicating their pitch-oriented musical preference. In contrast, they find the subsymbolic Spks with data mapped to Senv to be only high in complexity, but not intelligible nor musical. Designer E (P/E, sound-first, timbre-oriented) experiments with fewer but more various (non-repeated) selections of Spks models and mappable shapes, providing higher musical preference to subsymbolic models while not creating any overly complex snippets.

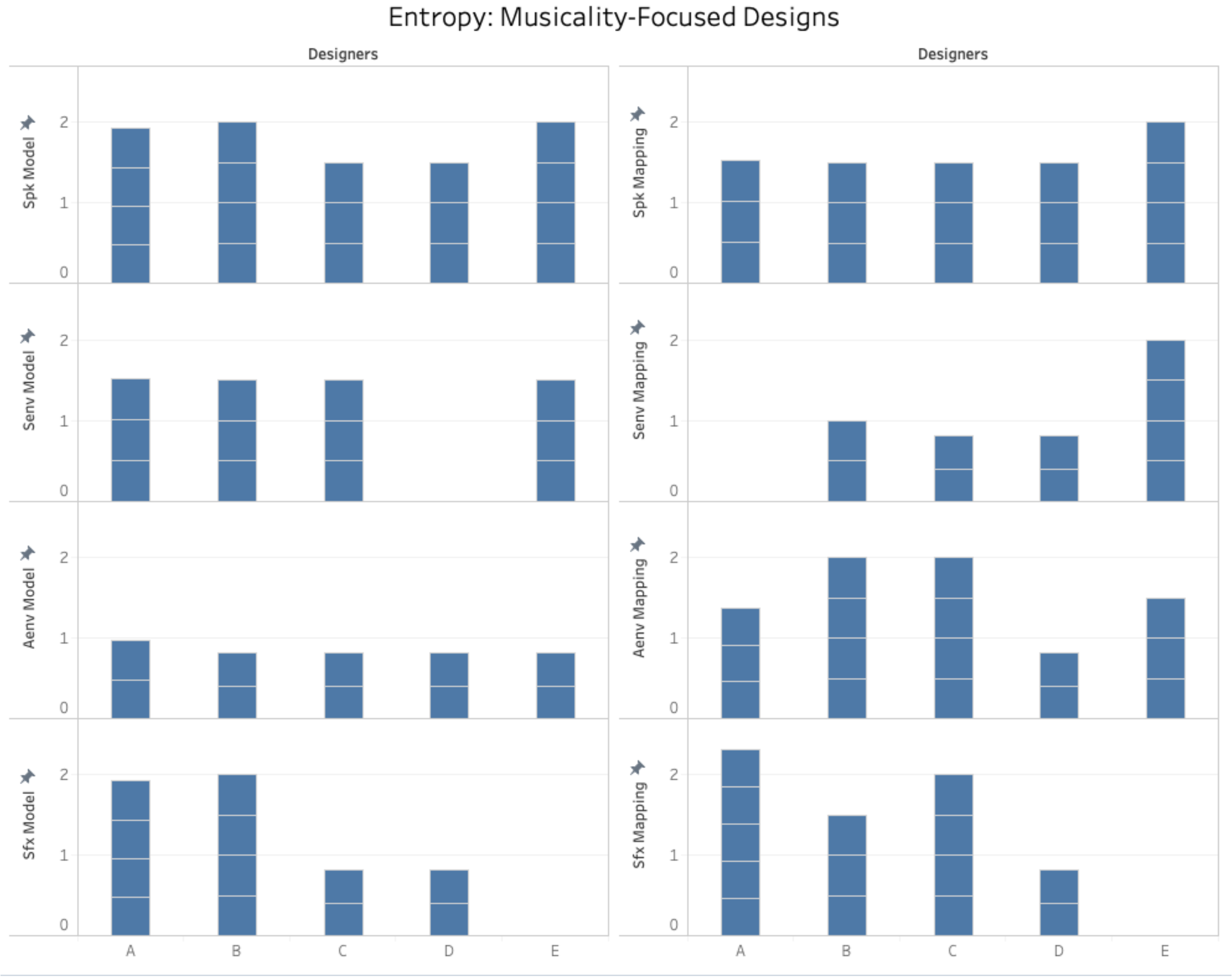


Figure 8.3 The entropy of musicality-focused snippets for each designer, grouped by model selections (left half), mapping selections (right half), and synthetic dimensions (rows).

The variety metrics (Figure 8.3) show that all the designers display relatively high entropies in Spks for both models and mappings, indicating an active exploration with this synthetic dimension. For the other dimensions, however, there is some level of contrast between the T/E and P/E designers. The T/E designers (A, B, and C) achieve relatively high entropies for the Aenv and Sefx dimensions, experimenting with various models and shapes. The P/E proponents (designer D and E), on the other hand, are noticeably more focused on the areas of exploration, that designer

D mostly experiments with Spks and some Sefx settings, while designer E explores the non-Sefx dimensions.

8.5.2 Compositional and Pedagogical Potentials

As part of assessing the musical appeals, the designer acknowledges the creative potentials of the snippets for certain parts of their composition. While they are only 'potentials' that have not been formally examined in actual compositions, they may characterize some of the design intentions guiding their workflow. For example, as designer B reviews the previous snippets in the browser, they recognize the polyphonic expressiveness of different snippet combinations. They confirm, "I wanted to [use polyphony] the whole time, playing different sounds with the [other] sounds together." For all the designers, at least some of the snippets seem to be able to inspire compositional ideas. Asked if they can picture a bigger musical piece with the snippets, designer B agrees that "yeah … this is one component that tells like the data and the other parts just support, musical aspect." Another common feedback is their interest in arranging the snippets over time. As designer A inquires, for example, "with this [environment / framework], do you have a way where you can piece the sounds together and put them all together?"

Besides the created snippets, designers A and D recognize the synthetic methodologies afford a new way of sonic expressions. Designer D confirms the merit of the spectral model- and mapping-based workflow that "it could be useful especially as a component for programming a synthesizer. Even more so for designing a sonification."

Designer A finds the listener-oriented perceptual analysis of design may add compositional as well as pedagogical values, explaining "I … like the analysis page. It really makes me think a little bit deeper about how this would connect to the data. In addition to this being really cool for composers, I feel this will also be cool educationally. I could definitely see this being used in a classroom. ... [The environment could be educational] for composers and also for musicians and even for maybe science students who are going to be using data and creating data sonification, so they can think a little bit more musically about how they can apply some different elements to it." While the analysis page and heuristics were originally created for the expert designer as the target user, this remark suggests the potential value of the page as a training tool for the non-expert listener.

8.5.3 Challenges in the Creative Process

Some designers indicate that the constraints imposed by the general methodology or environment affect their creative processes. Such observed constraints highlight how the framework elements interfere with some of the design goals. The constraints include the limited scope of data as well as the types of models and mappings provided.

First, despite that the mappable "data" are introduced to the designer as mere quantities, some (especially the data-first designers C and D) still find the structure of the data sequence to be a potential limiting factor of their aesthetic design. Designer A, for example, implies that the process of sound design may change according to the dimensionality and cardinality of data, asking the author "where do you get, [or] how do you decide what the data is?" and "[are you] going to make [the framework / environment] so the data can be more than seven points, or do you … like it being seven points?" Regarding the limitation of 'data,' Designer D shares a more specific view on the workflow of a sonification design. "When you're doing data sonification, you [often] think in terms of having a huge data stream, so dealing with just that little bit [snippets], if I was going to add another dimension to it, it would be some way to observe what happens over a longer period of time." The author interprets this meta suggestion that a data-first sound design may be informed

/ contextualized by larger time-scale trends of data, such as recurring patterns and the global contour, for creating better storytelling with the sound.

As the test environment deliberately simplifies the usage of SMSon, it also introduces various limitations. Several times, designer E indicates that their musical outputs are limited by the models and mappings (shapes) being provided. "I do think it might help if there were some more tonal not noise options [for spectral peaks], perhaps an unpitched tone. Like they're not necessarily notes in a scale or maybe instead of a scale of chromatic scale so they're all evenly spaced pitches."

In relation to the P/E workflow, however, designer C attributes the creative potentials of design to the problem-solving process, that "I feel like I can be creative and part of that is [because] there are constraints that I see [as] a challenge – like if I hear something in my head then I'll try and find the settings that I need to use. So, it is constrained, and I can't do everything I would want to do but that just makes me want to figure out other ways to do it."

Related to the constraints recognized with the framework / design environment, for several aspects of the design processes such as the synthetic timbral manipulations and selecting the models / mappings, the difficulty of verbal interpretations appears to affect the general design process as well. Designers A, B, and C, who often take the quick-paced T/E workflow, indicate the general difficulty of verbalizing the design thought processes in real time. For instance, designer B expresses that even associating the provided textual descriptors to the designed expressions can introduce a cognitive dissonance, that "it's super hard to describe [traditional] instruments as well, to describe the violin sounds that like without just being able to say, the 'violin.'" Designer D, on the other hand, finds a strong need for naming the designed snippets in order to (1) be able to remember and recall the previous sound designs and (2) characterize the often abstract designs so it can be better communicated to the listener, or can be a part of narrative-driven sonification soundscape. In terms of the interface, a common obstacle is in understanding the textual designations of the dimensions and models. Designer E, for example, recounts "the trickiest part [of the experiment] is knowing what the words mean under each pull downs, especially with spectral effects," suggesting that the textual organization of the models may hinder the predictability and planning of designs.

8.5.4 Case Studies of Musicality-Focused Design

To apply the observed patterns in actual designs, this section illustrates two examples of musicality-focused design process.

Table 8.6 The design process summary for a musically appealing snippet by designer C. Cells with a model name alone uses a constant modulation value of 0.5, while models with '| Source' are mapped with a modulation source. The ' 〃 ' denotes unchanged models / mappings from the previous step. An empty cell indicates that no models or modulations are selected. Lastly, the designer's comments are marked with double quotations, while the author's observations of the designer’s actions are marked with parentheses.

Step	Spks	Senv	Aenv	Sefx	Comments & Reactions
1	Minor Scale	Centroid | Saw-2		Phase Distortion | Data	"It's really hard to hear the data there."
2	Quasi Harmonics | Data	〃		〃	(Not intrigued)
3	Buzz | Data	〃			(Not intrigued)
4	Minor Chords | Data	〃			(Captivated with noticeable chord qualities)
5	〃	〃		Phase Distortion | Noise-Fast	"Subtle, but I do like it."
6	〃	〃	Whole | (noises)	〃	(Experiments with noise textures.)
7	〃	〃	Whole | Noise-Slow	〃	

For creating a musically appealing sonification, designer C attempts to take a systematic, or less-exploratory, approach with the data as the core / defining structure of the snippet (Table 8.6). They use the preceding snippet as the initial state, where the data is mapped to Spks:Minor Scale. First, they experiment with moving the data to various other dimensions / models such as Sefx:Phase Distortion, removing the motion from Spks (step 1). As the stationary harmonic model provides a high-pitched 'note' with subtle phase distortion, they find that "it's really hard to hear the data there." They explore other richer / denser Spks models as the primary timbral dimension with data (steps 2 to 4), but do not find them to be as musically appealing as Spks:Minor Chords. Satisfied with this simple and musical core structure, they resume the exploration of Sefx as well as Aenv modulations (steps 5 to 7). They attempt to decorate the timbre of the well-defined harmonic expression with subtle and less-structured motions, experimenting with different shapes (resolutions) of noises, as well as micro-adjusting the duration of the snippet.

Table 8.7 The design process summary for a musically appealing snippet by designer E.

Step	Spks	Senv	Aenv	Sefx	Comments & Reactions
1	Inharmonic Mix | Sq-2	Resonances		Phase Distortion | Data	"I should start with spectral peaks."
2	Inharmonic Mix | Data	〃		〃	(Experiments with duration)
3	Inharmonic Mix | Sq-4	〃			"Sounds like a weird telephone."
4	Quasi Harmonics | Sq-4	Spread | Saw-1			"Let's move this downward."
5	〃	〃	Segment: Hamming	Phase Distortion | Noise-Fast	"Very interesting."

Designer E, similarly, takes the approach of creating a core musical structure and adding subtle textural ornamentations (Table 8.7). However, unlike designer C, they seem to focus on timbral uniqueness that is not defined by the shape of data. They start by exploring the configurations of Spks (step 1)[60]. Then, they map Data to Spks, only to quickly discard for not providing a musically interesting timbral effect (steps 2 and 3). Remapping a square-shaped modulation to Spks, they experiment with the snippet duration and acknowledges that the speed

[60] This comes after assessing (multiple times) that Spks is the foundational part of the sound design in this environment. Designer E explains that they align the synthetic / timbral dimensions to the units in classical synthesizers, where Spks may correspond to the audio-rate oscillator.

affects the modulation shapes and the timbral characteristics in an unexpected way. They find an optimum point where the snippet has an extrinsic characteristic, as it "sounds like a weird telephone." While they adhere to the current motion characteristics of Spks, they also start experimenting with different combinations of Spks and Senv models (step 4). Finding a combination that yields a decaying resonance-like effect, they further adjust it with an overlapping decaying motion in Aenv.

8.6 Intelligibility-Focused Designs and Assessments

In this design challenge, the subject is asked to create three high-intelligibility snippets with low, mid, and high sound complexities. While the previous musicality-focused designs are rated subjectively for each design heuristic at the end of the design phase, this exercise encourages the designer to assess more from a listener's perspective even from the beginning of the design process.

The following observations explore both RQs 2-A and 2-B for how the designer manages the synthetic components while anticipating the receptions by the general untrained listener. The author examines the quantitative aspects of the snippets first, followed by the designer's interpretations of complexity and approaches to managing intelligibility.

8.6.1 Analysis of the Snippets and Variety Metrics

The submitted designs, marked as tasks 2-X in Appendix C, exhibit several common characteristics with the mappings and models. For example, for creating a 'high-intelligibility / low-complexity' sonification (denoted as task 2-A), every designer opts for mapping the data to Spks. Interestingly, the T/E proponents (designers A, B, and C) all select a single-harmonic-pitch model (Major Scale), while the P/E proponents (designers D and E) employ either non-pitched or more complex harmonic models. Also, all the designers select Senv:Centroid with a constant value (no modulation), except for designer A. None of them also employ a Sefx model.

Such a pattern in Spks applies for most of the mid-complexity snippets (task 2-B), while the high-complexity snippets (task 2-C) entail more diverging configurations. For the mid to high complexities, many favors to transition from Senv:Centroid to the Senv:Resonance model with non-constant modulations being applied. In addition, designers B, C, and E all transition from a symbolic Spks model to Spks:Noise-Buzz, and uniformly assess the decrease of musical appeals.

Moving from musicality-focused (task 1) to intelligibility-focused, interestingly, several designers (A, C, and D) appear to drastically change their design strategy. For example, designer D, first focused on symbolic Spks models in musical designs, shifts to subsymbolic Spks:Noise-Buzz for attaining data intelligibility. Designers B and C explore the mappings of data in a multiplicative way in musical designs but concentrates on the Spks dimension for intelligibility. Designer A, similarly, decides to shift the general mapping of data to Spks, instead of Senv in their musical exploration.

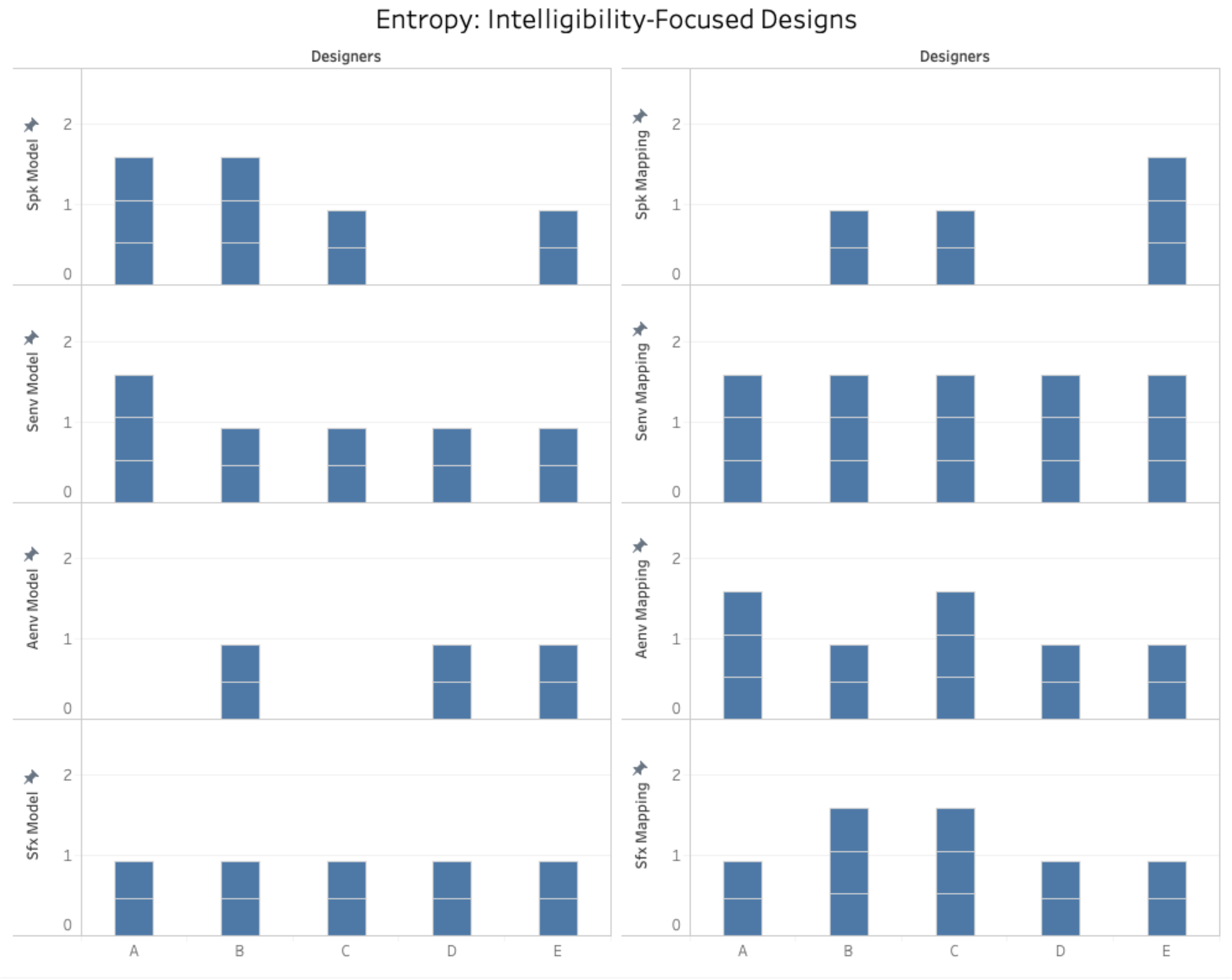


Figure 8.4 The entropy of intelligibility-focused snippets for each designer, grouped by model selections (left half), mapping selections (right half), and timbral dimensions (rows).

While this design task does not require the subject to explore musical varieties, the entropy of design configurations may provide a perspective to how they approach to increasing the complexity (Figure 8.4). There is, for instance, a noticeable contrast between designers A, D, and E. Designers A and D maintain the same mapping of data to Spks (resulting in zero entropy) but alternates the Senv models and mappings more actively. On the other hand, designer E maintains similar models while actively switching the mappings, including the data object, across all dimensions.

8.6.2 Elements of Sound Complexity

Through the design exercise of manipulating the 'sound complexity,' most participants reach a unique theory of what constitutes a complexity. Such theories may be grouped into one of the following types: (1) the physical combinations of models and mappings, and (2) the understandability of the design / organization process.

First, with the view that correlates the physical configurations of the models and mappings to a perceived sound complexity, the designer often appears to equate the complexity to the level of temporal motions in synthesis. Designers D and E argue that the complexity increases simply with the number of concurrent motions, while the others attribute it to particular mapping combinations. Designer E, for example, often notices concurrent timbral motions negatively affecting data intelligibility, that "it's very difficult to modulate more than two of these and to make it obvious, I'm finding. Modulating two... you *can* follow. Modulating one, it's very clear." However, designer E also identifies more effective types of complexity / concurrent motions that preserve the intelligibility of data. "I think there's a 'good' complexity. I was able to pull off with two, maybe three, of [the snippets] having a pattern that's different, although it would be hard if the patterns weren't regular – If you had some kind of semi-random data on spectral peaks and some different semi-random data on spectral envelope it might be trickier." Designer E also points out the inherent motions (or randomness) of Sefx models – "... but you can pull off complexity without even doing the patterns too. I was able to do it with just setting the static high value on jitter or these other spectral effects." There are several observations about particular combinations of temporal motions yielding different perceived complexity. Designer B describes, "the main [control of complexity] I did was balancing [a pattern] that's [very] rhythmical and stable, with another one that's more complex and different, mixing and balancing a big shape with a small one."

In the early phases of the design session, several designers interpreted 'sound complexity' to be largely about the clarity and understandability of a sound organization for the designer themselves, which may be tentatively called a 'process-focused' complexity. While resembling (1) the physical size of configuration aligned with timbral dynamics, designer A and D find the sheer number and behaviors of model / mapping combinations creating complexity with the mental model of a sound, rather than with auditioned results. Designer A describes, "the biggest thing that created the complexity was that there's so many different ways that you can change these sounds. Even with the amplitude envelope, it could be per segment, it could be the entire thing, but you have all these different shapes up here, and with the envelopes there's so many options and I feel like just the combinations is really what makes it complex." Designer B also tries to explain the 'process-focused' perspective, that "I interpreted [the sound complexity] as how hard it would be for me to recreate [the design] – how hard for me to know what it is exactly. Like a white noise would obviously be a white noise." When asked if this type of complexity causes the difficulty of extracting the data (intelligibility) for the listener, they respond that "not necessarily, I think." Designer E also provides another interpretation of complexity as a mixture of 'process complexity' and a perceived timbral intricacy. "I'm finding myself trying to come up with this word for the same concept every time [I audition], which is like, does it sound like all the worst parts of a synthesizer turned on at once like FM (frequency modulation) and oscillator sync, and noise versus something that's more melodic, whether it's like natural sounding or instrument sounding."

8.6.3 Managing Data Intelligibility

This section highlights some of the common strategies used to create higher data intelligibility. Such strategies include the use of pitch-oriented configurations, data-first organizations, and some considerations for the content and context of data. As the designer aims to control the sound complexity simultaneously with the intelligibility, both design heuristics are often intermixed in the description of intelligibility.

As already seen in common design patterns, for designing a 'high-intelligibility / low-complexity' snippet (task 2-A), four out of five designers opt for mapping the data to Spks with a symbolic model (e.g., major scale), while suppressing the motions in the other dimensions with static or non-oscillatory motions (see Appendix C). Designer A explains, "I know this is kind of just based off of the way I was trained with music. I just went directly to the major scale because I thought that would be easiest [for the listener to understand]. And then with the spectral effect, I was like, okay if I don't put anything on there. I know that would make it not as complex in the sound. And then, I was thinking about centroid because I feel like that centralized at least part of the sound, and then with the amplitude envelope sometimes just having it for the entire scheme, I felt would make it not as complex."

Mid- and high-complexity designs generally involve more synthetic dimensions, but with varying configurations of, e.g., where to map the data. Interestingly, designers B (sound-first and timbre-oriented) and C (data-first and pitch-oriented) take very similar design configurations for increasing the sound complexity, both with models and mappings (tasks 2-A through C). This may, however, highlight the slight difference between sound-first and data-first organization with data. Both designers appear to separate and prioritize the data dimension and musical ornamentations. Designer C (data-first) essentially reduces the complexity from musical explorations (task 1) by eliminating the non-primary (Spks and Senv) dimensions, then recreates a similar increase of complexity. Designer B (sound-first), on the other hand, lays out the general sound structure from scratch, explaining "I tr[ied] to balance stuff, so, amplitude had to be per segment, I think otherwise it would be very distracting but also not that intelligible if it was over the whole thing. I basically focus on either using spectral envelope or spectral peaks as the main. The main, like intelligible part, and the other one to try and make it sound complex and interesting."

Lastly, some designers find direct relationships between data and intelligibility. Designers D and E argue that establishing a semantic context of sonification by, for example, verbally explaining what the data are about, may help creating an intuitively understandable sonification, especially for untrained listeners. Designer D, for example, explains the elements of aesthetic decisions for a soundscape-type sonification, that "if you thought about, like maybe in a 911 call center, you might be playing the backlog like how many calls are in the queue or what's going on in terms of the crimes, so somebody could be designing the soundscape for that to either like more urgency or maybe to calm people when there's more going on maybe it's a reverse kind of thing yeah like when it's more intense, you just kind of like play some song do some background." Designer E, who typically takes a sound-first approach, also discusses a strategy for communicating the design to non-expert listeners by introducing a metaphoric mapping. Describing the 'high-intelligibility / low-complexity' snippet, "let's say we want to sonify weather – it's sunny versus rainy, rainy would be low. I would think you could hear this. And you would say, oh it got rainier and rainier and then got sunny." Listening to the 'mid-complexity' snippet, "I would say this is like intermittent rain sun rain sun, kind of." These perspectives may imply that the knowledge of data context may create an additional expectation, allowing the listener to predict and follow the trajectory of the data dimension more easily.

8.6.4 Case Studies of Intelligibility-Focused Design

Table 8.8 The design process summary by designer D for a "high intelligibility - mid complexity" snippet.

Step	Spks	Senv	Aenv	Sefx	Comments & Reactions
1	Noise-Buzz | Data	Mid-Side | Sine-2		Mirror	"I'm looking for interestingness..."
2	〃	〃		Phase Distortion	"The more interesting the sound is, the less intelligible."
3	〃	Resonances | Sine-1			"A pretty regular shape... Definitely more interesting."
4	〃	Resonances			(Now more confident about how the data behaves)
5	〃	Resonances | Sine-2			(Tries making it interesting... unexpected result)
6	〃	Resonances | Sine-4			"Sometimes it's inverted... but can tell the shape."

Designer D attempts to create a 'high-intelligibility / mid-complexity' snippet with some level of aesthetic appeals, taking a minimalistic but also somewhat exploratory approach (see Table 8.8). They indicate the element of complexity and musical appeals in terms of 'interestingness' or, as the author interprets, unexpectedness. Such unexpectedness in the design process appears to compete with data intelligibility and the P/E design workflow.

From the previously designed snippet, they find the Data mapped to Spks:Noise-Buzz to be highly intelligible. Having this mapping as the baseline, they start with an exploration of new combinations to increase the complexity (steps 1 and 2). They share their thought process, "I'm just looking for 'interestingness,' I guess, focusing on data intelligibility." With the added complexity and unpredictability of Sefx models, they describe "it's almost like, the more that you try to get an interesting sound, the harder it is to know what the effect of data is having on it." They then fall back to a more familiar / predictable configuration (step 3), explaining "I like resonances, and [Sine-1] is a pretty regular shape, so..." They find the result to be expressive and "more interesting," but obscure the presence / behavior of the data. They attempt to simplify the sound structure by removing the mapped shape from Senv:Resonances, and shuffle the data to verify the baseline intelligibility (step 4). After gaining some confidence / familiarity, they bring back the previous modulation of Senv:Resonances with faster sinusoidal shapes (steps 5 and 6). Although the result appears to be a little drastic and still surprising, they are satisfied with the balance that "I think I can almost always tell there's the shape of the data. Sometimes it's ‘inverted’ though."

Table 8.9 The design process summary by designer E for a "high intelligibility - high complexity" snippet.

Step	Spks	Senv	Aenv	Sefx	Comments & Reactions
1	Quasi Harmonics	Resonances	Whole | Sine-Decay		
2	〃	〃	Segment | Sine-Decay		(Recognizes rhythms)
3	〃	〃	Segment | Saw-2		"Now that's cool."
4	〃	〃	〃	Jitter | Data	"Let's go crazy with data... Works better than I thought."
5	〃	Resonances | Data	〃	〃	"So cool, sounds so much like a percussion."
6	〃	Resonances | (various)	〃	〃	"I can't tell which is higher or lower."
7	〃	Resonances | Const-X	〃	〃	(Alternates between different Const shapes)
8	Quasi Harmonics | Decay	Resonances | Triangle-2	〃	〃	"For more complexity, add a pattern on every single thing." |
9	Inharmonic Mix | Sq-1	〃	〃	〃	"Now it's harder to tell where the data is."
10	Noise-Tonal | Rise	〃	〃	〃	(Recognizes harmonic stability)
11	Noise-Buzz | Const-High	〃	〃	〃	"I guessed [the data] right, but upside-down."

With a clear plan-and-execute approach, designer E demonstrates a design process for 'high-intelligibility / high-complexity' snippet (Table 8.9). This example highlights (1) an effective approach of ensuring a complex musical structure as a baseline before encoding data, and (2) the issue of an implicit directionality in relation to data intelligibility and sound complexity.

In steps 1 through 3, they establish a rhythmic expression as a baseline musical structure, without using the data anywhere. They particularly focus on Senv:Resonances that they found to have a balance of musicality and salience. In step 4, they strategize a high-complexity and low-intelligibility configuration by mapping the data to Sefx:Jitter, finding that the intelligibility is somewhat retained unexpectedly. They then (step 5) duplicate the data mapping to Senv:Resonances, yielding a highly salient and musically complex expression that they describe as percussion-like.

Attempting to increase the complexity further, however, they realize the limitation with implicit directionality for the models and mappings. For instance, not being able to experiment with the data mapped to Senv:Resonances, they inquire the author about creating vertically inverted variations, that "I wonder if you want to have an option for negative data. Because all the other [shapes] have a negative version." They then experiment with different motional combinations (step 6), further realizing that the negative impact of directional uncertainty of Sefx:Jitter on intelligibility, stating "so this one, I can't tell which is higher or lower. It's just the nature of effect. I can't tell between the three different jitter sizes what the order is."

After simplifying the mapping and gaining some confidence with the directionality of Sefx:Jitter (step 7), they resume combining different mapping shapes. "Really, the way to add complexity, I think, is to add these little pattern[s] on every single thing." Switching to different Spks model / mapping (step 9) appears to decrease the intelligibility, as they find an unexpected data shape after revealing them. They analyze, "it's more complex, but I feel like it is harder to tell where the data is."

In step 10, they encounter a more simplistic tonal structure with Spks:Noise-Tonal, finding a unique musical appeal. They inquire if there is an opposite, sawtooth-like tonal structure, settling with the Spks:Noise-Buzz at a constant high frequency (step 11). While they are content with this final configuration for the intelligibility-complexity balance, they find the impression of the data shape is again "upside-down."

8.7 Discussion: Combinations of Design Procedure and Sound Organization Strategies

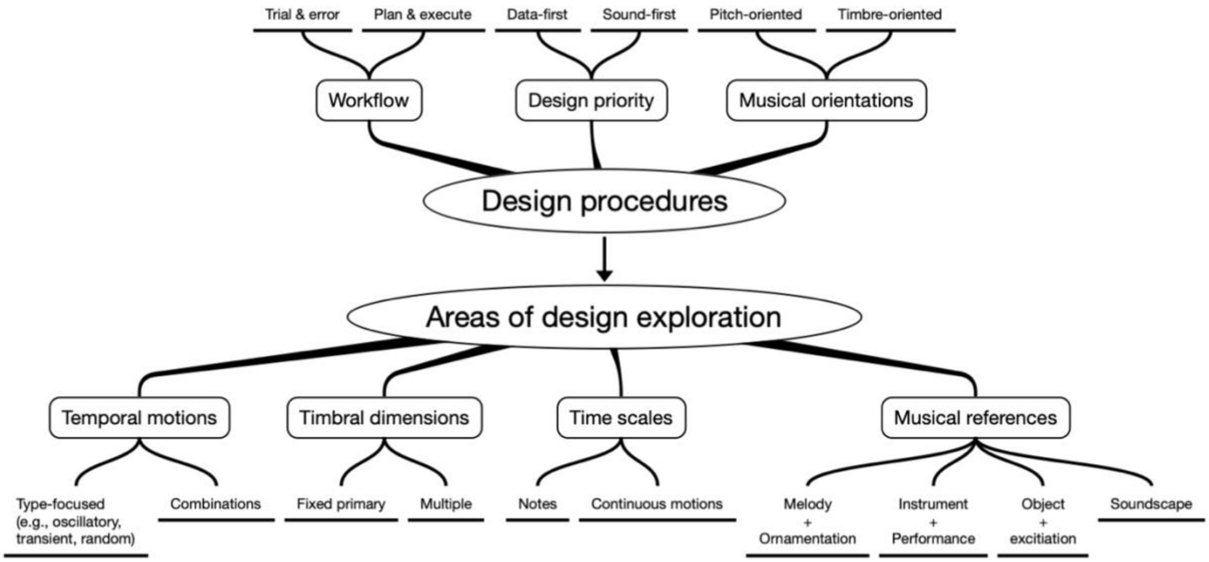


Figure 8.5 A consolidated overview of design procedures and sound organizations. Each designer selects and combines specific approaches (the leaf nodes with underlines) under different areas of focus (the rectangular nodes), creating a unique design pattern.

From the analysis of the snippets and verbal responses, several contrasting design patterns have emerged that are interwound and multimodal (Figure 8.5). The choice of design approaches is often affected moment-by-moment by the goals and stages of a design, as well as the designer's interpretations and strategies for satisfying certain design heuristics. In order to consolidate the findings to address RQ 2, this discussion aims to re-frame them, such that if SMSon and its implementation support the 'variability of design' principle (RQ 2-A) and the independent control of the listener-focused design heuristics (RQ 2-B).

With regard to RQ 2-A, there appears to be several common selections / combinations of framework elements (i.e., models, etc.), but also different scopes of how to use or explore with the elements. For example, for a high-intelligibility / low-complexity design, the use of a symbolic Spks model with Senv:Centroid is overwhelmingly popular. However, as the target complexity increases, the interpretations and design strategies diversify, creating several unique scopes of exploration. Designers C and D, for instance, attempt to add musical ornamentations utilizing additional synthetic dimensions, while designers B and E experiment with the combinations of temporal motions to increase the complexity. For musical explorations (task 1), generally, there are several directions of exploration focused on, e.g., oscillatory / structured patterns (designer A) or random motions (designer B), and mapping targets (designers B and C). Perhaps most interestingly, between tasks 1 and 2, all the designers displayed some amount of transition in such scopes from one mode to another, not necessarily in the same direction. The author considers that observing such multimodality, or interchangeability, within a single designer supports the 'variability' principle of SMSon.

For RQ 2-B, the assessment of design with heuristics also proved to be complicated with different types of designers offering contrasting interpretations. Notably, timbre-oriented designers (A and B) tend to align musical appeal with complexity, while pitch-oriented designers (C and D) generally find it positively correlate with data intelligibility (and negatively with complexity). This contrast also seems to reflect on the sound-first and data-first organizations as well, especially during the musicality-focused exploration (task 1). Designers with opposing workflows or orientations sometimes offer surprisingly similar responses. For example, the data-first designers (C and D), as well as designer A (sound-first), indicate that the structure and/or source of data may influence their musical / compositional direction. Designers D (pitch-oriented) and E (timbre-oriented) also share the idea that the communication of data context to the listener would be beneficial for establishing an intelligible sonification with a "predictable" musical structure. When focused on the presentation of the sound, the distinction between intelligibility and complexity remains somewhat unclear for most designers. Some expressed that process-focused complexity for the designer themselves may correspond to the perceptual response by the listener. This unclarity of perspective requires further investigation.

This study, therefore, revealed the complex multiplicity and interdependence of design techniques, design heuristics, and musical preferences. Regarding the independent and satisfactory control of both personal aesthetics and functionality of a sonification, the results indicate the possibility with only specific and limited approaches (e.g., a timbre-oriented combination of motion shapes, or data-first melodies with musical ornamentations). However, it has identified various concrete cases of how a musical idea may connect to the elements of SMSon and may be assessed and improved from the potential listener's perspective.

CHAPTER 9. DISCUSSIONS AND CONCLUSIONS

This thesis, in summary, investigated the general incompatibles between functional and aesthetic approaches to creating a sonification. Such incompatibilities were exemplified as the limited range of creative designs usable for functional sonifications, and the difficulty of communicating / understanding unique sound organizations (see chapters 1.2 and 3.1). To address them mainly from a design-methodologies perspective, the author developed spectromorphing sonification (SMSon). SMSon set four functional-and-aesthetic (F/A) goals, representing the properties of a sonification that were previously difficult to attain simultaneously (see chapter 3.5.1). Two user studies assessed how SMSon-based design was / was not able to address the aforementioned incompatibilities in relation to the F/A goals. The following discussions recapitulate some of the findings for potentially resolving the incompatibilities. In addition, several notable types of SMSon design with their distinct advantages and disadvantages will be presented as design guidelines. Lastly, the thesis concludes with a brief discussion of future work for SMSon and user experiments.

9.1 Closing the Gaps Between Functional and Aesthetic Values

9.1.1 Broadening the Range of Expressions

The user studies showed some promising results, especially in terms of expanding the range of aesthetic expressions with novel and more complex sound organizations. This was previously hampered in order to ensure the transparent display of data in sound. In this regard, the designs with SMSon attained a variety of designs as well as considerably intricate expressions with timbre without noticeably sacrificing the measurability of data.

For example, the listener's statistical responses indicated that neither changing the spectral models (algorithms) while retaining the mappings nor adding auxiliary temporal motions to unused timbral dimensions diminish the estimation accuracy (see chapter 7.4.2). This suggests flexibility for the designer to experiment and interchange the baseline models or auxiliary motions to create more unique expressions. Some listeners, however, pointed out the lack of perceived contrasts within individual stimuli, highlighting the limited range of expression in some spectral models (see chapter 7.5.5.1).

The design study, on the other hand, yielded a wide variety of snippet designs overall, reflecting the range of different musical ideas. Some process-focused designers expressed that the combinatorial options of models and mappings enabled a complex design exploration with many unexpected outcomes (see chapter 8.6.2). In addition, all the designers spontaneously displayed a dynamic shift of methodologies for different tasks, such as between pitch- and timbre-oriented designs and the areas of focus with model / mapping types (see chapter 8.7).

SMSon, therefore, has a strong potential not only with the variability of design and more complex expressions but also the flexibility for applying multiple design methodologies.

9.1.2 Connecting the Design Intentions and Receptions

In terms of facilitating better communication and/or understanding of bespoke designs, the aural exploration of SMSon-based designs appeared to be effective to a limited extent. The user studies revealed (1) cases where focused and qualitative analysis by the listener contributing to their quantitative estimation of data, and (2) various design attributes (i.e., changeable but characteristic properties of sound; see chapter 3.4.3) that may be similarly recognized by the listener and the designer. The response patterns were rather complex and diverse, preventing direct comparisons between users. However, there appear to be some similarities in how both the listener and designer narrow down and focus their aural explorations, creating a combination of variables such as different time scales, motional types, and harmonicity. These attributes were explored extensively by the same user from multiple angles, while other variable elements were relatively untouched / unnoticed.

For the listener, exploratory scopes often resembled the temporal focal length (see chapter 4.4) with a unique placement of analytical attention to the details of a sound. Some listeners, for example, tended to only recognize high-level impressions about stimuli, such as the similarity to a real-life sounding object or a single characteristic (e.g., 'fluctuating') applied to the whole of a stimulus. Others explored further into how such a sound object behaved, identifying its inner-structural characteristics (chapter 7.5.6.2). With creative uses of descriptors, some focused on the plurality or ambiguity of timbral dimensions, some analyzed more of the general frequency-domain contents, while some explored the temporal dynamics of timbre in different time scales (chapter 7.5.5.1).

For the designer, explorations seemed to be directed and bounded by task criteria (e.g., high intelligibility) as well as their musical orientations. Similar to the listener's aural analysis, design explorations generally entailed a focused experiment of certain sound structural elements. For example, timbre-oriented multidimensional designs typically experimented with the shapes, rates, and combinations of either oscillatory or non-structured temporal motions, while the types of spectral models were kept relatively unaltered (chapter 8.4.3). Designs with a core musical structure (e.g., a melodic shape) and ornamentations generally iterated over similar mapping sources (shapes) and/or spectral models to achieve a nuanced expression and a desired balance. Designs inspired by traditional synthesizers aimed to recreate the iconic effects such as frequency modulation with fewer components (chapter 8.5.4). Unlike most listeners, the designer also displayed the tendency of shifting the scope of design exploration from one to another depending on the main design goal (e.g., intelligibility-focused or musicality-focused).

In short, there appear to be certain groups of design attributes heavily explored by the listener and designer. While such attribute groups vary for different users, anticipating and aligning the focus of aural exploration may lead to better design communications. The following section discusses several design patterns for guiding and aligning the designer's intentions and listener's expectations.

9.2 Design Guidelines and Applications

Considering the potential overlaps between design and listening explorations, this section highlights three design patterns in SMSon. These patterns were observed or mentioned in both user studies, and may provide distinct advantages or disadvantages for, e.g., design communication, aural decomposition, and potentially creating larger-scale musical compositions. They are namely source-motion, melody-ornamentation, and mixed-voice patterns.

9.2.1 Source-Motion Approach

The first design approach is based on a spectromorphological view of timbral structure, employing distinct motional types that are easier to recognize and classify. A source-motion expression, e.g., a physical instrument-like sound excited (performed) by dynamic and characteristic motions (rather than melodic modulations), is relatively straightforward to implement with the model-mapping structure of SMSon. For example, inharmonic spectral peaks with a transient amplitude envelope may provide an impression of a metallic percussion struck with various intensities (see chapter 7.5.6.2).

Physical object-inspired designs may provide the listener some level of aesthetic relatability (familiarity), while drawing their attention towards subsymbolic temporal elements. A deliberate selection of distinct temporal motions, such as an oscillatory pattern mixed with a subtle random motion, may also ease the process of recognizing and describing the timbral dimensions (also chapter 7.5.6.2).

Even though such a design should strive for structural transparency, when presented in a real-life situation, some of the design peculiarities may not be self-explanatory enough with the auditory expression itself, requiring verbal communication. A verbal explanation could employ a source-motion analogy with common (non-technical) motional terms. For example, following the descriptive patterns with extrinsic references (chapter 7.5.5.2), the designer might introduce a sonification articulating the similarity to an existing instrument or object – "Listen for the sound qualities of this virtual string instrument. It automatically tunes itself up and down like a bouncing spring. The 'data' is hitting the body of the spring with different strengths." Alternatively, however, the author suggests the utilization of shorter 'hints' similar to the textual descriptors employed in the experiments. For example, "this sound has multiple motions. Can you find the vibrating / fluctuating / falling / etc. motions?" Rather than presenting a version of complete imagery, an abstract / incomplete description may potentially (1) incite a spontaneous aural exploration that is crucial for recognizing the inner (subsymbolic) structures of a sound (chapter 7.5.4.3) and (2) encourage the use of idiosyncratic perspectives / terms to fill the details (chapter 7.5.5.1). After all, no single verbal description can fully convey the design intentions to every listener with a unique aesthetic perspective and sound literacy.

9.2.2 Melody-Ornamentation Approach

In contrast to the previous timbre-focused model, a melody-ornamentation pattern is largely confined to traditional symbolic concepts such as pitch and note. The mapping scheme is also often constrained to data being mapped to a spectral-peaks model. Nevertheless, this was noticeably a popular way of how the listener interpreted (see chapter 7.5.6.1) and how the designer formulated the designs for high-intelligibility and low-complexity sonifications (chapter 8.6.1).

As a distinct advantage, such a conventional mapping provides a self-explanatory quality for design communication. In addition, this approach may naturally guide a design exploration as it establishes clear primary (i.e., melody, rhythm, etc.) and secondary (i.e., ornamentations) roles in a sound. This allows the designer to apply a technique similar to musical data moves to create subtle variations (see chapter 5.3.6), where replacing the secondary parameters creates a smooth transition. Several data-first and plan-and-execute type designers demonstrated the use of a primary structure, typically data mapped to a harmonic spectral-peaks model, as an anchor, while gradually changing the secondary effects with less distinct (e.g., non-oscillatory) motions and models (chapter 8.5.4). Designer C suggests that the spectral effects and amplitude envelope dimensions are suitable for such fine-tune adjustments (chapter 8.4.2).

Melody-based designs also have a clear downside of obscuring the timbral nuances for the listener, as frequently observed (sections 7.5.6.1 and 7.5.6.5). When data were mapped to timbral dimensions other than spectral peaks, it would possibly hamper the data estimation accuracy. On the other hand, in order to direct the listener’s attention to non-pitched elements of timbre, it may be useful to fix and emphasize a stationary pitch, as described by a listener (chapter 7.5.6.1; Table 7.17.8). Such a design technique may be effective for the source-motion as well as the mixed-voice models discussed next.

9.2.3 Mixed-Voice Approach

Lastly, the mixed-voice design approach extends the earlier models toward more dynamic (with multiple timbral characteristics) and/or polyphonic expressions. This view of complex sound was shared by several listeners who perceived the stimuli as a mixture of different sound sources or instruments. These listeners also generally attained higher estimation accuracy (see chapter 7.5.6.5).

As a major benefit, a mixed-voice design addresses the lack of contrast between timbral states (e.g., ‘high’ and ‘low’) that some listeners experienced and possibly aid their data estimation (see section 9.1.1). One way of creating such an effect is by mapping data / quantities to multiple synthetic dimensions, thus modulating various aspects of timbre in synchronization. This one-to-many mapping enhances the contrast between ‘low’ and ‘high’ state, especially with non-pitched (timbre-oriented) configurations, as demonstrated by several designers (chapter 8.4.3). However, it can also result in a static timbre at any given moment with fewer auxiliary temporal motions being utilized.

Another approach is to simply layer two distinct snippets / stimuli driven by the same data. The snippets, or voices, cross-fade from one to the other with the data modulating the amplitude envelope. This ‘polyphonic’ approach requires the designer to work toward creating unique and complex timbres that stand out from the others. A designer explains the strategy of combining and balancing the shapes of auxiliary motions, in terms of different time scales (e.g., whole snippet vs. segment), structures (e.g., oscillatory vs. random), and periodicities (chapter 8.6.2).

Beyond creating a mixed voice with high contrast, unique and discernible snippet designs also provide compositional possibilities. Several designers expressed a strong interest in creating a polyphonic musical piece featuring data-focused voices as well as aesthetics-focused voices (chapter 8.5.2). For the latter type, one may even omit using the data altogether and only map auxiliary shapes, creating a characteristic background rhythm or texture, as demonstrated by a designer (chapter 8.5.4). Ultimately, a polyphonic piece could employ all three types of exploratory designs discussed so far, with the designer carefully balancing and ensuring the contrast and discernibility of the data-focused voices.

9.3 Future Work

The user studies of SMSon revealed numerous factors that were potentially significant. However, these factors were only indirectly comparable between the two studies, mainly because of the asymmetry of the experimental designs. For example, the listener showed limited speculations of design processes as they were not informed with the synthetic methods, which the designer heavily explored. The designer, on the other hand, focused less on describing the traits of intelligibility as they spent less time on formal evaluations (i.e., the description and estimation tasks) but more on the elaboration of the sound structures and intentions.

Another notable difference was the level of attention towards the visual references, which the listener tended to gravitate toward, but the designer rarely acknowledged. In the listening test, the visual representation of data discreetly provided the directionality and scaling references in the estimation task as well as a generic 'segmented' time-domain shape of a sound object and playback positions[61]. The design test, in contrast, employed an informal estimative method that did not include fixed referential segments. The designer would show or hide the data sequence while randomizing the sequence or browsing the previous snippets in quick succession. In such a

[61] This inclusion was a necessary compromise for the accuracy-oriented listening tasks. The presence of such references may be questionable in terms of generalization, as they are likely not available in real-life applications of sonification.

condition, they reported that sometimes the general shape of their estimation was correct but was vertically inverted against their intuition. This suggests the importance of somehow communicating the polarity of data and timbral motions, especially when a visual reference is not available to the listener. Future user studies, therefore, may require the rebalancing of design and evaluative tasks as well as synthetic, aural, and visual explorations, to acquire more symmetric and comparable user responses.

The SMSon framework also has areas for further development. Some designers, for example, expressed interest in extending the ‘data’ view from the current seven points to a much larger size (e.g., 30), and even facilitating the use of real-life and multidimensional data (see chapter 8.5.3). Such extensions of data also imply a larger time-scale sound organization, instead of the current short sound-object designs. Ultimately, these extensions may lead to a general compositional environment that facilitates the multitrack arrangement of data and sonification voices. Some critical requirements for this include the interface design for reconfiguring the models and mappings within the timeline. Although it might be sufficient to only use and rearrange prerendered sonification snippets, the design and exploration of timbral sonification involve iterative adjustments of various parameters, as observed in chapter 8. Also, a designer suggested that the inclusion of evaluative tools (similar to the analysis page; see chapter 8.2.3) may be beneficial for organizing design intentions (chapter 8.5.2), even for a larger scale multitrack environment.

In addition, as another designer pointed out, the creative process with SMSon in the study was limited to a finite permutation of predefined models and mappings, constraining the range of expression (chapter 8.5.3). A future iteration of SMSon should allow the designer to modify or create new spectral models (algorithms) as well as temporal shapes for auxiliary mappings. This may be approached in two ways that Sonar already partially supports: (1) providing the user a programming or graphical interface to easily generate and transform spectral audio contents; and (2) allowing the use of existing audio samples as the source of spectral materials. For example, the designer might wish to feature an inharmonic expression that resembles a particular bell sound. For this, they could generate a spectral peaks model either by (1) micro-adjusting individual peaks with the Mono.comb, phase, etc. functions in Sonar (chapter 6.2.2) or (2) importing, analyzing, and normalizing (flattening) slices of a bell sample to create spectral wavetables (chapter 6.2.3). Such a custom model, however, must ensure that any input data can create a wide range and dynamically evolving timbral expressions.

9.4 Conclusion

This thesis explored the challenges in combining functional and aesthetic goals in sonifications, with a new design framework inspired by electroacoustic theories of sound organization. Two user studies assessed the framework, revealing not only various factors signifying the complexity of the challenge but also the potentials for novel design solutions. In conclusion, the author believes that the findings of this research will serve as valuable information for constructing new sonification practices and environments with more flexibility, clarity, and innovative opportunities.

APPENDIX A. MUSICAL DATA MOVES

Table A.1 The attributes of the periodic-table data set employed in the Musical Data Moves examples (chapter 5.3).

Attribute	Values Summary
AtomicNumber	1–112
Name	112 values
AtomicWeight	1.01–277 amu
Symbol	112 values
MeltingPoint	−272–3500 ° C
BoilingPoint	−268.6–5660 ° C
Density	0–22.5 g/cc
Year_Discovered	1669–1996
IonizationEnergy	3.89–24.59 eV
Table_Row	1–9
Table_Column	1–18
ChemicalSeries	10 values
AtomicRadius	0.3–2.7 Å
NaturalState	4 values
ThermalConductivity	0–429 W/m
SpecificHeat	93.62–14300 J/kg
ElectronAffinity	0–3.61 kJ/mole
NumberOfIsotopes	1–29
Classification	3 values
CrystalStructure	9 values
Discoverer	74 values
Cost	71 values
NaturalForming	6 values
PrimaryUse	84 values

APPENDIX B. LISTENING TEST SURVEY QUESTIONS

B.1 Survey on the Experiment

1. In general, were the word pairs in **sound-description** tasks helpful to analyze the sound, or rather misleading? Please provide reasons to the best you can.
2. Did the order of tasks (i.e., **description-first** or **estimation-first**) affect the difficulty of the **value-estimation** tasks? If yes, in what ways?
3. What were the most confusing or challenging elements in this listening test (besides the task criteria)?
4. Do you think the types of sound in this experiment could be used in a musical context? If yes: in what ways? If no: how so?
5. Please share with us any other feedback about the experiment!

B.2 Goldsmith Musical Sophistication Index

B.2.1 Questions

1. I spend a lot of my free time doing music-related activities.
2. I sometimes choose music that can trigger shivers down my spine.
3. I enjoy writing about music, for example on blogs and forums.
4. If somebody starts singing a song I don’t know, I can usually join in.
5. I am able to judge whether someone is a good singer or not.
6. I usually know when I’m hearing a song for the first time.
7. I can sing or play music from memory.
8. I’m intrigued by musical styles I’m not familiar with and want to find out more.

9. Pieces of music rarely evoke emotions for me.
10. I am able to hit the right notes when I sing along with a recording.
11. I find it difficult to spot mistakes in a performance of a song even if I know the tune.
12. I can compare and discuss differences between two performances or versions of the same piece of music.
13. I have trouble recognizing a familiar song when played in a different way or by a different performer.
14. I have never been complimented for my talents as a musical performer.
15. I often read or search the Internet for things related to music.
16. I often pick certain music to motivate or excite me.
17. I am not able to sing in harmony when somebody is singing a familiar tune.
18. I can tell when people sing or play out of time with the beat.
19. I am able to identify what is special about a given musical piece.
20. I am able to talk about the emotions that a piece of music evokes for me.
21. I don't spend much of my disposable income on music.
22. I can tell when people sing or play out of tune.
23. When I sing, I have no idea whether I'm in tune or not.
24. Music is kind of an addiction for me I couldn't live without it.
25. I don't like singing in public because I'm afraid that I would sing wrong notes.
26. When I hear a piece of music I can usually identify its genre.
27. I would not consider myself a musician.
28. I keep track of new music that I come across (e.g. new artists or recordings).
29. After hearing a new song two or three times, I can usually sing it by myself.

30. I only need to hear a new tune once and I can sing it back hours later.
31. Music can evoke my memories of past people and places.
32. I engaged in regular, daily practice of a musical instrument (including voice) for 0 / 1 / 2 / 3 / 4-5 / 6-9 / 10 or more years.
33. At the peak of my interest, I practiced 0 / 0.5 / 1 / 1.5 / 2 / 3-4 / 5 or more hours per day on my primary instrument.
34. I have attended 0 / 1 / 2 / 3 / 4-6 / 7-10 / 11 or more live music events as an audience member in the past twelve months.
35. I have had formal training in music theory for 0 / 0.5 / 1 / 2 / 3 / 4-6 / 7 or more years.
36. I have had 0 / 0.5 / 1 / 2 / 3-5 / 6-9 / 10 or more years of formal training on a musical instrument (including voice) during my lifetime.
37. I can play 0 / 1 / 2 / 3 / 4 / 5 / 6 or more musical instruments.
38. I listen attentively to music for 0-15 min / 15-30 min / 30-60 min / 60-90 min / 2 hrs / 2-3 hrs / 4 hrs or more per day.
39. The instrument I play best (including voice) is _____.

B.2.2 Responses

1. Completely Disagree
2. Strongly Disagree
3. Disagree
4. Neither Agree nor Disagree
5. Agree
6. Strongly Agree
7. Completely Agree

B.3 The Statistics of Listening Task Order and Musical Background

Table B.3.1 ANOVA type III (unbalanced groups) tests with ‘combined’ efficiency t-score as the dependent variable. The response to a Gold-MSI question (Q22) appears to have no significance but a potential relationship to the effect of listening task order.

	Df	Sum Sq	F value	Pr(>F)	Significance
MSI Q22	5	3.98	2.931	0.39074	
Order	1	4.95	8.213	0.01097	*
Q22:Order	5	9.94	2.786	0.02350	*
Residuals	1288	982.07			

Contrarily to the hypothesis for Research Question 1-B, that a preceding qualitative listening should increase the accuracy and/or efficiency of the following quantitative estimation of data, t-tests indicate no significant effect by the order of listening tasks. Similarly, the individual responses to the Gold-MSI musical background survey (see Appendix B.2) generally show no significant effect on the estimation accuracy. However, the ‘combined’ efficiency (the overall speed) of description-estimation tasks may have some patterns when sub-grouped by the listening task order and several of the Gold-MSI responses, namely questions 19, 20, 22, and 26. For example, with regard to question 22, the sensitivity to the pitch of singing may relate to the effect of listening order on the estimation speed (Table B.3.1). Subgroupings with the other questions and order show a similar statistical response.

These statistical effects are, however, difficult to interpret as posthoc analyses indicate no linear relationships among the factors. At present, the author speculates that the listener’s musical orientation has an influence on both the qualitative and quantitative analytical listening behaviors.

APPENDIX C. THE DESIGN TEST SNIPPETS DATA

The table on the next page shows the final configurations and subjective ratings of the snippets created by the participants in the design study (chapter 8). Task 1 snippets are musicality-focused, while task 2-X's are intelligibility-focused. Model / mapping cells with the red background indicate unique selections within subject. Unused properties are indicated with 'null' with yellow backgrounds.

Designer	Task	Spk Model	Senv Model	Aenv Model	Sfx Model	Spk Mapping	Senv Mappin	Aenv Mappir	Sfx Mapping	Duration	Intelligibility	Complexity	Musicality	Descriptors	Notes
A	1	major_chord	reson	segment	null	rise	data	hamm	null	3.99735363	0.95215179	0.55215179	0		
A	1	major_chord	reson	segment	mirror	rise	data	rise_logistic	sq_2	2.53833117	0	0.76415179	0		
A	1	quasiharm	spread	segment	odd_even	rsteps_8_1	data	rise_logistic	decay_exp	2.53833117	0.22015179	0.93615179	0.98415179	[Gentle-Metallic,Close-Dis	
A	1	parallel_cho	stria_1	whole	mirror	sin_2	data	rise_logistic	sin_decay	2.53833117	0	0	0		
A	1	two_random	stria_1	whole	scan	sin_2	data	tri_2	rand_8	2.53833117	0	1	0.93215179		
A	2-A	major_scale	centroid	segment	null	data	hamm	decay_exp	null	2.53833117	1	0.28815179	1	[Transition,Downbeat,Rele	
A	2-B	quasiharm	reson	segment	jitter	data	decay_4	tri_4	low_mid	2.53833117	0.90815179	0.48815179	0.14415179	[Portamento / carrying,Vib	
A	2-C	noisetonal	stria_2	segment	null	data	rsteps_8_1	rand_48	null	2.53833117	0.63215179	1	1		
B	1	noise_buzz	reson	whole	phase_dist	data	rand_24	rsteps_8_3	rsteps_8_1	4.19071806	0.82015179	0.38015179	0.78815179		
B	1	parallel_cho	stria_2	segment	scan	data	rand_24	rand_48	data	4.19071806	0.14815179	0.72815179	0.76415179		
B	1	major_chord	stria_3	whole	jitter	rand_24	data	data	data	0.5	0	0.30415179	0.40015179		
B	1	noisetonal	stria_3	whole	null	sin_rise	data	rise_logistic	null	5	1	0.34415179	0.60415179		Rhythmic bui
B	2-A	major_scale	centroid	whole	null	data	low_mid	low_mid	null	2.99537435	1	0.16415179	0.02815179	[Clarinet]	
B	2-B	minor_chord	reson	segment	phase_dist	data	rsteps_8_3	sq_4	hamm	2.99537435	1	0.61615179	0.86415179		
B	2-C	noise_buzz	reson	segment	phase_dist	rand_24	data	sq_4	sq_1	2.99537435	1	0.79615179	0.68815179		
C	1	minor_scale	spread	segment	phase_dist	sin_rise	data	null	rise_logistic	2.20507813	0.776	0.588	0.736		Musical appe
C	1	minor_scale	spread	segment	phase_dist	data	data	tri_1	data	2.20507813	0.308	0.54	0.796	[Pleasant-Ur	Sweet meloc
C	1	minor_chord	centroid	whole	phase_dist	data	decay_2	rand_8	rand_48	1.8359375	0.88	0.82	0.78	[Trumpet,Gentle-Metallic,	
C	1	noisetonal	stria_1	segment	scan	rsteps_8_1	data	rise	hamm	2.1171875	0.632	0.736	0.28	[Smooth-Rou	To me this is
C	2-A	major_scale	centroid	segment	null	data	mid	null	null	2.64453125	1	0.08	0.212		Sounds borin
C	2-B	major_scale	reson	segment	phase_dist	data	rand_8	rand_48	rise_logistic	2.57421875	0.988	0.484	0.612		Sounds almo
C	2-C	noise_buzz	reson	segment	phase_dist	rand_8	data	sin_1	sin_4	1.80078125	0.864	0.904	0.38	[Smooth-Rou	Borders on n
D	1	major_scale	reson	segment	null	data	decay_4	sin_2	null	4.64776124	0.88015179	0.10815179	0.82815179	[Clarinet,Legato / tied tog	
D	1	major_scale	reson	segment	null	tri_4	decay_4	null	null	4.64776124	0	0	0		
D	1	minor_scale	reson	segment	phase_dist	data	decay_4	null	hamm	4.64776124	0.76015179	0.33215179	0.82815179		
D	1	quasiharm	reson	whole	null	sin_rise	data	null	null	4.64776124	0.28415179	0.86815179	0.16015179		80's Sci-fi
D	2-A	noise_buzz	centroid	whole	null	data	null	null	null	2.88990285	0.72015179	0.65615179	0.46815179	[Natural-Mechanic,Prolong	
D	2-B	noise_buzz	reson	whole	null	data	sin_4	null	null	2.22191666	0.51615179	0.76815179	0.24415179	[Steady-Fluctuating,Colora	
D	2-C	noise_buzz	reson	segment	phase_dist	data	rsteps_8_1	rand_24	sin_decay	3.66336054	0.70015179	0.92015179	0.66415179	[Thin-Rich,Gong]	
E	1	noise	centroid	segment	null	null	rise	null	null	3.5	0	0.13615179	0.55215179	[Narrow-Broad,Soft-Hard,	
E	1	minor_scale	reson	segment	null	data	data	sin_4	null	4.20829664	0.94415179	0.37615179	0	[Natural-Unnatural,Thin-R	
E	1	quasiharm	spread	segment	null	sq_4	decay_1	hamm	null	4.66533982	0	0.40415179	0.80015179	[Natural-Mechanic,Gentle	
E	1	noisetonal	spread	whole	null	hamm	decay_2	hamm	null	4.66533982	0	0.39615179	0.46415179	[Gentle-Metallic,Close-Dis	
E	2-A	minor_chord	centroid	whole	null	data	null	null	null	3.62820337	1	0.21615179	0.16815179		
E	2-B	minor_chord	centroid	whole	null	hamm	data	null	null	3.62820337	0.89215179	0.61615179	0.50815179		
E	2-C	noise_buzz	reson	segment	jitter	high	tri_2	decay_2	data	3.62820337	0.56015179	0.84015179	0.40015179		

REFERENCES

[1] Alexander, R.L., Gilbert, J.A., Landi, E., Simoni, M., Zurbuchen, T.H. and Roberts, D.A. 2011. Audification as a Diagnostic Tool for Exploratory Heliospheric Data Analysis. (Jun. 2011).

[2] Barrass, S. and Vickers, P. 2011. Sonification Design and Aesthetics. *The Sonification Handbook*. Logos Verlag.

[3] Bech, S. and Zacharov, N. 2007. *Perceptual Audio Evaluation - Theory, Method and Application*. John Wiley & Sons.

[4] Bennett, C.H. 1988. Logical Depth and Physical Complexity. *The Universal Turing Machine: A Half-Century Survey* (1988), 227–257.

[5] Blackburn, M. 2011. The Visual Sound-Shapes of Spectromorphology: an illustrative guide to composition. *Organised Sound*. 16, 1 (Apr. 2011), 5–13. DOI:https://doi.org/10.1017/S1355771810000385.

[6] Bonebright, T.L. and Flowers, J.H. 2011. Evaluation of Auditory Display. *The Sonification Handbook*. Logos Verlag. 111–144.

[7] Bostock, M., Ogievetsky, V. and Heer, J. 2011. D3: Data-Driven Documents. *Visualization and Computer Graphics, IEEE Transactions on*. 17, 12 (2011), 2301–2309.

[8] Bracewell, R. 1965. *The Fourier Transform and Its Applications*.

[9] Bregman, A.S. and Pinker, S. 1978. Auditory Streaming and the Building of Timbre. *Canadian Journal of Psychology/Revue canadienne de psychologie*. 32, 1 (1978), 19.

[10] Cairo, A. 2012. *The Functional Art: An Introduction to Information Graphics and Visualization*. New Riders.

[11] Camilleri, L. and Smalley, D. 1998. The Analysis of Electroacoustic Music: Introduction. *Journal of New Music Research*. 27, 1–2 (Jun. 1998), 3–12. DOI:https://doi.org/10.1080/09298219808570737.

[12] Chen, C. and Yu, Y. 2000. Empirical Studies of Information Visualization: A Meta-Analysis. *International Journal of Human-Computer Studies*. 53, 5 (Nov. 2000), 851–866. DOI:https://doi.org/10.1006/ijhc.2000.0422.

[13] Choi, S.H. and Walker, B.N. 2010. Digitizer Auditory Graph: Making Graphs Accessible to the Visually Impaired. *CHI '10 Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA, 2010), 3445–3450.

[14] Choisel, S. and Wickelmaier, F. 2006. Extraction of Auditory Features and Elicitation of Attributes for the Assessment of Multichannel Reproduced Sound. *Journal of the Audio Engineering Society*. 54, 9 (2006), 815–826.

[15] Collins, N. Live Coding and Machine Listening.

[16] Collins, N., McLEAN, A., Rohrhuber, J. and Ward, A. 2003. Live Coding in Laptop Performance. *Organised Sound*. 8, 3 (Dec. 2003), 321–330. DOI:https://doi.org/10.1017/S135577180300030X.

[17] Composing Timbre Spaces, Composing Timbre in Space: An Exploration of the Possibilities of Multidimensional Timbre Representations and Their Compositional Applications - ProQuest: 2013. *https://search.proquest.com/openview/539797b38254c42e209e602437a074ee/1?pq-origsite=gscholar&cbl=51922&diss=y*. Accessed: 2020-11-20.

[18] Degara, N., Nagel, F. and Hermann, T. 2013. Sonex: An Evaluation Exchange Framework For Reproducible Sonification. (Jul. 2013).

[19] Deutsch, D. 1999. Grouping Mechanisms in Music. *The Psychology of Music (Second Edition)*. D. Deutsch, ed. Academic Press. 299–348.

[20] dhas, M.D.K. and Priyatharshini, P. 2018. Analysis and Synthesis of Audio Signal using Integer MDCT with KBD window. *2018 4th International Conference on Electrical Energy Systems (ICEES)* (Feb. 2018), 522–525.

[21] Donnadieu, S. 2007. Mental Representation of the Timbre of Complex Sounds. *Analysis, Synthesis, and Perception of Musical Sounds: The Sound of Music*. J.W. Beauchamp, ed. Springer. 272–319.

[22] Dubus, G. and Bresin, R. 2013. A Systematic Review of Mapping Strategies for the Sonification of Physical Quantities. *PLOS ONE*. 8, 12 (Dec. 2013), e82491. DOI:https://doi.org/10.1371/journal.pone.0082491.

[23] Elliott, T.M., Hamilton, L.S. and Theunissen, F.E. 2013. Acoustic Structure of the Five Perceptual Dimensions of Timbre in Orchestral Instrument Tones. *The Journal of the Acoustical Society of America*. 133, 1 (Jan. 2013), 389–404. DOI:https://doi.org/10.1121/1.4770244.

[24] Endrich, A. 1997. Composers' Desktop Project: a musical imperative. *Organised Sound*. 2, 1 (Apr. 1997), 29–33.

[25] Erickson, T., Wilkerson, M., Finzer, W. and Reichsman, F. 2018. Data Moves. *Technological Issues in Statistics Education (Submitted)*. (2018).

[26] Essl, G. 2010. *UrMus-an Environment for Mobile Instrument Design and Performance*. Ann Arbor, MI: MPublishing, University of Michigan Library.

[27] Fiebrink, R., Trueman, D. and Cook, P.R. 2009. A Metainstrument for Interactive, On-the-Fly Machine Learning. *In Proc. NIME* (2009).

[28] Finzer, W. 2014. Common Online Data Analysis Platform. *Computer Software](CODAP). Concord, MA: The Concord Consortium. https://codap. concord. org/releases/latest/static/dg/en/cert/index. html*. (2014).

[29] Flowers, J.H. 2005. Thirteen Years of Reflection on Auditory Graphing: Promises, Pitfalls, and Potential New Directions. *Faculty Publications, Department of Psychology*. (2005), 430.

[30] Furnham, A. and Boo, H.C. 2011. A Literature Review of the Anchoring Effect. *The Journal of Socio-Economics*. 40, 1 (Feb. 2011), 35–42. DOI:https://doi.org/10.1016/j.socec.2010.10.008.

[31] Grey, J.M. 1977. Multidimensional Perceptual Scaling of Musical Timbres. *The Journal of the Acoustical Society of America*. 61, 5 (May 1977), 1270–1277. DOI:https://doi.org/10.1121/1.381428.

[32] Grond, F. and Berger, J. 2011. Parameter Mapping Sonification. *The Sonification Handbook*. (2011), 363–397.

[33] Grond, F. and Hermann, T. 2012. Aesthetic Strategies in Sonification. *AI & SOCIETY*. 27, 2 (May 2012), 213–222. DOI:https://doi.org/10.1007/s00146-011-0341-7.

[34] Grosshauser, T., Bläsing, B., Spieth, C. and Hermann, T. 2012. Wearable Sensor-Based Real-Time Sonification of Motion and Foot Pressure in Dance Teaching and Training. *Journal of the Audio Engineering Society*. 60, 7/8 (Aug. 2012), 580–589.

[35] Hajda, J.M. 2007. The Effect of Dynamic Acoustical Features on Musical Timbre. *Analysis, Synthesis, and Perception of Musical Sounds*. Springer, New York, NY. 250–271.

[36] Hankinson, J.C.K. and Edwards, A.D.N. 1999. Designing Earcons with Musical Grammars. *SIGCAPH Comput. Phys. Handicap.* 65 (Sep. 1999), 16–20. DOI:https://doi.org/10.1145/569306.569307.

[37] Harvey, J. 2000. Spectralism. *Contemporary music review*. 19, 3 (2000), 11–14.

[38] Hermann, T. 2011. Model-Based Sonification. *The Sonification Handbook*. 399–427.

[39] Hermann, T. 2008. Taxonomy and Definitions for Sonification and Auditory Display. *Proceedings of the 14th International Conference on Auditory Display (ICAD 2008)* (2008).

[40] Hermann, T., Krause, J. and Ritter, H. 2002. Real-Time Control of Sonification Models with a Haptic Interface. *Proceedings of the international conference on auditory display (ICAD)* (2002), 82–86.

[41] Hermann, T. and Ritter, H. 1999. Listen to your Data: Model-Based Sonification for Data Analysis. *189–194, Int. Inst. for Advanced Studies in System research and cybernetics* (1999), 189–194.

[42] Hirst, D. 2011. From Sound Shapes to Space-Form: investigating the relationships between Smalley's writings and works. *Organised Sound*. 16, 1 (Apr. 2011), 42–53. DOI:https://doi.org/10.1017/S1355771810000427.

[43] Hoos, H.H., Hamel, K.A., Renz, K. and Kilian, J. 1998. *The GUIDO Notation Format – A Novel Approach for Adequately Representing Score-Level Music*.

[44] Hugill, A. 2012. *The Digital Musician*. Routledge.

[45] Hunt, A. and Wanderley, M.M. 2002. Mapping Performer Parameters to Synthesis Engines. *Organised Sound*. 7, 02 (Aug. 2002), 97–108. DOI:https://doi.org/10.1017/S1355771802002030.

[46] Huron, D.B. 2006. *Sweet Anticipation: Music and the Psychology of Expectation*. MIT press.

[47] Imago.: 2002. *http://www.trevorwishart.co.uk/publ_rec.html*. Accessed: 2019-02-18.

[48] Jayaram, P., Ranganatha, H.R. and Anupama, H.S. 2011. Information Hiding Using Audio Steganography–a Survey. *The International Journal of Multimedia & Its Applications (IJMA) Vol*. 3, (2011), 86–96.

[49] Jehan, T. 2005. *Creating Music by Listening*. Massachusetts Institute of Technology.

[50] Jehan, T. and Schoner, B. 2001. *An Audio-Driven Perceptually Meaningful Timbre Synthesizer*.

[51] Jeon, M., Hosseini, S.M.F., Landry, S. and Sterkenburg, J. 2016. Tutorial on In-vehicle Auditory Interactions: Design and Application of Auditory Displays, Speech, Sonification, &amp; Music. *Adjunct Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (New York, NY, USA, Oct. 2016), 225–228.

[52] Kalman, R.E. 1960. A New Approach to Linear Filtering and Prediction Problems. *Journal of basic Engineering*. 82, 1 (1960), 35–45.

[53] Knaflic, C.N. 2015. *Storytelling with Data: A Data Visualization Guide for Business Professionals*. John Wiley & Sons.

[54] Kramer, J.D. 1973. Multiple and Non-Linear Time in Beethoven's Opus 135. *Perspectives of New Music*. 11, 2 (1973), 122–145. DOI:https://doi.org/10.2307/832316.

[55] Kramer, J.D. 1988. *The Time of Music: New Meanings, New Temporalities, New Listening Strategies*.

[56] Kupper, L.L. 1967. Ties in Paired-Comparison Experiments: A Generalization of the Bradley-Terry Model AU - Rao, P. V. *Journal of the American Statistical Association*. 62, 317 (Mar. 1967), 194–204. DOI:https://doi.org/10.1080/01621459.1967.10482901.

[57] Landy, L. 2007. *Understanding the Art of Sound Organization*. Mit Press.

[58] Lerch, A. 2012. *Audio Content Analysis: An Introduction*. Wiley.

[59] Lima, M. 2009. Information Visualization Manifesto. *Visual Complexity*. (2009).

[60] Malt, M. and Jourdan, E. Zsa.Descriptors: a library for real-time descriptors analysis. *ResearchGate*.

[61] Mathews 1969. *The Technology of Computer Music*.

[62] Mcadams, S. 1999. Perspectives on the Contribution of Timbre to Musical Structure. *Comput. Music J.* 23, 3 (Sep. 1999), 85–102. DOI:https://doi.org/10.1162/014892699559797.

[63] McAdams, S. and Bregman, A. 1979. Hearing Musical Streams. *Computer Music Journal*. 3, 4 (1979), 26–60.

[64] McCartney, J. 2002. Rethinking the Computer Music Language: SuperCollider. *Computer Music Journal*. 26, 4 (Dec. 2002), 61–68. DOI:https://doi.org/10.1162/014892602320991383.

[65] Meyer, L.B. 1957. Meaning in Music and Information Theory. *The Journal of Aesthetics and Art Criticism*. 15, 4 (1957), 412–424. DOI:https://doi.org/10.2307/427154.

[66] Munzner, T. 2014. *Visualization Analysis and Design*.

[67] Murail, T. 2005. Spectra and Sprites. *Contemporary Music Review*. 24, 2–3 (Apr. 2005), 137–147. DOI:https://doi.org/10.1080/07494460500154806.

[68] Murail, T. 2005. The Revolution of Complex Sounds. *Contemporary Music Review*. 24, 2–3 (Apr. 2005), 121–135. DOI:https://doi.org/10.1080/07494460500154780.

[69] Neuhoff, J.G. 2011. Perception, Cognition and Action in Auditory Displays. *The Sonification Handbook*.

[70] Pauletto, S. and Hunt, A. 2006. The Sonification of EMG Data. (Jun. 2006).

[71] Peeters, G. 2004. A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project. (Jan. 2004).

[72] Polli, A. 2004. Atmospherics/Weatherworks: A Multi-Channel Storm Sonification Project. *ICAD* (2004).

[73] Risset, J.-C. and Wessel, D.L. 1982. Exploration of Timbre by Analysis and Synthesis. *Psychology of Music*.

[74] Roads, C. 2015. *Composing Electronic Music: A New Aesthetic*. Oxford University Press.

[75] Roads, C. 2001. *Microsound*. MIT Press.

[76] Roads, C. 2001. Time Scales of Music. *Microsound*. MIT Press.

[77] Rohrhuber, J., de Campo, A. and Wieser, R. 2005. Algorithms Today - Notes on Language Design for Just in Time Programming. *context*. 1, (2005), 291.

[78] Ross, J., Irani, L., Silberman, M.S., Zaldivar, A. and Tomlinson, B. 2010. Who Are the Crowdworkers? Shifting Demographics in Mechanical Turk.

CHI ’10 Extended Abstracts on Human Factors in Computing Systems (New York, NY, USA, Apr. 2010), 2863–2872.

[79] Rowe, R. 1992. *Interactive Music Systems: Machine Listening and Composing*. MIT Press.

[80] Sawe, N., Chafe, C. and Treviño, J. 2020. Using Data Sonification to Overcome Science Literacy, Numeracy, and Visualization Barriers in Science Communication. *Frontiers in Communication*. 5, (2020), 46.

[81] Sayood, K. 2012. *Introduction to Data Compression*. Newnes.

[82] Schaeffer, P. 2004. Acousmatics. *Audio Culture: Readings in Modern Music*. (2004), 76–81.

[83] Schafer, R.M. 2004. The Music of the Environment. *Audio Culture: Readings in Modern Music*. A&C Black.

[84] Schuett, J.H. 2019. Measuring the Effects of Display Design and Individual Differences on the Utilization of Multi-Stream Sonifications. (Jul. 2019).

[85] Schwarz, D. 2007. Corpus-Based Concatenative Synthesis. *IEEE Signal Processing Magazine*. 24, 2 (Mar. 2007), 92–104. DOI:https://doi.org/10.1109/MSP.2007.323274.

[86] Schwarz, D., Beller, G., Verbrugghe, B. and Britton, S. 2006. Real-Time Corpus-Based Concatenative Synthesis with CataRT. *9th International*

Conference on Digital Audio Effects (DAFx) (Montreal, Canada, Sep. 2006), 279–282.

[87] Serafin, S., Franinović, K., Hermann, T., Lemaitre, G., Rinott, M. and Rocchesso, D. 2011. Sonic Interaction Design. *The Sonification Handbook*.

[88] Serra, X. and Smith, J. 1990. Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic Plus Stochastic Decomposition. *Computer Music Journal*. 14, 4 (1990), 12–24.

[89] Shannon, C.E. 1951. Prediction and Entropy of Printed English. *Bell system technical journal*. 30, 1 (1951), 50–64.

[90] Siedenburg, K., Fujinaga, I. and McAdams, S. 2016. A Comparison of Approaches to Timbre Descriptors in Music Information Retrieval and Music Psychology. *Journal of New Music Research*. 45, 1 (Jan. 2016), 27–41. DOI:https://doi.org/10.1080/09298215.2015.1132737.

[91] Smalley, D. 1986. Spectro-morphology and Structuring Processes. *The Language of Electroacoustic Music*. S. Emmerson, ed. Palgrave Macmillan UK. 61–93.

[92] Smalley, D. 1997. Spectromorphology: Explaining Sound-Shapes. *Organised Sound*. 2, 02 (Aug. 1997), 107–126. DOI:https://doi.org/10.1017/S1355771897009059.

[93] Smalley, D. 1996. The Listening Imagination: Listening in the Electroacoustic Era. *Contemporary Music Review*. 13, 2 (Jan. 1996), 77–107. DOI:https://doi.org/10.1080/07494469600640071.

[94] Soong, F. and Juang, B. 1984. Line Spectrum Pair (LSP) and Speech Data Compression. *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '84.* (Mar. 1984), 37–40.

[95] Spiegel, L. 1986. Music Mouse^{TM}-An Intelligent Instrument. *Internet: http://retiary. org/ls/programs. html*. (1986).

[96] Stasko, J. 2014. Value-Driven Evaluation of Visualizations. *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization* (2014), 46–53.

[97] Terasawa, H., Slaney, M. and Berger, J. 2005. Perceptual Distance in Timbre Space. (Jul. 2005).

[98] The Goldsmiths Musical Sophistication Index (Gold-MSI): *https://www.gold.ac.uk/music-mind-brain/gold-msi/*. Accessed: 2019-10-18.

[99] Thoresen, L. and Hedman, A. 2007. Spectromorphological Analysis of Sound Objects: An Adaptation of Pierre Schaeffer's Typomorphology. *Organised Sound; Cambridge*. 12, 2 (Aug. 2007), 129–141. DOI:http://dx.doi.org/10.1017/S1355771807001793.

[100] Truax, B. 1990. Composing with Real-Time Granular Sound. *Perspectives of New Music*. (1990), 120–134.

[101] Tsuchiya, T. Data-To-Music API: Real-Time Data-Agnostic Sonification With Musical Structure Models. *Student Think Tank at the 21st International Conference on Auditory Display* 13.

[102] Tsuchiya, T. and Freeman, J. 2018. A Study of Exploratory Analysis in Melodic Sonification with Structural and Durational Time Scales. (2018).

[103] Tsuchiya, T., Freeman, J. and Lerner, L.W. 2016. Data-Driven Live Coding with DataToMusic API. *Proceedings of the 2nd Web Audio Conference (WAC-2016), Atlanta* (2016).

[104] Tufte, E.R., Goeler, N.H. and Benson, R. 1990. *Envisioning Information*. Graphics press Cheshire, CT.

[105] Vaidyanathan, P.P. 2007. The Theory of Linear Prediction. *Synthesis Lectures on Signal Processing*. 2, 1 (Jan. 2007), 1–184. DOI:https://doi.org/10.2200/S00086ED1V01Y200712SPR003.

[106] Verma, T.S. and Meng, T.H.Y. 2000. Extending Spectral Modeling Synthesis with Transient Modeling Synthesis. *Computer Music Journal*. 24, 2 (Jun. 2000), 47–59. DOI:https://doi.org/10.1162/014892600559317.

[107] Vickers, P. and Hogg, B. 2006. Sonification Abstraite/Sonification Concrete: An “Aesthetic Persepctive Space” for Classifying Auditory Displays in the Ars

Musica Domain. *Proceedings of the 12th International Conference on Auditory Display (ICAD 2006)* (Jun. 2006).

[108] Vogt, K. 2011. A Quantitative Evaluation Approach to Sonifications. (Jun. 2011).

[109] Wall, E., Agnihotri, M., Matzen, L., Divis, K., Haass, M., Endert, A. and Stasko, J. 2019. A Heuristic Approach to Value-Driven Evaluation of Visualizations. *IEEE transactions on visualization and computer graphics*. 25, 1 (2019), 491–500.

[110] Web Audio API: *https://webaudio.github.io/web-audio-api/*. Accessed: 2016-04-08.

[111] Wessel, D. and Wright, M. 2002. Problems and Prospects for Intimate Musical Control of Computers. *Comput. Music J.* 26, 3 (Sep. 2002), 11–22. DOI:https://doi.org/10.1162/014892602320582945.

[112] Wessel, D.L. 1979. Timbre Space as a Musical Control Structure. *Computer Music Journal*. 3, 2 (1979), 45–52. DOI:https://doi.org/10.2307/3680283.

[113] Williams, D. 2016. Utility Versus Creativity in Biomedical Musification. *Journal of Creative Music Systems*. 1, 1 (Sep. 2016). DOI:https://doi.org/10.5920/jcms.2016.02.

[114] Wishart, T. 1988. The Composition of "Vox-5." *Computer Music Journal*. 12, 4 (1988), 21–27. DOI:https://doi.org/10.2307/3680150.

[115] Yi, S., Lazzarini, V. and Costello, E. 2018. WebAssembly AudioWorklet Csound. *4th Web Audio Conference, TU Berlin* (2018).

[116] Zacharov, N. 2018. *Sensory Evaluation of Sound*. CRC Press.

[117] Zattra, L. 2005. Analysis and Analyses of Electroacoustic Music. *Sound and Music Computing (SMC05), Salerno, Italy*. 36, (2005).