Fakultät I
Institut für Sprache und Kommunikation
Fachgebiet Audiokommunikation

A Framework for Audio-Tactile Signal Translation

Maximilian Weber

Primary Supervisor:	Prof. Dr. Stefan Weinzierl
Secondary Supervisor:	Dr. Charalampos Saitis Queen Mary University of London
[redacted]	[redacted]

A thesis submitted to Technische Universität Berlin
in partial fulfilment of the requirements for the degree
Master of Science in Audiocommunication and -technology.

Technische Universität Berlin
Fakultät I - Geistes- und Bildungswissenschaften
Institut für Sprache und Kommunikation
15th July, 2020

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt gegenüber der Fakultät I der Technischen Universität Berlin, dass die vorliegende, dieser Erklärung angefügte Arbeit selbstständig und nur unter Zuhilfenahme der im Literaturverzeichnis genannten Quellen und Hilfsmittel angefertigt wurde. Alle Stellen der Arbeit, die anderen Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind kenntlich gemacht. Ich reiche die Arbeit erstmals als Prüfungsleistung ein. Ich versichere, dass diese Arbeit oder wesentliche Teile dieser Arbeit nicht bereits dem Leistungserwerb in einer anderen Lehrveranstaltung zugrunde lagen.

Titel der schriftlichen Arbeit

A Framework for Audio-Tactile Signal Translation

Verfasser

Weber, Maximilian, [redacted]

Betreuende Dozenten

Prof. Dr. Stefan Weinzierl,
Dr. Charalampos Saitis (Queen Mary University of London)

Mit meiner Unterschrift bestätige ich, dass ich über fachübliche Zitierregeln unterrichtet worden bin und verstanden habe. Die im betroffenen Fachgebiet üblichen Zitiervorschriften sind eingehalten worden. Eine Überprüfung der Arbeit auf Plagiate mithilfe elektronischer Hilfsmittel darf vorgenommen werden.

Ort, Datum	Maximilian Weber

Abstract

To enable a transition away from primitive, buzzing vibrations towards an new generation of wideband vibrotactile display systems requires strategies, standards and tools for designing, storing and transmitting tactile stimuli signals and patterns. Complicating the matter is a variety of different tactile display technologies and resulting display system variances due to missing industry standards. A lack of standardization, and loose use of fancy marketing terms poses a similar issue, as the origins of the “HiFi” standard for loudspeakers in the 1960s.

Due to the early sensory integration of both the auditory and vibrotactile modalities, and the resulting perceptual similarities, it appears to be feasible to translate auditory to vibrotactile stimulus signals and use ubiquitous audio material as a starting point for the design of tactile stimuli. This insight might become useful, as auditory perception is well researched, and methods for handling, editing and displaying audio material are widespread. Therefore, these methods might prove to be useful for handling vibrotactile signals aswell — effectively enabling the potential for a transfer of domain knowledge from the audio to the tactile domain.

On this basis, this work aimed to validate a novel audio-tactile signal translation method by measuring the *coherence* of both unimodal and bimodal vibrotactile stimuli towards their (non-musical) auditory sources, while expanding on signal processing methods discussed in previous works. The parametric format, used to describe the tactile stimulus signal, is designed to interface with existing tactile APIs, and is able adapt to various tactile actuator technologies during signal synthesis.

Zusammenfassung

Trotz technologischer Fortschritte werden taktile Vibrationen heute vornehmlich von einfachen Aktuatoren mit begrenzten Kapazitäten erzeugt. Um einen Wandel hin zu modernen Aktuatoren zu ermöglichen, und um eine starke Fragmentierung des Marktes zu vermeiden, bedarf es langfristig einer Strategie zur Standardisierung der Prozesse für die Gestaltung, Speicherung und Übertragung von vibrotaktilen Signalen. Die Vielfalt vibrotaktiler Technologien droht diesen Sachverhalt dabei zu einem komplizierten Unterfangen zu machen: Sowohl die Forschung, als auch die Industrie, benötigen daher eine Standardisierung der Verfahren zur Messung und Angabe relevanter Systemkennwerte vibrotaktiler Systeme. Diese Situation erinnert an den Ursprung der Standardisierungsvorgaben für “HiFi”-Lautsprecher zum Ende der 60er-Jahre.

Aufgrund der frühen Integration von vibrotaktilen und akustischen Stimuli und deren Ähnlichkeiten in der Wahrnehmung scheint es plausibel akustisches Klangmaterial in vibrotaktile Stimuli zu übersetzen. Eine valide Übersetzungsmethode ist unter anderem wünschenswert, um die Fülle an bereits existierendem Audiomaterial für den vibrotaktilen Sinn verwenden zu können. Die akustische Wahrnehmung ist fundiert erforscht und erlaubt daher die Frage, welche Erkentnisse aus dem Repertoire der Akustik und der akustischen Datenverarbeitung auf den vibrotaktilen Sinn übertragbar sind?

Diese Arbeit diskutiert bereits bekannte Übersetzungs- und Signalzerlegungsmethoden und formuliert darauf aufbauend eine neuartige, audio-taktile Übersetzungsmethode. Das damit einhergehende, parametrische Datenformat für vibrotaktile Signale ist so gestaltet, dass es sich den variablen Systemkennwerten diverser Technologie anpassen, und an bereits existierende Programmierschnittstellen ankoppeln kann. Zur Validierung dieser Methode wurde anhand eines empirischen Versuchs das Maß an “Kohärenz” zwischen vibrotaktiler Stimuli und deren akustischen Ursprungssignal gemessen.

Acknowlegdements

I would like to thank the Audio Communication Group, and especially Dr. Charalampos Saitis for providing me with guidance, knowledge and experience throughout this work. Thank you, for being available even though we were in different time zones and about 1000 km apart.

Further, I would like to thank the entire team at Lofelt for being patient, supportive and inspiring throughout the time I was working on this thesis. I am thankful for the convenient access to their actuator technology, professional equipment and guidance from experts of various domains.

This work would not have been possible without the loving support of my family and friends. I am blessed to have such a supportive network of people that show kindness and understanding for each other. A special thanks goes to William, for taking the time to proofread the entire thesis and discussing digital signal processing (DSP) topics early on.

Thank you Sarah, for allowing me to shut off my mind and being patient with me — even during the prolonged period of COVID-19 related home office.

Table of Contents

List of Figures

List of Tables

Acronyms

A2VT Audio-Vibrotactile Signal Translation 41, 42, 50

API application programming interface 3, 4, 75

BLT bilinear transform 31

codec coder-decoder algorithm pair 7, 26, 35, 37, 41, 48, 49, 55, 75

DFT discrete Fourier transform 88

DOF degree of freedom 33

DSP digital signal processing IX, 6, 25, 27, 33, 35, 41, 42, 51, 61, 67

DUT device under test 24

EAP electroactive polymer actuator 16, 17

ERM eccentric rotating mass 3, 15–17, 20, 21

ESS exponentially swept sine 22

FIR finite impulse response 30

GUI graphical user interface 16

HILN Harmonic and Individual Lines and Noise 37

HMI human-machine interface 3

HPSS harmonic percussive source separation 38–40

HRTF Head-Related Transfer Function 69

IID Interaural Intensity Difference 69

IIR infinite impulse response 31

IR impulse response 22, 23, 28, 29

ISTFT inverse short-time Fourier transform 40

ITD Interaural Time Difference 69

JND just-noticeable difference 10, 12, 41, 50, 71, 74

JSON JavaScript Object Notation 41

LRA linear resonant actuator 16, 17, 20, 21, 46

LS least squares 30, 31

NLTK Natural Language Toolkit 60, 61

STFT short-time Fourier transform 38, 47

ToA time of arrival 69, 74

VCA voice coil actuator 15–17, 25, 27, 33

Contribution of authors

This thesis, and the research to which it refers, is the candidate's own original work except for commonly understood and accepted ideas or where explicit reference to the work of other people, published or otherwise, is made. The thesis is formatted as a monograph comprising six chapters and includes contents from the following (forthcoming) conference publications:

- Chapter 3: Weber, M., Saitis, C. (**2020**). "Analysing and countering bodily interference in vibrotactile devices introduced by human interaction and physiology", in *Proc. EuroHaptics Conf.* (Leiden, Netherlands).
- Chapter 1, 3 and 4: Weber, M., Saitis, C. (**2020**). "Towards a framework for ubiquitous audio-tactile design", in *Proc. International Workshop on Haptic and Audio Interaction Design (HAID)* (Montreal, Canada).

The candidate was responsible for every step involved in designing and carrying out all experiments mentioned in this thesis, as well as analyzing collected data and preparing manuscripts for all the publications listed above. Charalampos Saitis, this thesis secondary supervisor, provided guidance on the experimental setups, data analysis, the interpretation of the results and structure of both the manuscripts listed above and this thesis. The company Lofelt GmbH in Berlin provided tactile actuators, measurement equipment and office space to conduct the measurements presented in chapter 3. Fabian Brinkman, a research assistant at the audio communication department of TU Berlin, helped to arrange the time and space required in the media lab for the perceptual evaluation presented in chapter 5.

1 Introduction

This chapter gives an introduction to the field of audio-tactile research and motivates it's relevance towards industry applications in Section 1.1. Next, the overall structure of the thesis is outlined in Section 1.2 and related work is discussed in Section 1.3. Finally the methodology for this research effort is outlined in Section 1.4.

1.1 Motivation

"Sound is touch at a distance." This quote by Stanford's Anne Fernald comes from a podcast about her research on the cognitive response of infants towards their mothers voice[1]. In this instance, the word "touch" is attributed metaphorically towards an emotional response induced by sound [19]. Being emotionally "touched by sound", especially by music, is a common experience many people can relate to. Explaining this emotional or cognitive response to sound is a core research topic in the field of music psychology [48, 36].

For the scope of this thesis "being touched by sound" is meant in a more literal sense: Sound, as a propagating vibration, can not only be sensed by the human ear, but also by mechanoreceptors in the skin, and thus evoke the sensation of touch. The sensation of touch can be induced by various mechanical forces, such as pressure fluctuations, shearing forces and vibrations applied to the skin [33]. Therefore, some vibrations can be perceived by both the auditory and the vibrotactile sense, as the sensitivity to vibrations overlap in a frequency range from 30 to 1000 Hz [102, 42]. To better understand what that means lets consider an example: Given, that the sound of an acoustic event transports enough vibrational energy, either air- or structure-borne, to allow for a mechanical deformation of the skin, sound can not only be heard (auditory system) but also felt through the skin's mechanoreceptors (somatosensory system). This experience is common at concerts, when perceiving a car engine's vibrations or while manipulating an object with our hands. The effects of integrated auditory-tactile sensations have been researched in recent years and cross-modal effects, for example, on loudness perception and the perceived quality of musical reproduction have been discovered [57, 58].

[1] WNYC RadioLab Podcast "Sound As Touch", September 24th 2007
https://www.wnycstudios.org/story/91514-sound-as-touch

The role of joint (bimodal) auditory-tactile perception has mostly been explored in a musical context [56, 57, 73, 72, 85]. This work, however, takes a more generalized approach by enabling research on similarities and differences between both modalities by using both unimodal and bimodal audio-tactile stimuli presentation. The vibrotactile stimuli signals were derived from arbitrary, *non-musical* audio sources. In a perceptual evaluation, both auditory and vibrotactile stimuli are therefore presented consecutively (unimodal) and simultaneous (bimodal) to investigate the perceived coherence between both modalities, and to validate the audio-tactile translation method proposed in this work.

The goal of this work is therefore to revisit, and expand on existing signal processing methods for inter-modal stimuli signal translation, and to validate a novel method that aims to coherently convert audio signals to vibrotactile stimuli signals. It further proposes a parametric signal-decomposition and re-synthesis method to enable a flexible authoring and editing scheme, that is further suspected to work well across various vibrotactile display technologies. The main research questions can thus be formulated as:

- If and how can audio signals be used as a source to generate vibrotactile stimuli?
- Which perceptual and system dependant aspects could have an influence on the design of a cross-modal signal translation framework?
- How can we validate a proposed signal translation framework?

Addressing these questions will help better understand important issues at the intersection of the auditory and tactile modalities:

- Do we use the same language (semantics) to describe a cross-modal translated stimuli?
- What sort of domain knowledge can we translate to the tactile from the comparably better researched audio domain?
- How can these findings inform design choices of audio and vibrotactile related applications in both unimodal (auditory or tactile) and bimodal (auditory and tactile) scenarios?

Industry Applications

Beyond research, applications in medicine, entertainment, and mobile communications can be envisioned: A better understanding of integrated audio-tactile experiences is beneficial for authoring and curating tactile content for future applications using wideband vibrotactile feedback. Domain knowledge around integration workflows, transmission, processing and reproduction of audio assets could potentially inform similar processes for tactile content curation.

Progress in audio-tactile translation can be utilized to aid people with hearing impairment in various applications by translating acoustic cues to touch and thus help them navigate through daily life [45, 77], transport immediate warnings [105] and benefit other means of affective computing [17]. Many applications in the realm of human-machine interface (HMI), such as teleoperation for industrial, medical or end user purposes benefit from vibrotactile feedback in addition to (kinaesthetic) force-feedback to convey more immediate and meaningful feedback to a remote operator. This has the benefit for a HMI to off-load vital information to the more proximate and adequate tactile sense instead of occupying other sensory channels. Such a delegation of information to the tactile sense essentially frees up perceptual capacity from the highly loaded visual and auditory sense [2]. Off-loading information to other sensory modalities is highly desirable in today's information driven world: In his book “The user illusion: Cutting consciousness down to size” Tor Nørretranders estimates the information bandwidth of touch to be ten times higher than for hearing, which gives a hint towards the untapped potential of this modality [69].

Further, use cases can be found in the entertainment sector, such as music, movies, telepresence and video games: Various mobile and desktop applications utilizing virtual and augmented reality technologies can benefit from additional modalities by providing a heightened sense of immersion to the user. For example, drawing or writing with a brush or pen on a touchscreen could display different tactile sensations in accordance with the current virtual tool in use. This could be achieved by varying the texture (vibrations) and resistance for different forms of virtual interactions. In the past, a majority of hardware systems integrating tactile displays used dull, low-bandwidth actuators, such as a eccentric rotating mass (ERM) to provide haptic feedback. Due to the need in all industry segments to evolve away from these legacy solutions and to provide truly wideband high fidelity tactile feedback, methods for curating, storing and reproducing vibrotactile signals will become mandatory.

A complexity arising from the industry is the need for a platform- and technology agnostic framework to integrate and display curated stimuli in a reliable and coherent way, despite differences introduced by various display technologies. A platform-agnostic solution would avoid, that tactile assets would require re-authoring and re-implementation for each specific device and allow the developers to author an experience once and deploy it to many different systems. The need for a unified format, agnostic to hardware capabilities (described in Section 3.1), while staying within the limits of the provided transmission bandwidth across platforms (such as desktop, automotive, wearable, mobile) in a fragmented ecosystem brings problems that are yet to be addressed, but are partially discussed in this thesis. Just recently, the company Apple has taken a first step in this direction by releasing a general purpose haptic application programming interface (API) called “CoreHaptics”[2] for their devices. One embodiment, discussed towards the end of this work, makes use of this API by using the audio-tactile framework proposed in this work.

Since we are apparently only consciously aware of an estimated 0.7 % of the total sensory information processed by our body [114], making informed design choices for all modalities (including touch) is an important aspect and will certainly be crucial for any application design with the user experience in mind.

1.2 Structure of the Thesis

The introductory Chapter 1 is meant to provide a motivational background for this work, present related work, outline the methodology and highlight the contributions for this field of research. Chapter 2 gives an overview on perceptual aspects of hearing and feeling (i.e., auditory and tactile perception) by disambiguating terminology used throughout this work, introducing perceptual aspects of both senses and discussing the multimodal integration of natural vibration events.

The hardware and software used for this work are discussed in Chapters 3 and 4. Variations of tactile display technologies and the interference on vibrotactile displays induced by the human physiology are discussed as factors that need to be considered for the design of the audio-tactile translation framework. Furthermore, the setup to measure the technology used in this work is presented together with the resulting data. A (modified) wristband, chosen to be used for the perceptual evaluation of the framework, is presented — together with a strategy to linearize the output. Lastly, the requirements for the framework are collected and design considerations for the implementations of the decomposition and resynthesis algorithms are discussed.

The perceptual evaluation of the framework is presented in Chapter 5 by showcasing the chosen stimuli signals, the procedure of the experiment and the results of the perceptual user test. Furthermore, the entire work is summarized and final conclusions are drawn in Chapter 6. Finally, the literature cited in this work is listed and accompanying data and visualizations are made available in the Appendix A.

1.3 Related Work

Given the ubiquity of high quality sounds from recordings and digital synthesis, as well as the already present implementation of such material in various applications, it is desirable to create vibrotactile stimuli from these readily available and information rich sources. Using an existing audio asset to derive a vibrotactile stimuli is especially

[2] Apple’s “CoreHaptics API”, last sighted 7th July 2020
https://developer.apple.com/documentation/corehaptics

desirable given the possibility for the resulting tactile stimuli to intrinsically *match* the temporal, dynamical and spectral progression of the audio asset to form a *coherent* percept from the integration of both modalities.

Early experiments in auditory-tactile translation were conducted in the 1920s by Gault in an experiment using a 14 feet long tube that was pressed against the palm of a subjects hand. Through multiple training sessions a subject was able to correctly identify up to 34 words together with sentences constructed by those words in various combinations, hinting at the potential of acoustic vibration cues for tactile stimulation to transmit and perceive information [27]. Due to the complexity in temporal and spectral modulation of human speech, understanding entire sentences through mere sound-induced vibrotactile stimuli is rather impressive and hints at the capabilities of the human skin in reliably differentiating vibrational patterns, despite the comparably low sensory resolution [102, 31, 24, 55].

Translating audio to tactile stimuli has been the subject of previous works researching mostly the joint (i.e., bimodal) display of audio-tactile stimuli. Within these works, various methods for audio-tactile signal translation are explored. As the perceptual frequency ranges of auditory and vibrotactile stimuli overlap, the most straight-forward method merely requires enough energy to be present in the tactile sensitivity range from 30 to 1000 Hz in a PCM encoded signal, such as the one contained in a WAVE file. This signal can then be downsampled and low-passed at approximately 1 kHz to be played back by the actuator, given that the latter is capable of recreating such a wideband and potentially complex, non-monophonic signal.

Making sure that the signal content within the tactile perceptual range matches the desired tactile experience, for example, a click, impact, or a more complex texture — is the focus and expertise of a tactile designer. Similar to a sound designer using either recordings, synthesis methods or other audio content, the designer can define a sample that suits the requirements for a tactile event. If audio assets are already present in an application, it is feasible to re-use, augment or transform the audio content to be used as a vibrotactile asset. As this step inherently requires an aesthetic choice to be made, a potential tactile coding-decoding schema (codec) or storage container should not transform or modify the intended content after the designer has concluded the work on the designed stimulus. In previous works this process has primarily been investigated for musical applications [73, 57, 6, 74, 21] or the bimodal auditory-tactile implications of walking on various surfaces [103, 70, 67].

If the audio source lacks meaningful content in the tactile perceptual range, there are various ways to augment or transform the signal. One option is to pitch-shift the signal downwards until a desired effect is achieved. This method was used in an experiment researching the influences of vibrotactile stimulation on musical (rhythm and melody) perception of subjects with cochlear implants [73]. This method works well if the content of the pitched-down signal is representative for the rest of the signal content and reflects the intended experience. Otherwise further filtering, editing and augmentation is most likely required.

To augment the low frequency range of an audio source and to achieve a higher level of parametric control over the temporal trajectory of the stimulus, a combination of an envelope follower (i.e., a smoothed curve of the temporal trajectory of the signals amplitude) and a signal generator was used in a study emulating the effect of vibrotactile cues in a seated concert setting [57]. Here, both the parameters of the envelope follower, and the pitch of the signal generator can be controlled independently until a satisfying result is achieved. Depending on the audio source material, and the fidelity of the reproduction system (i.e., tactile actuator), the signal synthesized this way can then be combined with the original audio source by adding the synthesized perceptual frequency content if needed, while not compromising on more complex timbral information from the audio source signal. The tactile signal representation using an envelope follower is inherently monophonic and therefore can't model the entire information within the tactile perceptual range sufficiently. It is feasible to track the temporal energy trajectory of a signal this way but it neglects changes in the frequency domain over time.

A series of three DSP principles, namely transposing, modulating, and filtering, have been tested in an experiment exploring the best method for tactile identification of environmental sounds (f.e. doorbell, vehicles, weather) on hearing impaired subjects [77]. Here, both the algorithms and the subjects significantly affected the results and showed large differences between individuals, regarding which algorithm worked best. The most promising approaches from this experiment have proven to be two different transposing algorithms, and an amplitude modulating algorithm when compared to equalizing, or using the unaltered source signal. Even though the task here was the identification of the signal sources for navigational purposes of the hearing impaired, this finding indicates the necessity for a signal augmentation or transformation, instead of using an unaltered audio signal to drive a tactile actuator.

For an experiment in speech recognition, a vocoder approach utilizing 16 solenoid actuators was utilized. The 16 channel filter bank of the vocoder ranged from 200 to 8000 Hz in third octave spacing. Each solenoid was driven by a 100 Hz square wave modulated by the energy of each filter channel [10]. Through this method, an abstract imprint of a speech signal's spectrum was created while an actual wideband reproduction was not utilized. This might have not been feasible due to technological limitations in actuator technology at the time — which today is still a comparably rare technology to find on the market.

By utilizing the knowledge around the four channel theory of touch [7], a more complex translation method catering towards the four individual types of mechanoreceptors has been proposed [6]. Here a set of audio analysis features [50] are mapped to a set of dynamic synthesis parameters: the spectral centroid of the audio signal was mapped to the pitch of a signal generator in a tactile perceivable range from 40 to 400 Hz. The spectral flatness of the signal was mapped to an equal power cross fade between a sine wave and a square wave. More tonality in the audio source was therefore represented with a richer harmonic spectrum in the vibrotactile domain by dynamically fading to the square wave signal. Finally the amplitude of the tactile signal is modulated by a envelope generator with an adjustable decay. This method illustrates a way to reflect timbral changes in the tactile range by utilizing information on the spectral envelope of the audio source while still allowing for a design choices by adjusting the analysis and synthesis parameters. This method doesn't capture the exact momentary spectral content by reflecting spectral change in the audio source merely by the flatness of the spectral envelope and approximating the spectral shape in the tactile domain by introducing harmonic content. This method was originally intended to add vibrotactile feedback to an electroacoustic instrument and would probably require further testing on how well it generalizes for other, non-musical source signals.

1.4 Methodology

The process of audio-tactile translation proposed in this work is primarily informed by related work, and will expand on those methods based on state of the art knowledge on tactile perception and current industry requirements. Inspiration on how to improve on previous methods was primarily drawn from established methods in the audio domain. Especially signal decomposition and resynthesis methods, as found in an audio coder-decoder algorithm pair (codec) development and musical applications were investigated.

This work aimed to validate the proposed audio-tactile translation method by using an exploratory perceptual experiment. Both a numeric rating on the coherence of auditory and tactile stimuli, and a psycholinguistic analysis of verbal descriptions have been conducted. The collective results were discussed as a measure of coherence between the audio source and respective tactile stimulus, as well as an indicator on the feasibility of using audio sources as a starting point for (audio-)tactile experience design.

2 Auditory and Tactile Perception

This chapter outlines psychophysical properties of both the auditory and the tactile sense. The functionality of the ears (hearing), and the sensory functionality of the skin (feeling) are introduced, to provide a foundation for the decision processes in the hard- and software layers later in this work. Even though both the auditory and tactile modality are sensitive to vibrations, they exhibit anatomical, functional and perceptual differences. The following sections are meant to give a brief overview of similarities and differences between both sensory modalities. Further, overlaps in the sensitivity of both modalities and the natural occurrence of audio-tactile events are discussed.

2.1 Tactile Terminology

The sense of touch is commonly referred to as the tactile modality, which together with the proprioceptive (the sense of position) and kinaesthetic (the sense of movement) modalities constitute what is known as "haptics" or haptic perception. In the industry, both the term haptic and tactile are often used synonymously, while in this work a clear definition of each term is desirable. As early as 1851 it has been argued that haptic encompasses an interactive, exploratory act, while tactile is a passive experience of touch [106]. This fortifies the definition of tactile being a subset of sensory information that remains when removing *active* participation, as well as proprioceptic and kinaesthetic attributes. It is important to note, that a tactile percept can be formed both when active movement of a subject is present, but also when the movement is evoked by an external process. Both are common experiences, for example, when probing an object's surface with our hands and perceiving the object's material texture (active), and when being touched by a different person (passive).

2.2 Auditory Anatomy and Perception

Hearing - referred to as the auditory sense - is the ability to perceive sound by detecting vibrations with the ear. The visible part of the ear, composed of the pinna and the ear canal, is referred to as the outer ear. The function of the pinna is primarily to act as a funnel which assists in directing sound further into the ear canal. At the other end of the ear canal, the eardrum (a.k.a. the tympanic membrane) separates the outer ear from the middle ear and the so called tympanic cavity. Sound as pressure fluctuations in the air is transmitted through the ear canal and excites the ear drum. Through small bone structures called ossicles, that are attached to the (excited) ear drum, sound is mechanically transmitted to the inner ear via the membrane-covered oval window. The other side of the oval window is a fluid-filled, spiral-shaped cavity called the cochlea. Sound is mechanically transmitted through the oval window and excites the fluid within the cochlea. It is important to note that this is not the only way the fluid in the cochlea can be excited: Structure-borne vibrations through bones and tissue within the body can also excite the fluid of the cochlea and form an auditory percept.

Inside the spiral-shaped cochlea energy is transmitted via traveling pressure waves of the fluid on the basilar membrane. The basilar membrane along the snail-shaped cochlea is wide at the opening and more narrow towards the center. It works similar to a frequency analyzer as the traveling waves in the fluid of the cochlea excite a frequency specific area via sensitive hair cells called Stereocilia along the organ of Corti. The displacement of the hair cells induces a change of electric conductance of the inner hair cell membranes. This change of conductance induces transmitters to be released to nerve endings and information to be transferred to the brain stem. The evoked action potentials travelling towards the central nervous system and the auditory cortex of the brain contain all of the temporally coded acoustical information. Within the auditory cortex and higher-level brain regions a percept of the acoustic event is formed.

Sound captured by the ear can be heard from about 20 Hz to 20 kHz. Next to inherited predisposition the age of a subject affects the upper hearing limit of a human subject. The most sensitive range of hearing is between 300 and 7000 Hz, while being less sensitive for lower or higher frequencies. The total number of perceivable pitch steps is estimated to be approximately 1400 [71] with the just-noticeable difference (JND) for frequencies being smaller (<1 Hz) at low frequencies and becoming increasingly larger with higher frequencies [109, 65].

2.3 Tactile Anatomy and Perception

A haptic percept is an integrated, multimodal experience that is built up by a variety of proprioceptic, kinaesthetic and other sensory aspects. Various physical properties of an object, such as the shape, orientation, hardness, warmth conductivity and surface roughness can be sensed through exploratory action [80, 33]. These properties can be picked up by an integration of specialized somatosensory receptors, enabling us to sense object attributes, such as location, stiffness, temperature and texture. The integration of these attributes allow us to to form a unified percept of an object or event. The purely tactile components thereof mainly contribute to the sensation of mechanical forces, warmth conductivity and pain reception. Being able to passively sense mechanical forces by the skin's mechanoreceptors (while leaving out pain and thermal properties) is what we call the tactile sense. It allows us to sense perceptual properties, such as contact, shearing, pressure and vibrations. A tactile percept can also be formed by actively probing an object, but since active movement inherently influences the speed, location and pattern of exploration this extended active exploration process is commonly defined as haptics, while tactile remains a subset of haptics. This disambiguation is important due to the fact that, once integrated, sensory, kinaesthetic and proprioceptic information can not be separated and researched independently.

The mechanoreceptors responsible for the sense of touch are commonly classified according to their adaptation properties and morphology. Hairless (glabrous) parts of the skin contain four different types of receptors: Merkel's receptors (SA–I); Ruffini's corpuscles (SA–II); Meissner's corpuscles (RA–I); and Pacinian corpuscles (RA–II). Slowly adapting receptors (SA-I and SA-II) evoke action potentials as long as pressure on the skin is present. The firing rate of these receptors is proportional to the intensity of the applied force. The rapidly adapting receptors (RA-I and RA-II) mainly react to movement of the skin, for example in form of sheering or vibrational forces. The numerals I and II indicate the sizes of the corresponding receptive fields. Receptors marked with the numeral I lie close to the surface of the skin and have small receptive fields. Receptors deeper in the tissue have larger receptive fields and are labeled with the numeral II. An overview of the different properties of the mechanoreceptors mentioned above is provided in Table 2.1. In addition to the adaptation characteristics and the sizes of the receptive fields, mechanoreceptors differ regarding the minimum amount of force that is necessary to evoke a sensation, the density of the receptors, and the sensitive frequency range.

Each of the listed mechanoreceptors has a specific function for the sense of touch. The slowly adapting Merkel's receptors are active when applying static pressure (indentation) to the skin. Due to their small receptive fields, they are able to detect fine contours, such as borders and edges. Ruffini's corpuscles are specialized in detecting sheering forces, such as stretching of the skin. Meissner's corpuscles, which are only present in hairless skin areas, detect the speed of skin deformation at comparably slow rates. This enables them to detect low frequency vibrations. Last but not least, the Pacinian corpuscles are specialized in detecting the speed of skin deformation. Compared to the Meissner's corpuscles they are able to detect a larger frequency range and encompass the largest receptive fields [87].

Due to the ability of the Pacinian corpuscles (RA-II) of detecting frequencies between 40 Hz and 1000 Hz (peak sensitivity between 225 and 275 Hz) they are considered to be the most important receptors for this thesis, next to the Meissner corpuscles (RA-I) with peak sensitivity between 25 and 40 Hz [6, 7]. Consequently the human tactile perception ranges from 25 to 1000 Hz. Vibrations with a frequency lower than 25 Hz are perceived as motion rather than a continuous vibration. Even though touch as a

Receptor	Type	Frequency Range (peak sensitivity)	Threshold skin deformation on hand (median)	Receptive field (median)	Receptor density at fingertip (palm)
Merkel's receptors	SA-I		$7 - 600\mu m$ ($56.5\mu m$)	$2 - 100mm^2$ ($11mm^2$)	$70/mm^2$ ($8/mm^2$)
Ruffini's corpuscles	SA-II		$40 - 1500\mu m$ ($331\mu m$)	$10 - 500mm^2$ ($59mm^2$)	$9/mm^2$ ($15/mm^2$)
Meissner's corpuscles	RA-I	5 - 200 Hz (25 - 40 Hz)	$4 - 500\mu m$ ($13.8\mu m$)	$1 - 100mm^2$ ($12.5mm^2$)	$140/mm^2$ ($25/mm^2$)
Pacinian corpuscles	RA-II	40 - 1000 Hz (225 - 257 Hz)	$3 - 20\mu m$ ($9.2\mu m$)	$10 - 1000mm^2$ ($101mm^2$)	$21/mm^2$ ($9/mm^2$)

Table 2.1: Properties of the hairless (glabrous) skin mechanoreceptors. Information derived from Treede and Russo et al. [96, 82]

vibratory sensor is in many ways inferior to hearing, there are striking resemblances to the traits obtained on auditory pitch perception [24]. Previous experiments on tactile frequency discrimination conducted on the forearm and hand reported JNDs ranging from 4 to 100 Hz in a stimulus range from 25 to 250 Hz [81, 31]. More recently the tactile ability on frequency discrimination was explored for full-body vibrations further validating that the JNDs increased with increasing frequency (e.g., approximately 7 Hz at 20 Hz and 66 Hz at 90 Hz) [55]. It is important to note that both the bodily position and contact conditions between the skin and a source of vibration have been reported to affect the results on JNDs and absolute thresholds of tactile sensitivity [37, 101]. The sensitivity may vary due to inheritance, sex and usually decreases when we get older towards higher frequencies — similar to decreased hearing sensitivity with age.

Similar to the auditory pathway, present frequency and magnitude information is coded into time-varying patterns of action potentials before being transmitted to the sensory nervous system. High intensity vibrations evoke multiple action potentials for each cycle, whereas for low intensity vibrations not every cycle period results in the release of an action potential [87].

2.4 Multimodal Integration: Natural Audio-Tactile Events

To form a coherent percept of the environment, an object or event, our brain combines information from various senses [92, 47]. For an auditory-tactile experience the integration of both modalities occurs early and close to primary sensory areas, as experiments using functional magnetic resonance imaging (fMRI) scans of primate brains have shown [47]. Such auditory-tactile sensory integration can be illustrated by an experiment named the “parchment-skin illusion”: This experiment is an easy to reproduce illusion illustrating how the auditory system can alter our perception of a haptic event, such as rubbing our hands together or tapping on a surface by modifying the corresponding auditory stimuli [46]. In other experiments, cross modal effects on the perception of roughness [34] and the perceived tactile distance have been reported [95]. These observations are evidence of a strong and early integration of both auditory and tactile sensory information. The commonality of both senses being sensitive to physical vibrations, the overlapping sensitivity ranges and the cross modal effects all support the notion of designing tactile experiences in conjunction with auditory cues, as they are also found in nature. As audio content is a widely available and sound can be considered an information rich source, it supports the notion of using audio material as a basis for tactile experience design. Furthermore, an emergent hypothesis is that there exists a supramodal representation of temporal frequency for the integrated auditory-tactile experience of exploring a surface using the haptic modality [112]. For example, a series of psychophysical experiments have shown evidence of a perceptual link between both the somatosensory and auditory frequency channels by a systematic interference on the perception of tactile frequency [113].

Joint audio-tactile percepts are a common and natural phenomenon. If we consider an acoustic event, it is possible to not only sense the acoustic waves propagating from the event with our ears, but also experience a skin deformation (i.e., a tactile stimulus) if enough energy is present. Sound waves can also propagate through structures, such as the ground or other objects we are in contact with and thus lead to a tactile stimulus. Events experienced this way are completely *passive*, meaning that no active participation or action is required by a subject to experience them. Common examples are the vibrations experienced at a concert [57] or while driving a car.

For interactive events, such as probing an object’s geometry, contours, and texture, we integrate kinaesthetic, tactile and proprioceptic information into what is considered to be an *active* haptic event. Here, forming a vibrotactile percept requires active participation by the subject while the velocity and direction of the movement have an influence on the characteristics of the resulting auditory and vibrotactile stimulus, thus integrating both the proprioceptic information on velocity and the resulting multimodal stimulus to form a percept. For both passive and (inter-)active tactile events it has been shown that an integration of both auditory and tactile information play a significant role in forming a percept by displaying various effects on the cognition or perceived quality of an event [85, 83, 84, 46, 113, 95, 30, 54].

When designing a vibrotactile stimulus for a virtual event, it is important to consider the consequences for both *passive* and *active* feedback and how strongly the virtual, or mediated interaction mirrors a *natural* event. Pressing a virtual button, for example, requires only a short (active) interaction which enables a single event call to trigger a corresponding tactile stimulus simulating the natural tactile response of a button. On the other hand, drawing in a virtual paint application, for example, enables a continuous event by simulating the interaction between pen and surface that is virtually drawn on. Such an event could not merely be triggered by a sole binary event, but would require a continuous synthesis of the desired tactile stimulus, until the contact with the surface is broken. In any case, using sound as a source for vibrotactile stimuli design seems feasible for many virtual interactions, as the resulting vibrations can conceptually not only be sensed by the mechanoreceptors, but also propagate to the ear to form an integrated percept of the interaction (and vice versa). This doesn't necessarily mean, that the information transmitted to the ear and to the skin are equal, but that they share a lot of common features, as they originate from the same source in many cases.

It thus becomes apparent that not only the design of the stimuli itself is important but also the plausibility of the interaction that triggers it (cause and effect). Enabling an environment to effectively design audio-haptic or multimodal interactions and the corresponding vibrotactile stimuli has yet to be developed. Even if tactile stimuli signals are designed similarly to audio counterparts, the fidelity of the interaction, transmission and reproduction of the stimuli needs to be warranted to make sure the intention of the tactile designer is transported sufficiently to the recipient.

The strong and integrated connection between auditory and tactile stimuli is a concept getting increased attention in the design of musical interfaces, and has been urged to be improved due to the "veil of tactile paralysis" between the musician and the sound source [52, 6, 72]. While digital instruments have been reported to be "lifeless" and "cold" compared to their analog counterparts, the addition of an artificial bodily resonance to simulate natural tactile events has indicated to be a remedy for the missing "warmth" in digital musical interfaces [79]. This could have the potential to improve the tightly interlocked feedback loop between virtuous musicians and their (digital) instrument — especially when the musician is in a loud environment and can't rely on the auditory response of the instrument alone. Recently, digital synthesizers, such as the OP-Z by Teenage Engineering, have been countering the lack of tactile feedback by giving life to their product using a so-called "Rumble module"[1]. The same principles, as described for musical instruments, can be thought of for many other (digitally mediated) interactive applications.

[1] TE's "Rumble Module": tactile feedback for the OP-Z synthesizer
`https://teenage.engineering/products/op-z/modules/rumble`

3 Hardware

Reproducing a vibrotactile stimulus requires a systematic approach and dedicated hardware. In this chapter, state of the art vibrotactile actuation technologies, their design and characteristics are discussed to inform design choices for the audio-tactile translation framework. Further, the equipment utilized to profile (i.e., measure) actuators and the design choices leading to a wristband form factor are discussed. Lastly, a method for linearizing a voice coil actuator (VCA) driven system, that seems robust against bodily induced interference, is presented.

3.1 Vibrotactile Stimulation Technologies

In the last two decades various actuator technologies have been brought to market and were used in previous experiments, while some actuation methods are still being researched. The following section gives an overview over the variety of vibrotactile technologies and discusses their shortcomings and benefits. Discussing these variations is important, as any of the technologies on the market can appear in an application or a device. One of the requirements of the proposed framework of this work was to be able to adapt to the fidelity of the technology at hand, while maintaining the audio-tactile correspondence as good as possible and to use each technology to it's full potential.

3.1.1 Common Actuator Technologies

This section gives an overview of commonly used actuator technologies, also called tactors, found in consumer electronics, such as smartphones, wearable technology and game controllers today.

Eccentric Rotating Mass Actuators

The most widely spread actuator type today is the ERM actuator. It is composed of an eccentric mass attached to the axis of a DC controlled motor. The speed of the DC motor controls both strength and frequency simultaneously, restricting the motor from creating a truly wideband frequency response. The centrifugal force of the mass can be felt as a buzzing vibrational force by the mechanoreceptors. ERM actuators are cheap and can be found in most mobile phones and peripheral hardware, such as game controllers today.

They are also found in some of the earliest mobile applications, such as pagers in the l980s.

Linear Resonant Actuators

In more recent times the linear resonant actuator (LRA) technology has become more commonplace in modern smart phones, due to their enclosed form factor, energy efficiency and easier control. Functionally, these devices share a high degree of similarity with voice coil drivers found in loudspeakers. The efficiency of the LRA is due to the high quality factor (Q-factor) of the electrodynamic frequency response, which reduces the amount of power required to run the actuator close to it's electro-mechanical resonance frequency. On the other hand, the high Q-factor restricts these actuators from being truly wideband, as they are designed to operate in a very narrow frequency range. A set of exemplary frequency responses measured from smartphones using various LRAs can be seen in Figure 3.1. Compared to ERM actuators they have shorter rise- and fall times which makes them more suitable for recreating short, impulse-like stimuli like clicks for, as they are used in graphical user interface (GUI) interactions.

Piezoelectric and Electro-Active Polymer Actuators

Both piezoelectric and electroactive polymer actuator (EAP) actuators consist of electroactive materials. When a voltage signal is applied, the material bends as one side shrinks and the other side expands, thus creating a flexing motion. A common base material for piezoelectric actuators is a set of ceramics called lead zirconate titanate (PZT), which are brittle in their raw form. Similarly, the material used for EAP applications is a type of polymer (i.e., plastic) that exhibits a change in size or shape when exposed to an electric field. An advantage of these actuators is their fast response time and the ability of the material to be set and held at a deflection position or vibrate. Unlike ERM and LRA technologies, both the amplitude and the frequency of deflection can be controlled independently. Piezo and EAP actuators are available in small form factors that enable them to be embedded into mobile applications. While both technologies can work well for close proximity vibration, such as finger-tip touch surfaces, they are mostly inadequate for creating more robust vibrations needed for other devices, such as headphones or handheld controllers. One downside to both EAP and piezoelectric actuators is that the driving signal is required to be at a relatively high voltage at around 200 V, compared to other actuator technologies.

Voice Coil Actuators

As of today, the most promising actuator technology for providing high-fidelity wide band tactile feedback is the VCA technology. Similar to the LRA, the VCA actuators share

Signal Characteristic	Actuator Type ERM	 LRA	 EAP & piezo	 VCA
Variable Amplitude	o	+	+	+
Variable Frequency	o	o	+	+
Monophonic[1]	-	o	+	+
Polyphonic[2]	-	o	o	+
Complex Waveform[3]	-	-	-	+
Rise- & Fall Time	-	o	+	o

Table 3.1: This table illustrates the capabilities of actuator technologies towards the reproduction of waveforms of increasing complexity. The ratings in the table are based on the arguments provided in the sections above and range from *bad* (-) over *neutral* (o) to *good* (+).

functional similarities to a loudspeaker driver: They contain a moving mass permanent magnet, flexible membranes and a voice coil. By applying an alternating voltage to the voice coil, the permanent magnet is accelerated which makes the mass oscillate and induce a vibration. The moving mass is held in equilibrium by the membranes, which act as mechanical springs. Tuning the moving mass, membrane stiffness and voice coil properties can allow these actuators to have a wider frequency response when compared to LRA or ERM actuators. Due to their inherent physical properties, VCAs can also display a strong fundamental resonance frequency, at which they are most energy efficient. Operating a VCA outside of it's resonant frequency is feasible, with the drawback of it being less energy efficient in these frequency ranges. Depending on the design, a VCA motor design can achieve short rise- and fall times, minimum distortion and a high signal fidelity overall.

All actuation technologies presented in this section profit from a well tuned system control loop to optimize their driving efficiency and to compensate for non-linear behaviours. Similar to a loudspeaker system, the goal of an optimal driving solution would be to allow the acceleration (or force) trajectory of an actuator to trace an input waveform as close as possible.

[1] Monophonic here means, that a single sinusoidal with varying frequency *and* amplitude can be reproduced. This is proposed to make a clear distinction from the capability of an system being able to reproduce *either* a desired frequency *or* a desired amplitude, but not both simultaneously.

[2] Polyphonic here means, that multiple sinusoidal with varying frequency *and* amplitude can be reproduced simultaneously. This rating is based on the approximate bandwidth with regards to the tactile sensitivity range for each technology, as well as the amount of expected distortion.

[2] Complex waveform is meant to describe a signal that is hard to approximate with a reasonable set of deterministic components, such as sinusoidals.

3.1.2 Advancements in Vibrotactile Technologies

Next to the actuator technologies described above, there is a range of technologies not commonly found in consumer electronics today. These technologies are either still in research, too expensive for consumer grade products or have not yet to be adopted in the industry due to technological or cost limitations. This overview is meant to give an idea on the most promising technologies that could achieve wide spread market adoption in the future.

Friction Modulation

When it comes to interactive (haptic) vibrotactile reproduction there are methods that enable a modulation of the already existing friction between the skin and a surface. This method, commonly referred to as “friction modulation”, can be achieved by electrostatic modulation of a display or by inducing a modulated ultrasonic carrier vibration into the interaction surface. Using these methods, vibrotactile actuated touchscreen prototypes have previously been built [5, 62, 111] and companies like “Hap2U”[1] are trying to bring this technology to market.

From personal experience, these devices are still in research and don’t provide the type of high fidelity response that would be required to validate audio-tactile translation methods, while keeping the structure of the stimuli coherent between both modalities. These displays also require additional interaction by the user which introduces further kinaesthetic and proprioceptic modalities to the audio-tactile reproduction which is not desirable for this research.

Ultrasonic Phased Arrays

The most experimental and seemingly science-fiction method for vibrotactile reproduction are ultrasonic displays [64]. This technology is composed of a grid array of ultrasonic transducers and a software framework, allowing for gesture tracking and phase array control — comparable to beamforming (i.e., spatial filtering) or wave field synthesis technology. Current limitations in the precision and resolution of these displays are estimated to be due to the low density of transducers in the array and the physical limit of their individual size to pack them closer together. Applications of this technology reach from rendering 3D displays using Styrofoam pebbles that are levitated by concentrated pressure zones in the wave field, simulating tactile textures in mid-air and also modulating the ultrasonic carrier with audio to enable object-based sound positioning. Next to the company “Ultrahaptics” (now: “Ultraleap”, as they fused with the company “Leap

[1] Hap2U: Ultrasonic Friction Modulation Technology, last visited 12th May 2020
http://www.hap2u.net/

Motion") [12] many publications in this field come from a joint EU-funded research project called "Levitate"[2].

3.1.3 Variability of Vibrotactile System Properties

Due to the system dependant differences between the aforementioned vibrotactile display technologies, and individual actuator model designs, defining a universal data format and transmission protocol, that allows for a coherent experience across multiple technologies poses a current industry problem. This is because no single company wants to rely on an end-to-end solution that only works for a single component supplier, but ideally wants to be flexible in their product design, while maintaining a single entry point for content curation. This is especially true for system platforms in gaming (f.e., Sony Playstation, Microsoft Xbox) or on mobile (f.e., Android, Apple iOS) as the content creators and developers would ideally want to design a tactile experience once, and not have to think about the issue of how these experiences translate to various platforms and hardware implementations.

Platform and system dependant differences in available bandwidth, frequency response, and bodily induced interference should ideally be solved by software and not pose a problem that needs to be addressed by experience- and product-designers. Defining a ubiquitous platform-agnostic format or system, that provides each technology with the necessary information to reproduce a target tactile stimuli is a key component for enabling a market shift towards the next generation in vibrotactile feedback. A recent publication proposed the term "vibrator transparency" in reference to a control system, that "absorbs" the difference in a vibrator environments' frequency characteristics, to achieve a device-agnostic stimuli reproduction. The proposed system therefore enables the design of a vibrotactile signal on a tactile display, then compensates for the characteristics of the display used during the design, then finally adapts the designed signal to match the same output on a different tactile display [97]. A similar strategy is proposed in this work, but instead of requiring the transfer function for both the senders' (designer) and receivers' (end user) tactile devices, this work makes use of the parametric signal representation properties, that are used in the proposed tactile format. In combination with display specific information, such as the available bandwidth and dynamic range, the proposed method may achieve a simpler and more intuitive method to compensate between different vibrotactile displays ad hoc (i.e., "vibrator transparency"). This claim has yet to be verified, as it was only conceptualized for a future work beyond this thesis efforts. Initial experiments have shown promising results when prototyping this functionality on the set of smartphones documented in the measurements below (see Figure 3.1).

[2] "Levitate" publications, last visited on 21st April 2020
https://www.levitateproject.org/publications

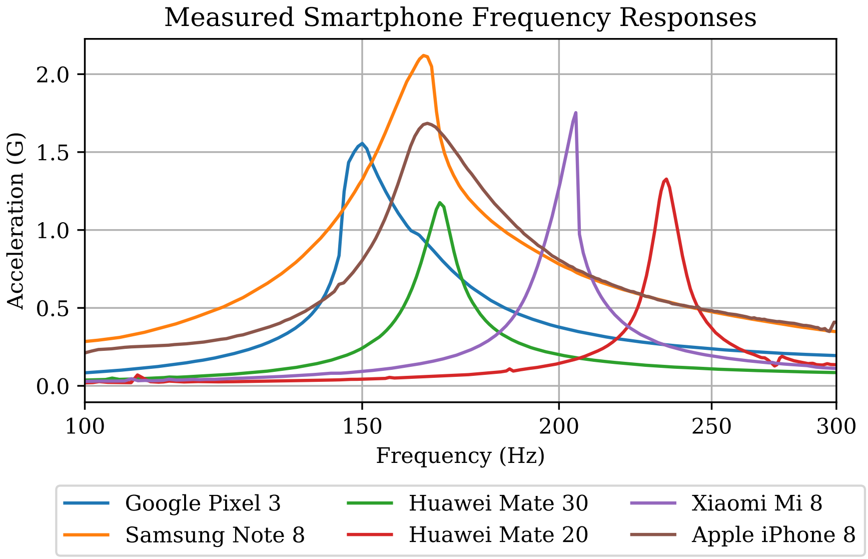


Figure 3.1: Frequency response measurements of various smartphone models conducted in a test jig using an accelerometer. All devices were measured using the total acceleration of all axis (x, y and z), to make sure any off-axis energy is contained in the measured acceleration. For reference $1G \overset{\text{def}}{=} 9.81\frac{m}{s^2}$. The absolute tactile sensitivity threshold (depending on actuator size and position) is around $0.05\frac{m}{s^2} \approx 0.5G$ [56].

As an example for the variability of vibrotactile displays, a set of smartphones that are currently on the market, were measured using the test jig and accelerometer setup described in Section 3.2. The frequency response measurements are illustrated in Figure 3.1. We can observe differences in resonant frequency, available bandwidth and system quality (Q-factor) influenced by the smartphone design, actuator positioning, the choice of the actuator technology and the actuator model. Smartphone manufacturers are beginning to realize the benefits of more wide band actuator technologies. Most modern smartphones use LRA actuators as the (legacy) ERM actuators are being replaced as the new de facto standard. These actuators are often mounted in a way to either induce a shearing force on the skin by orienting the actuator parallel to the screen, or by inducing a vibration orthogonal to the screen — both variations depend on the smartphone design, actuator model and the mounting orientation of the actuator.

Initial steps towards standardizing and evaluating haptics (including vibrotaction) have been made in parts of ISO 9241 [40, 39, 41]. While these parts of the ISO standard provide high-level guidelines for the integration of haptic modalities (i.e., tactile and kinaesthetic [100]), a concept for a full-stack solution achieving these standards is yet to be formalized and evaluated. Furthermore, the ISO standard at it's root mainly portraits an ergonomic perspective on haptics with a focus on how to design user-initiated interactive task primitives and interaction elements [40]. While care has been given to ensure perceptual, information encoding, and systematic parameters are addressed [39] there is a lack of detail on how the information inherent to a tactile stimulus is intended to be composed and how this information is accurately reproduced across various tactile display technologies. This issue is especially important during a transitional period, in which the integration of adequate high performance hardware is slowly adapting and most devices still run on legacy ERM or LRA actuators. The actuator models used for these applications mostly appear to be designed for power efficiency instead of the quality of experience.

3.1.4 A Theoretical Model for Bodily-induced Interference

When using a vibrotactile device the form factor affords a range of user interactions [68]: A vibrotactile device can be designed to be attached to the human body in form of a wearable device, while other devices afford various grasp interactions. Such interactions can dynamically change the bodily interference on the vibrotactile system [49]. This section formulates a theoretical model that aids understanding bodily interference from a systematic perspective.

The force needed to create a vibrotactile stimuli in a device can be modeled using an approximate linear spring-mass system for most actuator technologies. A spring-mass system can induce a force vector $\vec{F_0}$ by accelerating a moving mass m_0 following Newton's second law of motion ($\vec{F_0} = m_0\vec{a}$). The way the acceleration is induced depends on the actuators design: For a voice coil actuator (as used in this embodiment) the mass m_0 consists of a set of permanent magnets which in turn are set into motion by inducing a magnetic field by applying an alternating current to a voice coil surrounding the magnets.

The force of the actuator is opposed by both the application device $\vec{F}_{app}$ and the skin $\vec{F}_{skin}$. If a rigid connection between the actuator and the device is ensured the stiffness k_{app} and dampening d_{app} coefficients vanish — leaving only the device mass m_{app} to be considered as the vibrations propagates through the device to the human skin.

By applying the resulting force from the device to the skin, i.e., setting the skin tissue into motion (note the acceleration vector $\ddot{x}_{skin}$ in Figure 3.2) the mechanoreceptors are excited and allow the somatosensory cortex and higher level brain regions to form a tactile percept. The region of the skin that is set into motion can be modeled by a mass m_{skin}, an elastic stiffness k_{skin} and viscous dampening d_{skin} that counteracts the force of the vibrating device. For measurements an accelerometer is often attached directly to the device housing. The acceleration measured this way is illustrated by the acceleration vector $\ddot{x}_{mes}$.

It could be argued that understanding these aspects is important to make an application inclusive and resistant to gender or physiology induced biases. A technology should ideally be designed around all possible users and minimize the amount of negative experiences for individuals due to technological shortcomings, which could be induced by a lack of physiological variety in user tests, for example.

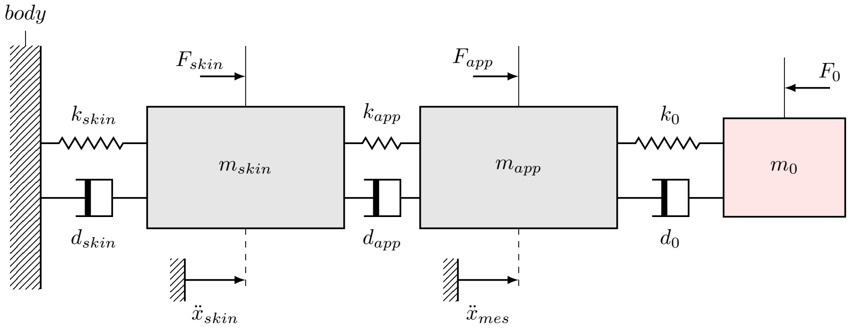


Figure 3.2: Schematic representation of the body-device-actuator mechanics. The mass m_0 is the moving mass within the actuator generating the perceivable force F_0.

3.2 Actuator Profiling and Linearization

To ensure an accurate reproduction of a stimuli signal throughout the entire vibrotactile playback system it is desirable to compensate for non-linear behaviour and to maintain a flat frequency response. A theoretical solution for linearizing a system requires finding the inverse of the transfer function of the system H^{-1}, which applies the vibrotactile stimulus to the skin. Pre-processing a target signal with the inverse system response this way can rectify a set of undesired effects and flatten the resulting frequency response. In a previous study, a haptic (i.e., vibrotactile) display was dynamically compensated in a bilateral teleoperation experiment, reporting that the use of the dynamic compensation "vastly outperform traditional position-position control at conveying realistic contact accelerations" [53]. In this work, instead of using a feedback path and dynamic compensation, we will try a simpler solution to compensate the actuators non-linearities and bodily-induced interference by running system identification measurements across various subjects to derive a static inverse filter. A static filter design has also shown promising results in a previous work, which used an auto-regressive parameter estimation for the equalization of a vibrotactile system [11].

A common way to identify a systems transfer function H is by measuring the systems output $y(t)$ towards a known input $x(t)$ and by doing so try to determine a mathematical relation between them without going into the details of what is actually happening inside the system. This approach is called system identification and can be conducted using prior knowledge about the system (grey box, i.e., having an approximate model) or no prior knowledge at all (black box). When assuming that the system is linear and time-invariant (LTI), it is possible to equalize the frequency response of the system using a linear filter and achieve a satisfying result without having to deal with minor non-linear behaviour. A well understood measurement procedure uses an impulse response (IR) measurement procedure. Most common an exponentially swept sine (ESS) is used across the desired frequency range — which in this case would be the perceptual bandwidth up to 1000 Hz.

Knowing the input signal $x(t)$ to the system allows us to find the optimal inverse frequency response using deconvolution, with $h(t)$ being the systems IR and $y(t)$ the measured output of the system, as seen in Equation 3.1. Using properties of the Fourier transform enables a deconvolution of the measured spectrum $Y(i\omega)$ and the input signal spectrum $X(i\omega)$ by division — with $\omega = 2\pi f$ being the angular frequency we get:

$$\begin{aligned} x(t) \circledast h(t) &= y(t) \\ \iff \mathfrak{F}\{x(t)\} \cdot \mathfrak{F}\{h(t)\} &= \mathfrak{F}\{y(t)\} \\ \iff X(i\omega) \cdot H(i\omega) &= Y(i\omega) \\ \iff H(i\omega) &= \frac{Y(i\omega)}{X(i\omega)} \\ \iff h(t) &= \mathfrak{F}^{-1}\{\frac{Y(i\omega)}{X(i\omega)}\} \end{aligned} \tag{3.1}$$

By calculating the inverse Fourier transform, we end up with the IR $h(t)$ of the measured system $H(i\omega)$. The IR $h(t)$ obtained this way can be stored for further analysis, and can be used as a target to find an inverse filter for the system.

3.2.1 System Identification Setup

The system identification procedure described above was implemented in the software framework Max/MSP using the “HISS Impulse Response Toolbox” [38]. The toolbox allows for a flexible environment to generate excitation signals and also retrieve an impulse response measurement by automatically deconvolving the recorded acceleration signal with the excitation signal. An exponentially swept-sine (ESS) excitation signal [18] ranging from 10 to 1500 Hz — exceeding the perceptual bandwidth — with a total duration of 20 seconds was used for each sweep. The duration was chosen long enough to allow the actuator to reach a steady state for each frequency component.

To measure the acceleration profile of the actuator a custom circuit board using a ADXL325 3-axis accelerometer by Analog Devices was attached to the actuator housing or device enclosure. The acceleration signals of all three axis were individually recorded. As the three axis are orthogonal, the total magnitude of the acceleration a_{total} was calculated by the square root of the sum of squares ($\sqrt{a_x^2 + a_y^2 + a_z^2} = a_{total}$). This was done to avoid losing any off-axis energy in the process as the single axis operation of the actuator can’t be assumed to be perfect. Each IR measurement throughout this work was individually rendered to a WAVE file for further analysis, filter design and plotting in a Python script using the SciPy open-source ecosystem.

The hardware setup used for the measurements in this work was calibrated using a Siglent SDG1010 signal generator and a Rigol DS4014 oscilloscope. The calibration ensured a controlled input and output voltage throughout the entire signal chain. The measurement software (described above) was running on a Apple MacBook Pro 15" (mid 2015) using a Motu 624 USB audio interface for signal playback and retrieval of the accelerometer signals.

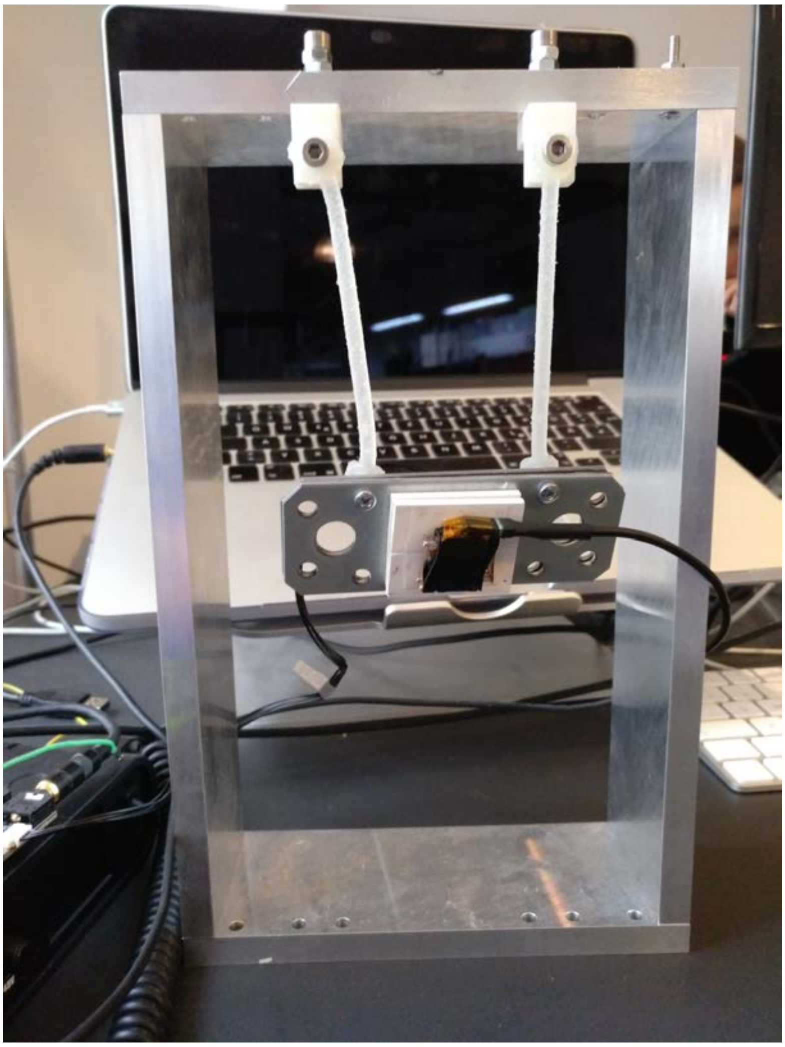


Figure 3.3: The image shows the test jig used for the system identification procedures of actuators. It uses visco-elastic silicone straps to attach the actuator and attached mass in a approximately “free-floating” manner. The accelerometer is attached to the back of the metallic plate seen in the image to avoid electromagnetic interference from the device under test (DUT).

3.2.2 The L5 Voice Coil Actuator

The L5 actuator used for the experiments in this thesis was designed by the company Lofelt GmbH in Berlin. The actuator was designed to overcome shortcomings of other actuator types on the market (see Section 3.1). Covering the entire vibrotactile perceptual bandwidth from low frequencies up to 1000 Hz (see Section 2.3) could not be achieved with state of the art devices. Balancing the device size, cost and power consumption against a desired vibrational force, impulse fidelity and frequency response is an ongoing pursuit for (mechanical) engineers throughout the industry.

The L5 actuator is a type of VCA actuator. It is made up of an arrangement of permanent magnets that act as the moving mass. These permanent magnets are held in equilibrium at the center of the actuator by two flexible copper beryllium (CuBe) membranes, which are in turn attached to the actuator housing. A copper wire, making up the voice coil, is wrapped around the chassis of the mass-membrane configuration (for reference, see Figure 3.5). To drive the actuator, an AC voltage of the desired frequency (or waveform) is applied to to the voice coil. For each period of the AC cycle, the resulting magnetic induction of the coil will apply a force on the permanent magnets and move them away from equilibrium, when the stiffness of the membranes is overcome. The moving mass then oscillates back and forth, according to the input signal and membrane stiffness, resulting in a vibrational force. The force of this movement can either be calculated by measuring the distance the moving mass travels over time, multiplied by the mass ($\vec{F_0} = m_0\vec{a}$), or by attaching a accelerometer to the device.

The L5 actuator has an impedance of $8.8\,\Omega$ and was driven using a voltage of $1\,V_{rms}$. To ensure the actuator was driven with enough power a custom amplification circuit board was used. The frequency response of the L5 actuator measured on a suspended test jig (see Figure 3.3) is illustrated in Figure 3.4. Note that this measurement does not reflect the behaviour of the L5 in a use case scenario, as the true device mass and bodily induced interference is not present. The resonance peak at around 67 Hz is prominent and allows the actuator to induce a comparably strong acceleration force, even at low frequencies. The resonance peak also indicates, that the L5 actuator is an underdamped system and therefore benefits from either mechanical dampening, electronical- or DSP-equalization to ensure a flat frequency response within the range of operation (i.e., tactile sensitivity range).

The resulting force of the actuator and the required voltage to drive it depend on the mass or device properties it is attached to. For example, for industry applications this could be a mobile phone (150 g), game controllers (200 g), headsets (250 g), head-mounted displays (500 g) or wristbands (40 g). The actuators performance is further altered by the dampening and stiffness of the human tissue, making ad hoc or application specific motor control ideal.

3.2.3 The Vibrotactile Wristband Display

To Perform a user tests that reflects a realistic application a wearable wristband form factor was chosen. Various devices such as smart watches, health trackers and fitness bracelets use a similar form factor and are thus widely spread in the consumer market.

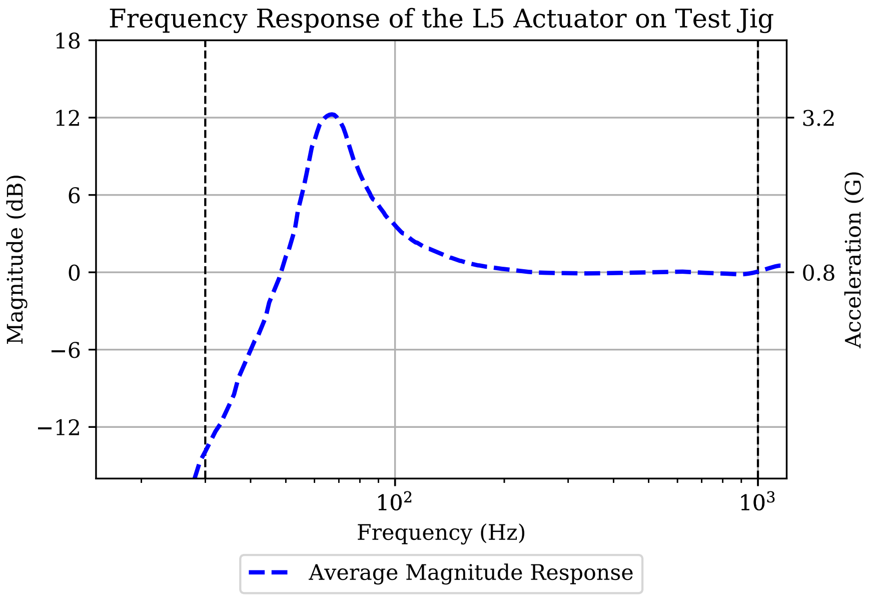


Figure 3.4: The average frequency response across ten measurements conducted in the actuator acceleration test jig (described below in Section 3.2) with 120 g attached mass. The plot was normalized by the mean across all frequency bins. The resonant peak at 67 Hz with a mean acceleration $3.2\,G_{rms}$ is clearly visible, as well as the flat response from 200 to 1000 Hz with $0.8\,G_{rms}$. For reference $1\,G \overset{\text{def}}{=} 9.81\,ms^{-2}$.

The main inspiration for this use case was the wearable "subwoofer" wristband formerly sold as *Basslet* by the company Lofelt.

For the perceptual evaluation of the audio-tactile translation framework, two Basslet wristbands were modified by removing the original Bluetooth connection to avoid any interference during transmission and signal modifications induced by the Bluetooth codec. To achieve this, all components except the actuator were removed from inside the wristband. A cable was soldered directly to the actuators coil and a connection was made available from outside the wristband case. Using these modifications the wristband was operable using an external driver board, allowing more control over the signal driving the actuator. For images of the original, wireless wristband and the modified, cable bound version please see Figure 3.6 and 3.5 respectively.

As humans are more sensitive towards vibrotactile stimuli in hairless (i.e., glabrous) skin areas, the wristband used in this work was advised to be worn with the actuator in contact with the inner arm, rather than the outside — like rotating a wristband watch by 180°, as depicted in Figure 3.6.

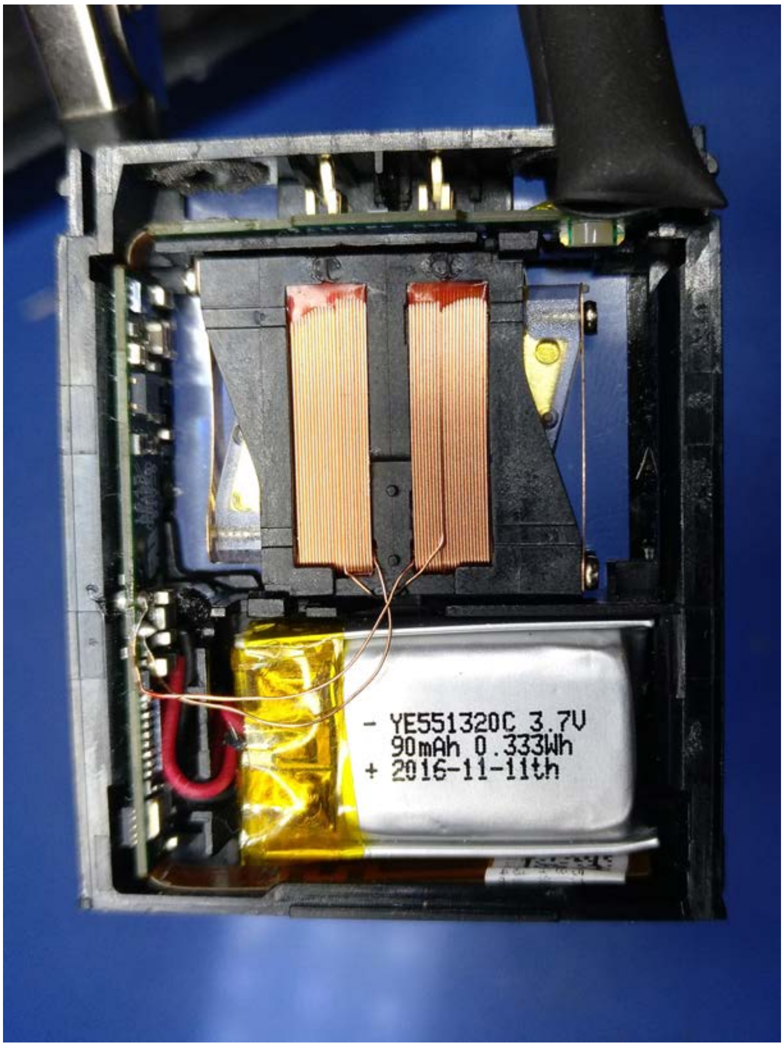


Figure 3.5: This image shows the L5 actuator in the original "Basslet" configuration: On top we can see the L5 VCA actuator. Notice, that the voice coil and membranes are visible. Above the actuator and to the left we can see the PCB housing the bluetooth antenna, DSP microcontroller and a driver board. On the bottom we can see the battery of the device. For the scope of this thesis all components except the actuator were removed and a cable was installed to drive the actuator, as seen in the Figure 3.6.

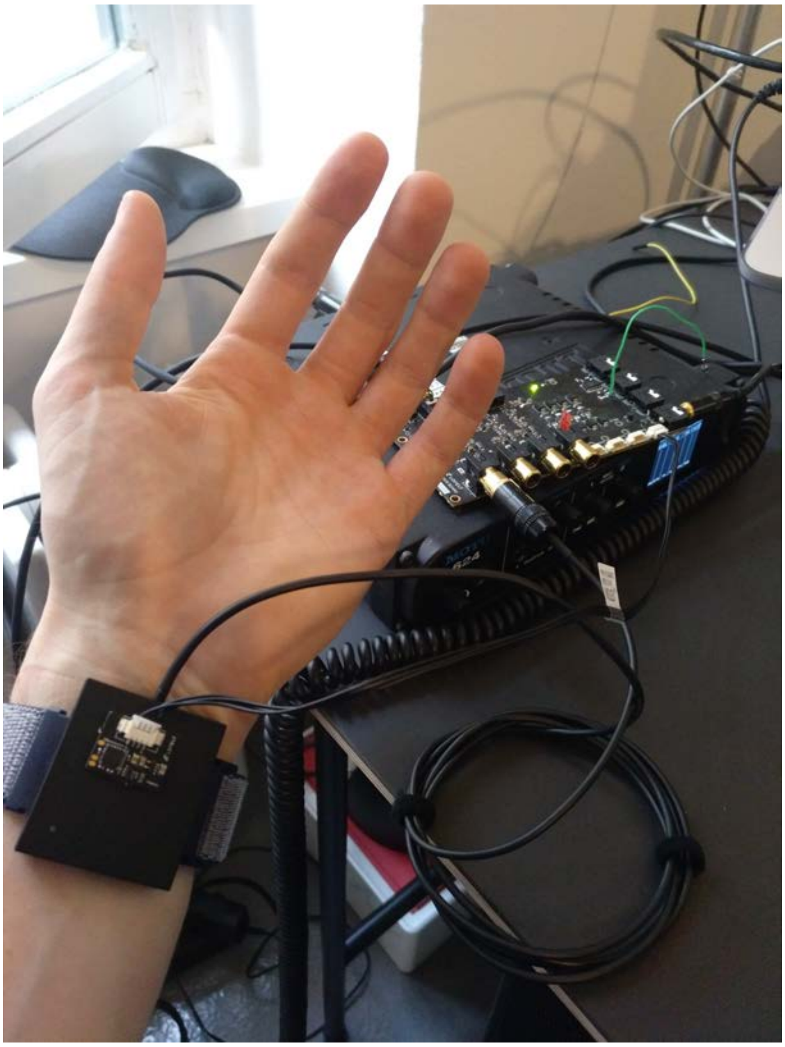


Figure 3.6: This image shows the vibrotactile wristband attached to the anterior of the left wrist of a subject. In the background the Motu 624 USB audio interface and the custom driver board can be seen. The small PCB attached to the housing of the wristband is a 3-axis accelerometer, which in turn is connected to the custom driver board to sample the acceleration data with the audio interface. Notice the cable bound configuration compared to the original Bluetooth version shown in Figure 3.5.

3.2.4 System Linearization

To achieve a controlled stimuli reproduction it is desirable to both measure and counteract actuator-, device- and bodily-induced interference on the tactile display. Using the IR measurement procedure outlined in Section 3.2 an approximate inverse frequency response $H_{wrist}^{-1}(\omega)$ for the wristband display was obtained. This was achieved by conducting IR measurements across 14 subjects (4 females, 10 males; average age = 34.5, SD = 4.6; range = 29 - 47), mainly composed of office colleagues from various professional backgrounds. All subjects were advised to attach the wristband comfortably, once on

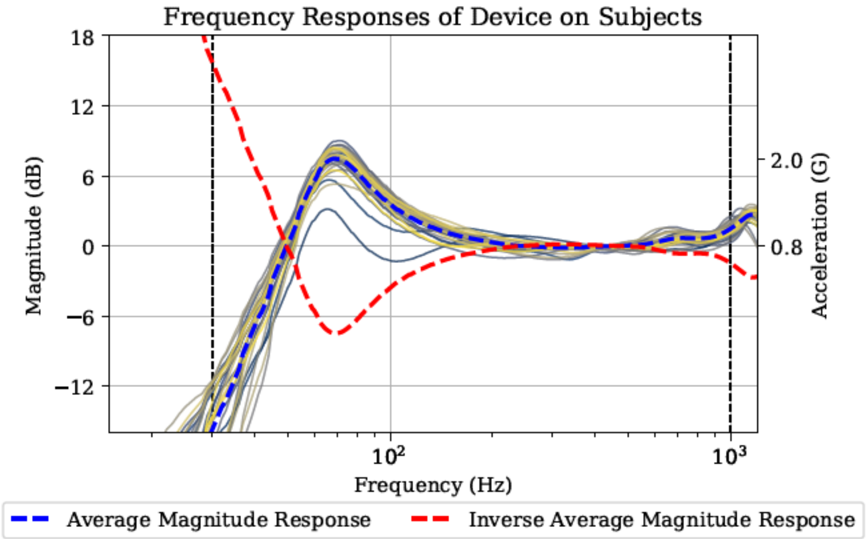


Figure 3.7: Frequency response plot derived from 28 impulse response measurements of 14 subjects using a wristband form factor with the L5 actuator on the inner and outer forearm. A prominent resonance peak of the system at 76 Hz is noticeable with a peak acceleration of $2\,G = 19.6\,ms^{-2}$.

the inner and once on the outer forearm — resulting in a total of 28 individual impulse responses.

The frequency response of each individual IR was filtered using third-octave band smoothing. The frequency responses were vertically aligned to each other by normalizing each measurement by the respective mean energy. Next, using Equation 3.2 with $N = 14$ the average magnitude frequency response $H_{avg}(\omega)$, and the inverse $H_{avg}^{-1}(\omega_k)$ were calculated. The results of this process can be seen in Figure 3.7.

We can observe, that the resonance peak from the measurements of the L5 actuator in the test jig (see Figure 3.4) was dampened by about 6 dB when measured in the wristband on subjects. This is mainly reasoned to be due to the dampening effect of the skin, as discussed in Section 3.1.4. The variance in the individual frequency responses indicate that the bodily induced interference on the system due to varying physiology and the subjective opinions on a “comfortable fit” have an effect on the system response. Next to two observable outliers in the individual measurements, this effect doesn’t seem to be very strong though as most measurements line up nicely with the calculated average response.

$$H_{avg}(\omega) = \frac{1}{N}\sum_{n=1}^{N}|H_n(i\omega)| \tag{3.2}$$

Inverse Filter Design

A common solution for equalizing the frequency response of a system is to design a filter (or set of filters) to approximate the target inverse frequency response $H^{-1}(\omega)$. This is done to attenuate overly present, resonant frequency bands and amplifying weaker frequency bands to achieve an approximately flat frequency response.

There are various ways to design a compensating filter. One approach is to try and approximate the target frequency response with an equation using the Yule-Walker algorithm. Using the simplified transfer function obtained this way allows the design of a filter by simply inverting the retrieved transfer function. This method has been shown to work for vibrotactile systems in a previous study [11].

Another common optimization method utilizes recursive least squares (LS) to approximate an optimal finite impulse response (FIR) filter with the filter coefficients $\boldsymbol{h}_{opt}$ by minimizing the (weighted) error function $E(\boldsymbol{h}, \omega_k)$ (see Equation 3.3). The error function $E(\boldsymbol{h}, \omega)$ is used as a measure of deviation between a target frequency response $D(\omega)$ and the frequency response of a filter design $H_d(\omega)$. The weighting matrix $W(\omega)$ is applied to emphasize frequency regions deemed more important in the regression. Even if the equation is not solved analytically by calculating the derivative and finding the minimum of the error function for $\boldsymbol{h}$, so that $\nabla_h E(\boldsymbol{h}, \omega)|_{h=h_{opt}} \overset{!}{=} 0$, the error function $E(\boldsymbol{h}, \omega)$ is still useful to obtain a numeric result on the performance of a filter towards a design target.

$$\underset{E \to 0}{\text{minimize}}\, E(\boldsymbol{h}, \omega_k) = \underset{E \to 0}{\text{minimize}} \sum_{k=\omega_{min}}^{\omega_{max}} \|W(\omega_k)[D(\omega_k) - H_d(\boldsymbol{h}, \omega_k)]\|_2^2 \tag{3.3}$$

A FIR filter, that was obtained by using the LS method, can result in quite a high order (i.e., have many coefficients) and thus increase the computational cost, and, depending on the phase response, might result in a increased group delay. This can induce a significant delay on a playback system. This can be a problem for time-crucial applications where audio, video and tactile stimuli should remain synchronized. If feasible, this can be avoided by using delay compensation methods, such as delaying all media assets by the same amount of time.

For the process of linearizing the actuator system response in this work, we are trying to find an approximation for the target spectrum $D(\omega) = H^{-1}(\omega)$ by minimizing the error through optimizing the design of the filter $H_d(\omega)$ while keeping any induced delay as low as possible. The target for $H_d(\omega)$ is illustrated as the blue, dashed line in Figure 3.7.

Another common filter design method is to manually (i.e., heuristically) tune a set of parametric filters until a satisfying result is achieved. In this scope, this approach seemed the most feasible as the overall frequency response is quite flat and the target bandwidth is comparably small (30 to 1000 Hz). After each iteration the error between the desired, flat spectrum and the resulting average spectrum across all filtered impulse responses was calculated. This improves the reliability of the visual tuning done by looking at the filter frequency response plots in correspondence with the target, inverse spectrum as can be seen in Figure 3.8.

A parametric infinite impulse response (IIR) filter can be designed by using an initial analog prototype filter $H_a(s)$, which is then digitized using the bilinear transform (BLT). This design method allows direct control over key parameters, such as the sampling rate F_s, the quality Q, the relative gain G in decibel and the center frequency f_0 of the filter [9]. The process of tuning each filter was iterated until a satisfying match towards the target spectrum with $H_d(\omega_k) \approx H_{wrist}^{-1}(\omega_k)$ was achieved. To be able to track improvements on each iteration, the error function described in the above LS process was used. To get a baseline value, the deviation of the measured spectrum from a flat spectrum was calculated: The target spectrum $H_{wrist}(\omega_k)$ had an original mean deviation from a flat spectrum of $\mu(E(H_{flat} - H_{wrist})) = 0.559$ (SD = 1.870), which was aimed to be minimized which each iteration of the filter design process. The filter resulting from the iterative process $H_d(\omega_k)$ had a final mean deviation of $\mu(E(H_{wrist} - H_d)) = 0.381$ (SD = 0.236) towards the target spectrum along all frequencies on the target interval $[\omega_{min}, \omega_{max}]$ with $\omega_k = 2\pi f_k$ and f_k between 30 and 1000 Hz.

The final parameter settings for each filter in the filter cascade approximating H_{wrist}^{-1} are documented in Table 3.2 and the resulting frequency responses of each filter, as well as their sum are illustrated in Figure 3.8. The plot shows an additional high-pass filter, that was initially used to compare the behaviour in the drop-off towards lower frequencies but wasn't used in the final implementation. The resulting frequency response when applying the set of filters to the measured data can be seen in Figure 3.9. It can be observed, that the remaining deviation of each measurement from zero (i.e., a flat spectrum) is within a margin of about $\pm 4\,dB$. The two outlier measurements discussed above can be identified below the average frequency response graph and, as expected, maintain a stronger deviation (i.e., remain outliers) after filtering.

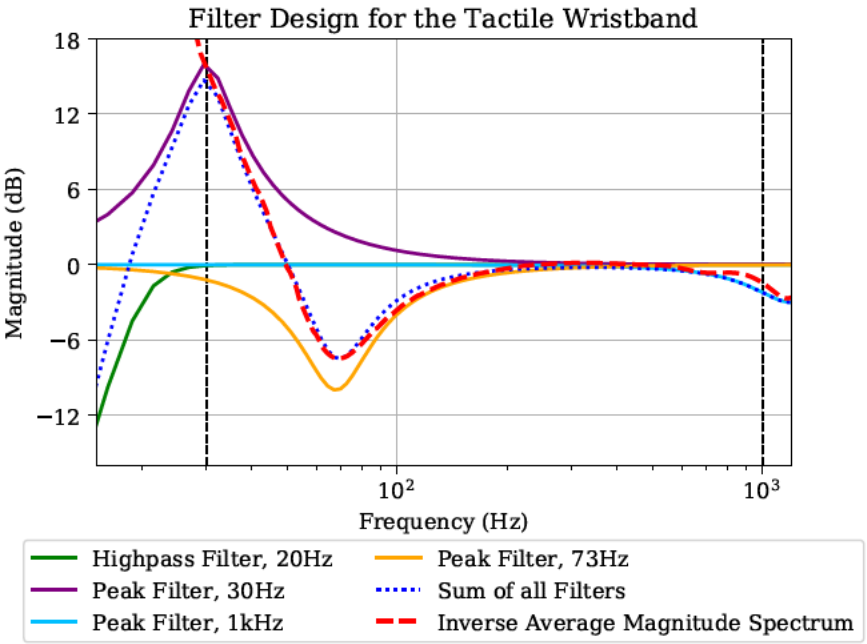


Figure 3.8: This figure illustrates the frequency responses of the designed filters to equalize the vibrotactile wristband frequency response. The deviation between the resulting, designed inverse filter and the target inverse frequency response can be seen by observing the blue dotted plot and the red dashed plot.

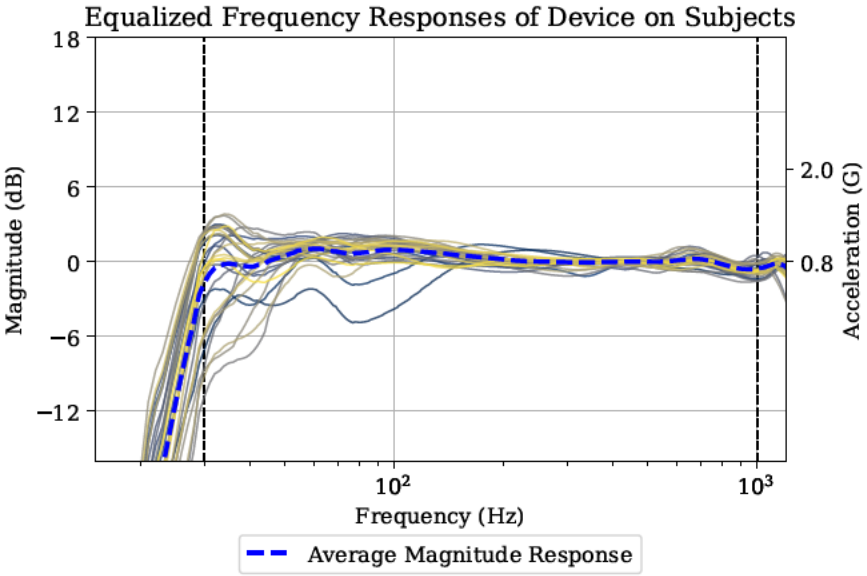


Figure 3.9: The frequency responses of the filtered impulse responses using the filters depicted in Figure 3.8. The dashed blue line depicts the average of all measured frequency responses.

Filter Number	Type	Order	Frequency (Hz)	Gain (dB)	Quality Q
1	Bandpass	n/a	30	16.0	1.5
2	Bandpass	n/a	68	-10.0	1.6
3	Bandpass	n/a	1200	-3.0	1.5
4	Highpass	5	20	n/a	n/a

Table 3.2: Parametric filter settings used to approximate the target inverse average frequency response. The result on the measurements can be seen in Figure 3.9.

Conclusion

In this section a set of measurements and a linearization process for the L5 VCA actuator used in this work was presented. Application specific system changes, such as added dampening and mass through the wristband form factor and bodily induced interference by the human physiology (i.e., skin) were discussed. Even though the human somatosensory system is sensitive to vibrations lower than 30 Hz it is increasingly hard to reproduce these frequencies in a 1-degree of freedom (DOF) VCA while maintaining a flat response at higher frequencies. To this date this still poses a challenge for the mechanical design of vibrotactile actuators. Trying to fix this problem with DSP alone can lead to problems with overheating the coil. This is due to the high amount of gain required to compensate the attenuation below the natural drop-off of the actuator at lower frequencies. It is therefore wise to high-pass the signal driving the motor at the lowest reasonable frequency without trying to force the actuator to perform in frequency bands it was not designed for. The inverse filtering conducted in this work mainly aimed to dampen the strong resonance peak and provide a overall flat frequency response. This is why filter number 1 (see Table 3.2) was not used during the final implementation of the perceptual user test and the natural drop-off of the actuator below 50 Hz was tolerated, while maintaining the high-pass filter (number 4) around 20 Hz just to be safe and remove any remaining DC bias.

4 Audio-Tactile Signal Translation

With the perceptual characteristics of both the auditory and tactile modalities, the constraints of vibrotactile actuators and the audio-tactile design workflow in mind, this chapter documents the decision making and implementation specific details of the DSP signal chain used in the proposed audio-tactile translation framework. Inspiration was drawn from audio codec design, data science methods and audio processing concepts used in musical applications. The overarching goal of the proposed framework was to design a set of analysis, storage and synthesis methods that provide a starting point for the design of tactile stimuli from audio assets. A flowchart illustrating the proposed workflow is illustrated in Figure 4.1.

A tactile experience designer or content creator would ideally utilize the final audio assets curated for a application and encode them using an analysis tool chain (i.e., encode the data). The data obtained from this process is saved in a custom data model. The data model can then be parsed and displayed by an authoring tool, which enables a preview of the stimulus and allows a comfortable way to edit the tactile stimulus. If the designer is satisfied, the data can be stored, transmitted or integrated in a target application.

On the playback side a product owner needs to make a decision for a tactile actuator and integrate it in the hardware application (f.e., smartphone). As the actuator has specific properties and limitations, this information is saved in a actuator specific configuration file on the device. The information on the actuator includes the resonance frequency, a frequency range with a defined dynamic range, a step response (i.e., rise and fall time) and information on the capability of the actuator being able to display multiple frequencies at once (i.e., monophony or polyphony).

When a tactile event is triggered on the target device, the encoded data is parsed and handed to the synthesis engine. Together with the actuator specific configuration a prioritization takes place, suiting the actuators capabilities. This can be decisions on the amount of voices (i.e., amount of individual sinusoidals) being played, the synthesis method being used or a re-scaling of the output frequency range.

4.1 Requirements

Various requirements were considered for the design of the audio-tactile signal translation framework: Not only is a coherent perceptual translation desirable, it is also important

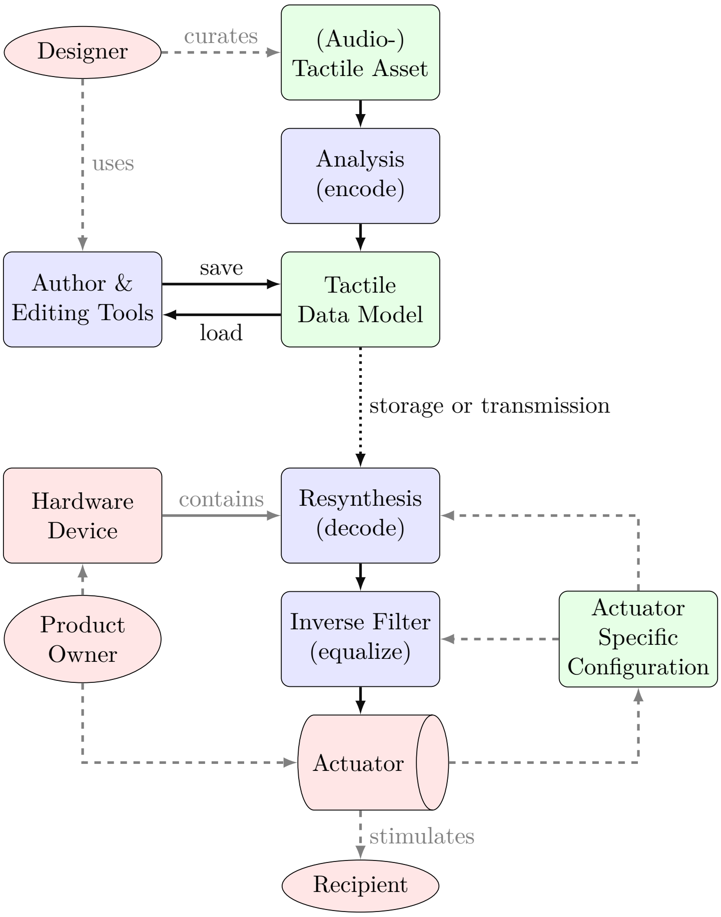


Figure 4.1: Flowchart of a proposed framework integration and tactile design workflow. Human resources involved in the process are highlighted in the red ovals. Blue rectangles illustrate a process in the information flow. Green rectangles illustrate data files (i.e., information). The actuator hardware is illustrated as a cylinder.

to ensure that the intermediate data model is useful for both content creators (the designers of a tactile experience) and developers (the programmers implementing the tactile stimuli) and ensures a consistent stimulus reproduction across various tactile devices. For a content creator it is important that the intended tactile experience is played back as close to the original design as possible on the recipients device. As noted before, this process is complicated by the variability between various actuator embedded in an tactile capable devices, as discussed in Section 3.1.3. For a developer it is desirable to easily integrate tactile media content into any software or hardware application — thus requiring the tactile data format to be easily integrated into the development workflow.

The most important aspects that influenced the design of the proposed framework are listed below:

1. The framework is able to produce a coherent translation between audio and vibrotactile stimuli.

2. The framework enables authoring, editing and manipulating (transposing, stretching) the resulting vibrotactile stimuli pattern.
3. The framework allows for adjustable data reduction (i.e., quality degradation)
4. The framework allows for adaptive degradation, which means the resynthesis is able to adapt to the fidelity of a target reproduction system.
5. The framework is computationally inexpensive on the decoding (re-synthesis) side to reduce computational cost during playback.
6. The framework allows for real time parameter control over the synthesis engine, so that dynamic panning, intensity control and delay shifts are possible.
7. The framework works both "offline" on an entire audio asset, and in real time (i.e., run with low, imperceptible delay during run time)

4.2 Parametric Signal Representation

To best address the requirements above, a parametric signal representation seemed to be the most suitable solution. Depending on the output of the exact method in use, control over various signal properties in an intuitive and predictable way can be provided. To achieve a parametric representation a source signal is analysed and important features are extracted (i.e., the signal needs to be decomposed). Next, the components retrieved this way need to be matched with a corresponding synthesis method to ensure an approximate reproduction of the source signal. An approximation of the original signal is proposed to be suitable, as the sensitivity and resolution of the skin appears to be lower when compared to the sensitivity and fidelity of our ear.

Using a sparse, model-based or parametric representation to render vibrotactile stimuli has been previously explored in various works, ranging from approximating recorded material texture signals by source-filter models [35, 13] and approximating texture signals from sinusoidal components [3].

4.2.1 Spectral Modelling

Next to other methods that achieve a parametric signal representation, the main inspiration for the signal decomposition method used in this embodiment comes from the "spectral modelling" approach described by Xavier Serra. This method retrieves both modal (deterministic sinusoidals) and residual noise components (filtered stationary stochastic noise), which can be saved and later used to additive reconstruct the source signal [91]. Further inspiration was drawn from investigations into various parametric audio codecs, such as AAC, Opus and Harmonic and Individual Lines and Noise (HILN) [1, 98, 76].

In a later work a conceptually similar method known by harmonic percussive source separation (HPSS) was proposed to isolate harmonic, percussive and residual components by analysing the spectrogram $S(m, \omega)$ of a target signal [20]: First, a spectrogram matrix $S(m, \omega)$ with a time frame m and angular frequency ω of a discrete time source signal $x[n]$ is calculated using a segmented short-time Fourier transform (STFT) with a sliding window function $w[n]$ of length N, as described in Equation 4.1. The window function is set to 0 anywhere outside the interval $[0, N-1]$, thus effectively computing a STFT for each successive input data frame m with M samples distance. This results in an approximate time-frequency representation of the input signal, which can be tuned by varying the window size N, hop size M and amount of zero padding added to each frame.

$$\begin{aligned} spectrogram\{x[n]\}(m, \omega) &= |S(m, \omega)|^2 \\ S(m, \omega) &= \mathfrak{F}\{x[n]\}(m, \omega) \\ &= \sum_{n=0}^{N-1} x[n + m \cdot M] \cdot w[n] e^{-j\omega n \frac{2\pi}{N}} \end{aligned} \tag{4.1}$$

Using the calculated spectrogram matrix S, sinusoidal (harmonic) components are isolated by using median filtering along the horizontal time axis n to retrieve a matrix of filtered frequency traces $H(m, \omega)$. After further processing by using a time-frequency ridge detection the isolated frequency traces retrieved this way can be separated into individual pitch tracks. A similar routine is conducted to isolate percussive (i.e., transient) components in a spectrogram matrix $P(m, \omega)$. Here, median filtering is applied vertically along the frequency axis ω to emphasize short wideband events, which are suspected to correspond with percussive transient events. Further processing can aid in isolating the estimated amplitude and temporal position of suspected transient events. Finally, the temporal trajectory of the spectral shape of the remaining noise (i.e., residual components) $R(m, \omega)$ can be isolated by subtracting both the harmonic H and percussive spectrogram P from the spectrogram matrix S as seen in Equation 4.2.

$$R = S - (H + P) \tag{4.2}$$

If any data reduction is conducted at this point, or if the matrices are reconstructed from more abstract data the corresponding approximate component matrices of H, P and R are denoted as $\hat{H}$, $\hat{P}$ and $\hat{R}$ respectively, as seen in Equation 4.3.

$$\begin{aligned} \hat{H} &\approx H \\ \hat{P} &\approx P \\ \hat{R} &\approx R \end{aligned} \tag{4.3}$$

To reconstruct the signal, the harmonic, percussive and residual components can be recombined to retrieve an approximation of the input spectrogram $S(m, \omega)$ and then

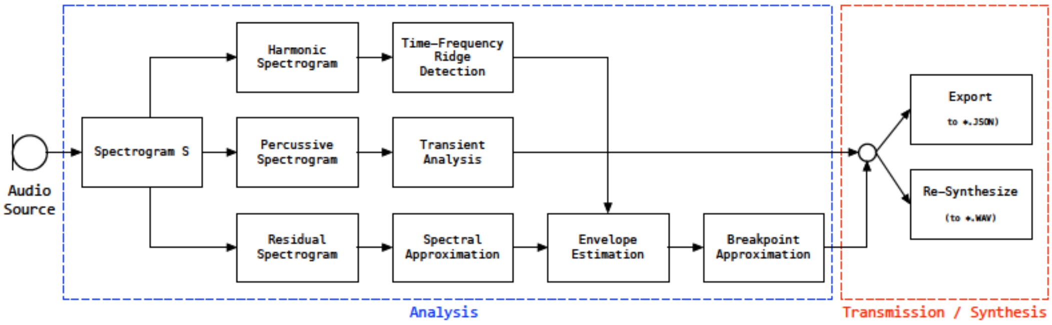


Figure 4.2: Flowchart of the full, computationally intensive signal decomposition and resynthesis method. First, a spectrogram matrix of the incoming audio source signal is computed. Next, HPSS is applied to isolate harmonic, transient and residual spectrograms. The individual spectrogram matrices are then further processed to retrieve a high-level representation of individual signal components, thus achieving a parametric representation.

transforming the spectrogram back to a discrete time signal $\hat{x}[n]$ by using the *inverse* of the segmented overlap-add method used to calculate the original spectrogram $S(m, \omega)$, as described in Equation 4.4. A flowchart of the entire process from analysis to resynthesis is illustrated in Figure 4.2.

$$\begin{aligned} \hat{S} &= \hat{H} + \hat{P} + \hat{R} \\ \hat{x}[n] &= \mathfrak{F}^{-1}\{\hat{S}(m, \omega)\}[n] \end{aligned} \tag{4.4}$$

The overall concept of decomposition is suitable for the framework proposed in this work, as the intermediate information on pitch, amplitude, spectral shape and the temporal evolution of a signal can be saved as individual components. This allows individual components to be modified and optimized towards the capabilities of the connected tactile display technology, as provided in Table 3.1. The benefits of such a decomposed format would be to allow for a technology- and hardware agnostic specification of the desired vibrotactile stimuli and allow high-level control on individual modal, stochastic and temporal components of the stimuli signal. Another key benefit of a re-synthesis method is the capability of introducing variance to repeating events, such as game and user interface assets, by randomly altering modal component amplitudes and applying random filters to the residual noise (see Section 4.3.3). This method can allow for a significant reduction of data that is usually stored and transmitted by an application. This is due to the possibility, that variations of the same interaction don't have to be stored and re-transmitted as separate events each time they are triggered, but instead slight variations of an event could be synthesized on the fly [51].

4.2.2 Reducing Computational Complexity

When initially prototyping the proposed method above, a subjectively inaudible reproduction of a sound stimulus could be achieved. This made the entire procedure feel like using a sledgehammer to crack a nut. Therefore, due to the evidently lower sensitivity and resolution of the tactile sense (see Chapter 2.3) a rougher approximation of the source signal was proposed to still enable an adequate reproduction of the stimulus.

Instead of the computationally intensive HPSS and time-frequency ridge tracking method described above, individual components could be approximated with a lower precision, as the tactile sense shows less sensitivity in distinguishing individual frequencies in a complex frequency spectrum [31, 81, 55]. It is proposed that an approximation of the spectral shape could be conducted using a computationally cheaper filter bank and a sinusoidal resynthesis instead of the costly inverse short-time Fourier transform (ISTFT) overlap-add method. This process works similarly to a phase vocoder, as a desired amount of individual frequency bands are retrieved and can be treated as individual frequency components. It is proposed, that a more precise extraction of individual pitch tracks is not required and can provide a perceptually indistinguishable reproduction of the source signal. Next, percussive (i.e., transient) components are proposed to be retrieved using a low-cost onset detection algorithm. Individual transient components can be captured this way and can also be treated as individual components. Finally, next to the filter bank analysis an additional estimation on the most dominant pitch of the source signal is proposed to be done by calculating the spectral centroid of the source signal, or on individual frequency bands of the filter bank.

Initial tests of the filter bank analysis and resynthesis approach, using a real-time prototype implementation in Max/MSP showed promising results between a few office colleagues and the main author. This was assessed by A/B testing between a set of low-passed (at 1000 Hz) versions of the original source signals and the filter bank approximated signals, using a mere set of 6 bands spread out between 30 and 1000 Hz. Depending on the sound source it was often impossible to tell by the participants which variant was currently being displayed. What type of sounds are feasible to be recreated this way and which types of sounds don't work as well was not explored in this work, but could be the topic of a future experiment using a more diverse group of participants and a perceptual AB/X experiment design.

All components, namely the total signal envelope; individual frequency band envelopes and corresponding pitches; transient positions and magnitudes; the spectral envelope of the residual noise; as well as the information on the spectral centroid can individually be saved and edited. Further, all time-value trajectories (envelopes and pitch value changes) can undergo a data reduction schema. It is thus far unclear how many frequency bands ensure a perceptually indistinguishable reproduction, or which of the signal components might be obsolete. This means that it remains an open question how big the difference between the source and the approximate reproduction can be to be noticed when being displayed to a set of subjects (i.e., finding the JND). This needs to be explored with further perceptual tests by exploring the parameter space of the codec process.

4.3 Implementation

The proposed Audio-Vibrotactile Signal Translation (A2VT) framework was implemented in Python, using the scientific open-source software ecosystem SciPy. The entire code is structured and packaged according to the Python documentation. This has the benefit, that the entire framework can be distributed and installed by following the standard installation procedure using the default Python package manager `pip` and to make the functionality accessible with the standard `import` command. Individual features of the framework, such as the DSP algorithms, compression, plotting and synthesis are maintained as individual file structures, thus keeping the code clean and easier to maintain. All algorithms and processes discussed in this section are described in more detail below.

After creating a A2VT instance with the desired parameter settings, a target audio file is loaded. The audio file is pre-processed by mixing the contained channels down to mono and band-limiting the resulting signal. Using the default settings the band-limiting is done with a high-pass filter at 5 Hz and a low-pass filter at 1800 Hz. This is done to remove information beyond the tactile sensitivity range, while maintaining some redundant information that can aid the analysis process. After normalizing the signal the amount of time above a default amplitude threshold of -75 dB is estimated. This is done because the line simplification algorithm (a.k.a. breakpoint approximation) during compression was used with a target number of “breakpoints per seconds” parameter and thus avoids using too many breakpoints, in case the audio signal contains silent segments that don’t require any coding.

After pre-processing the analysis is conducted on the signal. First, all required analysis parameters are loaded from the settings file and set within the analysis object. During analysis the overall signal envelope and trajectory of the spectral centroid are calculated. Next, the signal is run through both the filter bank and the transient detection algorithm. All time-value trajectories gathered this way are stored as arrays within the A2VT instance.

To allow for efficient storage, all time-value trajectories are compressed using a line simplification algorithm, which essentially removes irrelevant and redundant information. Instead of equal-spaced time-value arrays the line simplification algorithm returns non-equal spaced time-value pairs, that are referred to as breakpoints. The compressed data can then be re-synthesized or exported to a JavaScript Object Notation (JSON) file. All entries in the JSON file are explicitly named and can be identified by using

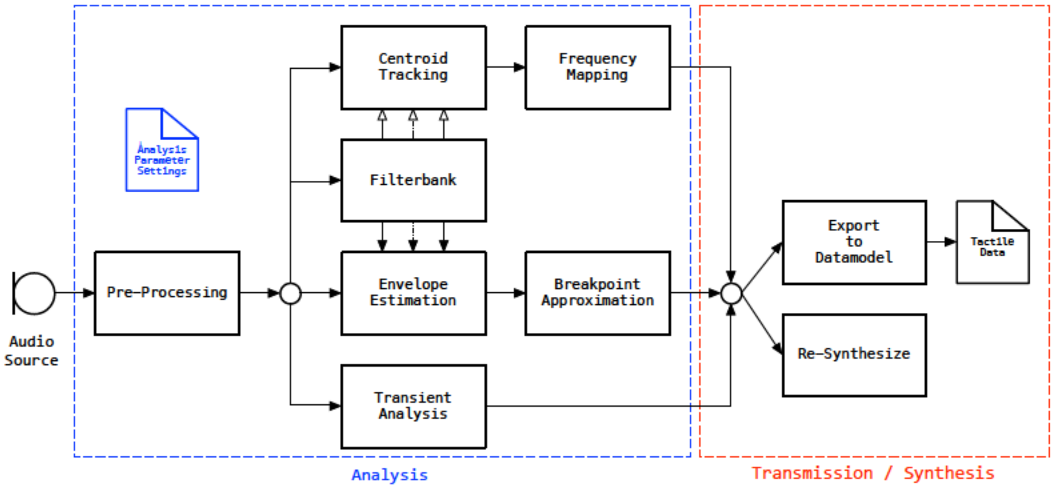


Figure 4.3: Flowchart of the proposed, approximate signal decomposition and resynthesis method with reduced computational cost. After pre-processing the audio source, various spectral and temporal components are extracted and compressed. The resulting data can then either be exported or used to re-synthesize the signal.

human-readable key values. Alternatively, the data can be plotted using a built-in plotting method. An overview of the entire process is illustrated in the flowchart in Figure 4.3.

4.3.1 Encoding: Analysis Algorithms

To retrieve the desired parametric signal representation, which is proposed to be segmented into various signal components, a set of DSP algorithms is required. These algorithms are used to analyse an audio source signal as part of the A2VT framework to enable a efficient encoding and comprehensible representation of the tactile signal, thus making authoring and editing feasible for content creators. All algorithms used in the A2VT pipeline are described below.

Envelope Follower

The envelope follower is designed to extract the momentary amplitude envelope of a signal in real time. It allows for parametric control over the *attack* and *release* phase of the signal, thus enabling separate control over the sensitivity of the envelope follower during rising and falling amplitude segments. From a theoretical perspective, the envelope follower is basically a state switching one-pole low-pass filter with different coefficients for attack c_a and release c_r. The coefficients determine the precision, or in other words, the decay behaviour of the envelope follower. To catch quick changes in amplitude, such as transients, the attack time should be small. Depending on the desired detail on the decay of the signal a comparably long decay time is suggested. The exact values are up to personal preference though.

The parameter expressing the amount of smoothing, or in other words, the length of the decay t_{ms} is defined in milliseconds. The corresponding amount of samples at a sampling rate F_s required to calculate the coefficients is calculated using Equation 4.5.

$$N_t = \lfloor \frac{t_{ms} \cdot F_s}{1000} \rceil \tag{4.5}$$

To calculate the coefficients for the one-pole filter either equation from 4.6 can be used. The equation for $c_{99\%)}$ yields a slightly steeper (faster) decay response when used in the one-pole filter in comparison to $c_{63\%}$.

$$\begin{aligned} c_{99\%} &= 0.01^{\frac{1}{N_t}} \\ c_{63\%} &= e^{\frac{-1}{N_t}} \end{aligned} \tag{4.6}$$

The z-transform for the implemented one-pole filter used in the envelope follower is shown in Equation 4.7.

$$H(z) = \frac{b_0}{1 - a_1 \cdot z^{-1}} \tag{4.7}$$

To get a varying response during the attack and release phase of the signal the decay coefficients for both the chosen attack- and release time are calculated using Equation 4.6. The filter coefficients for the one-pole filter are derived by setting $a_1 = c$ and $b_0 = 1 - c$. For each sample we calculate if the signal in on a rising (attack) or falling (release) slope by calculating the discrete derivative. Depending on the current slope we either chose the attack or the release one-pole filter to smooth the signal envelope.

A pseudo-code implementation can be seen in Algorithm 4.1. The algorithm in this implementation also features a hold count H that can be defined in milliseconds and expressed in samples using Equation 4.5. An exemplary result of the envelope follower can be seen illustrated with a blue dashed line in Figure 4.4.

An optional expansion of this algorithm uses an adaptive control over the attack- and release time on the basis of the momentary Crest factor of the signal and a forgetting factor [29]. The adaptive control was implemented but not used for the tactile stimuli of the perceptual evaluation.

Transient Detection

The transient (or onset) detection algorithm in this implementation uses a differential envelope of the signal. This means the delta between two instances of the envelope follower described in Section 4.3.1 is calculated for each sample. Using different attack- c_a, but the same hold H and release c_r coefficients yields differing results — especially during the rise time — for both envelope followers. The magnitude of this difference can

Algorithm 4.1 Envelope Follower

```
procedure TRACEENVELOPE($x, c_a, c_r, H$)
    $e \leftarrow 0$
    $h \leftarrow 0$
    while $x$ do                                    ▷ Iterate over samples of $x$
        $x \leftarrow |x|$                          ▷ Get current samples absolute value
        if $x > e$ then
            $e \leftarrow c_a(e - x) + x$           ▷ One-pole filter for attack
            $h \leftarrow 0$
        else
            if $h \geq H$ then                      ▷ Hold count reached, release in progress
                $e \leftarrow c_r(e - x) + x$       ▷ One-pole filter for release
            else                                    ▷ In hold phase
                $h \leftarrow h + 1$                ▷ Increment hold count
        return $e$                                  ▷ Return envelope value for current sample
```

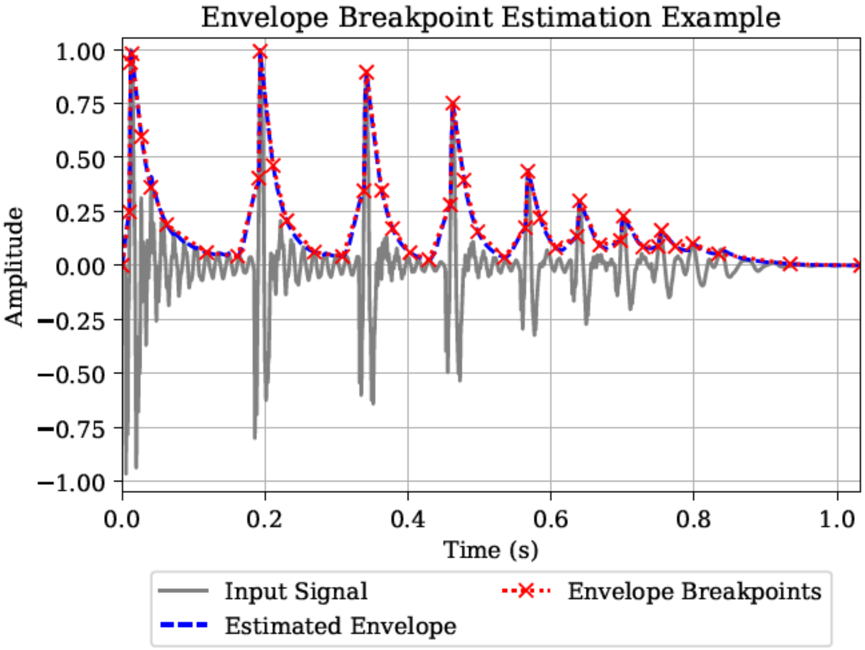


Figure 4.4: An exemplary plot of the time domain signal (grey) and the output of the envelope follower (blue dashed). The plot further illustrates the result of the line simplification schema (a.k.a. break point approximation). The red crosses illustrate the data that is saved. The red dotted line shows a linear interpolation of said data points. We can observe, that the reduced set of data maintains a small error towards the original envelope data.

be calculated by the absolute delta $\Delta e = |e_f - e_s|$ between the fast e_f and the slow e_s envelope follower. The delta Δe is what we call the "transient score" in this embodiment. Further, a threshold value v_{min} is compared to the transient score, yielding a *binary* transient score — which is equivalent to a set of rectangular windows in the time domain for each onset that was detected.

Finally, to retrieve a single transient point $T[n]$ (time value and amplitude) for each transient detected this way we look for the peak value within the rectangular window defined by the binary transient score. A pseudo-code mock up of the transient detector implementation can be seen in Algorithm 4.2. To reduce the amount of false positives and irrelevant transient points a mask in the time and amplitude range is applied, so that points too close to each other or transients with a relatively low, imperceivable amplitude are dropped automatically.

Algorithm 4.2 Transient Detector

```
procedure DETECTTRANSIENTS($x, v_{min}, c_{a,f}, c_{a,s}, H, c_r$)
    while $x$ do                                          ▷ Iterate over samples of $x$
        $e_f \leftarrow$ TRACEENVELOPE($x, c_{a,f}, H, c_r$)      ▷ Calculate fast envelope
        $e_s \leftarrow$ TRACEENVELOPE($x, c_{a,s}, H, c_r$)      ▷ Calculate slow envelope
        $d_e \leftarrow |e_f - e_s|$                              ▷ Calculate transient score
        if $d_e \geq v_{min}$ then                          ▷ Derive binary score by threshold
            $b_e \leftarrow 1$
        else
            $b_e \leftarrow 0$
        return $b_e$                                      ▷ Return binary transient score
```

Filter bank

As the ability of the tactile sense to discriminate between frequencies is quite limited [55, 81, 31] a relatively coarse filter bank design is proposed to approximate the spectral envelope of the the source signal over time. The filter bank was constructed using a combination of second-order cascaded biquadratic band-pass filters [75]. Only the first and last filter of the filter bank were constructed using a low-pass and high-pass filter of the same biquadratic form, respectively. The amount of bands of the filter bank can be set by a parameter in the settings file for further experiments or as a variable for data reduction (i.e., quality adjustment).

The cut-off frequencies of the individual filters are derived by equally spacing them between the minimum and maximum frequency defined in the settings file. The default setting for this range is set from 30 to 1000 Hz. The center frequencies of the band-pass filters and the cut-off frequencies of the low-pass and high-pass filters are log-spaced by default, while maintaining an overlap close to the -6 dB mark inspired by a Linkwitz–Riley

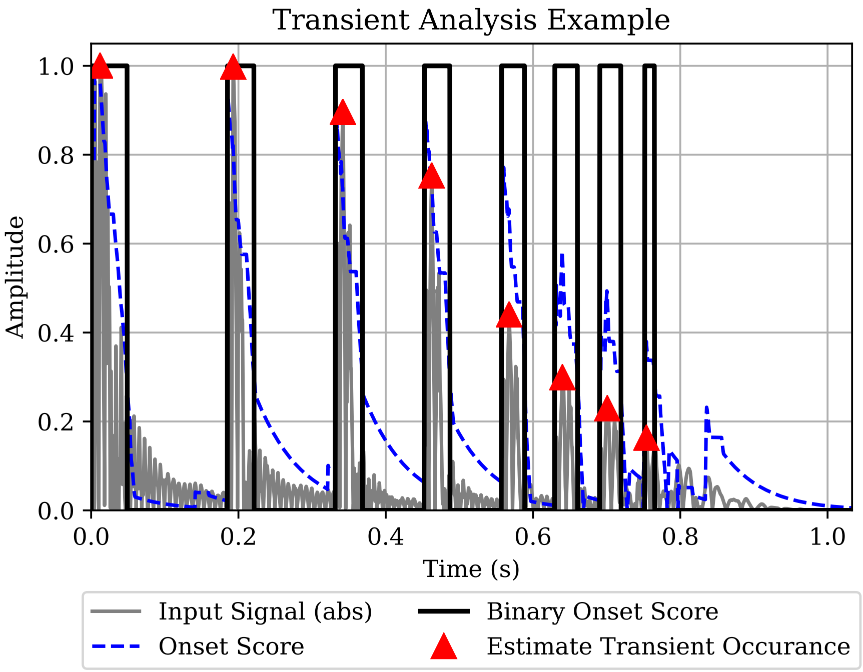


Figure 4.5: A plot illustrating the absolute value of an exemplary input signal (grey). The onset score derived from the delta between the slow and fast envelope followers (dashed blue) and the thresholded binary score (black). The final position and intensity of the estimated transients using the peak value within each binary frame are marked with the red triangles.

crossover design. Other filter arrangements, such as a linear layout can also be defined. A frequency response of a the proposed filter bank design using six bands, as well as the sum of all bands is illustrated in Figure 4.6.

As discussed in Section 3.1 most tactile actuators on the market are not designed to reproduce a complex or wide-band frequency spectrum, but instead are optimized to operate in a narrow frequency range due to increased power and cost efficiency. Depending on a vibrotactile displays capabilities an actuator may not be able to re-synthesize the entire estimate frequency spectrum derived from the filter bank, but could use a subset of all filter bands or a custom frequency mapping instead. For very narrow band actuators such as a LRA a single band, either closest to the resonance frequency of the actuator, or the band declared to be the most important by the tactile designer could be used for a degraded but approximate resynthesis.

Spectral Centroid

The capabilities of most tactile actuators on the market are quite limited to a narrow frequency range. To operate in this range to their best potential additional frequency tracking or spectral approximation can be conducted. In a previous work the spectral

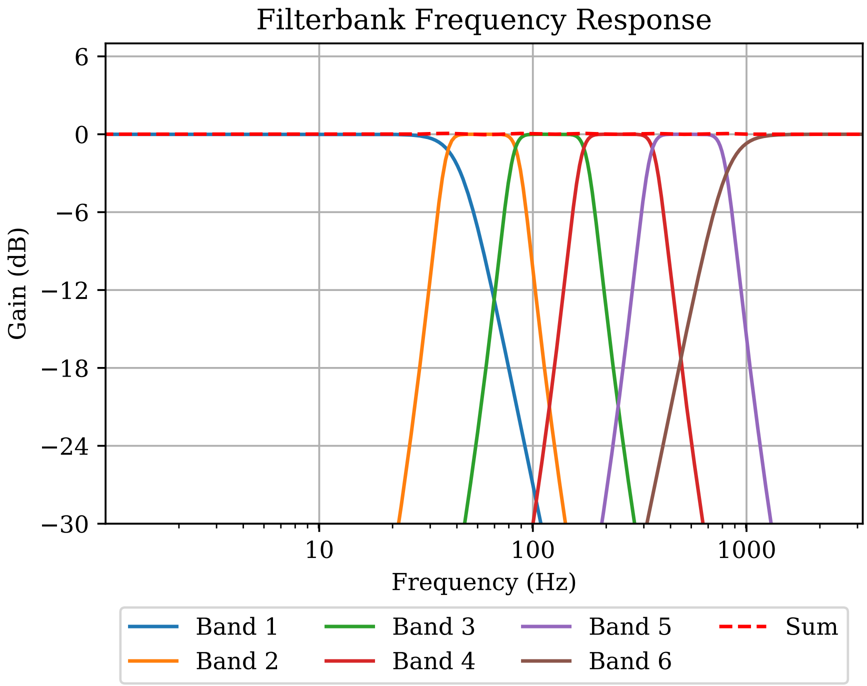


Figure 4.6: This plot shows the frequency response of an exemplary six band filter bank design, spanning a frequency range from 30 to 1000 Hz. Further, the sum of all bands is illustrated showing an approximately flat response with a maximum deviation of 0.07 dB gain and -0.03 dB attenuation.

centroid was utilized as an approximate measure for the brightness of a musical instrument signal, which in turn controller the pitch of a vibrotactile actuator embedded in the instrument [6].

Generally, the spectral centroid is a measure of the central tendency of a spectrum. In this implementation, the central centroid is estimated using the center of gravity of the spectral energy, which is performed on a segmented (i.e., block-wise) STFT on the time domain (audio) signal, as described in Equation 4.8. It is defined as the ratio between the frequency-weighted sum of the power spectrum and its unweighted sum [50]. Perceptually, this spectral feature has indicated a connection with the brightness of a sound [32, 89]. In this work the centroid is first calculated and as a measure of data reduction irrelevant and redundant values of the temporal trajectory are removed using a line simplification algorithm [104]. An exemplary, linearly interpolated temporal trajectory of such values is illustrated in Figure 4.7.

$$v_{SC}(n) = \frac{\sum_{k=0}^{\kappa/2-1} k \cdot |X(k,n)|^2}{\sum_{k=0}^{\kappa/2-1} |X(k,n)|^2} \tag{4.8}$$

To limit the output of the spectral centroid v_{SC} to a reasonable range, we can preprocess the audio source signal by applying a low-pass filter to restrict the upper limit of the

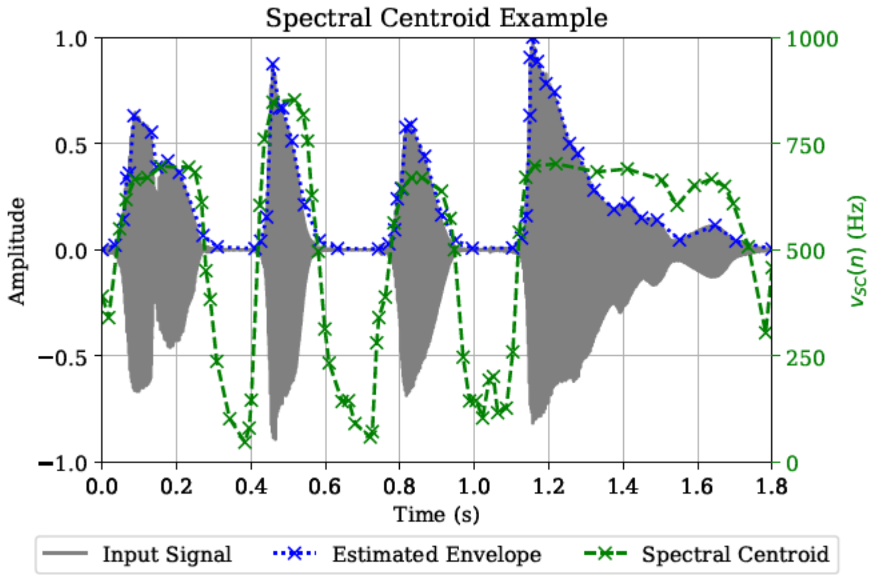


Figure 4.7: An exemplary plot of the spectral centroid on a recording of a clarinet playing a sequence of four distinct note pitches. The four notes can be observed as distinct changes in the values of the spectral centroid v_{SC} (dashed). The signals envelope, using the envelope follower is also illustrated.

spectral centroid. For this embodiment the upper limit of the human tactile sensitivity (see Section 2.3) at 1000 Hz was chosen. Further, we can scale the resulting range of the spectral centroid v_{SC} to any arbitrary range using equation 4.9 with i_{max} and i_{min} being the upper and lower limit of the source input range respectively, and equivalently with o_{max} and o_{min} for the upper and lower limit of the target output range. Scaling can be useful when a narrow band actuator is used for a tactile application. A narrow band actuator could use it's optimal range o_{min} and o_{max} to it's best potential, even though the original range captured by the spectral centroid i_{min} and i_{max} would have originally exceeded the actuators capacity.

$$scale(x) = \frac{(x - i_{min}) \cdot (o_{max} - o_{min})}{(i_{max} - i_{min}) + o_{min}} \tag{4.9}$$

4.3.2 Encoding: Data Model and Compression

When operating at a common sampling rate for audio (f.e., 44.1 kHz), deriving multiple envelopes N_B from the filter bank along the full length of the input signal initially scales the amount of information up and increases the amount data that needs to be stored. This is not desirable, as ideally the amount of data should be reduced to guarantee an efficient codec schema. To achieve a compressed representation of the envelope trajectories, a line simplification algorithm commonly used in cartography or GPS data storage was implemented [16, 104]. The line simplification algorithm (a.k.a. breakpoint

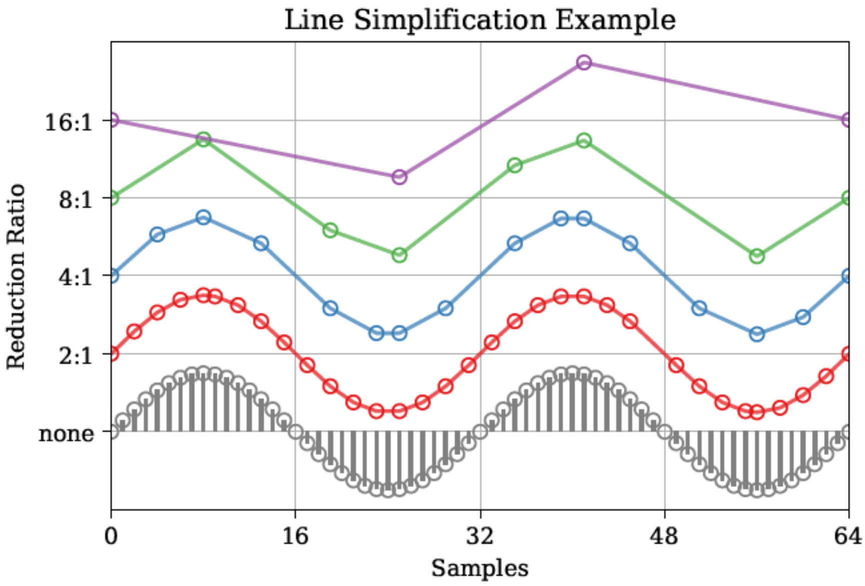


Figure 4.8: This figure provides an overview over various levels of data reduction using the implemented line simplification algorithm [16, 104]. Note that the algorithm is generally applied to the signal envelope and other higher level features derived from the analysis and not the signal itself. The sine wave shown here was only meant as a demonstrative example.

approximation) allows to define a target number, ratio or error margin to influence the resulting amount of breakpoints (i.e., time-value pairs). These breakpoints can then be re-interpolated using linear or cubic interpolation before resynthesis to restore the approximate shape of the temporal trajectory of a parameter value changes derived from the analysis process. For this embodiment a linear interpolation was chosen. The target amount of breakpoints per second N_P was set to 40, but can be scaled up or down depending on bandwidth and reproduction quality. Results of the line simplification method, as well as the re-interpolated trajectories can be seen in the exemplary Figure 4.8. Further plot illustrating practical use cases for envelope tracing and for simplifying a spectral centroid trajectory can be seen in Figures 4.4 and 4.7 respectively.

A theoretical estimation on the total compression of the frameworks codec can be achieved by calculating the ratio between the amount of information (per second) initially sent into the encoder and the resulting information from the output. First, the amount of Bit per second used by an exemplary source WAVE file can be calculated by multiplying the sampling rate F_s by the Bit depth Q_s and amount of channels N_C for each sample as seen in Equation 4.10. To portrait a more realistic comparison the size of a pre-processed signal S_{ds} is additionally used, as a lot of information is discarded initially by band-limiting and downsampling the input signal to a target sampling rate of 3 kHz. This step ensures that all important information within the tactile sensitivity range up to 1 kHz is maintained. Additionally, from experiments using office colleagues, a difference between a 12 or 16 Bit depth on the quantization of a signal sample could not be perceived by the tactile sense, while a stronger decrease of the Bit depth became noticeable.

$$\begin{aligned} S_x &= F_s \cdot Q_s \cdot N_C \\ S_{in} &= 44100 \cdot 16 \cdot 1 \quad = 705600\, b/s = 88200\, B/s \\ S_{ds} &= 3000 \cdot 12 \cdot 1 \qquad = 36000\, b/s = 4500\, B/s \end{aligned} \tag{4.10}$$

To estimate the amount of resulting data after the encoding schema, we need the amount of bands N_B used in the filter bank, the average amount of break points per second N_P the quantization Bit depth Q_s for the break points and the average amount of transients per second N_T. For the example calculated below, we assume an average of 2 transients per second. Each Transient, as well as each envelope break point is composed of two floating point values describing the temporal position (timing) of the and the value at that point in time. This means we have to multiply all points by a factor of 2 (Eq. 4.11).

As the parameters N_B, N_P, Q_s, N_T are adjustable in the compression routine of the A2VT framework the strength of the compression is adjustable at the expense of the resulting signal quality. The values given in Equation 4.11 are derived from the settings used in the experiments of this work. Further studies using a AB/X routine to find the strongest imperceivable compression or JND are planned in future experiments.

$$\begin{aligned} S_{\hat{x}} &= (N_B \cdot N_P \cdot Q_s + N_T \cdot Q_s) \cdot 2 \\ S_{\hat{x}} &= (6 \cdot 20 \cdot 12 + 2 \cdot 12) \cdot 2 \qquad = 5808\, b/s = 726\, B/s \end{aligned} \tag{4.11}$$

Finally, we can calculate the compression ratio C towards the input signal S_{in}, the more realistic down-sampled signal S_{ds} and the encoded signal S_{out}.

$$\begin{aligned} C_{in,\hat{x}} &= \frac{S_{\hat{x}}}{S_{in}} \quad = \frac{726\, B/s}{88200\, B/s} \approx 0{,}0082 \quad \approx 0{,}823\% \\ C_{ds,\hat{x}} &= \frac{S_{\hat{x}}}{S_{ds}} \quad = \frac{726\, B/s}{4500\, B/s} \approx 0{,}1613 \quad \approx 16{,}13\% \end{aligned} \tag{4.12}$$

It is important to note here, that we assumed an average of 2 transients per second for this estimation. The true number of transients per second may vary between signals. Each detected transient would sum up to another $24\, b$ of information that are required to be saved or transmitted.

4.3.3 Decoding: Re-synthesizing a Signal from Data

As soon as a tactile event is triggered, either through an interaction by the user or by a virtual event in the application, all the information contained in the tactile data model is passed to the synthesis (i.e., decoding) engine. After parsing the data and retrieving both the parameter break points and the transient information from the data container, the full envelopes $a_m[n]$ are reconstructed from the break points using linear interpolation at the target DSP sampling rate f_s. Next, a set of N_B sinusoidal is generated for each envelope $a_m[n]$, using the accompanying center frequency f_m of the corresponding band. Alternatively, the frequencies can be modulated using the values derived from the spectral centroid analysis $v_{SC}[n]$ and a desired harmonic relationship for each band. The initial phase of the sinusoidal φ_m can be set if desired but hasn't made a strongly perceptually noticeable difference in subjective tests. The resulting, full scale sinusoidal is then scaled using the current time value from the interpolated envelope.

$$\hat{x}[n] = T[n] + \sum_{m=1}^{N_B} a_m[n] \cdot sin(2\pi \frac{f_m}{f_s} \cdot n + \varphi_m) \tag{4.13}$$

Additionally, the transients $T[n]$ are synthesized in parallel. The transient synthesis positions a predefined waveform for each point in time defined by the transient time values and scales it with the accompanying amplitude value. For the implementation of this work a single cycle of a square wave at the resonance frequency of the target actuator was used. This is proposed because at it's resonance frequency the actuator is more efficient and accelerates faster. The square wave is proposed to push more energy into the system in a shorter time span, due to the harmonic content of this waveform. In an application the information on the resonance frequency of the actuator would ideally be stored in an actuator specific configuration file, as proposed the flowchart displayed in Figure 4.1. The actuator specification is derived from a set of actuator profiling tests including the measurements described in Section 3.2. The ad hoc actuator information is important due to the variability in actuators described in Section 3.1.3 and is assumed to be a key component in achieving a good stimuli reproduction.

Creating ad hoc Variations of a Stimulus Signal

As most parameter values of the proposed format can be manipulated during authoring, storage or during re-synthesis of the signal it, is possible to create ad hoc variations of a stimuli. To achieve this, either each break point amplitude value or the full envelope $a_m[n]$ of a corresponding sinusoidal can be scaled by a random value $p(x)$ drawn from a uniform distribution $F \in [0.0, v_{max}] \cup [0.0, 1.0]$.

The desired amount of the resulting variation can be controlled by the upper limit v_{max} of the uniform distribution range, resulting in the random variable $p(x)$ with $p(x) \sim F$ as illustrated in Equation 4.14. By applying this operation to each break point amplitude value before interpolation, we can add variety to a synthesized stimulus by slightly scaling the amplitude values for each band and the resulting temporal trajectory. Applying the scalar to the entire envelope vector after the break point interpolation will create variance in amplitude for each individual frequency band. This process imitates the modal behaviour modeled from physical object recordings [51].

$$\hat{a}_m[n] = a_m[n] \cdot |p(x, v_{max})| \tag{4.14}$$

This process was inspired from an implementation done for video games by adding variance to each footstep or impact event: Each surface has different properties that need to be reflected in the tactile response. But not each footstep should be the same as this would be perceived as unnatural. In sound design practice, it is common to use a semi-random or round robin process to pick a footstep sound out a set of variations that fit the current event — such as a footstep on a specific surface. Using the method illustrated above requires only one media asset to be saved, which effectively reduces the amount of storage required [51]. As this process is proposed to be done during decoding (i.e., resynthesis) it can be done on the recipient side and therefor decreases the amount of bandwidth required when streaming data to the vibrotactile device, such as a peripheral hardware controller.

5 Perceptual Evaluation

To validate the audio-tactile translation method proposed in this work and more broadly the use of audio sources to synthesize tactile stimuli, a perceptual evaluation was conducted. The validation had to be of a perceptual manner, as the relationship between measurable physical properties of a stimulus signal, and the perceived characteristics rely on the respective sensory (auditory and tactile) modalities, which cannot be accessed in a direct manner. To otherwise approximate the perceived quality and cognitive representation of a stimulus, a *standardized* perceptual model would ideally be utilized. Since no such standard model exists to date, an empirical evaluation using voluntary participants was chosen. This chapter describes the entirety of the perceptual evaluation process and discusses the results.

5.1 Method

To perform the perceptual evaluation of the audio-tactile signal translation process, a set of materials and processes is required. The stimuli sources, the testing procedure and the hardware required for the application of the stimuli are documented in the sections below. The empirical test contained two tasks made up of both a qualitative, linguistic evaluation on verbal descriptions and a numeric evaluation using stimulus coherence ratings on an unipolar scale. The resulting data from both tasks, as well as information on the participants, was examined to assess the perceived *coherence* between auditory and tactile stimuli pairs and to validate the audio-tactile translation process.

5.1.1 Participants

Accompanying the perceptual test was a questionnaire to collect both demographic and musicality information of the participants (see Appendix A.3). A total of 33 subjects participated in the user test (5 females, 28 males; average age = 27 years, SD = 2.71 years; range = 23 - 35). All participants reported to be German residents and fluent in the German language, which was important for the psycholinguistic analysis. Three participants reported to be of a non-German speaking heritage. From the participants, 82% were recruited from the M.Sc. program in Audio Communication and Technology at TU Berlin; 12% were colleagues from departments unrelated to audio and haptics; and 6%

were work colleagues, that have their profession in the haptics industry. All participants gave written informed consent in accordance with the Declaration of Helsinki.

To assess the musicality of the subjects, the German version of the "Goldsmiths Musical Sophistication Index" (Gold-MSI) questionnaire was used [66, 86]. From the participants, 55% reported to have at least 2 years of engagement in regular, daily practice of a musical instrument (average years = 3.9, SD = 3.3, range = 0 - 10+), while reportedly practicing an average of 1.2 hour per day on their primary instrument (SD = 0.6, range = 0 - 2). Further, the participants reported to have an average of 4 years of formal training on a musical instrument (SD = 3.04, range = 0 - 10+) and an average of 1.3 years of formal training in music theory outside of the regular school curriculum (SD = 1.8, range = 0 - 7+).

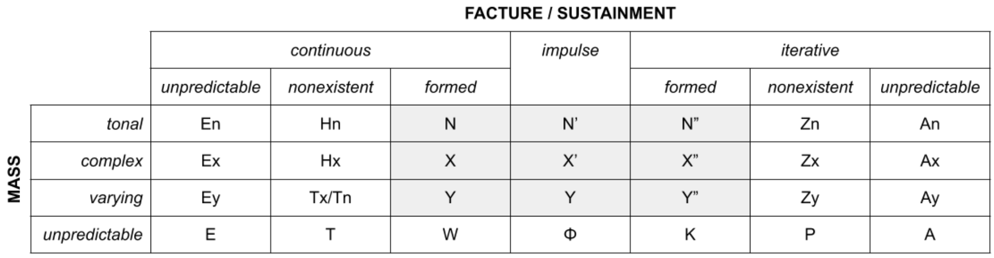
5.1.2 Audio Source Material and Tactile Stimuli Generation

Due to the huge variety of available audio material — ranging from natural recordings, to synthesized sounds and processed signals — it is an elaborate task to gain confidence, that an algorithm is able to generalize it's functionality along all possible types of sounds. For this reason, it was initially important to find a corpus of audio material that contains a variety of sounds. Previous works on audio-tactile translation had a strong focus on musical signals (see Section 1.3). For this work, the proposed inter-modal translation process was explored for more general, non-musical sound sources [61, 60]. A useful way of collecting such sounds is through electroacoustic music, where sound identities appear intentionally obscured or unconnected to their source. Researchers at the PRISM laboratory in Marseille have previously compiled a collection of 200 electroacoustic sounds representative of the nine balanced sound classes of Schaeffer's typology of *acousmatic* sounds — sounds experienced by attending to their intrinsic morphology and not to their physical cause [59]. These sounds are based on three profiles of temporal energy envelope (continuous, impulse, iterative) and three profiles of spectral content (tonal, complex pitch, varying pitch) [14]. This classification system offers an objective tool to obtain a sound corpus representative of most sound morphologies. An overview of Schaeffer's typology classes can be seen in Table 5.1.

Since the corpus of 200 sounds came without class labels, the author and second supervisor independently classified the sounds and achieved an agreement of 86% (172 of 200) — while the agreement on the classes was 100% for all sounds chosen for the procedure of the experiment. To ensure a reasonable duration for the user test, a subset of ten sounds were selected from the initial 200. This subset of sounds, representing various classes from Schaeffer's typology, is meant to give initial confidence in the translation method to generalize across a modest range of sounds, while maintaining the same parameter settings in the translation process. From the ten sounds, two sounds were attributed to the N'' class (iterative; definite pitch; formed iteration), one sound to the

X class (held; complex pitch; formed sustain), two sounds to the X'' class (iterative; complex pitch; formed iteration) and five sounds to the Y'' class (iterative; slight variable mass; formed iteration). The set of selected sounds had an average duration of 1.92 seconds (SD = 0.96, range = 0.8 - 3.4) and was made up of various content, such as artificial drone sounds, rich textures, impacts, plucking sounds, natural recordings and iterative processes. Spectrograms of the chosen sounds can be found in the appendix in Figures A.1, A.2, A.3 and A.4.

The audio material was not altered from the original corpus source and showed deviations in loudness according to a standardized loudness measurement [44, 93]. While 7 out of the 10 sounds measured -20 LUFS (±3) three of the sounds are considered outliers, being either louder or quieter at -30, -13, -15 LUFS. When normalizing the audio sources to the same LUFS the loud sounds (at -13 and -15) would be so strongly attenuated, that signal details were lost due to being too quite. From subjective tests, all sounds were clearly audible and no participant reported on a stimulus to be too quite or to loud, consequently no countermeasures were conducted. It is further reasoned that normalizing loudness is more important in *comparative* listening tests, whereas in this embodiment the ratings and qualitative descriptions were given individually for each stimulus and no influence of loudness on these measures is expected to be present. To further reduce the influence of loudness levels between stimuli, the order of stimuli presentation was randomized.



		FACTURE / SUSTAINMENT						
		continuous			*impulse*	*iterative*		
		unpredictable	*nonexistent*	*formed*		*formed*	*nonexistent*	*unpredictable*
MASS	*tonal*	En	Hn	N	N'	N"	Zn	An
	complex	Ex	Hx	X	X'	X"	Zx	Ax
	varying	Ey	Tx/Tn	Y	Y	Y"	Zy	Ay
	unpredictable	E	T	W	Φ	K	P	A

Table 5.1: Schaeffer's typology of sound objects. The highlighted cells illustrate the desired types used for this study. The table is derived from *Des Objets Sonores* [14].

The tactile stimuli signals were rendered from the audio sources using the audio-tactile framework described in Chapter 4. The analysis and resynthesis (i.e., codec) parameters were kept the same for each audio sample and respective tactile signal resynthesis. The exact set of parameter settings used in the codec process is included in the attached digital media in (`a2vt_main/settings.json`). The tactile stimuli were rendered to a WAVE file and both audio and tactile WAVE files were embedded in the user test file structure for automated playback during the test procedure.

5.1.3 Procedure

To validate the coherence of the sensations derived from the inter-modal audio-tactile translation method an empirical user test in three groups (A, B and C), yielding direct and indirect, qualitative coherence measures was conducted. For this experimental evaluation a total of 11 subjects per group, resulting in a total set of 33 individual test subjects participated. A total of 10 audio-tactile stimuli pairs were selected from the aforementioned corpus (section 5.1.2). The general test procedure for group A was conducted in two separate tasks, as follows:

1. The participants were asked to listen to the (10) sound stimuli consecutively and instructed to write down the first words that come to mind for each stimuli, i.e., freely describing their subjective perception and cognition of each stimulus. The participants were specifically asked to focus on the associations (i.e., concepts) evoked by the sounds without trying to identify the physical sources or events that produced them.
   The exact phrase used was: *"Bitte beschreibe deinen sensorischen Eindruck, ohne auf den Ursprung des Stimulus einzugehen. Versuche dabei abstrakte Merkmale, die den Ablauf und Inhalt beschreiben, zu verwenden."* (English: *"Please describe your sensory impression, without trying to identify the origin. Try to use abstract features, that describe the process and content."*
2. The participants were presented audio and vibrotactile stimuli consecutively (unimodal) and were asked to rate the "coherence" between both stimuli on an unipolar, continuous scale from 0 to 10.
   The exact phrase used was: *"Bitte bewerte die Kohärenz bzw. Stimmigkeit der aufeinanderfolgenden Stimuli zueinander. Die Skala dafür geht von 0 (schlecht) bis 10 (sehr gut)."* (English: *"Please assess the coherence, or consistency of the consecutive stimuli to each other. The scale for this goes from 0 (bad) to 10 (very good)."*

Group B carried out the same two tasks, but task 1 instead involved the corresponding vibrotactile stimuli.

Finally, participants in group C were first presented with audio-tactile stimuli pairs simultaneously (bimodal) and were asked to freely describe the combined sensation (i.e., similar to the first task of groups A and B), and secondly asked to rate the perceived coherence on a unimodal continuous scale on the same stimuli pairs.

Coherence here is meant to describe the level at which temporal and spectral (i.e., timbral) characteristics of the original (audio) signal are maintained throughout the signal translation process — while providing a comparable sensation across both modalities that we aim to measure.

The entire perceptual test procedure was automated using a Python implementation, which allowed the participants to replay the current stimulus any number of times, and then asked for a qualitative, verbal description of the sensory impression (i.e., verbal associations) or the coherence rating respectively, as described above. The group assignment was done semi-random by using a pre-populated set of options (11 subjects for each group) to make sure each group will be equally represented. Also, the sequence of the ten presented stimuli, both for the verbal descriptions and the ratings, was randomized for each participant to avoid any bias induced by the order.

Test Setup and Hardware

The user test was conducted in a small, quiet room at the Technical University Berlin[1] to allow for an undisturbed trial. The subjects were advised to position themselves in a seat in front of a desk with a screen and a keyboard. The acoustic stimuli were applied using Beyerdynamics DT 990 Pro headphones. The headphones were measured to output a mean value of 82 dBSPL across all ten sound stimuli (SD = 5.97 dB) using a miniDSP EARS headphone jig[2] and the Room EQ Wizward acoustic measurement software[3]. The measurement routine was calibrated using the default calibration method as advised by the user manual. This procedure required the microphone calibration files, that could be assessed by using the serial number of the EARS headphone jig. The tactile stimuli were applied using the modified wristband (see Section 3.2.3). Each participant was asked to tighten the wristband strap to their own, subjective comfort while making sure the armband is not too lose, but maintains good skin contact. This procedure is thought to mirror a realistic scenario of a wristband fit in daily use.

Although the L5 actuator used in the wristband is nearly inaudibly silent it can come to occasional auditory artifacts due to the physiological variability introduced by the participants. This can lead to an accidental, short overdrive of the actuator due to a lack of force countering the actuators acceleration (see Section 3.1.4), which can result in the moving mass hitting the actuator housing. A white noise stimulus intended to mask such undesired, auditory cues was therefore played back during the unimodal, tactile stimuli display. Applying such a auditory masking stimulus is a practice that has been previously reported in tactile perception experiments [4, 110] and in auditory-tactile experiments involving musical instruments to mask any undesired auditory cues [22]. Due to the overlapping, perceptual sensitivity regarding the frequency ranges for the auditory and tactile modalities, a tactile stimulus can still be heard, even with a completely silent actuator, due to structure born transmission such as bone conduction.

[1] Room H2001d MediaLab

[2] miniDSP EARS product page, last viewed 20th April 2020:
https://www.minidsp.com/products/acoustic-measurement/ears-headphone-jig

[3] Room EQ Wizard software, last viewed 20th April 2020:
https://www.roomeqwizard.com/

The sound level of the audio stimuli was set to a default value before starting the test. Using a set of test sounds (not included in the user test) the subjects were asked if they were comfortable with the perceived volume of the sound. No participant reported discomfort and denied the option to reduce the volume of the stimuli. The same procedure was conducted for the intensity of the tactile stimulus. After this initial comfort check the test program was started and the participant was left alone with the door to the room shut to further improve the ability to concentrate and to improve the acoustic isolation of the subject.

To display the vibrotactile stimuli to the subjects, a wristband form factor was used (see 3.2.3 and Figure 3.6). The location for applying the tactile stimuli was chosen to be the anterior of the underarm (similar to wearing a watch the wrong way around). This location allows for stimuli exposure to hairless (glabrous) skin and therefore stimulation of the underlying Meissner (RA-I) and Pacinian (RA-II) corpuscles [96].

Using a wearable device is beneficial compared to any form factor that requires active participation of the user, such as a touch display. This is due to the fact that interference from factors, such as the applied pressure, bodily position and exposed surface area for a tactile display are easier to control between test subjects. To further reduce the bodily interference on the tactile display between subjects, which is induced by non-linear actuator behaviour and the factors described above, the actuator was linearized as described in Section 3.2.4.

Data Analysis

After collecting all test data, we looked at qualitative comparisons across the three group conditions as a measure of *coherence* of the inter-modal audio-tactile sensations based on the proposed signal translation method. To analyse the collected qualitative attributes, a psycholinguistic approach was utilized [84]. A similar experiment for word-sound relations instead of tactile-sound relations has previously been reported and results were discussed in terms of similar conceptual processing networks [88]. The quantitative data based on the bipolar, discrete *coherence* scale underwent statistical analysis to determine the overall rating, to compare conditions and to find possible corner cases.

5.2 Results

This section presents the analytical results and discussion on the data gathered by the perceptual evaluation (i.e., user test) described above. Since the perceptual evaluation was composed of three groups, conducting two tasks respectively, the resulting analysis is split among the two tasks. First, a statistical analysis of the numeric coherence ratings was conducted, and secondly an exploratory psycholinguistic analysis was conducted on the qualitative, verbal descriptions.

5.2.1 Coherence Ratings

The coherence rating test yielded a total of 330 individual ratings from 3 groups of 11 participants, each rating 10 stimulus pairs. The raw coherence rating data can be found in the appendix in Table A.1. The mean and standard deviation for each stimulus pair and group can be seen in Table 5.2. Further, the distribution of the individual stimuli ratings for each group are illustrated in Figure 5.1. A series of independent two-sample t-tests were conducted to compare individual sample ratings across the test conditions (i.e., groups). A series of two-sample F-tests across all stimulus-group combinations showed that all stimulus ratings were found to be of equal variance. The results of these tests can be found in Table A.2, which is found in the appendix. Significant results were discovered for stimulus pairs with the ID 1, 9 and 10.

The stimulus pair with ID 9 resulted in mean ratings of 5.18 and 8.18 between groups A and C respectively [independent samples t(20) = -3.19, p = .005, equal variance]. Between the same conditions (A and C) stimulus ID 1 also resulted in a significant difference in the mean of the ratings with 5.27 and 7.18 respectively [independent samples t(20) = -2.837, p = .01, equal variance]. Equivalent observations were made between the conditions B and C for stimulus pair ID 9, resulting in mean ratings of 6.36 and 8.18 respectively [independent samples t(20) = -2.104, p = .048, equal variance]. Under the same conditions, stimulus pair ID 10 resulted in a significant difference between the mean ratings of 6.55 and 8.82 [independent samples t(20) = -3.662, p = .002, equal variance]. Finally, a significant difference for the stimulus pairs with ID 9 and 10 in the ratings for the constructed group condition {A+B} (unimodal) when compared to C (bimodal) were found. The mean rating for ID 9 resulted in 5.77 for {A+B} and 8.18 for C [independent samples t(31) = -2.97, p = .006, equal variance]. Similarly, for stimulus pair with ID 10 we a significant difference between 7.05 and 8.82 for the conditions {A+B} and C was found [independent samples t(31) = -3.09, p = .004, equal variance].

Since the coherence rating task was equal for both groups A and B, both exposed to an unimodal, consecutive stimuli display, a combined group, formed by the union of both groups {A+B} was constructed. As expected, group A and group B undergoing the same, unimodal, consecutive stimuli presentation, showed no significant difference in the coherence rating [independent samples t(18) = -0.78, p = .45, equal variance].

A two-sample F-test across all group combinations showed that all groups were found to be of equal variance. The coherence rating was significantly lower in group A than in group C with average value = 6.69 and 7.75 respectively [independent samples t(18) = -2.55, p = .02, equal variance]. A similar, significant observation was made between groups B and group C with average value = 6.98 and 7.75 respectively [independent samples t(18) = -2.35, p = .03, equal variance]. Consequently, the coherence ratings were also lower for the union of groups {A+B} compared to group C with average value = 6.84 and 7.75 respectively [independent samples t(18) = -2.62, p = .02, equal variance]. The

ID	Group A		Group B		Group C		All Groups	
	Mean μ	SD σ	Mean μ	SD σ	Mean μ	SD σ	Mean μ	SD σ
1	5.27	1.95	6.45	2.02	7.18	1.08	6.30	1.08
2	5.55	2.42	7.00	2.45	6.09	1.92	6.21	1.92
3	7.27	2.00	7.64	2.54	8.45	1.69	7.79	1.69
4	7.09	1.51	6.27	2.15	7.55	1.21	6.97	1.21
5	8.09	1.87	8.09	2.07	8.73	1.19	8.30	1.19
6	7.18	1.89	7.55	1.86	7.64	1.12	7.45	1.12
7	7.09	1.45	6.82	2.40	7.36	1.29	7.09	1.29
8	6.64	1.86	7.09	1.45	7.45	1.75	7.06	1.75
9	5.18	2.44	6.36	2.11	8.18	1.94	6.58	1.94
10	7.55	1.63	6.55	1.57	8.82	1.33	7.64	1.33
Total	6.69	2.08	6.98	2.08	7.75	1.63	7.14	1.99

Table 5.2: Statistical results for the coherence ratings for each test group and each stimulus pair respectively. The first column (ID) displays the stimulus pair ID used throughout this work. Both groups A and B were presented with the audio and tactile stimulus pairs consecutively (unimodal). Group C was exposed to both audio and tactile stimulus pairs simultaneously (bimodal).

Samples	t	DF	SD	p
A vs. B	-0.78	18	0.84	0.45
A vs. C	-2.55	18	0.92	**0.02**
B vs. C	-2.35	18	0.73	**0.03**
{A+B} vs. C	-2.62	18	0.78	**0.02**

Table 5.3: This table displays the results of a two-sample (independent) t-test, measuring the two group samples shown in the first column. Bold p-values indicate if the level of significance ($p \leq .05$) has been met to reject the H_0 hypothesis.

results of an independent two-sample t-test between the groups in a tabular form can be seen in Table 5.3. Overall the results show, that a simultaneous bimodal stimuli display achieved a significantly higher coherence rating compared to the consecutive unimodal stimuli display.

5.2.2 Verbal Descriptions

The dataset collected through the first, free verbalization task of the user test (see Section 5.1.3) is composed of 330 individual, verbal descriptions. Spelling mistakes were manually corrected to ensure that finding similar word stems during analysis and during the scripted, exploratory analysis using the Natural Language Toolkit (NLTK) was less prone to errors and mismatches. A sum of 1120 individual concepts were identified among the 1618 total words (including stop words). Removing stop words by using the default German stop word lexicon of the NLTK resulted in a set of 1229 words. It is important

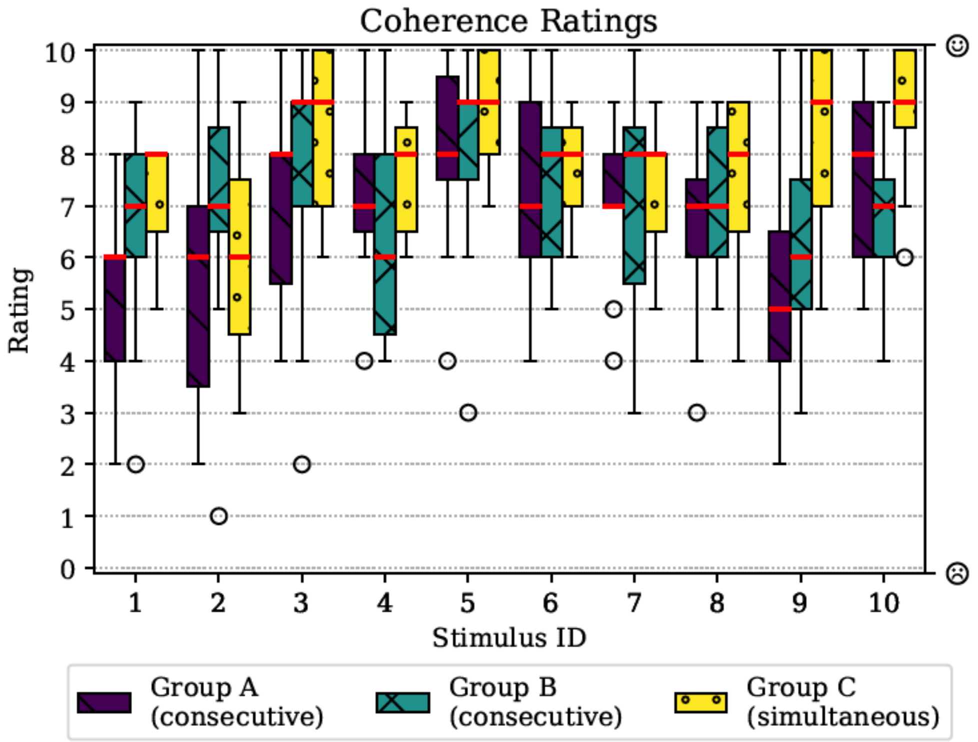


Figure 5.1: This plot illustrates the coherence ratings for the unimodal (consecutive) stimuli representations of groups A and B, and the bimodal (simultaneous) stimuli representation of group C. On average the bimodal stimuli representation (group C) achieved a significantly higher mean score, as can be seen in Table 5.2. The median rating for each individual stimulus ID and group conditions is illustrated with a red dash. Outliers in the ratings are illustrated using circles.

to note that the analysis described below was done manually. A bigger lexicon or sample size would be required to get reliable results from an automated analysis using the NLTK. The NLTK was therefore only used for pre-processing of the verbal descriptions and to gather an estimate on the total vocabulary (i.e., word stems) as described above. The processed, verbal descriptions (i.e., extracted microconcepts) for each stimulus of groups A, B and C can be found in the appendix in Tables A.3, A.4 and A.5 respectively.

To analyse the qualitative descriptions an exploratory, psycholinguistic analysis was conducted using the so-called constant comparison method, which in turn is based on grounded theory [94]. This method has previously been utilized in a perceptual evaluation of verbal descriptions used to describe violin quality by experienced musicians [84]. The goal of the exploratory analysis was to first identify emerging key concepts and semantic categories, and to then find indications of varying distributions of said semantic groups among the three test conditions (A, unimodal, auditory; B, unimodal, tactile; C, bimodal, auditory and tactile). Similarities and differences between the groups might indicate the relevance of certain perceptual aspects and can inform the DSP methods used in the translation process (see Section 4). Further, the analysis might infer abstract, perceptual similarities and differences between the auditory and tactile senses and thus support and inform the notion of joint audio-tactile experience design due to the early integration of both senses.

The constant comparison method is composed of a set of, not necessarily sequential, data coding steps: *open coding*, used to identify key concepts; *axial coding*, used to link concepts based on semantic proximities and thus forming semantic categories and intercategorical associations; *theoretical sampling* and *selective coding*, used to map new data with the emerging conceptual framework in mind and to potentially improve it; and *theoretical saturation*, which concludes the coding steps when no further categories or concepts emerge even when presented with new data.

First, word groups indicating a similar, descriptive quality (i.e., key concepts) were identified when analysing the data of the first group (A). These concepts were then classified amongst emerging semantic categories (open coding, axial coding). Next, the descriptions from the second group (B) were examined (theoretical sampling) and new concepts were identified. Consequently, the emerging categorical framework was updated (selective coding). The coding routine was extended by examining the descriptions from the third group (C), using the existing framework (theoretical sampling) and was consequently concluded, as no further concepts emerged and theoretical saturation had been achieved.

From the emerging semantic categories various cognitive objects of reference (i.e., What has been described?) were identified. There were three primary, distinct objects that were identified from the corpus of descriptions, namely: properties of and the *stimulus* itself; the physical *source* or events that produced the stimulus; and the *subject*, describing evoked, subjective feelings and sensations. Even if the participants had been advised to answer in full sentences, many of them confined their answers to short, incomplete sentences and keywords. This occasionally made it hard to evaluate what the object of reference was. If the interpretation of a description was impossible, the data point was omitted. Omitting data was only necessary for 4 out of the 330 unique descriptions. Most of the time, the conjugations and (German) grammatical cases gave a clear indication on the object of reference, even when analysing a sequence of seemingly disconnected words. Full sentences were easier to code, as they indicated clearer answers on *what* and *how* the evoked percepts were described. Concepts and their antonyms were group together, as the direction for each semantic dimension didn't seem to be important for this experiment.

The emerging semantic categories of the underlying verbal corpus used to describe perceptual attributes of the presented stimuli can be summarized as temporal change (e.g., short, rising, fast), temporal structure (e.g., rhythm, pulsing, chaotic), spectral property (e.g., low, hollow, bright), onomatopoeia (e.g., hum, crackle, snap), intensity (e.g., strong, big, soft), origin (e.g., machine, ball, drum), source movement (e.g., vibrate, fall, jump), and affective reaction (e.g., pleasant, stimulant, tense). Note that the raw data is composed of German descriptions and the examples given for each group underwent a translation to give the (non German speaking) reader an approximate idea for each category. Each semantic category primarily described a distinct object of reference, thus temporal change, temporal structure, intensity and spectral property were attributed to the *stimulus*. The categories origin and movement, which primarily describe the physical source and event of a stimulus, were attributed to the *source* as an object of reference. Onomatopoeic concepts were used both for describing the *source* and the *stimulus*. As the descriptions of the affective reaction neither describe the stimulus itself, nor the physical source, this category was attributed to the *subject*. The emerging semantic categories and concepts are shown in Tables 5.4, 5.5 and 5.6. The raw, unprocessed data and the helper scripts are contained in the accompanying, digital medium.

Evidently, the amount of microconcepts extracted from each group varies, with 427 microconcepts extracted from the descriptions of group C, followed by group B with 385, and group C with 308 microconcepts; summing up to a total of 1120 microconcepts extracted across all groups. To compare the frequency of occurrence of the semantic concepts between groups, the frequency is expressed in percentage for each group. Additionally, the total frequency of occurrence of each category across all groups is also expressed in percentage. The occurrences for each group, and the total calculated this way are illustrated in the bar graph of Figure 5.2.

Semantic category	Microconcepts (Group A)	Object of reference
Temporal Change (TC)	moduliert (3); oszillierend (2); kurz; abflachen; nachlassen; nachhallen; hallig; rising;	stimulus
Temporal Structure (TS)	Rhythmus, rhythmisch (7); perkussiv (7); Impuls; impulsiv; chaotisch; sequenziell; wechselnd; variant;	stimulus
Intensity (In)	laut (4); groß (2); viel; hart; wenig; leicht; klein; dynamisch;	stimulus
Spectral Property (SP)	rauschen (8); scharf (8); tief (6); Bass (7); dumpf (5); organisch (4); warm (4); sphärisch (4); rau (3); metallisch (3); verzerrt (2); blechern (2); synthetisch (2); hohl (2); hell (2); tonal (2); dissonant (2); glitchy (2); flach (2); melodisch (2); kalt; verstimmt; maschinell; spitz; hölzern; dunkel; dark; hoch; breites Spektrum; noisy; brennen; diffus; monoton; gefiltert; harmonisch;	stimulus
Onomatopoeia (On)	klappern (2); knacken (2); knistern (2); rieseln (2); klirren; brubbeln; knacken; brummen; wobbeln; quietschen; klicken; wummern; flimmern; dröhnen; röhrend; surren; chirpen; boing;	stimulus, source
Origin (Or)	Tischtennis (3); Basketball (3); Ball (3); drum (3); Club (3); Techno (3); Maschine (2); (Gitarren-)Saite (3); Motorrad (2); Motor (2); engine (2); Kiste (2); Sport (2); (Gummi-)Band (2); Herzschlag; Nadeln; Bienensummen; Wasser; synthesizer; Stift; Ping Pong; Musikinstrument; Kettensäge; Presslufthammer; Sprungfeder; Propeller; Raster; Blech; Schlagbohrer; Strom; Hi-Hat; Ride; Kugel; Tier; Nashorn; Werkzeug; Baustelle; Küchengerät;	source
Movement (Mo)	fallen (6); kratzen (4); rollen (4); schwanken (4); in Bewegung (3); schütteln (3); wühlen (2); schieben (2); aufräumen; abprallen; hüpfen; streifen; zupfen; springen; schaben; atmen; ziehen;	source
Affective Reaction (AR)	unangenehm (5); angenehm (3); lästig (2); spannend; verspielt; bedrohlich; stechend; lustig; entschleunigend; Unruhe; Angst; nervig; energetisch; aggresiv;	subject

Table 5.4: Emerging semantic categories from evoked perceptual descriptions for group A. Brackets indicate the amount of occurrences for each microconcept.

5.2.3 Discussion

The results of the coherence rating show a significant effect (higher rating) on the simultaneous, bimodal stimuli display when compared to the ratings conducted on the same stimulus pair, but in a consecutive, unimodal manner. Generally, the median rating given in the bimodal (i.e., simultaneous) test configuration was always higher, except for the stimulus pair with ID 2, which also scored the lowest average rating in total (see Figure 5.1 and Table 5.2). The highest average rating was scored by stimulus pair ID 5 with a total mean rating of 8.30 [SD = 1.19]. Overall, the higher rated, and therefore infered perceived coherence of simultaneous auditory-tactile stimuli display, supports the notion of a joint audio-tactile experience design. This notion is in agreement with previous studies showing early integration across the modalities [26, 23, 47, 113, 63]. It further supports the notion of using audio source material for the synthesis of tactile

Semantic category	Microconcepts (Group B)	Object of reference
Temporal Change (TC)	kurz (10); lang (9); schnell (7); [an-, auf-] steigen (6); schwingend (6); anschwellen (4); mittellange (3); abebben (3); Steigerung (3); abklingen (3); ausklingen (2); abfallen (2); stoppen (2); zunehmen (2); release (2); abrupt (2); vorbeiziehen (2); abnehmen; [stärker, tiefer] werdend (2); anhaltend; aufschaukelnd; Ende; Abbruch; spontan; plötzlich; einschwingen; ausschwingen; ausgedehnt; langsam;	stimulus
Temporal Structure (TS)	Pulse (9); Impuls (8); gleichmäßig (5); durchgehend (4); im Takt (4); instabil (4); wiederholend (4); pulsierend (3); Rhythmus (3); rhythmisch (3); pulsieren (3); Peaks (2); triolisch (2); schlagend (2); perkussiv (2); am Ende (2); konstant (2); gleichbleibend; statisch; periodisch; stabil; stolpernd; koordiniert; definiert; deutlich; zufällig; am Ende [ändert sich etwas]; sequenziell; Abfolge [von etwas]; chaotisch; überlegt; Stakkato; beständig; strukturiert;	stimulus
Intensity (In)	stark (19); schwach (6); leicht (6); klein (5); sanft (3); hohe [Intensität, Energie] (2); gering (2); kräftig (2); groß; hart; leise; laut; zaghaft; verhalten; bescheiden; bestimmt; prägnant; druckvoll; durchdringend; subtil; intensiv; zurückhaltend; [kaum] wahrnehmbar;	stimulus
Spectral Property (SP)	tieffrequent (6); tief (4); mittelfrequent (3); hochfrequent (2); hoch (2); natürlich (2); hell (2); rauschen (2); scharf; trocken; kalt; dunkel; robotisch;	stimulus
Onomatopoeia (On)	zittern (3); beben (2); blubbern; klopfen; ruckeln; rütteln; holpern;	stimulus, source
Origin (Or)	Vibration (14); Schwingung (6); Motor (3); Wind (3); [Tischtennis-]Ball (3); Herzton (2); Club; Nadelstiche; Bodenwellen; Maschine; Bohrmaschine; Elektroschocks; Vibrationsalarm; Wassertropfen; Morse code; Meer; Katze; Handy; Baustelle; U-Bahn;	source
Movement (Mo)	vibriert (2); tastend (2); vorbeiziehen (2); vorbeifahren (2); fallen; springen; wackeln; Gallopieren; Autofahren; hüpfend; schütteln; Beben;	source
Affective Reaction (AR)	angenehm (4); entspannend (2); Aufmerksamkeit (2); Vorsicht; nervös; Gefahr; anstoßend; antreibend; gruselig; erschreckend; insistent; dringend; aufrüttelnd; aufrührend; anregend; Frech; wachsam;	subject

Table 5.5: Emerging semantic categories from evoked perceptual descriptions for group B. Brackets indicate the amount of occurrences for each microconcept.

Semantic category	Microconcepts (Group C)	Object of reference
Temporal Change (TC)	schnell (5); kurz (4); schwingen (4); nachhallend (4); abklingen (4); abfallend (2); beschleunigen (2); [Änderung] gegen Ende (2); nachschwingen; ansteigend; langsam; auspendeln; attack [Zeit]; zügig; lange;	stimulus
Temporal Structure (TS)	Schläge (7); perkussiv (5); rhythmisch (4); transient (4); [schnelle] Abfolge; konstant (3); pulsierend (3); Rhythmus (2); einmalig [auftretend] (3); definiert (3); chaotisch (2); präzise (2); deutlich; Puls; gleichmäßig; loop-haft; kontinuierlich; im Takt; variierend; ungeordnet; geordnet; permanent; Wiederholung; geplant;	stimulus
Intensity (In)	leicht (7); klein (6); hart (4); stark (3); schwer (2); fest (2); laut (2); kräftig (2); groß (2); fein; wuchtig; subtil; heavy; winzig; weich;	stimulus
Spectral Property (SP)	hell (11); hohl (10); dumpf (8); hoch, hoher [Ton] (7); rauschen (5); metallisch (6); tief (5); mechanisch (5); unnatürlich (4); rund (4); inharmonisch (3); organisch (3); dunkel (3); hochfrequent (3); rau (3); warm (3); atonal (3); klar (3); geräuschhaft (3); digital (3); blechern (2); rauschig (2); natürlich (2); Bass (2); vielschichtig (2); tonal (2); breitbandig (2); motorisch (2); scharf (2); wenig [Frequenzen, Töne] (2); tieffrequent; fein; harmonisch; boxy; echt; gefiltert; noise; gedämpft; kalt; kühl; luftig; schmales [Spektrum]; diffus; abstrakt; schrill; artifiziell; holzig; physisch[er Ton];	stimulus
Onomatopoeia (On)	brummen (5); klirren (4); knistern (4); knattern (4); rumpeln (2); dröhnen (2); zittern (2); bouncen (2); hämmernd; holpern; rasseln; quitschen; klimpern; wummern; knacken; schnarren; gluckern; schnurren; gurgeln; boomend; rattern; wabern; grummeln; beben; wellig;	stimulus, source
Origin (Or)	Motor (6); Trommel (5); elektrisch (4); Gummi[-band] (4); Ball (4); Tischtennis (3); Club (3); Maschine (3); Saite (3); Synthesizer (2); drums (2); Drone (2); Schublade (2); metallenes [Objekt, Dach] (2); Flugzeug[-turbine] (2); Glas (2); Basketball; Katze; Motorrad; Ventilator; Ping Pong; Presslufthammer; Lichtschwert; Aufschlag; Reißverschluss; Sticks; Techno; Fluss; [Roboter-]schwein; Keller; Auto; Computerplatine; Tamburin; Becken; Tropfen; Eimer; U-Boot; Tiefsee; Unterwasser; Tauchgang; Dose; Membran; Resonanzkörper; Wind; tierisch; Zahnarzt;	source
Movement (Mo)	vibrieren (7); fallen (4); kratzen (4); zupfen (3); [näher] kommen (3); springen (3); klopfen (2); bouncen (2); wühlen (2); rumräumen; mitschwingend; werfen; schaben; abprallen; dribbeln; Aufprall; tropfend; Au	source
Affective Reaction (AR)	unangenehm (11); angenehm (8); direkt (2); Spannung [steigernd, erzeugend, aufbauend] (5); Drucken; Adrenalin; treibend; schmerzhaft; ansprechend; nervös; wohlklingend; motivierend; elektrisierend; Wohlempfinden; beschützt; aggressiv; brachial; stressig;	subject

Table 5.6: Emerging semantic categories from evoked perceptual descriptions for group C. Brackets indicate the amount of occurrences for each microconcept.

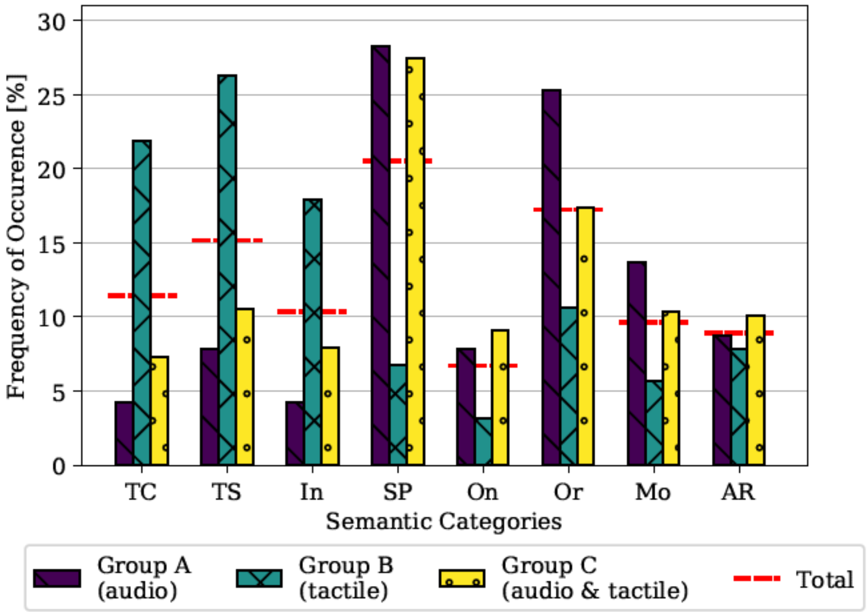


Figure 5.2: This bar plot illustrates the frequency of occurrence for each semantic category among the test groups respectively. The bars for each group were normalized by the total count of microconcepts and sum up to 100% respectively. For reference the frequency of occurrence for each semantic category across all groups (total) is illustrated with a dashed line.

stimuli, not only due to the early integration of both modalities, but also the inherent temporal and spectral coherence of the both signals that can be provided this way. The results also suggests that the DSP framework used for the audio-tactile translation works "well enough" for the subjects to give positive ratings on average (mean = 7.14, SD = 1.99, range = 0 - 10) across all 330 ratings.

When comparing individual audio-tactile stimulus pairs, the biggest difference in the mean rating between group conditions can be observed between groups A and C on stimulus pair ID 9 with an absolute difference of $|8.18 - 5.18| = 3.0$, which also shows the highest absolute difference in mean ratings between conditions, as can be seen illustrated in Figure 5.3. The stimulus pair ID 10 shows a slightly weaker, but similar deviation relative to the other stimulus pairs and will also be discussed in more detail below. The smallest mean coherence rating difference between group conditions can be observed for stimulus pair ID 5 between groups A and B with an absolute difference of 0.0, while stimulus pair ID 6 indicates the smallest, total difference between all group conditions with a mean coherence rating sum of 0.92. This comparably small difference indicates an indifference between group condition, as the mean ratings are comparably similar with a mean rating of A = 7.18 [SD = 1.89], B = 7.55 [SD = 1.86] and C = 7.64 [SD = 1.12] for each group respectively and this stimulus pair. Overall, the results in mean differences from Figure 5.3 could be discussed as an indicator on the level of similarity (or agreement)

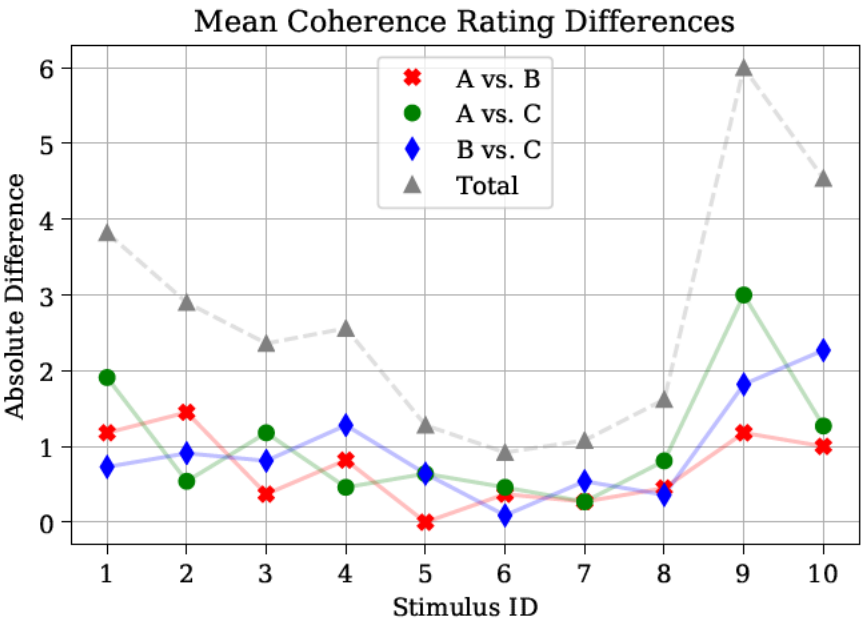


Figure 5.3: This plot illustrates the absolute difference between mean group ratings of individual stimulus pairs (by ID). The grey plot (triangle markers) indicates the sum of the absolute differences between groups for each stimulus pair. The mean of the total differences is $\mu = 2.7$ [SD = 1.66].

between the coherence ratings of the group conditions, which means that a small value indicates a high similarity between group ratings — while a big difference indicates a low level of similarity between the ratings of the three group conditions. It also shows that the differences in mean ratings for each stimulus between group conditions are relatively equal, except for the outlier case (ID 9 and ID 10) discussed above.

The results of the psycholinguistic analysis showed that descriptions of the temporal trajectory and intensity of the signal were more dominantly present for unimodal, tactile stimuli display (B) compared to the other two groups (A, C), which both contained exposure to the auditory stimuli. On the other hand, concepts describing spectral features of a stimulus were found more frequently when auditory stimuli were presented. This observation supports the results of previous, perceptual tests that indicated a lower frequency resolution of the tactile sense when compared to audition (see Chapter 2) and suggests that such high level, descriptive concepts are evoked less frequently when exposed to a tactile stimulus alone. This observation expands to more abstract, cognitive concepts, such as the origin (i.e., source) of the stimulus and it's behaviour (i.e., movement). These semantic categories were extracted more frequently when an audio stimulus was present. Similarly, onomatopoeic verbal associations where found more frequently when the auditory stimulus was present. This observation could be explained by the nature of onomatopoeia, as it is defined as "imitating or suggesting the sound that it describes". Nevertheless, the onomatopoeic descriptions of the unimodal tactile stimuli (B) describe physical attributes that could be related to touch directly, such as to tremble (zittern), to quiver (beben), to knock (klopfen), shake (rütteln), to stutter (ruckeln), or to jolt (holpern). The onomatopoeic concepts extracted from the unimodal auditory descriptions (A) however, are richer in vocabulary describing auditory concepts, while both auditory and touch-related concepts appear in the auditory-tactile group (C). A more thorough investigation on which concepts are common between the auditory and tactile modalities, and what type of signal properties evoke the same concepts across different modalities remains a topic for future experimentation.

The audio stimulus ID 2 (mentioned above) sounds like a recording of a set of random tools or objects being shuffled or moved around in a wooden box in an unorganized, but somehow rhythmic manner. A more daring hypothesis on the origin of this sound would be, that a corpus of such sounds, as described above, was used to replicate a spoken phrase using a sample-mosaic technique [90] to generate this “layered” sound, due to the perceived, rhythmic pattern and obscured sound-source relation. Next to the associations describing the source of this audio stimulus, as what can be summarized to “Tools being moved in a (wooden) box”, which are mainly found in groups A and C, there are subjective descriptions of it being perceived as “chaotic, dense, annoying and restless”. A set of concepts from group A, that were also extracted from the unimodal tactile descriptions (B), remain to be the ones describing a lack of structure, f.e. “irregular, unstable, changing”. The tactile stimulus, when lacking the audio context, appears to be perceived as quite random as it is missing all the important, contextual information gathered from the rich spectrum (i.e., timbre) to evoke the rich associations given for the equivalent audio source. Interestingly, there is a cluster of concepts to be found in the tactile descriptions (B), that suggests transfer of information (f.e. Morse Code, Instructions, communication, thoughtful [structure]) — which could be due to the rhythmic, spoken nature of the sound. Retrospectively, listening back to the ten chosen sounds, it becomes apparent, that all other sounds (except ID 2) seem to have less total acoustic events happening at the same time. Due to this inadvertent underrepresentation of multi-layered sounds — i.e., sounds with multiple, overlapping sonic events — effects of this “layered” trait remain to be explored. Due to our ability to localize and separate sonic events with our two ears (and brain) it seems to be a more complex task to find out under which configuration our tactile sense is also able to localize and separate sound sources [43]. The localization of sounds using our ears is found to be due to both spectral cues, as well as the time of arrival (ToA), Interaural Time Difference (ITD) and Interaural Intensity Difference (IID), which can be modeled using the Head-Related Transfer Function (HRTF), such theory for tactile localization of distant or competing sound sources is a topic that has been previously explored but to date isn’t fully understood [25, 78, 8].

The biggest median difference in the coherence ratings between the consecutive, unimodal display (A,B) and the simultaneous, bimodal display (C) was observed on the stimulus pair ID 9. This stimulus sounds like a recording of an object being dragged across a raster in a linear manner that could be compared to a zipper being closed. When playing back the tactile stimulus it seems like the distinct transients of the “raster” were smeared quite heavily and therefore do not reflect the original dynamic and sequence of impulses of the audio source very well. This example is a good indicator, illustrating that the translation process will require further exploration of the parameter space, to make sure it generalizes nicely across a wide range of audio signals — especially when it comes to transient analysis and synthesis. This does not necessarily mean that all parameters need to be optimized and fixed at some point, but to figure out which parameters might be useful for a tactile designer to manipulate to achieve a desired result faster and avoid undesired results.

A similar, but not quite as bad reproduction can be observed with the stimulus pair ID 10. The audio source of this stimulus-pair sounds like a wooden ship or rope creaking in the waves. Due to the rapid sequence of broadband impulses, which lend the creaking sound it’s perceived roughness, the transient analysis and resynthesis also displays a problematic smearing of the individual stick-slip events, which are assumed to be the reason for the “creaking impulses”. A quick succession of events, like in these examples, can also bring a tactile actuator to the limits of it’s capabilities due to the inertia of the moving mass. In this case though, after inspecting the resynthesized signals more closely, it seems like the transients of each impact are properly detected, but the resulting transient synthesis of each event is too long, making the distinct events overlap into each other. This means the resynthesis of transients should be revisited to make sure quick successions of impacts are still displayed properly. An easy fix for this issue would be to set both the fundamental frequency of the square wave signal used for transient synthesis to a higher frequency and the number of cycles for each event lower — thus shortening each synthesized events duration and avoiding overlaps. The settings used for this embodiment used the low fundamental resonance frequency of the actuator, which achieves a very efficient way to drive the actuator, but shows it’s drawbacks in this case. Furthermore, the transient synthesis could be optimized by filtering out transient events by proximity rules derived from previous experiments on temporal masking: Temporal masking effects have indicated, that the minimum detectable gap in a sequence of two vibrotactile events was found to be around 8 ms, while the minimum detectable gap increases for lower intensity levels (similar to auditory temporal maskers) [99, 28]. The higher rating in the bimodal display of stimulus ID 9 could be explained by the importance of the audio signal, essentially making up for the lack of fidelity in the tactile signal. Exploring the importance of the audio signal towards a tactile event (and vice versa) goes back to the idea of exploring the parameter space of the translation in more detail. A proposal would be to first use an ABX-test or a “two-up – one-down” experiment, to find the parameter settings at which a quality degradation from a tactile source signal towards a resynthesized tactile signal becomes noticeable in an unimodal display condition (i.e., identifying the JND for unimodal tactile display). The same experiment could then be conducted on a bimodal display to see, if the JND and therefore the quality of the tactile reproduction changes if the subject is exposed to the added auditory modality.

The stimulus pair with the smallest differences in mean ratings between group conditions and the highest overall rating, is stimulus pair ID 5. This stimulus sounds like a basketball being dropped from a few centimeters above ground, which creates a sequence of hollow, rubbery impact sounds. From the linguistic analysis it becomes apparent, that the association with a “ball being dropped” appears equally in all three group conditions. First off, this could be reasoned to be due to the predictable, rhythmic pattern of the impacts created by the bouncing object, which doesn’t require a lot of spectral information to be recognized. This means that, due to the temporal resolution of the skin, such a predictable and rhythmic trait can be picked up more easily by the skin when compared to a more static, complex texture with a lack of a predictable temporal structure. This effect would be in accordance with previous experiments on temporal sensitivity for vibrotactile stimuli, which support the importance of onsets and showed an increased sensitivity towards amplitude-modulated sinusoidals, when compared to broad- or narrow-band noise carriers, which could be found naturally in complex signals [107, 15]. Secondly, a bouncing ball is argued to be a natural experience people might be familiar with from playing sports or other ball activities. The vibrations emitted from dribbling a ball could theoretically not only be heard, but also be felt through the contact medium (floor or table).

Coincidentally, a similar temporal pattern can be found in the sound source ID 6. This recording sounds like a table tennis ball (i.e., ping-pong or whiff-whaff ball) being played back and forth once. Subjectively, this rhythmic sequence is easily recognizable and has quite an iconic quality. Even though in the unimodal tactile condition (group B) this stimulus wasn’t identified, due to a lack of descriptions found in the semantic *Origin* category, the stimulus was described as weak (schwach), fast (schnell) and careful (vorsichtig). These descriptions could be due to the lack of low frequency content in this sound sample, which in turn is considered to be connected to the lower mass of the bouncing object (ping-pong ball) and therefore a weaker structural vibration resulting from each impact event — especially when compared to the previous basketball example. Nonetheless, both stimulus pairs ID 5 and 6 show an above average rating and small differences in the ratings between group conditions. This could be reasoned to be due to the more predictable rhythmic nature of both stimulus pairs when compared to more static or chaotic counterparts, such as the pairs with ID 1 and 2 (which both show a below average rating in total).

Overall, the findings from the experiment seem to indicate a sweet spot for vibrotactile stimuli between clearly identifiable, predictable rhythmic patterns and completely static textures. The most complex type of signal for tactile perception seems to be an unpredictable and overlapping sequence of events, which is easier for our ears to tell apart (as discussed above) but seems to overwhelm the tactile sense when trying to form a coherent percept from the sensory information. For a tactile designer, this means that applications containing many tactile events a priority or clever mixing needs to take place — similar to the practice of mixing a studio album and positioning sounds in a stereo panorama. Although, for vibrotactile assets, this would require multiple actuators, as it appears to be hard for our sensory nervous system and the higher level brain functions to tell multiple tactile events apart, based on the information that is passed on by the mechanoreceptors. Experiments exploring mixing of multiple vibrotactile assets and the ability of the human sensory system to identify (i.e., separate) or form a coherent percept from multiple events is proposed to be the topic for a future experiment and has been discussed in previous works, although in slightly different contexts [15, 108].

Finally, the white noise stimulus, intended to mask any unwanted auditory cues during the unimodal, tactile stimuli display, has been reported to be (initially) distracting by eight participants. Some of those reported that they quickly got used to the audio masker, even though it “required more concentration” and were happy that they could repeat each stimulus as often as desired. Depending on the position of a tactile display, providing a sole, tactile stimulus with no additional auditory stimulus can pose a complex issue due to the possibility of bone-conducted cues finding their way to the cochlea. Providing a clean separation between auditory and tactile cues is a challenging task, but also illustrates how fundamentally connected both senses are. Discussions around audio maskers in tactile experiments have been conducted in previous works, which focused on tactile perception [4, 110, 22].

6 Conclusion

The term “haptics” encompasses a variety of sensory modalities, including tactile, kinaesthetic, thermosensory and proprioceptic components. When we talk about touch we mainly refer to the tactile modality, which is able to sense mechanical forces, such as shearing, pressure and vibrations. The vibrotactile modality, which mainly contributes to sensing ridges and textures through vibrational excitation, is the main focus of this work. More specifically, due to an overlap in the sensitive frequency range of both the auditory and tactile modalities and the natural occurrence of integrated audio-tactile events, this work proposes to use audio material as a starting point for vibrotactile stimulus design.

To enable an audio-tactile design workflow, this work proposes a novel framework to design vibrotactile stimuli based on audio source material. The framework’s core data model is composed of a parametric representation, which is derived by extracting various features from a source audio signal. This representation both compresses the signal and makes it flexible in adapting to various tactile actuation technologies during re-synthesis. Previous works used a simpler set of features to synthesize a (mostly) monophonic tactile stimulus from an audio source, while the method proposed in this work aims to maintain a richer set of spectral information. This process achieves a perceptual approximation of the target stimulus and therefore ensures a higher fidelity reproduction — especially to display finely nuanced textural information on wide-band vibrotactile actuators, which have only become available in recent years.

An investigation into current vibrotactile display technologies showed, that an adaptive (parametric) synthesis method is suitable to ensure an approximate reproduction, even within the narrow bandwidth of most actuators found on the market today. Furthermore, the form factor and interaction with a tactile device was investigated regarding the issue of bodily-induced interference, which is a trait that needs to be considered during the design of a tactile application. These systematic and bodily-induced interferences on the tactile display’s frequency response were measured in an exemplary use case of a custom built vibrotactile wristband. Furthermore, strategies to measure and counteract these interferences using system identification methods and inverse filtering were proposed. To summarize, current technological limitations for the implementation of high fidelity vibrotactile displays in everyday devices are mostly due to a lack of platform-agnostic standards for the design, transmission and reproduction of tactile stimuli. The requirements for the design of the framework were therefore informed by these issues.

To validate the reproduction quality and coherence between the audio source and a derived tactile stimulus an exploratory perceptual evaluation using a customized vibrotactile wristband was conducted. In this experiment, subjects were exposed to either unimodal (audio or tactile), or bimodal (audio and tactile) stimulus pairs. For each exposure the subjects both numerically rated the experience and gave free verbal descriptions on the perceived experience. The rating of the bimodal (audio and tactile) condition was statistically significant and suggested, that a joint audio-tactile experience was preferred over the isolated unimodal display of either an audio or tactile stimulus. This supports the notion of a joint audio-tactile experience design, which is in agreement with previous studies showing early integration across the modalities.

Furthermore, a psycholinguistic analysis was conducted on the verbal descriptions using the so-called constant comparison method. The goal of this exploratory analysis was to first identify emerging key concepts and semantic categories, and to then find indications of varying distributions of said semantic categories among the three test conditions. After identifying eight semantic categories throughout all verbal descriptions, it became noticeable, that descriptions associated with temporal properties and the intensity of a stimulus were dominant among the unimodal tactile condition, whereas descriptions associated with spectral properties and the origin of the stimulus were found more dominantly as soon as the subjects were exposed to an audio stimulus. From this observation it was concluded, that information on the temporal progression and intensity of a stimulus are more important for the perceptual quality, when compared to the spectral resolution of a tactile stimulus. This observation supports the reasoning behind the comparably coarse frequency approximation used in this work, and is in line with findings from previous experiments on the frequency resolution of the tactile sense. Further investigations on the verbal descriptions used for the tactile sense, further exploring the audio-tactile semantic space, and identifying key descriptors for tactile signal attributes could not only aid the design of meaningful tactile synthesis engines, but also help understand perceptual aspects of touch for the design of applications better.

From the overall positive ratings derived from the perceptual study it was concluded that this initial proposal of the framework yields a promising direction. To get more conclusive results on the translation method it is proposed to further explore the parameter space of the analysis and synthesis process by running A/B-X tests with varying parameter settings. Alternatively, a test with varying compression (i.e. quality) levels to find the JND between the original source stimulus and the reproduction is proposed. Furthermore, the ability of distinguishing between multiple tactile events in a mix, using multiple actuators on the body and the ability to accurately reproduce tactile information from a virtual location (similar to amplitude- or ToA-difference panning in stereo audio) are topics for future experiments.

Going into this work, it was interesting to see how many issues can arise due to the lack of standardized procedures when dealing with tactile signals, as well as dealing with the variety of tactile actuation technologies. Since the start of this thesis, next to the efforts put into the ISO norm [40, 39, 41], various companies and associations in the emerging tactile industry have published drafts for a tactile standardization process[1]. It is pleasant to see, that the community is trying to work together by paving the way for industrial applications and to work towards a general agreement, that could also benefit and accelerate the research community. This also highlights the relevancy of the topics discussed in this work at the current time. I'm happy that the process of working on this thesis broadened and deepened my knowledge around various topics, such as mechanical engineering, electrical engineering, signal processing, software design, statistics, measurement design, perception and conducting empirical user tests.

Not all aspects, that have been considered for the design of the framework presented in this work (see Section 4.1) have been validated and will require more work in the future. Especially the functionality of the resynthesis engine while adapting to various actuator types and their capabilities hasn't been formally demonstrated or validated. One implementation of this framework, using the CoreHaptics API by Apple for resynthesis, is already available to use for developers and (tactile) designers[2]. This implementation uses a subset of the components extracted in the analysis to drive the Taptic engine actuator through the CoreHaptics API. It is a first step, that proves that the proposed codec works well, not only for "high fidelity" signals, but also with the limited capabilities of a monophonic playback, as provided by Apple's CoreHaptics API. As discussed before, the parameter space of the framework will require further investigation and, if proven to scale nicely across various scenarios, will require more work on the compression by introducing better ways to store the information of the extracted components.

[1] Lofelt "VT-1: A Specification Proposal for Realistic Vibrotactile Feedback",
https://lofelt.com/resources
Immersion "High Definition (HD) Actuator Selection and Testing Guidelines",
https://www.immersion.com/the-haptic-stack-hardware-layer/
Institute of Electrical and Electronics Engineers "P1918.1",
https://standards.ieee.org/project/1918_1.html

[2] Lofelt Composer: A tool aiding design for audio-tactile experiences
https://composer.lofelt.com/

Literature

[1] Allamanche, E. ; Geiger, R. ; Herre, J. ; Sporer, T. : MPEG-4 low delay audio coding based on the AAC codec. In: *Audio Engineering Society Convention 106* Audio Engineering Society, 1999, S.

[2] Allan, K. ; White, T. ; Jones, L. ; Merlo, J. ; Haas, E. ; Zets, G. ; Rupert, A. : Getting the Buzz: What's Next for Tactile Information Delivery? In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* Bd. 54 SAGE Publications Sage CA: Los Angeles, CA, 2010, S. 1331–1334

[3] Allerkamp, D. ; Böttcher, G. ; Wolter, F.-E. ; Brady, A. C. ; Qu, J. ; Summers, I. R.: A vibrotactile approach to tactile rendering. In: *The Visual Computer* 23 (2007), Nr. 2, S. 97–108

[4] Bensmaïa, S. ; Hollins, M. ; Yau, J. : Vibrotactile intensity and frequency information in the pacinian system: a psychophysical model. In: *Perception & psychophysics* 67 (2005), Nr. 5, S. 828–841

[5] Bernard, C. ; Monnoyer, J. ; Denjean, S. ; Wiertlewski, M. ; Ystad, S. : Sound and Texture Synthesizer. In: *International Workshop on Haptic and Audio Interaction Design (HAID)*, 2019, S. 0

[6] Birnbaum, D. ; Wanderley, M. M.: A Systematic Approach to Musical vibrotactile feedback. In: *ICMC*, 2007, S. 397–404

[7] Bolanowski Jr, S. J. ; Gescheider, G. A. ; Verrillo, R. T. ; Checkosky, C. M.: Four channels mediate the mechanical aspects of touch. In: *The Journal of the Acoustical society of America* 84 (1988), Nr. 5, S. 1680–1694

[8] Borg, E. ; Ronnberg, J. ; Neovius, L. ; Lie, T. : Vibratory-coded directional analysis: Evaluation of a three-microphone/four-vibrator DSP system. In: *Journal of rehabilitation research and development* 38 (2001), Nr. 2, S. 257–264

[9] Bristow-Johnson, R. : Cookbook formulae for audio EQ biquad filter coefficients. In: *on-line publication: http://shepazu.github.io/Audio-EQ-Cookbook/audio-eq-cookbook.html, last sighted on 11th November 2019* (2016)

[10] Brooks, P. ; Frost, B. J.: Evaluation of a tactile vocoder for word recognition. In: *The Journal of the Acoustical Society of America* 74 (1983), Nr. 1, S. 34–39

[11] Bukkapatnam, A. T. ; Depalle, P. ; Wanderley, M. : Autoregressive Parameter Estimation for Equalizing Vibrotactile Systems. In: *Proceedings of International Workshop on Haptic and Audio Interaction Design*, 2019, S. 9–16

[12] Carter, T. ; Seah, S. A. ; Long, B. ; Drinkwater, B. ; Subramanian, S. : UltraHaptics: multi-point mid-air haptic feedback for touch surfaces. In: *Proceedings of the 26th annual ACM symposium on User interface software and technology* ACM, 2013, S. 505–514

[13] Chaudhari, R. ; Çizmeci, B. ; Kuchenbecker, K. J. ; Choi, S. ; Steinbach, E. : Low bitrate source-filter model based compression of vibrotactile texture signals in haptic teleoperation. In: *Proceedings of the 20th ACM international conference on Multimedia* ACM, 2012, S. 409–418

[14] Chion, M. : Guide to sound objects: Pierre Schaeffer and musical research. In: *Trans. John Dack and Christine North), http://www.ears.dmu.ac.uk* (2009)

[15] Craig, J. C.: The role of onset in the perception of sequentially presented vibrotactile patterns. In: *Perception & psychophysics* 34 (1983), Nr. 5, S. 421–432

[16] Douglas, D. H. ; Peucker, T. K.: Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. In: *Cartographica: the international journal for geographic information and geovisualization* 10 (1973), Nr. 2, S. 112–122

[17] Eid, M. A. ; Al Osman, H. : Affective haptics: Current research and future directions. In: *IEEE Access* 4 (2015), S. 26–40

[18] Farina, A. : Advancements in impulse response measurements by sine sweeps. In: *Audio Engineering Society Convention 122* Audio Engineering Society, 2007, S.

[19] Fernald, A. : Intonation and Communicative Intent in Mothers' Speech to Infants: Is the Melody the Message? In: *Child Development* 60 (1989), Nr. 6, 1497–1510. http://www.jstor.org/stable/1130938. – ISSN 00093920, 14678624

[20] Fitzgerald, D. : *Harmonic/percussive source separation using median filtering.* 2010

[21] Fontana, F. ; Järveläinen, H. ; Papetti, S. ; Avanzini, F. ; Klauer, G. ; Malavolta, L. ; Musica, C. di ; Pollini, C. : Rendering and subjective evaluation of real vs. synthetic vibrotactile cues on a digital piano keyboard. In: *Proceedings of the Sound and Music Computing Conference. Maynooth, Ireland: SMC*, 2015, S.

[22] Fontana, F. ; Papetti, S. ; Järveläinen, H. ; Avanzini, F. ; Giordano, B. L.: Perception of vibrotactile cues in musical performance. In: *Musical Haptics*. Springer, Cham, 2018, S. 49–72

[23] Foxe, J. J.: Multisensory integration: frequency tuning of audio-tactile integration. In: *Current biology* 19 (2009), Nr. 9, S. R373–R375

[24] Franzén, O. ; Nordmark, J. : Vibrotactile frequency discrimination. In: *Perception & Psychophysics* 17 (1975), Nr. 5, S. 480–484

[25] Frost, B. ; Richardson, B. : Tactile localization of sounds: Acuity, tracking moving sources, and selective attention. In: *The Journal of the Acoustical Society of America* 59 (1976), Nr. 4, S. 907–914

[26] Fujisaki, W. ; Nishida, S. : Audio–tactile superiority over visuo–tactile and audio–visual combinations in the temporal resolution of synchrony perception. In: *Experimental brain research* 198 (2009), Nr. 2-3, S. 245–259

[27] Gault, R. H.: Progress in experiments on tactual interpretation of oral speech. In: *The Journal of Abnormal Psychology and Social Psychology* 19 (1924), Nr. 2, S. 155–159. http://dx.doi.org/10.1037/h0065752. – DOI 10.1037/h0065752. – ISSN 0145–2347

[28] Gescheider, G. A. ; Bolanowski, S. J. ; Chatterton, S. K.: Temporal gap detection in tactile channels. In: *Somatosensory & motor research* 20 (2003), Nr. 3-4, S. 239–247

[29] Giannoulis, D. ; Massberg, M. ; Reiss, J. D.: Parameter automation in a dynamic range compressor. In: *Journal of the Audio Engineering Society* 61 (2013), Nr. 10, S. 716–726

[30] Gillmeister, H. ; Eimer, M. : Tactile enhancement of auditory detection and perceived loudness. In: *Brain research* 1160 (2007), S. 58–68

[31] Goff, G. D.: Differential discrimination of frequency of cutaneous mechanical vibration. In: *Journal of experimental psychology* 74 (1967), Nr. 2p1, S. 294

[32] Grey, J. M. ; Gordon, J. W.: Perceptual effects of spectral modifications on musical timbres. In: *The Journal of the Acoustical Society of America* 63 (1978), Nr. 5, S. 1493–1500

[33] Grunwald, M. ; Beyer, L. : *Der bewegte Sinn: Grundlagen und Anwendungen zur haptischen Wahrnehmung*. Springer-Verlag, 2013

[34] Guest, S. ; Catmur, C. ; Lloyd, D. ; Spence, C. : Audiotactile interactions in roughness perception. In: *Experimental Brain Research* 146 (2002), Nr. 2, S. 161–171

[35] Guruswamy, V. L. ; Lang, J. ; Lee, W.-S. : IIR filter models of haptic vibration textures. In: *IEEE Transactions on Instrumentation and Measurement* 60 (2010), Nr. 1, S. 93–103

[36] Hallam, S. ; Cross, I. ; Thaut, M. : *Oxford handbook of music psychology*. Oxford University Press, 2011

[37] Harada, N. ; Griffin, M. J.: Factors influencing vibration sense thresholds used to assess occupational exposures to hand transmitted vibration. In: *Occupational and Environmental Medicine* 48 (1991), Nr. 3, S. 185–192

[38] Harker, A. ; Tremblay, P. A.: The HISSTools impulse response toolbox: Convolution for the masses. In: *Proceedings of the International Computer Music Conference* The International Computer Music Association, 2012, S. 148–155

[39] ISO: *Ergonomics of human-system interaction – Part 920: Guidance on tactile and haptic interactions*. International Organization for Standardization, 2009

[40] ISO: *Ergonomics of human-system interaction – Part 910: Framework for tactile and haptic interaction*. International Organization for Standardization, 2011

[41] ISO: *Ergonomics of human-system interaction – Part 940: Evaluation of tactile and haptic interactions*. International Organization for Standardization, 2017

[42] ISO/TC 43 Committee: *ISO 226:2003: Acoustics - Normal equal-loudness-level contours*. 2003

[43] Israr, A. ; Kim, S.-C. ; Stec, J. ; Poupyrev, I. : Surround haptics: tactile feedback for immersive gaming experiences. In: *CHI'12 Extended Abstracts on Human Factors in Computing Systems* ACM, 2012, S. 1087–1090

[44] ITU: *Algorithms to measure audio programme loudness and true-peak audio level*. 2015

[45] Johnson, L. A. ; Higgins, C. M.: A navigation aid for the blind using tactile-visual sensory substitution. In: *2006 International Conference of the IEEE Engineering in Medicine and Biology Society* IEEE, 2006, S. 6289–6292

[46] Jousmäki, V. ; Hari, R. : Parchment-skin illusion: sound-biased touch. In: *Current biology* 8 (1998), Nr. 6, S. R190–R191

[47] Kayser, C. ; Petkov, C. I. ; Augath, M. ; Logothetis, N. K.: Integration of touch and sound in auditory cortex. In: *Neuron* 48 (2005), Nr. 2, S. 373–384

[48] Kumler, K. : *Being Touched by Music: A Phenomenological-hermeneutical Approach to Understanding Transformational Music Experience*, Duquesne University, Diss., 2006

[49] Kyung, K.-U. ; Ahn, M. ; Kwon, D.-S. ; Srinivasan, M. A.: Perceptual and biomechanical frequency response of human skin: implication for design of tactile displays. In: *First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics Conference* IEEE, 2005, S. 96–101

[50] Lerch, A. : *An introduction to audio content analysis: Applications in signal processing and music informatics.* Wiley-IEEE Press, 2012

[51] Lloyd, D. B. ; Raghuvanshi, N. ; Govindaraju, N. K.: Sound synthesis for impact sounds in video games. In: *Symposium on Interactive 3D Graphics and Games*, 2011, S. 55–62

[52] Marshall, M. T. ; Wanderley, M. M.: Vibrotactile feedback in digital musical instruments. In: *Proceedings of the 2006 conference on New interfaces for musical expression*, 2006, S. 226–229

[53] McMahan, W. ; Kuchenbecker, K. J.: Haptic display of realistic tool contact via dynamically compensated control of a dedicated actuator. In: *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems* IEEE, 2009, S. 3170–3177

[54] Merchel, S. ; Altinsoy, M. E.: The Influence of Vibrations on Musical Experience. In: *J. Audio Eng. Soc* 62 (2014), Nr. 4, 220–234. http://www.aes.org/e-lib/browse.cfm?elib=17134

[55] Merchel, S. ; Altinsoy, M. E. ; Stamm, M. : Just-noticeable frequency differences for whole-body vibrations. In: *INTER-NOISE and NOISE-CON Congress and Conference Proceedings* Bd. 2011 Institute of Noise Control Engineering, 2011, S. 2234–2239

[56] Merchel, S. ; Altinsoy, M. E.: Auditory-tactile music perception. In: *Proceedings of Meetings on Acoustics ICA2013* Bd. 19 ASA, 2013, S. 015030

[57] Merchel, S. ; Altinsoy, M. E.: Music-induced vibrations in a concert hall and a church. In: *Archives of Acoustics* 38 (2013), Nr. 1, S. 13–18

[58] Merchel, S. ; Schwendicke, A. ; Altinsoy, M. E.: *Feeling the sound: audio-tactile intensity perception.* 2011

[59] Merer, A. ; Aramaki, M. ; Ystad, S. ; Kronland-Martinet, R. : Perceptual characterization of motion evoked by sounds for synthesis control purposes. In: *ACM Trans. Appl. Percept.* 10 (2013), S. 1–24

[60] Merer, A. ; Ystad, S. ; Kronland-Martinet, R. ; Aramaki, M. : Semiotics of sounds evoking motions: Categorization and acoustic features. In: *Computer Music Modeling and Retrieval. Sense of Sounds* (2008), S. 139–158

[61] Merer, A. ; Ystad, S. ; Kronland-Martinet, R. ; Aramaki, M. : Abstract sounds and their applications in audio and perception research. In: *International Symposium on Computer Music Modeling and Retrieval* Springer, 2010, S. 176–187

[62] Meyer, D. J. ; Peshkin, M. A. ; Colgate, J. E.: Fingertip friction modulation due to electrostatic attraction. In: *2013 world haptics conference (WHC)* IEEE, 2013, S. 43–48

[63] Mohebbi, R. ; Gray, R. ; Tan, H. Z.: Driver reaction time to tactile and auditory rear-end collision warnings while talking on a cell phone. In: *Human Factors* 51 (2009), Nr. 1, S. 102–110

[64] Monnai, Y. ; Hasegawa, K. ; Fujiwara, M. ; Yoshino, K. ; Inoue, S. ; Shinoda, H. : HaptoMime: mid-air haptic interaction with a floating virtual screen. In: *Proceedings of the 27th annual ACM symposium on User interface software and technology* ACM, 2014, S. 663–667

[65] Moore, B. C.: Frequency difference limens for short-duration tones. In: *The Journal of the Acoustical Society of America* 54 (1973), Nr. 3, S. 610–619

[66] Müllensiefen, D. ; Gingras, B. ; Stewart, L. ; Musil, J. J.: Goldsmiths Musical Sophistication Index (Gold-MSI) v1. 0: Technical Report and Documentation Revision 0.3. In: *London: Goldsmiths, University of London.* (2013)

[67] Nordahl, R. ; Berrezag, A. ; Dimitrov, S. ; Turchet, L. ; Hayward, V. ; Serafin, S. : Preliminary experiment combining virtual reality haptic shoes and audio synthesis. In: *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications* Springer, 2010, S. 123–129

[68] Norman, D. A.: *The psychology of everyday things.* Basic books, 1988

[69] Nørretranders, T. : *The User Illusion: Cutting consciousness down to size.* Viking, 1991

[70] Okamoto, S. ; Ishikawa, S. ; Nagano, H. ; Yamada, Y. : Spectrum-based vibrotactile footstep-display for crinkle of fragile structures. In: *2011 IEEE International Conference on Robotics and Biomimetics* IEEE, 2011, S. 2459–2464

[71] Olson, H. F.: *Music, physics and engineering.* Bd. 1769. Courier Corporation, 1967

[72] O'Modhrain, S. ; Gillespie, R. B.: Once more, with feeling: Revisiting the role of touch in performer-instrument interaction. In: *Musical haptics.* Springer, Cham, 2018, S. 11–27

[73] Papadogianni-Kouranti, M. ; Egermann, H. ; Weinzierl, S. : Auditive and Audiotactile Music Perception of Cochlear Implant Users. (2015)

[74] Papetti, S. ; Schiesser, S. ; Fröhlich, M. : Multi-point vibrotactile feedback for an expressive musical interface. In: *NIME*, 2015, S. 235–240

[75] Parks, T. W. ; Burrus, C. S.: *Digital filter design.* Wiley-Interscience, 1987

[76] Purnhagen, H. ; Meine, N. : HILN-the MPEG-4 parametric audio coding tools. In: *2000 IEEE International Symposium on Circuits and Systems. Emerging Technologies for the 21st Century. Proceedings (IEEE Cat No. 00CH36353)* Bd. 3 IEEE, 2000, S. 201–204

[77] Ranjbar, P. ; Stranneby, D. ; Erik, B. : Vibrotactile identification of signal-processed sounds from environmental events. In: *Journal of rehabilitation research and development* 46 (2009), Nr. 8, S. 1021–1036

[78] Richardson, B. ; Wuillemin, D. ; Saunders, F. : Tactile discrimination of competing sounds. In: *Perception & psychophysics* 24 (1978), Nr. 6, S. 546–550

[79] Rimell, S. ; Howard, D. M. ; Tyrrell, A. M. ; Kirk, R. ; Hunt, A. : Cymatic. Restoring the Physical Manifestation of Digital Sound Using Haptic Interfaces to Control a New Computer Based Musical Instrument. In: *ICMC*, 2002, S.

[80] Robles-De-La-Torre, G. ; Hayward, V. : Force can overcome object geometry in the perception of shape through active touch. In: *Nature* 412 (2001), Nr. 6845, S. 445

[81] Rothenberg, M. ; Verrillo, R. T. ; Zahorian, S. A. ; Brachman, M. L. ; Bolanowski Jr, S. J.: Vibrotactile frequency for encoding a speech parameter. In: *The Journal of the Acoustical Society of America* 62 (1977), Nr. 4, S. 1003–1012

[82] Russo, F. A. ; Ammirante, P. ; Fels, D. I.: Vibrotactile discrimination of musical timbre. In: *Journal of Experimental Psychology: Human Perception and Performance* 38 (2012), Nr. 4, S. 822

[83] Saitis, C. ; Fritz, C. ; Scavone, G. : Sounds like melted chocolate: How musicians conceptualize violin sound richness. In: *2019 International Symposium on Music Acoustics (ISMA)*, 2019, S.

[84] Saitis, C. ; Fritz, C. ; Scavone, G. P. ; Guastavino, C. ; Dubois, D. : Perceptual evaluation of violins: A psycholinguistic analysis of preference verbal descriptions by experienced musicians. In: *The Journal of the Acoustical Society of America* 141 (2017), Nr. 4, S. 2746–2757

[85] Saitis, C. ; Järveläinen, H. ; Fritz, C. : The role of haptic cues in musical instrument quality perception. In: *Musical haptics*. Springer, Cham, 2018, S. 73–93

[86] Schaal, N. K. ; Bauer, A.-K. R. ; Müllensiefen, D. : Der Gold-MSI: replikation und validierung eines fragebogeninstrumentes zur messung musikalischer erfahrenheit anhand einer deutschen stichprobe. In: *Musicae Scientiae* 18 (2014), Nr. 4, S. 423–447

[87] Schmidt, R. F. ; Lang, F. ; Heckmann, M. : *Physiologie des menschen: mit pathophysiologie*. Springer-Verlag, 2007

[88] Schön, D. ; Ystad, S. ; Kronland-Martinet, R. ; Besson, M. : The evocative power of sounds: Conceptual priming between words and nonverbal sounds. In: *J. Cogn. Neurosci.* 22 (2009), S. 1026–1035

[89] Schubert, E. ; Wolfe, J. : Does timbral brightness scale with frequency and spectral centroid? In: *Acta acustica united with acustica* 92 (2006), Nr. 5, S. 820–825

[90] Schwarz, D. : Corpus-based concatenative synthesis. In: *IEEE signal processing magazine* 24 (2007), Nr. 2, S. 92–104

[91] Serra, X. ; Smith, J. : Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition. In: *Computer Music Journal* 14 (1990), Nr. 4, S. 12–24

[92] Stein, B. E. ; Meredith, M. A.: *The merging of the senses*. The MIT Press, 1993

[93] Steinmetz, C. : *csteinmetz1/pyloudnorm: 0.1.0 (Version v0.1.0)*. https://github.com/csteinmetz1/pyloudnorm. Version: 2019

[94] Strauss, A. ; Corbin, J. : *Basics of qualitative research techniques*. Sage publications Thousand Oaks, CA, 1998

[95] Tajadura-Jiménez, A. ; Väljamäe, A. ; Toshima, I. ; Kimura, T. ; Tsakiris, M. ; Kitagawa, N. : Action sounds recalibrate perceived tactile distance. In: *Current Biology* 22 (2012), Nr. 13, S. R516–R517

[96] Treede, R.-D. : Das somatosensorische System. In: *Physiologie des Menschen*. Springer, 2010, S. 272–297

[97] Ujitoko, Y. ; Sakurai, S. ; Hirota, K. : Vibrator Transparency: Re-using Vibrotactile Signal Assets for Different Black Box Vibrators without Re-designing. In: *2020 IEEE Haptics Symposium (HAPTICS)* IEEE, 2020, S. 882–889

[98] Valin, J.-M. ; Maxwell, G. ; Terriberry, T. B. ; Vos, K. : High-quality, low-delay music coding in the opus codec. In: *arXiv preprint arXiv:1602.04845* (2016)

[99] Van Doren, C. L. ; Gescheider, G. A. ; Verrillo, R. T.: Vibrotactile temporal gap detection as a function of age. In: *The Journal of the Acoustical Society of America* 87 (1990), Nr. 5, S. 2201–2206

[100] Van Erp, J. B. ; Kyung, K.-U. ; Kassner, S. ; Carter, J. ; Brewster, S. ; Weber, G. ; Andrew, I. : Setting the standards for haptic and tactile interactions: ISO's work. In: *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications* Springer, 2010, S. 353–358

[101] Verrillo, R. T.: Effect of contactor area on the vibrotactile threshold. In: *The Journal of the Acoustical Society of America* 35 (1963), Nr. 12, S. 1962–1966

[102] Verrillo, R. T. ; Fraioli, A. J. ; Smith, R. L.: Sensation magnitude of vibrotactile stimuli. In: *Perception & Psychophysics* 6 (1969), Nr. 6, S. 366–372

[103] Visell, Y. ; Cooperstock, J. R. ; Giordano, B. L. ; Franinovic, K. ; Law, A. ; McAdams, S. ; Jathal, K. ; Fontana, F. : A vibrotactile device for display of virtual ground materials in walking. In: *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications* Springer, 2008, S. 420–426

[104] Visvalingam, M. ; Whyatt, J. D.: Line generalisation by repeated elimination of points. In: *The Cartographic Journal* 30 (1993), Nr. 1, S. 46–51. http://dx.doi.org/10.1179/000870493786962263. – DOI 10.1179/000870493786962263

[105] Vorapatratorn, S. ; Nambunmee, K. : iSonar: an obstacle warning device for the totally blind. In: *Journal of Assistive, Rehabilitative & Therapeutic Technologies* 2 (2014), Nr. 1, S. 23114

[106] Weber, E. H.: *Die Lehre vom Tastsinne und Gemeingefühle auf Versuche gegründet.* Friedrich Vieweg und Sohn, 1851

[107] Weisenberger, J. M.: Sensitivity to amplitude-modulated vibrotactile signals. In: *The Journal of the Acoustical Society of America* 80 (1986), Nr. 6, S. 1707–1715

[108] Weisenberger, J. M.: Vibrotactile temporal masking: Effects of multiple maskers. In: *The Journal of the Acoustical Society of America* 95 (1994), Nr. 4, S. 2213–2220

[109] Wier, C. C. ; Jesteadt, W. ; Green, D. M.: Frequency discrimination as a function of frequency and sensation level. In: *The Journal of the Acoustical Society of America* 61 (1977), Nr. 1, S. 178–184

[110] Wilson, E. C. ; Reed, C. M. ; Braida, L. D.: Integration of auditory and vibrotactile stimuli: effects of phase and stimulus-onset asynchrony. In: *The Journal of the Acoustical Society of America* 126 (2009), Nr. 4, S. 1960–1974

[111] Winfield, L. ; Glassmire, J. ; Colgate, J. E. ; Peshkin, M. : T-pad: Tactile pattern display through variable friction reduction. In: *Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07)* IEEE, 2007, S. 421–426

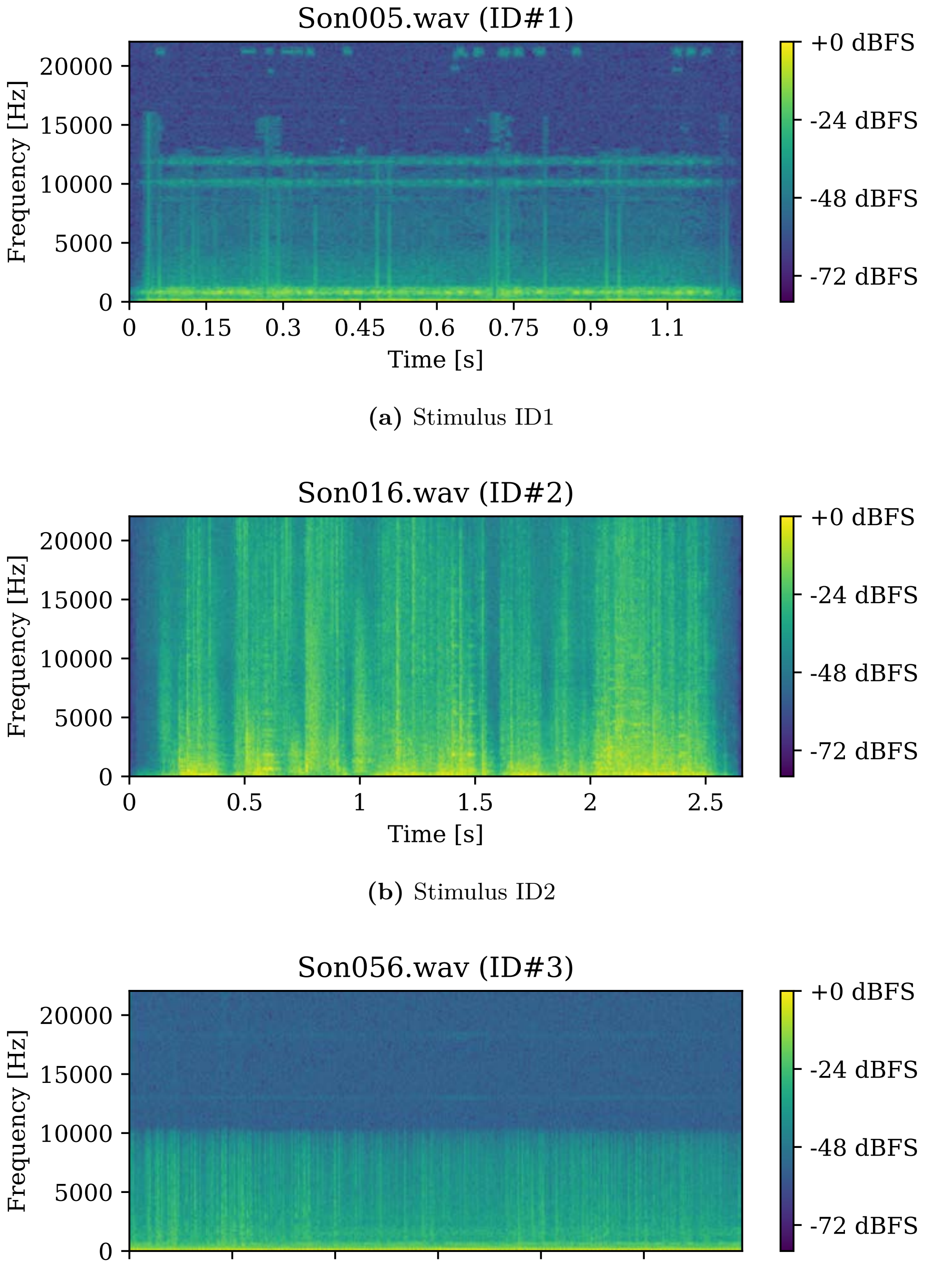
[112] Yau, J. M. ; Hollins, M. ; Bensmaia, S. J.: Textural timbre: the perception of surface microtexture depends in part on multimodal spectral cues. In: *Communicative & integrative biology* 2 (2009), Nr. 4, S. 344–346

[113] Yau, J. M. ; Olenczak, J. B. ; Dammann, J. F. ; Bensmaia, S. J.: Temporal frequency channels are linked across audition and touch. In: *Current biology* 19 (2009), Nr. 7, S. 561–566

[114] Zimmermann, M. : The nervous system in the context of information theory. In: *Human physiology.* Springer, 1989, S. 166–173

A Appendix

A.1 Spectrograms of Selected Stimuli

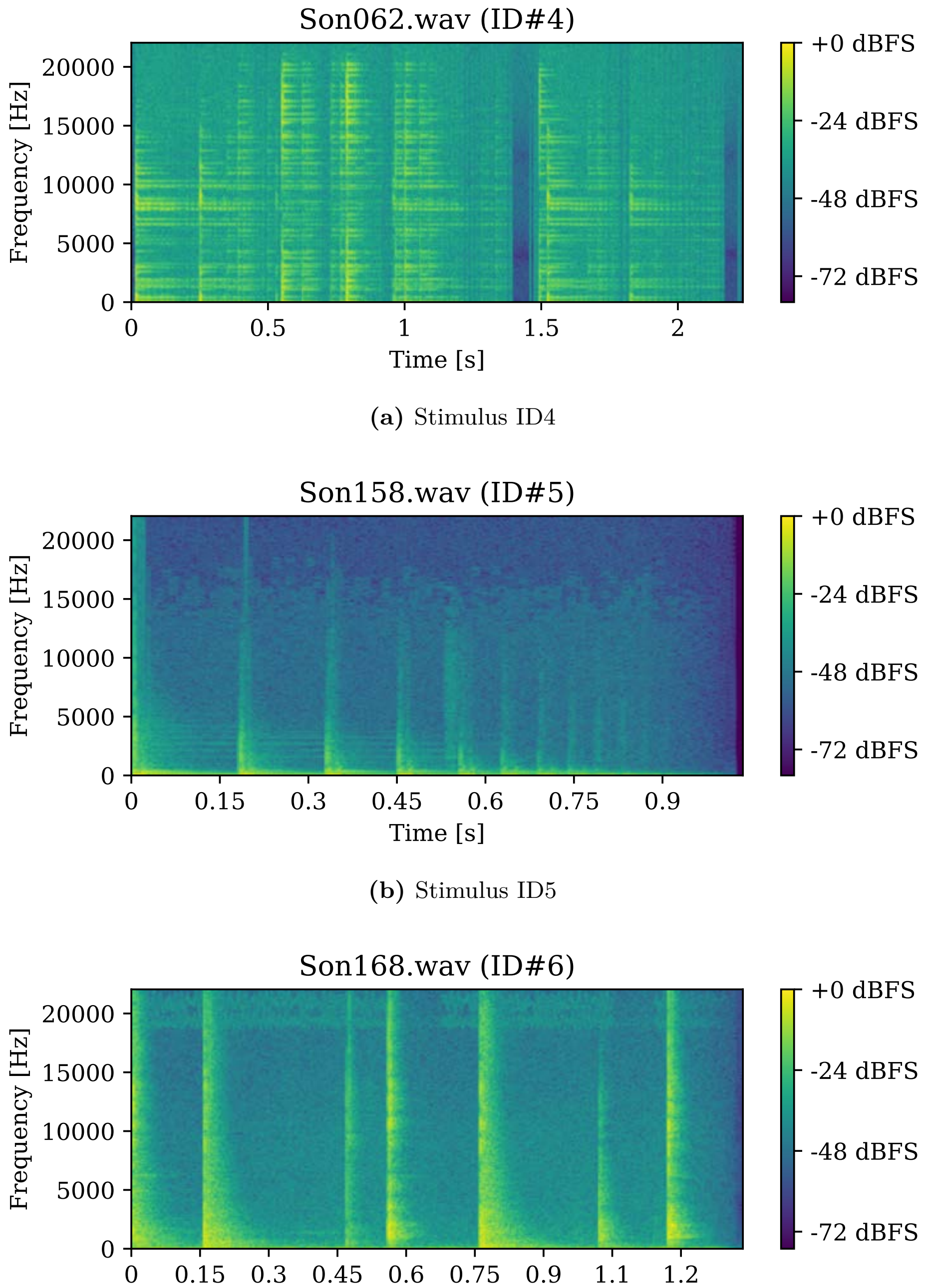


(a) Stimulus ID1

(b) Stimulus ID2

(c) Stimulus ID3

Figure A.1: Spectrograms of the audio stimuli with ID 1, 2 and 3.

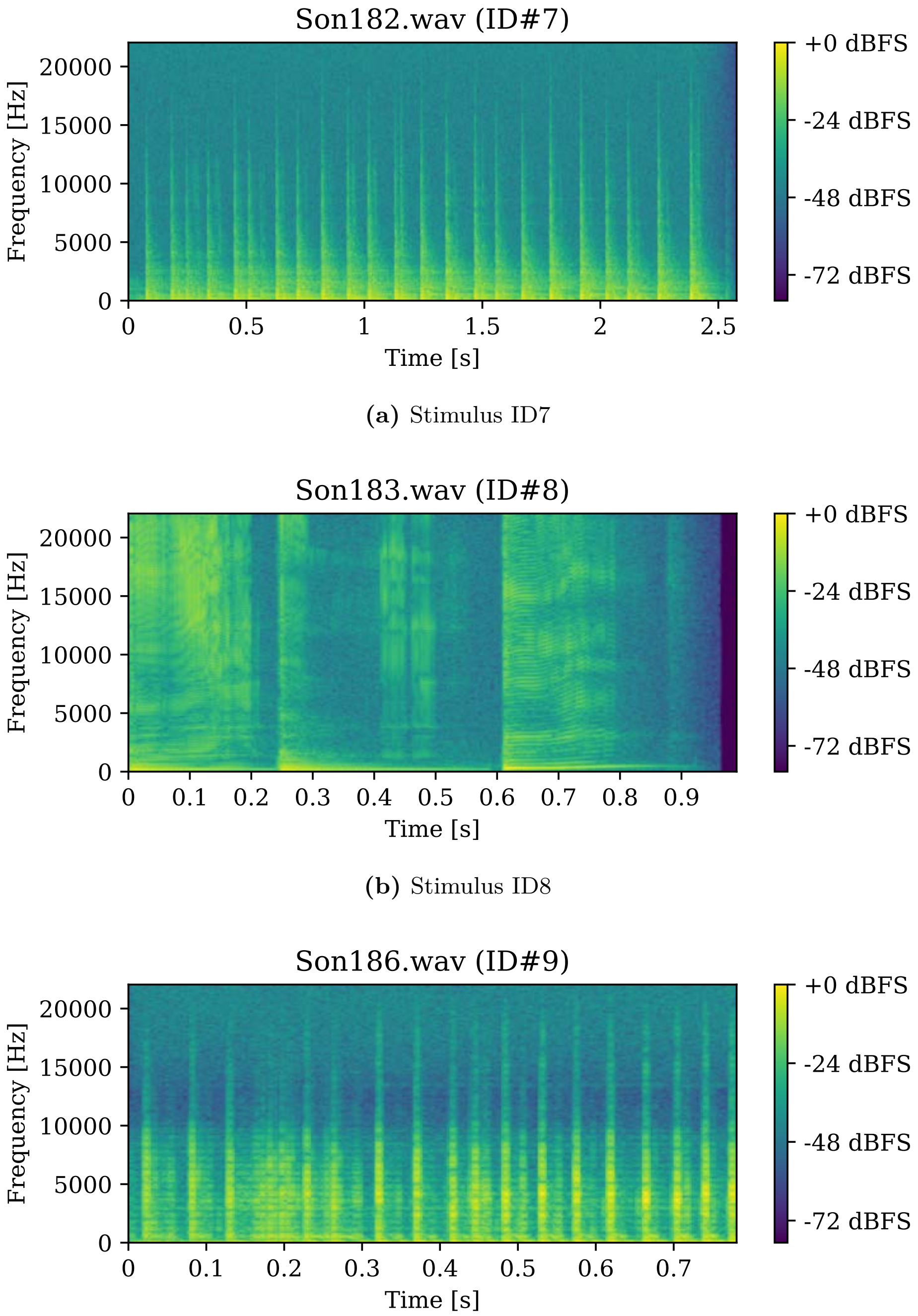


(a) Stimulus ID4

(b) Stimulus ID5

(c) Stimulus ID6

Figure A.2: Spectrograms of the audio stimuli with ID 4, 5 and 6.

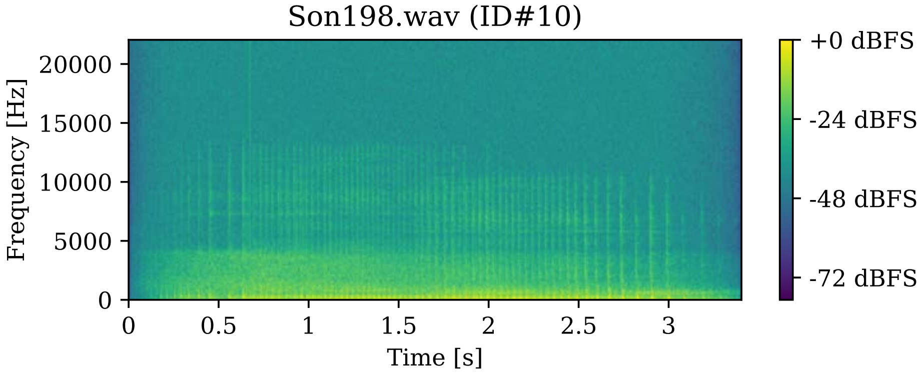


(a) Stimulus ID7

(b) Stimulus ID8

(c) Stimulus ID9

Figure A.3: Spectrograms of the audio stimuli with ID 7, 8 and 9.



(a) Stimulus ID10

Figure A.4: Spectrograms of the audio stimuli with ID 10.

The spectrograms of the WAVE (PCM) encoded audio source files all had a sampling rate of 44.1 kHz and used a single signal channel (mono audio). The parameters, used for the block-wise discrete Fourier transform (DFT) processing to create the spectrograms of the signals, were the following: window length = 512 samples; hop size = 128 samples; window type = hann; DFT bin size (zero padded) = 1024 samples. The spectrograms were created in Python, using the SciPy toolkit.

A.2 Raw Data: Coherence Ratings

Group	ID	Ratings											Mean μ	SD σ
A	1	3	6	8	5	6	6	3	8	5	6	2	5.27	1.86
	2	3	4	6	6	3	2	5	10	8	7	7	5.55	2.31
	3	5	8	8	5	8	4	10	10	8	6	8	7.27	1.91
	4	6	7	7	7	7	6	4	10	8	8	8	7.09	1.44
	5	6	8	10	7	8	10	9	10	4	8	9	8.09	1.78
	6	7	6	9	6	7	9	7	10	5	4	9	7.18	1.80
	7	4	7	8	7	8	8	8	9	7	5	7	7.09	1.38
	8	7	4	3	7	6	9	6	9	8	7	7	6.64	1.77
	9	8	2	2	4	5	7	5	10	4	6	4	5.18	2.33
	10	8	6	5	7	9	9	8	10	6	9	6	7.55	1.56
	Total												6.69	2.07
B	1	7	8	8	8	2	6	9	6	4	6	7	6.45	1.92
	2	7	9	7	8	8	10	9	1	6	7	5	7.00	2.34
	3	8	9	9	4	2	9	10	9	6	9	9	7.64	2.42
	4	4	9	6	5	4	10	8	5	4	8	6	6.27	2.05
	5	8	10	10	7	6	9	9	3	9	9	9	8.09	1.98
	6	5	8	8	10	5	10	7	8	8	9	5	7.55	1.78
	7	3	8	9	10	3	8	6	9	8	5	6	6.82	2.29
	8	6	9	9	7	6	7	6	6	9	8	5	7.09	1.38
	9	5	8	4	6	9	7	5	3	10	6	7	6.36	2.01
	10	6	4	7	7	7	9	6	6	4	8	8	6.55	1.50
	Total												6.98	2.08
C	1	8	8	5	7	8	6	7	8	8	8	6	7.18	1.03
	2	9	8	7	8	6	3	4	5	7	6	4	6.09	1.83
	3	10	10	6	6	8	7	9	10	10	10	7	8.45	1.62
	4	8	9	8	9	6	7	7	6	8	9	6	7.55	1.16
	5	9	10	8	7	8	10	10	7	9	10	8	8.73	1.14
	6	9	8	6	9	7	7	9	8	8	6	7	7.64	1.07
	7	7	9	7	8	9	8	8	8	6	6	5	7.36	1.23
	8	8	8	9	9	6	9	5	4	7	8	9	7.45	1.67
	9	10	9	8	5	7	9	10	5	10	10	7	8.18	1.85
	10	6	10	7	9	9	8	10	10	10	9	9	8.82	1.27
	Total												7.75	1.62

Table A.1: All coherence ratings for each of the ten stimulus pairs from 33 participants, equally distributed across the three groups A, B and C — summing up to a total of 220 individual coherence ratings on consecutive, and 110 on simultaneous stimuli display.

Raw Data: Coherence Rating t-Tests

Condition	ID	t	DF	SD	p
A vs. B	1	-1.395	20	1.986	0.178
	2	-1.400	20	2.436	0.177
	3	-0.373	20	2.288	0.713
	4	1.032	20	1.859	0.314
	5	0.000	20	1.973	1.000
	6	-0.455	20	1.876	0.654
	7	0.323	20	1.982	0.750
	8	-0.640	20	1.665	0.529
	9	-1.214	20	2.282	0.239
	10	1.462	20	1.604	0.159
A vs. C	1	-2.837	20	1.578	**0.010**
	2	-0.585	20	2.187	0.565
	3	-1.493	20	1.856	0.151
	4	-0.777	20	1.372	0.446
	5	-0.953	20	1.567	0.352
	6	-0.687	20	1.552	0.500
	7	-0.467	20	1.368	0.645
	8	-1.062	20	1.807	0.301
	9	-3.19	20	2.205	**0.005**
	10	-2.004	20	1.489	0.059
B vs. C	1	-1.054	20	1.618	0.304
	2	0.969	20	2.201	0.344
	3	-0.889	20	2.160	0.385
	4	-1.710	20	1.745	0.103
	5	-0.883	20	1.690	0.388
	6	-0.139	20	1.537	0.891
	7	-0.664	20	1.926	0.514
	8	-0.531	20	1.607	0.601
	9	-2.104	20	2.027	**0.048**
	10	-3.662	20	1.455	**0.002**
{A + B} vs. C	1	-2.00	31	1.78	0.054
	2	0.21	31	2.32	0.834
	3	-1.30	31	2.08	0.203
	4	-1.39	31	1.68	0.174
	5	-1.00	31	1.72	0.325
	6	-0.45	31	1.64	0.656
	7	-0.63	31	1.76	0.533
	8	-0.95	31	1.68	0.348
	9	-2.97	31	2.20	**0.006**
	10	-3.09	31	1.55	**0.004**

Table A.2: Statistical t-Test results of stimulus coherence ratings in different setup conditions (i.e. groups A, B and C). Bold p-values indicate if the level of significance ($p \leq .05$) has been met to reject the H_0 hypothesis.

A.3 Raw Data: Verbal Descriptions

ID	Microconcepts (Group A)
1	rauschen (5); rhythmisch (4); Bass (3); Club (3); moduliert; tieffrequent; energetisch; rau; Wasser; Blasen; Bewegung; rollend; schiebend; röhrend; dumpf; verzerrt; Techno; elektronisch; Wummern; wobbelig; atmosphärisch; organisch; dark; noisy;
2	aufräumen (3); wühlen (2); Kiste (2); chaotisch (2); lästig; scharf; unmelodisch; schwankend; variant; stechend; abwechselnd; viel; nervig; klappern; unruhig; perkussiv; hölzern; Werkzeug; Maschinenteile; mittlere [Frequenzanteile]; hell; dunkel; fallen;
3	Bass (4); atmosphärisch (3); knistern (2); Maschine (2); tief (2); monoton; unangenehm; Baustelle; laut; Küchengerät; elektronisch; warm; atmosphärisch; organisch; angenehm; dunkel; kratzen; Synthesizer; Strom; knacken; dröhnen; Motor; rieselnd; rauschen; harmonisch; rau;
4	fallen (2); fragil (2); leicht; verspielt; Nadel; Platte; angenehm; hell; klirren; unangenehm; springend; Ball; verzerrt; Impuls; rieselnd; flimmern; chirpen; perkussiv; synthetisch; dissonant; metallisch; scharf; Hi-Hat; Ride;
5	Basketball (3); fallen (3); springen (2); nachhallend (2); dumpf (2); Ball; Rhythmus; Sport; tief; hohl; perkussiv; weich; groß; boing; abflachend; schüttelnd; zur Ruhe kommend;
6	Tischtennis (3); scharf (3); perkussiv (2); unangenehm; klicken; wenig Druck; flach; spitz; aggressiv; impulsiv; Stift; abprallen; laut; schwankend; metallisch; scharf; dynamisch; Ping Pong; Wind; kurz; klappern; noisy; kalt; klackern; hoher Ton; Blech; blasen;
7	rhythmisch (2); Trommel (3); drum (3); rollend (2); Musikinstrument; musikalisch; Spannung; Unruhe; warm; hohl; angenehm; sequenziell; hallig; schwankend; melodisch; schüttelnd; flächig; oszillierend; perkussiv; tonaler Anteil; leicht; verstimmt; dissonant; hochgestimmt; ethnic; tribal; straf; hart;
8	Gitarrensaiten (3); Sprungfeder; Gummiband; gezupft; hüpfen; schwankend; lästig; moduliert; lustig; kratzend; ziehend; Bienensummen; tonal; perkussiv; ethnic; synthetisch;
9	scharf (2); kratzen (2); schaben; Tiergeräusch; Nashorn; schnarchend; rising; organisch; leicht bedrohlich; rollend; oszillierend; Angst; unangenehm; Brubbeln; über Raster streifen; Propeller; Schlagbohrer; Motorrad; engine; metallisch; Surren; rau;
10	rau (2); Motorrad; motorisch; engine; Bewegung; Stühle knarzen; Entschleunigung; gefiltertes Rauschen; verrauscht; unangenehm; Kettensäge; Presslufthammer; ausgeliefert; rollend; oszillierend; warm; dumpf; organisch; schüttelnd; atmend; tief; Quietschen; scharf;

Table A.3: This table contains the perceptual descriptions for the ten audio stimuli for Group A. The first column (ID) references the stimulus ID. A total of 308 microconcepts were extracted from this group.

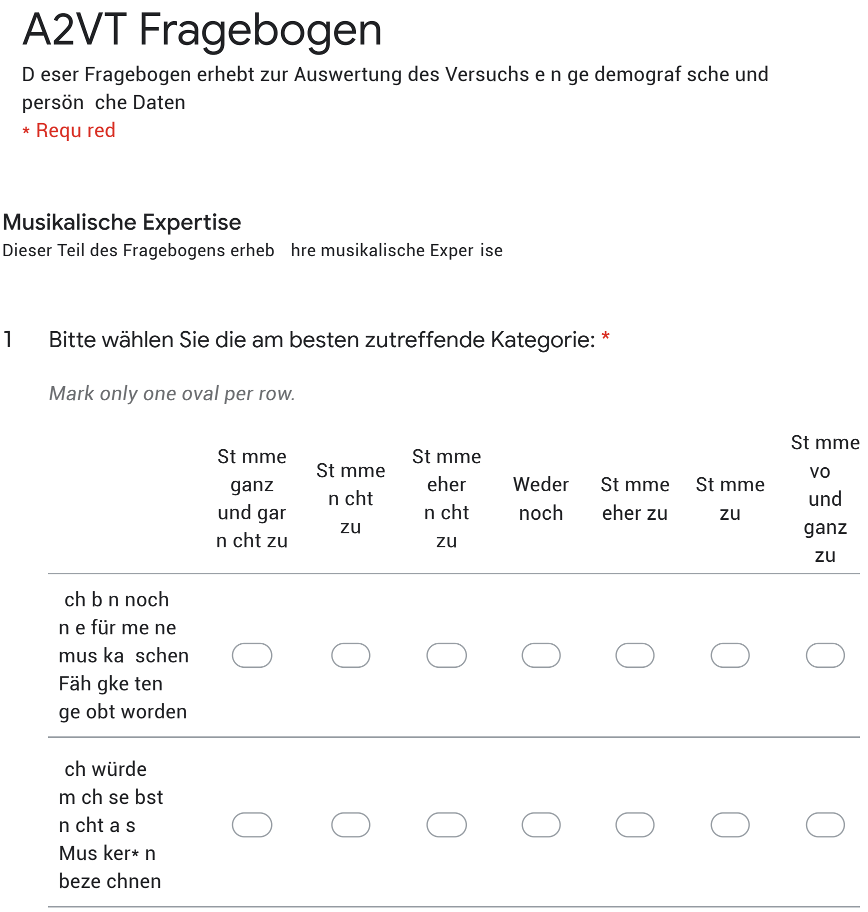
ID	Microconcepts (Group B)
1	schwingend (2); tief[frequent] (2); starten (2); anschwellen; mitschwingen; Konstanz; Beständigkeit; gleichmäßig; kurze Vibration; mehrere Pulse; angenehm; blubbern; wackeln; Gallopieren; sanft; aufrüttelnd; steigernd; abklingend; kurz; kräftig; Motor; holprig; tief; kalt;
2	unregelmäßig (2); instabil (2); wiederholend (2); rhythmisch (2); Impuls (2); Anleitung; Wind; ruckeln; rütteln; mitelllange Vibration; unterschiedlich stark; groovehaft; mittelfrequent; perkussiv; sanft; Echo; entfernt; abebben; anregend; [im] Takt; Morse Code; Kommunikation; kleine Elektroschocks; schwankend; niedriger Pegel; gleichmäßig schwingend; überlegt; aufrührend; wachsam; wellenartig; leichte Peaks;
3	Vibration (3); Maschine (2); lang [gestreckt] (2); stark (2); Motor; Pegel halten; abebben; laut; Baustellenlärm; wiederholend; kaum trennbare Impulse; Abfolge von 808 kick drums; durchdringend; konstant; druckvoll; rütteln; Durchbruch; Steigerung; schlafende Katze; tief; leicht; dunkel; stabil; pulsierendes Schwingen; hohe Energie; unregelmäßige Pulse; überlegt; antreibend; kräftig; wellenartig; lang; entspannend; Wind an Deck eines Schiffes;
4	leicht (3); sanft (2); tastend (2); schwach (2); Takt (2); zurückhaltend; kaum wahrnehmbar; tieffrequente Vibrationen; kürzere Pulse; eindeutiger Rhythmus; koordiniert; strukturiert; Aufmerksamkeit; Insistent; trocken; Berührung; angenehme Vibrationsstärke; kleine unterschiedlich starke zufällige Pulse; subtil; merkbar; natürlicher Charakter; Wassertropfen; verhalten; zaghaft; leise; bescheiden; Herzschlag; hoch; wiederholend;
5	kurz (3); schnell (3); [Tischtennis-]Ball (3); Impulse (3); Abklingen (2); fallen; hopsen; hüpfend; spontan; plötzlich; vorbeiziehen; tieffrequente Vibration; abfolgende Pulse; triolisch; längere release Zeit; ausklingen; abnehmend; ausschwingend; abfallend; frech; leicht gedämpfter Impact; schnell steigende Schwingung; abrupter Stopp; anstossendes Gefühl; hart;
6	Puls (5); schwach (2); schnell; vorsichtig; zittrig; nervös; tieffrequente Vibration; Impulse; abfolgenden Pulse; triolisch; längere release Zeit; ausklingen; pulsierend; kurz; körperlich; lebendig; Herzton; rhythmisch; angenehm; kleine unterschiedlich starke unregelmäßig erscheinende Schwingungen; leichter Windhauch; mittelstark; regelmäßig; unregelmäßig; aufsteigend; hell;
7	Rhythmus (2); vorbeifahrender Zug; anschwellen; einschwingen; ansteigend; stützend; stark; konsistent; gleichmäßig verteilt; kontrolliert; schwach; lange; tieffrequente vibration; Puls Abfolge; schnell; insistent; periodisch; hell; Motor-ähnlich; Stakkato; Geballer; holprig; ruckeln; hohe Intensität; Gefahr; Bodenwellen; pulsierend; robotisch; statisch; U-Bahn Fahrt; gleichbleibende Schwingung; niedriges Energie-Level; klein; subtil; unregelmäßig;
8	Puls (3); kurz (2); prägnant; perkussiv; schwach; tieffrequente Vibration; abfolgende Pulse; zeitlich getrennt; stolpern; stoppen; Ende; Abbruch; Ton-Abfolge; tief; hoch; Schlussakkord; tiefer werdend; groß; dreimaliges Vibrieren; näher zusammen; mittlere Stärke; scharf; diskret; ähnlicher Pegel; etwas unregelmäßig; kleine Nadelstiche; Klopfen an einer Tür; adequate Länge; angemessene Impulsstärke;
9	stark (4); durchgehend (3); Aufmerksamkeit; Konzentration; Start; zittern; Beben; vorbeiziehen; durchschütteln; intensiv; mittelfrequente Vibration; mittellange Dauer; kurz; schnell; hochfrequent; dringend; Handy-Vibration; Vibrationsalarm; anregend; gruselig; erschreckend; [stärker] werdend; anschwellen; Intensität steigernd; gleichmäßig; Die Sequenz vor dem Drop eines Tracks;
10	lange (4); angenehm; entspannend; Meer; langsam; durchgehend; mittelfrequente Vibration; geringe Intensität; entfernt; intensiv; Beben; instabil; anschwellen; abebben; Chaos; Rauschen; ansteigend; in die Länge gezogen; ausgedehnt; unregelmäßig stark; Autofahren; ruckeln; natürliche Schwankungen; Subwoofer in einem Club; gleichmäßig; Intensität nimmt gleichmäßig zu und ab; die Intensität fällt geringer ab; bzw. Intensität nimmt stärker zu; aufschaukelnd; instabil; zitternd; hochfrequent; [am Ende] stärkere Vibrationen;

Table A.4: This table contains the perceptual descriptions for the ten tactile stimuli for Group B derived from the corresponding audio stimuli used in Group A. The first column (ID) references the stimulus ID. A total of 385 microconcepts were extracted from this group.

ID	Microconcepts (Group C)
1	Club (3); pulsierend (2); Bass (2); dunkel (2); Puls; treibend; Techno; Adrenalin; angenehm; Keller voller Maschinen; mechanisch; warm; rau; motivierend; eindringlich; rauschig; schnell; dumpf; gedämpft; Unterwasser; Tauchgang; aggressiv; tanzbar; loop-haft; atonal; perkussiv; brachial; boomend; dröhnend; wummern; Kraft; grummeln; brummen; [Schlag] am Ende; brechend;übersteuert; rhythmisch pochend; zügig schreitender Ablauf;
2	unangenehm (3); Schublade (2); wühlen (2); Klirren (2); chaotisch (2); rumräumen; ungeordnet; organischer Rhythmus; vielschichtig; nervös; rumpelig; transient; rauschhaft; stressig; hektisch; dumpf; kratzen; schaben; holzig; hohl; chaotisch; hell; Ordnung; tief; mechanisch; Dose öffnen; Roboterschwein; tierisch; gurgelnd; brachial; diffus; grob; störend; inharmonisch;
3	Spannung (3); Drone (2); knistern (4); vibrierend (2); tief (2); brummen (2); Steigerung; elektrisierend; elektrisch; Presslufthammer; rau; angenehm; dumpf; abstrakt; tonal; wuchtig; konstante Spannung; mechanisch; dunkel; U-Boot; Tiefsee; Wohlempfinden; beschützt; kratzen; hohl; Synthesizer; dröhnend; wabern; großer Ventilator; Lichtschwert; maschinenartig; bebend; gleichmäßig; hell; zittrig;
4	metallisch (4); angenehm (2); klirren (2); elektrisch (2); digital (2); hoch[frequent] (2); metallisch; hohl; feste Schläge; kaputt; schmerzhaft in den Höhen; [angenehm] in den Mitten und Tiefen; dumpf; hell; schrill; definiert; kalt; kühl; unangenehm; klar; wenige Frequenzen; kurzlebig; einmalig; subtil; kleines Objekt fällt in ein Glas; klimpernd; unnatürlich; inharmonisch; perkussiv; gläsern; zerbrechlich; Drums; Zahnarzt; hallig; transient; präzise; atonal;
5	[Basket-]Ball (5); hohl (2); schwer (2); dumpf (2); fallendes [Objekt] (2); rund (2); abklingend (2); auspendelnd; natürlich; luftig; kurz; transient; angenehm; werfen; hörbarer Resonanzkörper; Aufprall; federnd; springend; bouncend; tieffrequent; gummiartig; physischer Ton; hart; hell; stark; laut; warm; definiert; hallend;
6	Tischtennis (3); unangenehm (3); leicht (3); blechern (2); hoch[frequent] (2); hart (2); Ping Pong; winzig; feste Schläge; unnatürlich; knacken; hell; nicht im Takt; fallende Dinge auf ein Metalldach; klopfen; hohl; springend; gefilterter Noise; kurze Attackzeit; drums; Aufschlag; federnd; vielschichtig; gewisser Rhythmus; klar; schmales Spektrum; schnell; klein; transient; präzise; rhythmisch; kräftig; stark;
7	trommeln (5); hohl (3); hell (2); Schlag (2); perkussiv (2); rhytmisch (2); Tamburin ohne Schellen; direkt; federnd; kontinuierlich; wohlklingend; Spannung erzeugend; wenig klare Tiefen; zeitlich abfallend; abprallend; organisch; leicht; weich; Wiederholung; geplant; schnelle Abfolge; Tropfen in einen Eimer; boxy; nicht tonal; echt; langsam beschleunigend; deutlich; leichte Steigerung; leichtes Crescendo;
8	gummiartig (3); gezupfte Saite (3); unangenehm (2); inharmonisch (2); hell (2); vibrierend; variierend; flexibel; elastisch; nah; hoch; zeitlich abklingend; einmalig; springender Synthesizer; tropfend; quietschend; hohl; lustig; kratzen; sehr definiert; schwingend; schnarrend; leicht; hoher Ton; eher hochfrequent; scharfe Höhen; tonal; rasselnd; kurz;
9	vibrierend (2); konstant (2); unnatürlich (2); dumpf (2); knattern (2); permanent; auf Dauer unangenehm; mechanisch; hämmernd; Spannung aufbauend; pulsierend; breiter Spektralanteil; schnurrende Katze; angenehm; knattern; rau; kratzen; holpern; hart; vibrierend; direkt; perkussiv; hoher Ton; kleiner digitaler Hall; ansteigend; Reißverschluss; start eines Motors; brummender Motor; Flugzeugturbine; hell; gluckern; rumpelnd; scharf gegen Ende; Überlagerungen;
10	Motor (5); angenehm (3); rauschend (3); näher kommend (3); motorisch (2); knattern (2); vibrierend (2); natürlich; organisch; metallisch; viel Raum; beschleunigend; einmalig; abfallend; artifiziell; ratternd; relativ lang; breitbandig; hell; halt machend; wellig; rund; brummend; klopfen; Nachhall; stinkend; penetrant; Start und Stop einer Maschine; Wind; abklingen; mechanisch;

Table A.5: This table contains the perceptual descriptions for the ten, bimodal audio-tactile stimuli pairs for Group C. The second column (ID) references the stimulus ID. A total of 427 microconcepts were extracted from this group.

A.4 Demographic & Musical Expertise Questionnaire



A2VT Fragebogen

D eser Fragebogen erhebt zur Auswertung des Versuchs e n ge demograf sche und persön che Daten

* Requ red

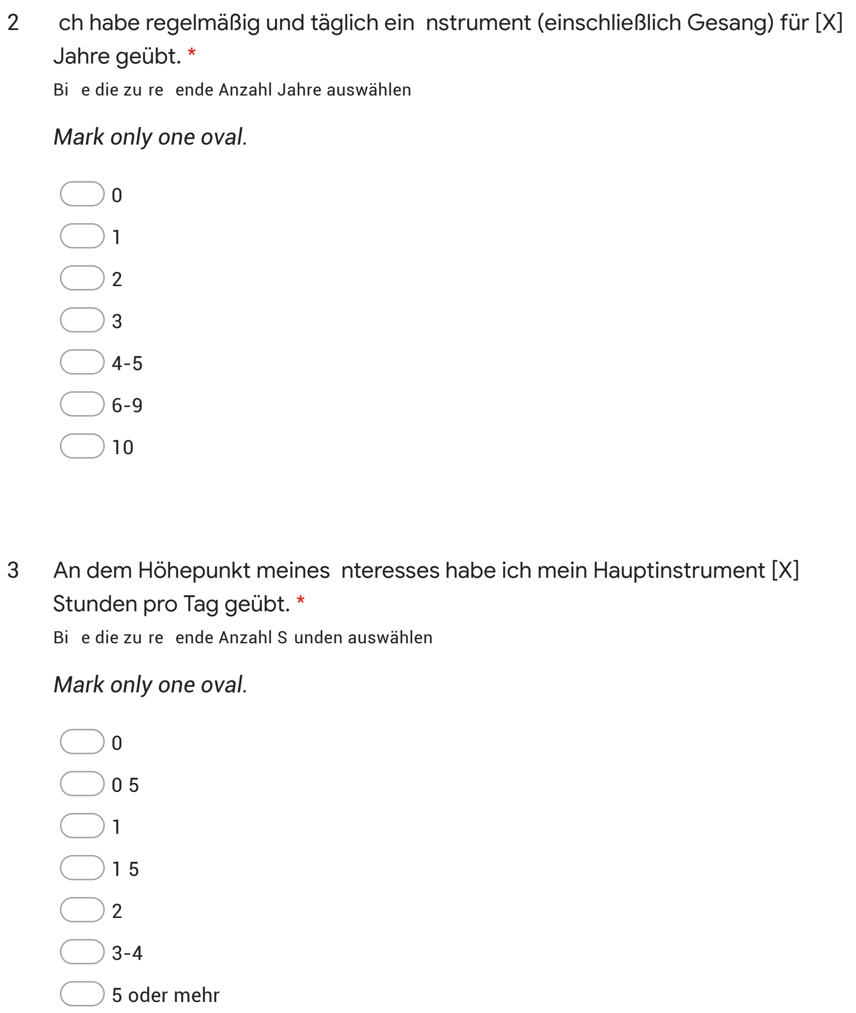
Musikalische Expertise

Dieser Teil des Fragebogens erheb hre musikalische Exper ise

1 Bitte wählen Sie die am besten zutreffende Kategorie: *

Mark only one oval per row.

	St mme ganz und gar n cht zu	St mme n cht zu	St mme eher n cht zu	Weder noch	St mme eher zu	St mme zu	St mme vo und ganz zu
ch b n noch n e für me ne mus ka schen Fäh gke ten ge obt worden	○	○	○	○	○	○	○
ch würde m ch se bst n cht a s Mus ker* n beze chnen	○	○	○	○	○	○	○



2 Ich habe regelmäßig und täglich ein Instrument (einschließlich Gesang) für [X] Jahre geübt. *

Bitte die zutreffende Anzahl Jahre auswählen

Mark only one oval.

○ 0

○ 1

○ 2

○ 3

○ 4-5

○ 6-9

○ 10

3 An dem Höhepunkt meines Interesses habe ich mein Hauptinstrument [X] Stunden pro Tag geübt. *

Bitte die zutreffende Anzahl Stunden auswählen

Mark only one oval.

○ 0

○ 0 5

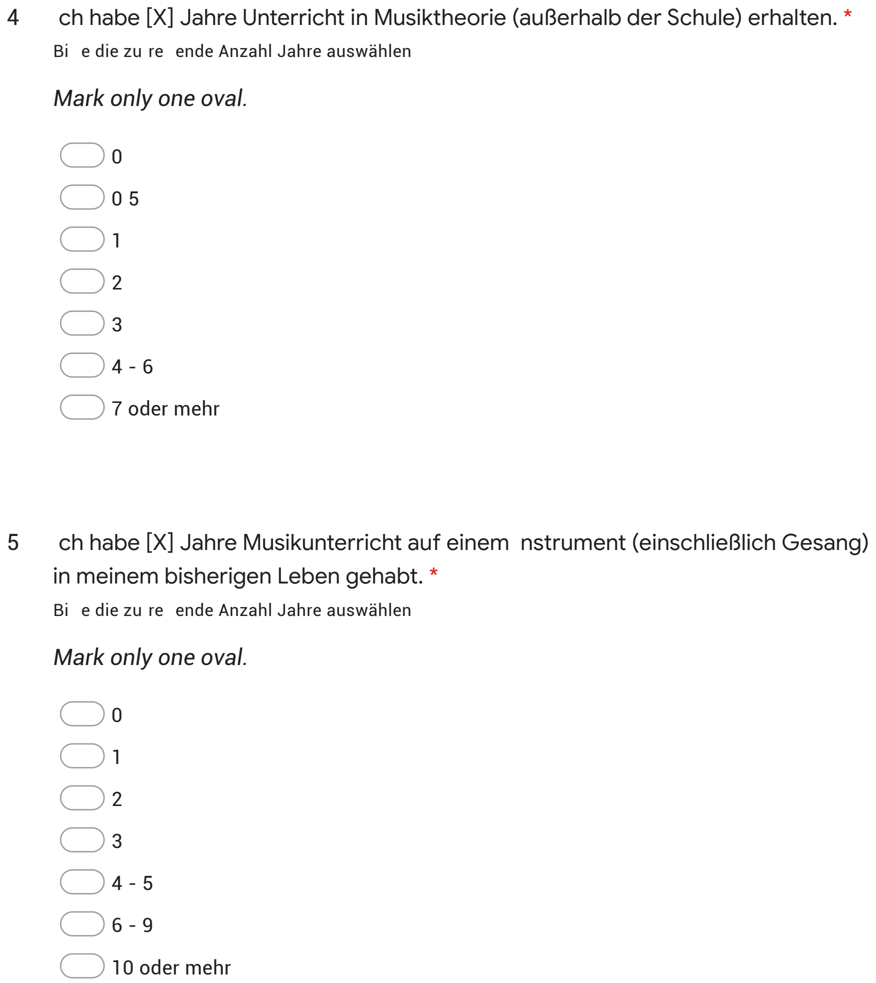
○ 1

○ 1 5

○ 2

○ 3-4

○ 5 oder mehr



4 ch habe [X] Jahre Unterricht in Musiktheorie (außerhalb der Schule) erhalten. *

Bitte die zutreffende Anzahl Jahre auswählen

Mark only one oval.

○ 0
○ 0 5
○ 1
○ 2
○ 3
○ 4 - 6
○ 7 oder mehr

5 ch habe [X] Jahre Musikunterricht auf einem nstrument (einschließlich Gesang) in meinem bisherigen Leben gehabt. *

Bitte die zutreffende Anzahl Jahre auswählen

Mark only one oval.

○ 0
○ 1
○ 2
○ 3
○ 4 - 5
○ 6 - 9
○ 10 oder mehr

6 ch kann [X] verschiedene nstrumente spielen. *

Bi e die zu re ende Anzahl ns rumen e auswählen

Mark only one oval.

○ 0

○ 1

○ 2

○ 3

○ 4

○ 5

○ 6 oder mehr

7 Berufsstatus *

Bi e den ak uell zu re enden Beru ss a us auswählen

Mark only one oval.

○ Schü er* n

○ Student* n

○ Vo ze t angeste t

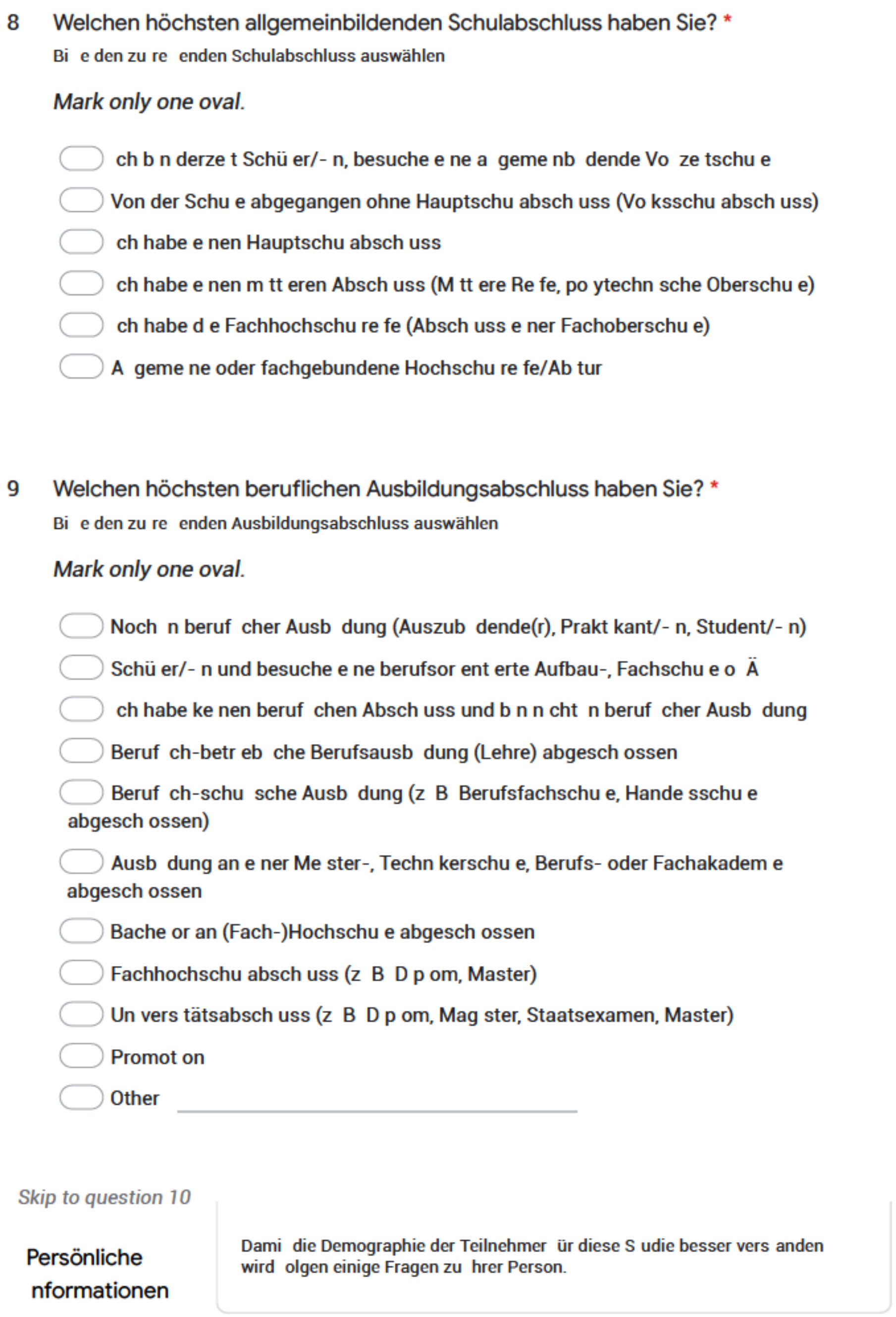
○ Te ze t angeste t

○ Fre beruf ch tät g

○ m E ternschutz

○ Arbe ts os

○ m Ruhestand (Rentner* n)



8 Welchen höchsten allgemeinbildenden Schulabschluss haben Sie? *

Bi e den zu re enden Schulabschluss auswählen

Mark only one oval.

○ ch b n derze t Schü er/- n, besuche e ne a geme nb dende Vo ze tschu e

○ Von der Schu e abgegangen ohne Hauptschu absch uss (Vo ksschu absch uss)

○ ch habe e nen Hauptschu absch uss

○ ch habe e nen m tt eren Absch uss (M tt ere Re fe, po ytechn sche Oberschu e)

○ ch habe d e Fachhochschu re fe (Absch uss e ner Fachoberschu e)

○ A geme ne oder fachgebundene Hochschu re fe/Ab tur

9 Welchen höchsten beruflichen Ausbildungsabschluss haben Sie? *

Bi e den zu re enden Ausbildungsabschluss auswählen

Mark only one oval.

○ Noch n beruf cher Ausb dung (Auszub dende(r), Prakt kant/- n, Student/- n)

○ Schü er/- n und besuche e ne berufsor ent erte Aufbau-, Fachschu e o Ä

○ ch habe ke nen beruf chen Absch uss und b n n cht n beruf cher Ausb dung

○ Beruf ch-betr eb che Berufsausb dung (Lehre) abgesch ossen

○ Beruf ch-schu sche Ausb dung (z B Berufsfachschu e, Hande sschu e abgesch ossen)

○ Ausb dung an e ner Me ster-, Techn kerschu e, Berufs- oder Fachakadem e abgesch ossen

○ Bache or an (Fach-)Hochschu e abgesch ossen

○ Fachhochschu absch uss (z B D p om, Master)

○ Un vers tätsabsch uss (z B D p om, Mag ster, Staatsexamen, Master)

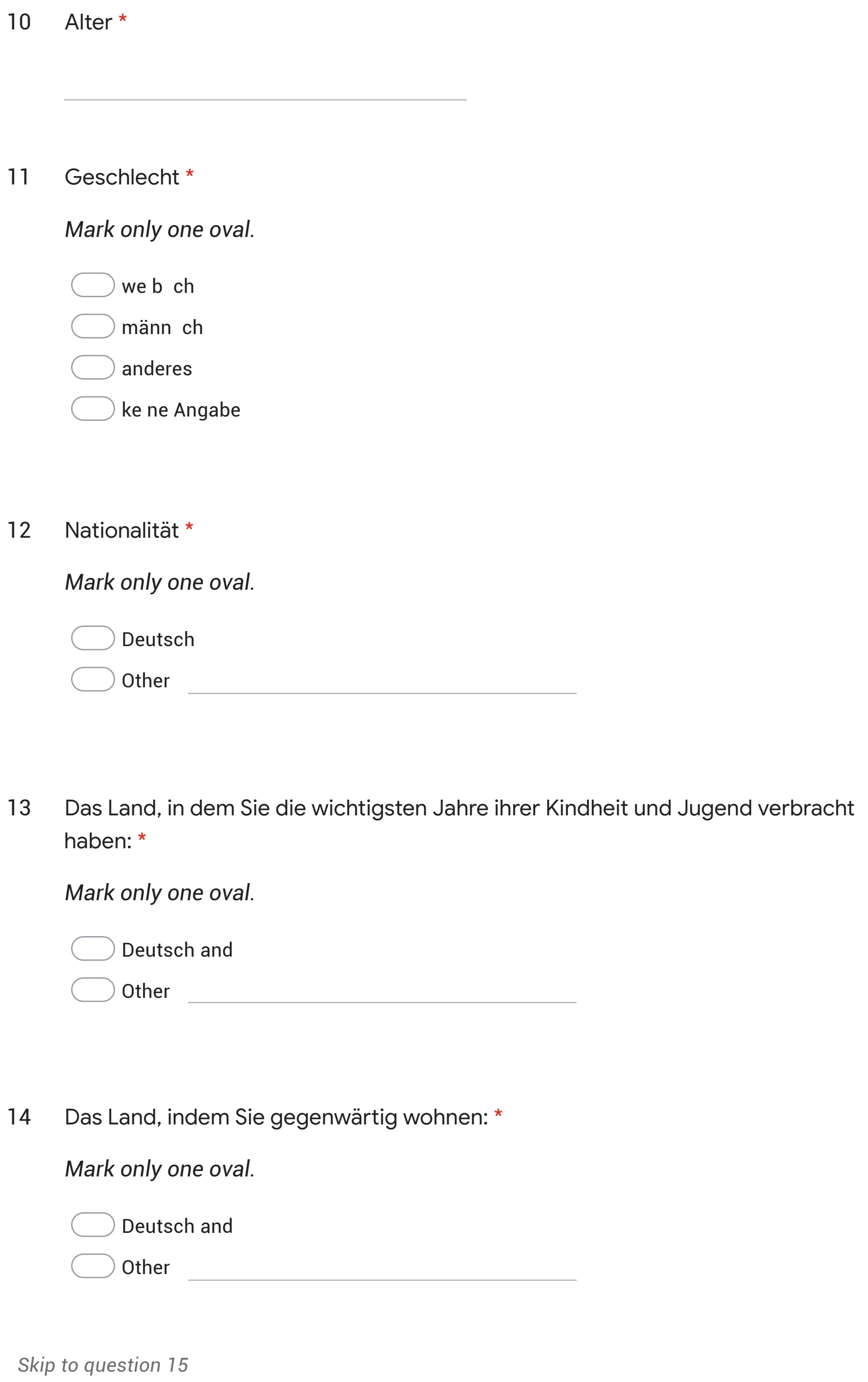
○ Promot on

○ Other ______________________

Skip to question 10

Persönliche nformationen

Dami die Demographie der Teilnehmer ür diese S udie besser vers anden wird olgen einige Fragen zu hrer Person.



10 Alter *

____

11 Geschlecht *

Mark only one oval.

○ we b ch
○ männ ch
○ anderes
○ ke ne Angabe

12 Nationalität *

Mark only one oval.

○ Deutsch
○ Other ______________________________

13 Das Land, in dem Sie die wichtigsten Jahre ihrer Kindheit und Jugend verbracht haben: *

Mark only one oval.

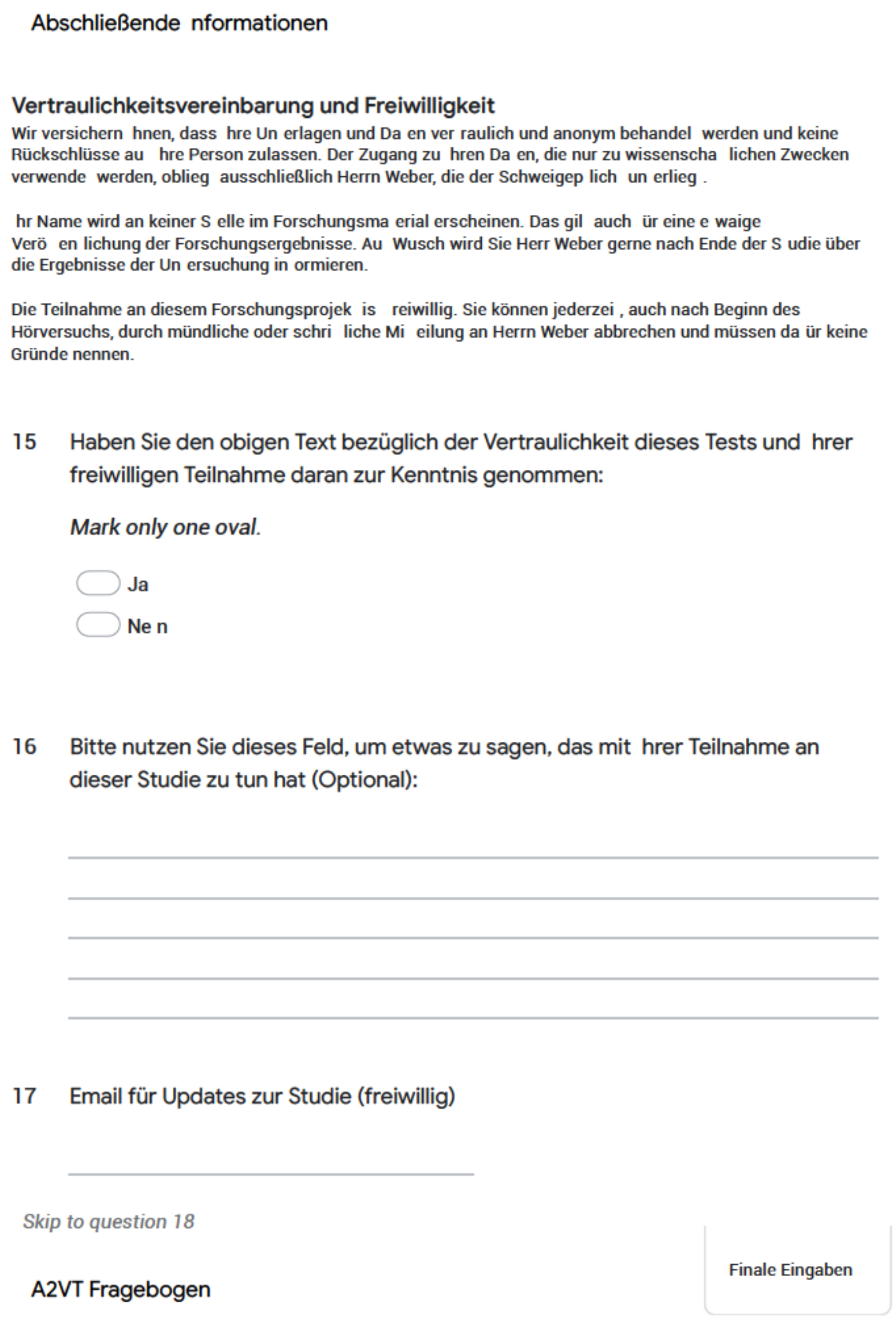
○ Deutsch and
○ Other ______________________________

14 Das Land, indem Sie gegenwärtig wohnen: *

Mark only one oval.

○ Deutsch and
○ Other ______________________________

Skip to question 15



Abschließende nformationen

Vertraulichkeitsvereinbarung und Freiwilligkeit

Wir versichern hnen, dass hre Un erlagen und Da en ver raulich und anonym behandel werden und keine Rückschlüsse au hre Person zulassen. Der Zugang zu hren Da en, die nur zu wissenschaftlichen Zwecken verwende werden, oblieg ausschließlich Herrn Weber, die der Schweigep lich un erlieg .

hr Name wird an keiner S elle im Forschungsma erial erscheinen. Das gil auch ür eine e waige Veröffen lichung der Forschungsergebnisse. Au Wusch wird Sie Herr Weber gerne nach Ende der S udie über die Ergebnisse der Un ersuchung in ormieren.

Die Teilnahme an diesem Forschungsprojek is reiwillig. Sie können jederzei , auch nach Beginn des Hörversuchs, durch mündliche oder schriftliche Mitteilung an Herrn Weber abbrechen und müssen da ür keine Gründe nennen.

15 Haben Sie den obigen Text bezüglich der Vertraulichkeit dieses Tests und hrer freiwilligen Teilnahme daran zur Kenntnis genommen:

Mark only one oval.

○ Ja

○ Ne n

16 Bitte nutzen Sie dieses Feld, um etwas zu sagen, das mit hrer Teilnahme an dieser Studie zu tun hat (Optional):

____

____

____

____

____

17 Email für Updates zur Studie (freiwillig)

____

Skip to question 18

A2VT Fragebogen

Finale Eingaben