Technische Universität Dresden

Auditory-Tactile Music Perception

Sebastian Merchel

von der Fakultät Elektrotechnik und Informationstechnik
der Technischen Universität Dresden

zur Erlangung des akademischen Grades eines

Doktoringenieurs

(Dr.-Ing.)

genehmigte Dissertation

Vorsitzender:	Prof. Dr. phil. nat. habil. Ronald Tetzlaff
Gutachter:	Dr.-Ing. Ercan Altinsoy Prof. Dr. rer. nat. Armin Kohlrausch
Tag der Einreichung:	30. 08. 2013
Tag der Verteidigung:	22. 05. 2014

Please cite this dissertation as:

Merchel, S. (2014). Auditory-Tactile Music Perception (ISBN: 978-3-8440-3161-4),
Shaker Verlag, Germany.

Preface

My research would not have been possible without the support of *Dr. Ercan Altinsoy*, who has been my supervisor in all things concerning auditory-tactile perception, experimental methodology, and creative thinking. Thank you! I'm very grateful to *Prof. Ute Jekosch* for believing in me and giving me the opportunity for this work. I would also like to thank *Prof. Armin Kohlrausch* for his expertise and support.

Thanks to all of my colleagues and students at the Institute of Acoustics and Speech Communication, Chair of Communication Acoustics. You created a wonderful atmosphere, and I truly enjoyed working with you! I am particularly thankful for lively debates with *Jürgen Landgraf, Maik Stamm,* and *Robert Rosenkranz*. A special thanks to *Anna Schwendicke*, who has written her Diploma thesis under my supervision on auditory-tactile loudness perception. Some of the results are included in Section 5.5. A special thanks to *Margitta Lachmann* for her support with the preparation of the images. Thanks to all of the *participants* in my sometimes weird experiments. *Anne Postler*, thank you for your thorough proofreading. Thanks to my *fellow students* in the 'Group for Academic Arithmetic Sunshine' for your encouragement and long-lasting fellowship.

To my parents: thank you for making me what I am. *To all of my friends*: thank you! I will be back in the real world, I promise.

To my wife and children: I love you, and I am grateful for your support and understanding. Being with you has been the best distraction of all!

This thesis was written in the Faculty of Electrical and Computer Engineering at the Dresden University of Technology, Chair of Communication Acoustics. I am truly grateful for the opportunity to work and teach in such an exciting scientific environment.

Dresden, Mai 2013 *Sebastian Merchel*

Abstract

Sound and vibrations are often perceived via the auditory and tactile senses simultaneously, e.g., in a car or during a rock concert. Even in a concert hall or a church, sound can excite suprathreshold vibrations in the ground or seats. If concert recordings are played back through headphones, this vibratory information is missing to date. The same holds true in the majority of cases for reproduction with multimedia or high-fidelity systems.

This thesis extends our understanding of the coupled perception of sound and vibration using the example of auditory-tactile music perception. The capabilities and limitations of both modalities are compared first. Unfortunately, particularly for the perception of vibrations at low levels, only limited knowledge exists to date. Therefore, the frequency discrimination and intensity perception of whole-body vibrations is investigated in several experiments. The most evident difference between both modalities is the dramatically reduced ability to distinguish between vibration frequencies in the tactile domain. Another important difference is the steeper growth of the perceived magnitude for touch compared to hearing. A new perceptually motivated measurement for the perceived vibration magnitude M is defined to represent human vibration intensity perception, comparable to auditory loudness N. Additionally, cross-modal effects are considered, e.g., the influence of whole-body vibrations on loudness perception. An auditory-tactile loudness illusion is proven.

In the second part of this work, it is investigated whether sound-induced whole-body vibrations influence the quality of a concert experience. Vibrations are found to play a significant role in the perception of music. The fundamental knowledge gained in the first part, is used to develop and evaluate various perceptually optimized approaches to generate vibrations from music sequences. The results can be applied to improve audio reproduction systems or even concert halls.

Please cite this dissertation as:

Merchel, S. (2014). Auditory-Tactile Music Perception (ISBN: 978-3-8440-3161-4), Shaker Verlag, Germany.

Contents

1

Introduction

1.1 General Introduction

Sound and vibration perception often occurs in the context of live music experiences. For instance, think of the bass during a rock concert or hearing and feeling a church organ while sitting on a wooden pew. Even in classical concerts, perceivable vibrations can be excited by low-frequency instruments, particularly if the instruments are played by a large orchestra. However, sound and vibrations are typically integrated into one multimodal event: the concert experience. Vibrations can be perceived at up to several hundred hertz through mechanoreceptors in the skin (tactile sense). Simultaneously, sound is heard via the auditory system down to a lower limiting frequency of approximately 20 Hz. The considerable overlap between the frequency ranges of both sensory systems is illustrated in Figure 1.1. Additionally, the fundamental frequencies of typical orchestral instruments are shown. Many of these instruments, e.g., double bass or kettledrum, can be both heard and felt. If concert recordings are played back through headphones, this vibratory information is missing. The same holds true in the majority of cases for reproduction with multimedia or high-fidelity systems, which may be due to low reproduction levels, vibration-attenuating furniture, or the limited frequency range of conventional loudspeakers. The question that arises is whether the perceived quality of a concert experience is influenced by the presence of body vibrations.

The audio signal on todays DVDs already contains an additional channel for low-frequency effects (LFE), which is intended for reproduction using a subwoofer. However, the generation of tactile components is still very restricted. Enhancement of such systems might be possible using a vibration actuator coupled to a surface in contact with the listener, e.g., an electrodynamic shaker mounted to a seat. The positive effects of custom-made seat vibrations have been already demonstrated for action-oriented DVD movies [184]. Furthermore, in some cases, sound and vibrations are applied in therapeutic scenarios, e.g., to increase the level of relaxation [13, 82] or to relieve pain [14]. Because there are no custom vibration tracks on any current media, one might ask whether it is feasible to generate music-related vibrations directly from an audio signal. How does the perception of the concert experience change if vibrations are added? Is such an audio-driven vibration generation approach applicable for different music genres, e.g., rock, pop, and classical? Which frequency ranges should be reproduced? Is there a preferred vibration level? What are the perceptual differences between the auditory and tactile senses, and should these differences be considered? Which requirements should be met by the vibration reproduction system?

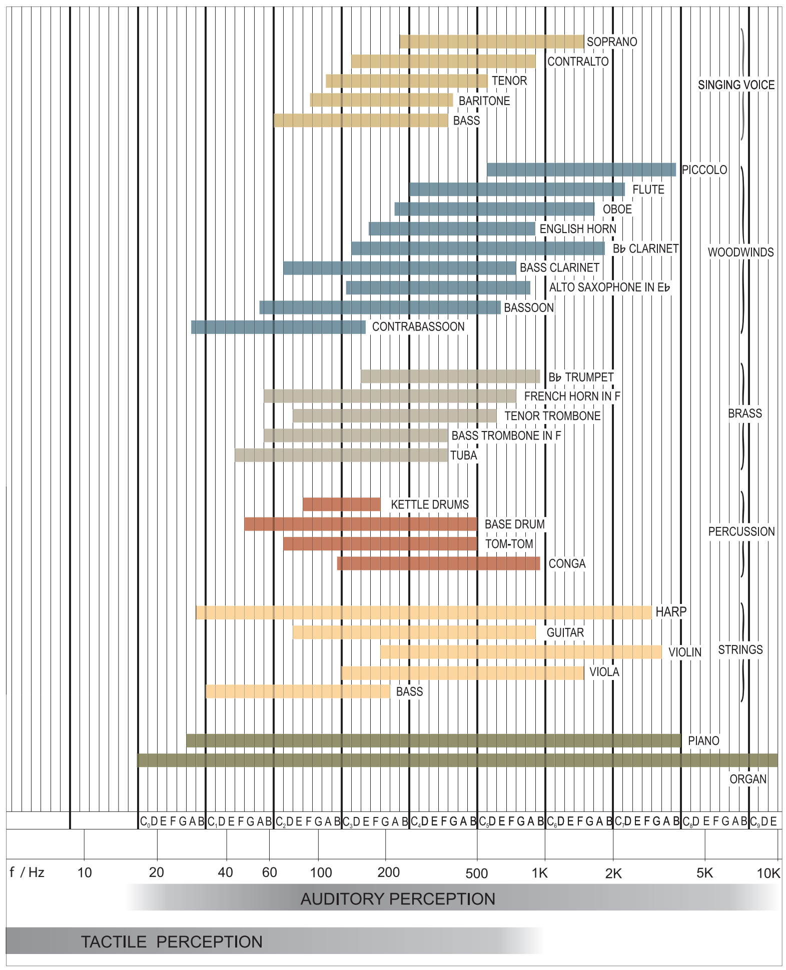


Fig. 1.1. Frequency ranges of auditory and tactile perception and fundamental frequencies of instruments and voices adapted from [1]. The frequency ranges for transients, noise, and harmonics are not included.

1.2 Objectives of the Thesis

The aim of this thesis is to gain a better understanding of the auditory-tactile perception of music. Through these insights, recommendations for the generation and reproduction of music-related vibrations will be possible. The experiments described in the following provide the necessary groundwork to create vibrations from music in many different applications, e.g., in cars, concert venues, cinemas or hi-fi setups at home. A schematic illustration of this audio-induced vibration generation approach is provided in Figure 1.2. An existing music mix, which might have been perceptually coded for data transmission, is used as a starting point. From this source signal, whole-body vibrations (WBVs) are generated using various approaches considering the perceptual capabilities and limitations of both modalities. Therefore, basic knowledge of the fundamental characteristics of the auditory and tactile sensory modalities and the interaction of both modalities is necessary.

Hearing has been studied quantitatively since the 19th century, and the literature has developed a vast knowledge base, ranging from physiological to psychophysical aspects. However, research on the sense of touch has been rather limited, particularly related to the perception of WBVs. Therefore, it is necessary to extend our understanding of the tactile sense, e.g., regarding the fundamental characteristics of frequency and intensity perception. Additionally, the different sensory modalities might interfere with each other. Therefore, broadening our knowledge of the combined perception of different sensory inputs is important.

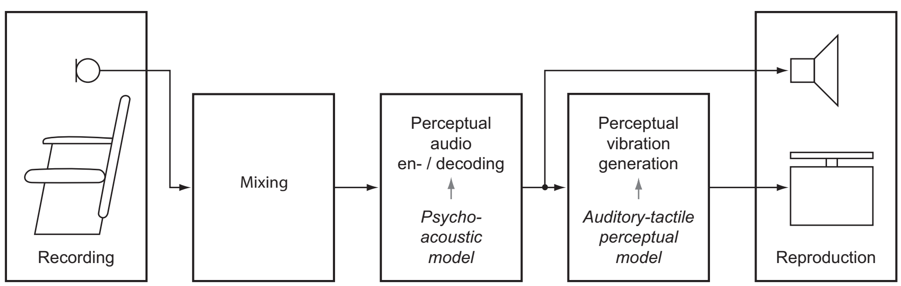


Fig. 1.2. Schematic illustration of audio-induced vibration generation. To develop perceptual vibration-generation approaches, fundamental knowledge of auditory and tactile perception and the interaction of both types of perception might be useful.

1.3 Organization of the Thesis

In the first part of this thesis, the capabilities and limitations of the auditory and tactile modalities are compared. An understanding of the similarities and differences between hearing and touch is fundamental for developing audio-based vibration generation algorithms. **Chapter 2** of this thesis starts with a short overview of the anatomical, physiological, and neurological aspects of both senses. Subsequently, basic psychophysical characteristics, related to absolute and differential sensitivity, are discussed in further detail. Because little or no experiments have been performed for WBVs, data for vibrations at the hand or forearm are often used for comparisons.

Chapter 3 describes the development and characteristics of the experimental system that will be used for further investigations.

In **Chapter 4** and **Chapter 5**, fundamental experiments are discussed that aim to provide insights into the perception of WBVs. Chapter 4 addresses the frequency resolution for WBVs. In Chapter 5, intensity perception is discussed. The growth of perceived vibration magnitude with level is evaluated. Curves of equal vibration intensity are determined, and an experiment is described regarding the cross-modal intensity matching of sound and vibrations. Finally, the influence of vibrations on loudness perception is investigated. The experiments described in these chapters have been partially published before [113, 116, 117, 118].

Chapter 6 discusses the existence of perceivable audio-induced vibrations in actual concert rooms using measurements in two exemplary venues: an opera house and a church. The data presented here are based on a journal publication by the author [114].

In **Chapter 7**, the fundamental knowledge gained previously is applied to the development and evaluation of music-related vibration-generation approaches. Different parameters used in these approaches are examined with regard to their perceptual consequences. In the beginning, various algorithms are assessed that mainly modify the frequency content of the vibration signal. Subsequently, approaches adapting the dynamic range and level of vibration are concerned. Finally, an outlook regarding processing of the temporal structure is provided.

The thesis concludes in **Chapter 8**, with a general discussion and the outlook of further research.

2

Comparison of the Auditory and Tactile Modalities

To design vibrations from acoustical signals, it is necessary to understand the basic characteristics of auditory and tactile perception. Therefore, this chapter will provide an introduction to the anatomical, physiological, and psychophysical properties of both modalities.

In Section 2.1, the structure and functionality of the human hearing organ will be compared to the histology and physiology of the mechanoreceptive system. Further, the neural processing in the somatosensory and auditory areas of the brain will be discussed.

An overview of the basic psychophysical capabilities and limitations of both modalities regarding the different parameters that influence the absolute perception threshold will be provided in Section 2.2. Section 2.3 will then compare the differential sensitivity for suprathreshold signals.

2.1 Anatomy and Physiology

2.1.1 Auditory

Hearing is the ability to perceive sound by detecting vibrations with the ear. Figure 2.1 provides a schematic drawing of a cross-section of the human ear. Pressure fluctuations in the air are transmitted through the ear canal via the eardrum and ossicles to the oval window of the fluid-filled cochlea. Alternatively, the cochlea can be excited by bone-conducted vibrations.

Inside the snail-shaped cochlea, energy is transmitted via a traveling wave on the basilar membrane. The basilar membrane, which is narrow at the base and widens toward the apex, works similarly to a frequency analyzer. Depending on frequency, the traveling wave excites a maximum at a specific place along the organ of Corti, which is drawn schematically in Figure 2.2. The outer hair cells, which are located along the organ of Corti, are responsible for the amplification of the traveling wave. For this purpose, they are

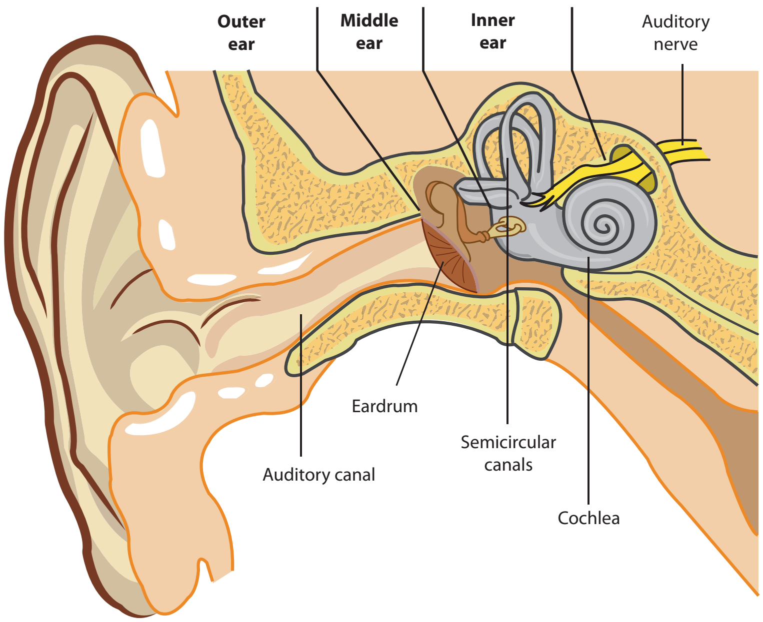


Fig. 2.1. Schematic drawing of a cross-section of the human ear, adapted from [150].

controlled via efferent nerves and can move actively. The resulting displacement changes the electric conductance of the inner hair cell membranes. In this manner, transmitters are released to the nerve endings and transferred to the brain stem. The evoked action potentials, which are transferred toward the central nervous system and auditory cortex, contain all of the acoustical information.

The frequency of the sound is directly linked to the anatomic placement of the corresponding neurons along the organ of Corti. Additionally, the time pattern of the neural pulses can transfer frequency information. There is considerable evidence that both mechanisms are important for pitch perception [37, 141], and a combination suggests itself [50]. Low frequencies, which are interesting in the context of this study, might be coded by a combination of place and temporal patterns.

Other physical parameters are encoded as well: intensity is coded by the discharge rate, time by the duration of excitation, and location by binaural differences or spectral patterns. However, these relationships will not be discussed further here.

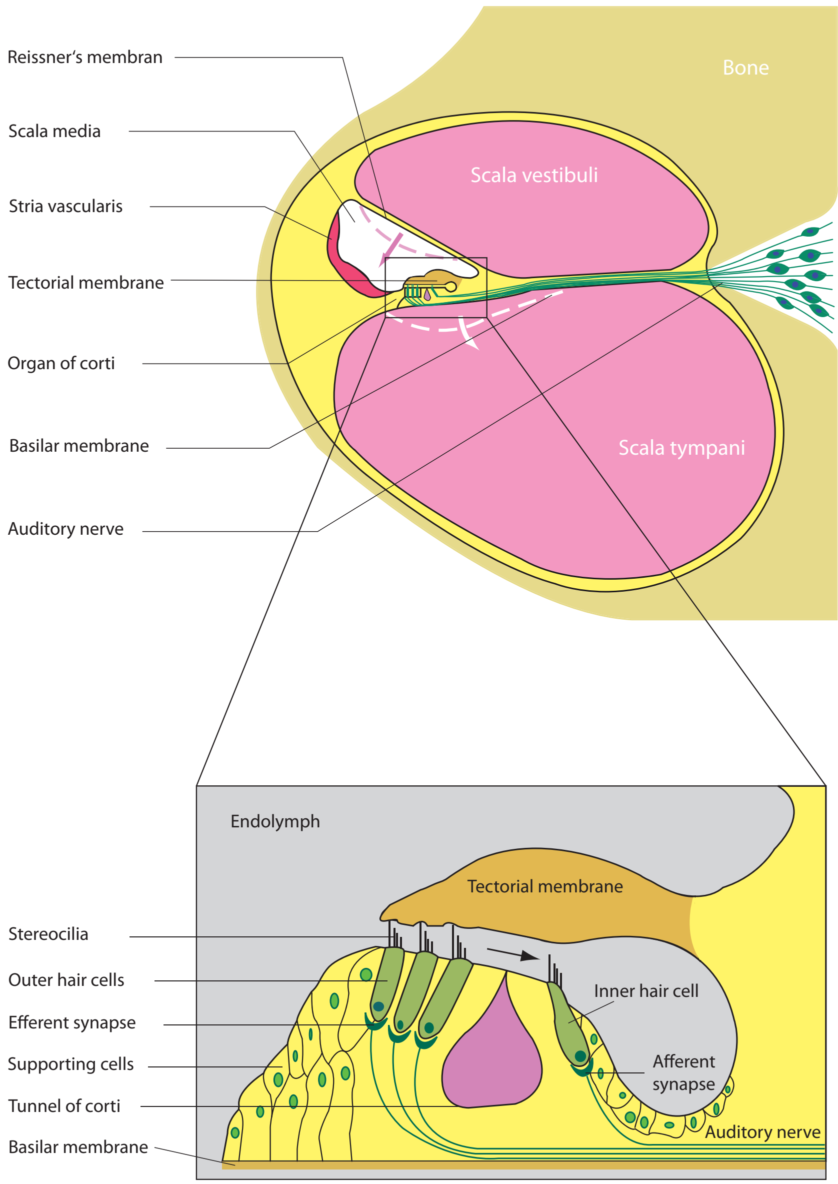


Fig. 2.2. Schematic drawing of a cross-section of the cochlear spiral and the organ of Corti, modified from [150].

2.1.2 Tactile

Vibrations can be perceived not only by the ear but also by our somatosensory system. This system consists of a variety of different receptors: nocireceptors (pain), proprioceptors (body position), thermal receptors (cold and warmth), and mechanoreceptors (e.g., vibration, pressure, stretching).

The mechanoreceptors in the skin can be categorized according to their morphology or adaptation properties. There are four morphologically different receptor types in the glabrous skin: Merkel's receptors (SA–I); Ruffini's corpuscles (SA–II); Meissner's corpuscles (RA–I); and (Vater-)Pacinian corpuscles (RA–II or PC). Figure 2.3 presents a cross-section of the glabrous and hairy skin and the location of the mechanoreceptors in the skin layers.

Slowly adapting receptors (SA–I and SA–II) evoke action potentials in the corresponding nerve fibers as long as an indentation of the skin is present. The firing rate is proportional to the stimulus intensity. Rapidly adapting receptors (RA–I and RA–II) react only to movement of the skin. The numerals I and II indicate the sizes of the corresponding receptive fields. Receptors marked with the numeral I lie close to the epidermis and have small receptive fields. Receptors deeper in the tissue have larger receptive fields and are labeled with the numeral II. The adaptation characteristics and receptive fields of all four receptor types are illustrated in Figure 2.4.

An overview of the different properties of mechanoreceptors is provided in Table 2.1. In addition to the adaptation characteristics and the sizes of the receptive fields, mechanoreceptors differ regarding the minimum skin indentation that is necessary to evoke perception, the density of the receptors, and the sensitive frequency range.

Table 2.1. Properties of the glabrous skin mechanoreceptors [5, 92].

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

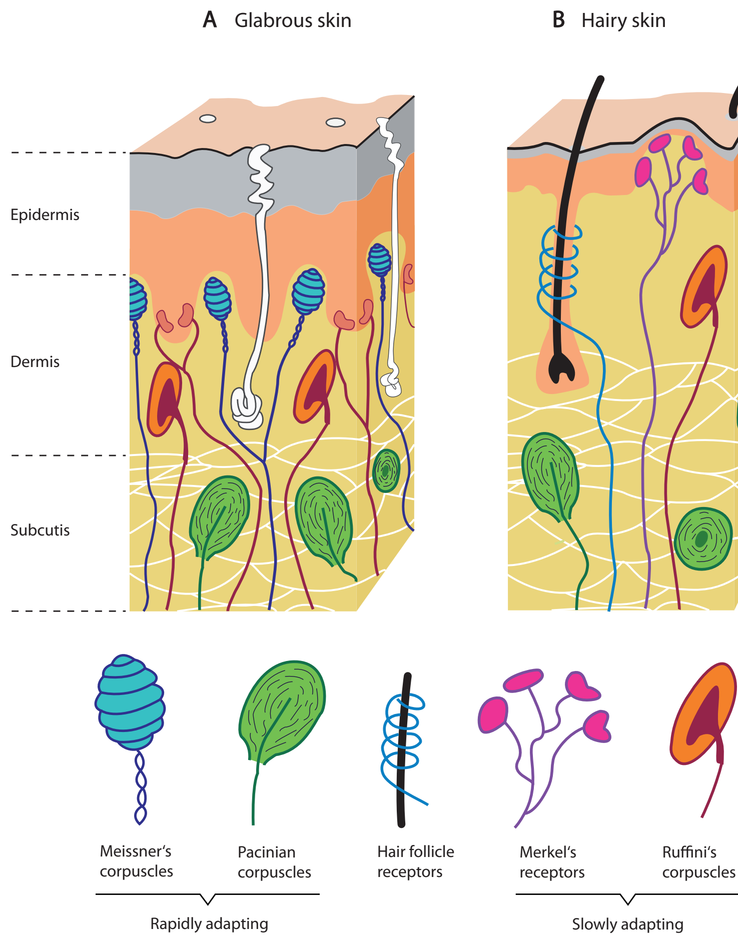


Fig. 2.3. Schematic illustration of the histology of mechanoreceptors in the hairy and hairless skin of primates. Merkel's receptors and Meissner's corpuscles can be found close to the surface. Pacinian corpuscles and Ruffini's corpuscles lie deeper in the tissue. Adapted from [150].

Different receptors have different functions. Merkel's receptors are activated by static indentation of the skin. Due to their small receptive fields, they are able to detect borders and edges. Ruffini's corpuscles react similarly to static deformation. They are specialized to perceive shearing or stretching of the skin. Meissner's corpuscles (only in the glabrous skin) primarily detect the speed of skin indentation at low frequencies. They are useful for the recognition of small bumps and ridges. However, they can also detect vibrations with low frequencies. Finally, Pacinian corpuscles are sensitive to changes in

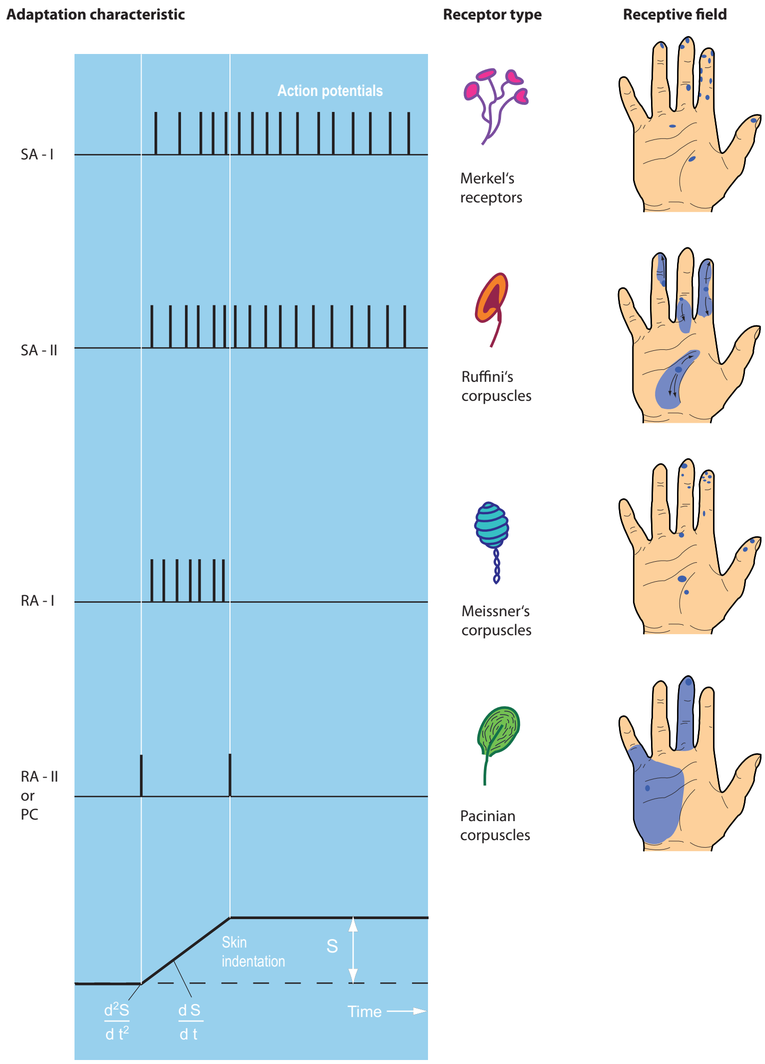


Fig. 2.4. Rate of adaptation and size of the receptive field for different mechanoreceptors. Adapted from [150, 164].

the speed (acceleration) of the skin during deformation. They cover the largest frequency range and have the largest receptive fields.

Music-induced vibrations (air- or structure-borne) often stimulate large areas of the skin at frequencies greater than 40 Hz; thus, the Pacinian channel might be the most important pathway in the context of this study. The sensitivity of the Pacinian channel depends on many factors, such as stimulus duration, area of excitation, or the age of the subject. This topic will

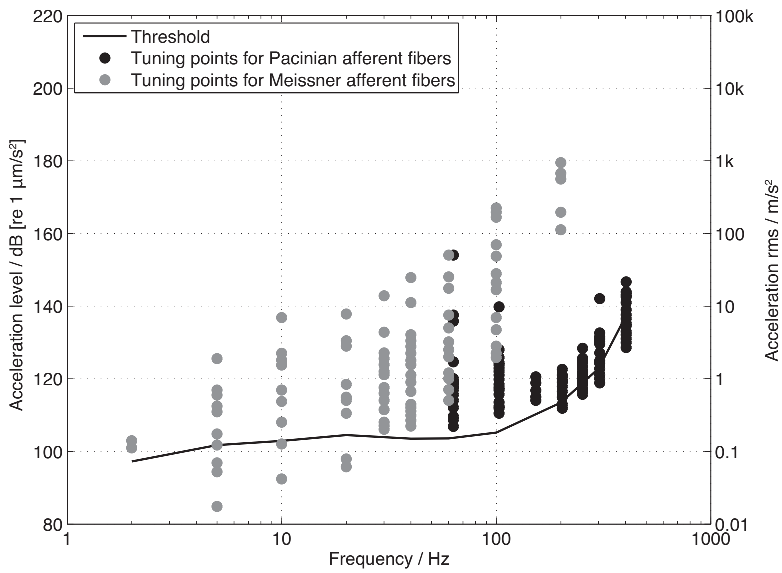


Fig. 2.5. Tuning points for Pacinian and Meissner afferent fibers in the glabrous skin of monkeys hands compared to the frequency-dependent perception threshold for vibrations at the same location. The threshold is correlated with the most sensitive receptors at low and high frequencies, with overlap in the range of 40 Hz. Data adapted from [135].

be discussed in detail in Section 2.2.1. However, Meissner's corpuscles also contribute to vibration perception at low frequencies.

Similar to audition, the physical properties of vibrations are coded into neuronal patterns. For instance, frequency and intensity are coded into the time pattern of excited action potentials. For low magnitudes, not every period of oscillation excites an action potential. For high vibration intensities, multiple action potentials are generated per oscillation period [150]. Figure 2.5 presents the tuning points for Pacinian and Meissner afferent fibers in the glabrous skin of monkeys hands, with data from Mountcastle [135]. The tuning point is the acceleration at a given frequency, which excites exactly one action potential per oscillation period. Additionally, the threshold of vibration perception is plotted for six monkey subjects. The Meissner corpuscles are most sensitive to vibration at less than approximately 40 Hz. At higher frequencies, the Meissner threshold rises and the Pacinian corpuscles dominate perception. No evidence for a frequency to place transformation or an active tuning mechanism, such as in the cochlea, was observed for tactile perception [135].

If vibrations excite the whole body, several additional sensory channels might be excited and contribute to perception: the vestibular system; proprioceptors in the muscles, tendons or joints; visceral receptors in the internal organs, vibrating at different resonance frequencies; or even vision at very low frequencies. The sensation of WBVs becomes a complex phenomenon. However, for low vibration intensities at intermediate and high frequencies, as expected for most music-induced vibrations, only receptors of the skin might be stimulated.

The term *whole-body vibration* is defined in VDI 2057 [168] as ‘mechanical vibrations within the frequency range of 0.1 to 80 Hz, which affect the whole body via the feet of the standing person, via the buttocks, feet and back of a seated person, or via the contact area of a person in a lying position.’ However, as understood in this study, the term will be extended to a broader frequency range and a more general excitation area: *WBVs are defined as mechanical vibrations that excite large parts of the body via sound waves or the vibrations of a contact surface.*

2.1.3 Neural Processing

The auditory and somatosensory cortices lie close to each other in the brain. Both modalities follow a consistent topological organization in their neuronal structures and corresponding receptive fields in their cortices. In audition, this organization scheme is referred to as *tonotopy*, which means that similar frequencies (or places along the basilar membrane) are represented in topologically neighboring regions in the auditory cortex. In contrast, for tactile perception, different regions of the body are represented by different cortical areas, termed *somatotopy*. However, the size of each cortical area is not proportional to the size of the body part; instead, it is proportional to the density of receptors at the specific body site and is thus also correlated with the spatial resolution of the tactile sense in this region. This correlation can be visualized using a so-called homunculus, as shown in Figure 2.6. The hand and the regions around the mouth are particularly enlarged, corresponding to regions with high peripheral innervation density and thus high spatial resolution.

In addition, many areas in the brain react to multi-sensory inputs. For instance, neurons in the superior colliculus respond to auditive, tactile, and visual excitation [157]. Other studies have detected auditory-tactile interaction effects in areas close to and in the primary auditory cortex [47, 95]. Further, excitation of the auditory cortex by tactile stimuli has been demonstrated [15, 46].

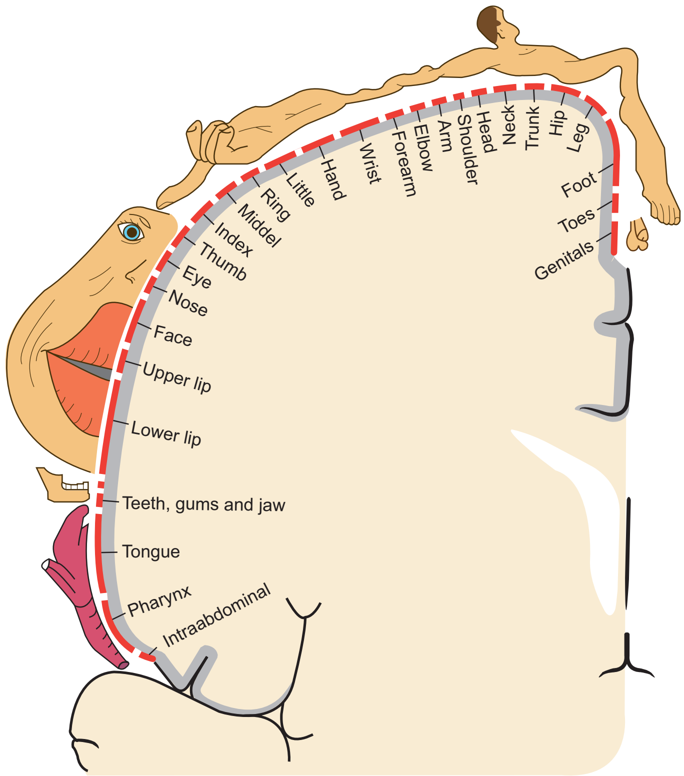


Fig. 2.6. Cortical homunculus showing the skin area as a function of the quantity of related neurons in the somatosensory cortex, adapted from [150]. The number of central neurons is proportional to the spatial resolution of the tactile sense in the corresponding region of the skin.

2.2 Absolute Sensitivity

The physiological and neurological mechanisms are important for understanding how the auditory and tactile systems work. In the context of this study, the psychophysical abilities and limitations of both modalities play more dominant roles and are discussed below. A direct comparison reveals similarities and differences. Knowledge of these similarities and differences is important for the design of sound-induced vibrations and of corresponding reproduction systems.

The perception of sound is a complex area that has been studied quantitatively since the 19th century. The basic physical attributes of sound (e.g., the intensity, frequency, or location of a sound source) have been correlated with perceptual attributes, such as loudness, pitch, or distance. Different effects, such as temporal acuity or masking, characterize the auditory system.

In comparison, the perception of vibration has not been studied as extensively. In contrast to our hearing, vibrations can be perceived by different parts of the body. Most studies have focused on vibrations transmitted via the hand and finger. Because sound-induced vibrations are likely to be perceived by the whole body, special attention is given below to literature relating to WBVs in the frequency range above 20 Hz. However, the principal receptors in the skin are similar at different body sites, and sound-induced vibrations are likely to stimulate mainly the Pacinian system (refer to Section 2.1.2). Thus, data from the hand are used for comparison if no other studies exist.

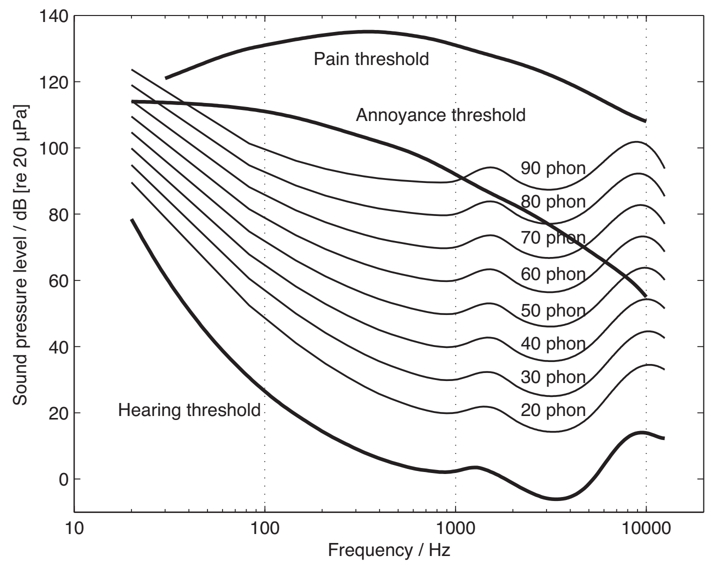
A fundamental characteristic of a sensory modality is the absolute perception threshold. Minimum and maximum perceivable levels for auditory and tactile perception will be discussed in this section. Basic effects, such as energy integration, masking, and adaptation, are compared.

2.2.1 Sensation Area

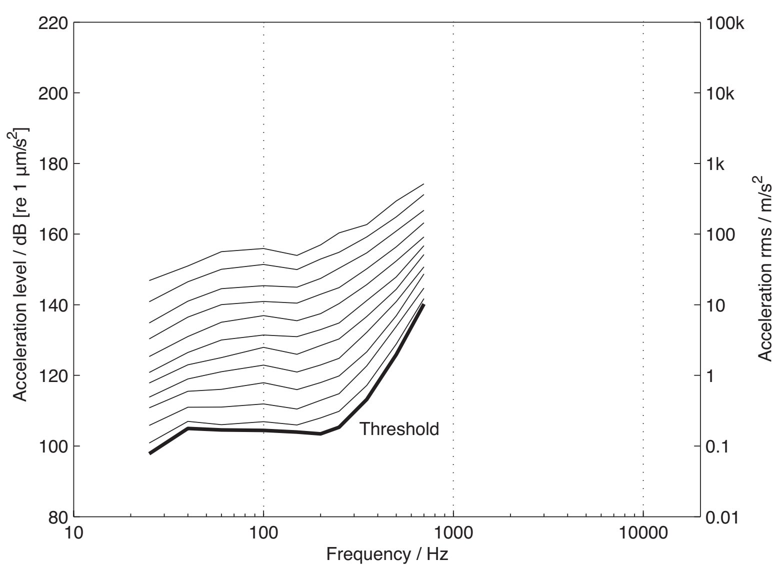
Auditory

Sound can be heard at between a few hertz and approximately 20 kHz. The upper frequency limit depends strongly on the age of the subject. Figure 2.7 (a) demonstrates that the hearing is most sensitive to sound pressure at between approximately 300 and 7,000 Hz. It becomes less sensitive for decreasing and increasing frequency. In addition, the figure provides estimates for the pain and annoyance thresholds for sinusoidal signals, according to Winckel [194]. The curves of equal subjective intensity (equal loudness contours) are plotted according to ISO 226:2003 [83]. They follow the threshold curve to some degree. These curves grow closer at lower frequencies. The auditory dynamic range is thus frequency dependent, from 50 dB to more than 100 dB.

The hair cells in the cochlea can be regarded as the most sensitive mechanoreceptors of the human body. The minimum perceivable sound pressure causes only 10^{-10} m of displacement in the inner ear, which corresponds roughly to the diameter of a hydrogen atom [150].



(a)



(b)

Fig. 2.7. Curves of equal subjective intensity, plotted as a function of frequency for **(a)** sounds (according to ISO 226:2003 [83] and Winckel [194]) and **(b)** vibrations of a 2.9 cm^2 contactor on the thenar eminence (adapted from Verrillo [179]).

Tactile

In comparison, the tactile sense is rather limited. Only frequencies up to approximately 1 kHz can be perceived via the mechanoreceptive system. Similar to the ear, the vibration sensitivity of the skin depends on frequency. Figure 2.7 (b) presents the frequency-dependent perception threshold on the thenar eminence, adapted from Verrillo et al. [179]. The glabrous skin becomes more sensitive to the acceleration of its surface with decreasing frequency. Similar results were reported for various regions of the body [67]. The sensitivity was found to depend on the distribution and density of the mechanoreceptors, with lower thresholds for areas with higher receptor density [101]. Hairy skin is approximately 10 dB to 20 dB less sensitive depending on the frequency [176].

The curves of equal subjective intensity follow the threshold to some degree. Again, a frequency dependence can be observed, with smaller dynamic ranges for frequencies greater than approximately 300 Hz. At frequencies less than 200 Hz, vibrations more than 40 dB to 55 dB above threshold become very unpleasant or painful [123]. The dynamic range can thus be quantified at between approximately 40 dB to 50 dB.

A common measurement unit for vibrations is the acceleration level L_{acc}. It is defined as the logarithmic ratio of the acceleration a and a reference value $a_0 = 1\,\mu\mathrm{m\,s}^{-2}$. In contrast to sound pressure level, 0 dB acceleration level is not related to the perception threshold as can be seen in Figure 2.7 (b).

$$L_{\mathrm{acc}} = 20\,log\frac{a}{a_0}\,\mathrm{dB}$$

Of particular relevance in this study are WBVs for seated subjects. Therefore, threshold curves for WBVs from various laboratories [10, 115, 121, 133, 139, 156] are summarized in Figure 2.8. The shapes of the threshold contours are similar. In terms of acceleration, an overall trend toward an increasing threshold with increasing frequency is observed in the range of 5 to 300 Hz. However, Miwa [121] and Parson and Griffin [139] measured lower thresholds, particularly at lower frequencies. Different body postures or body support might explain some of the variability between studies in this frequency range. For example, the surface of the seat used by Morioka and Griffin [133] was approximately half the size of the flat seats used in other studies, which themselves provided contact with the thighs. The smaller contact area and the absence of contact with the thighs might have reduced the sensitivity to vibrations. Differences between studies might also be partially explained by different psychophysical methods. For example, Miwa [121], Bellmann [10], Stamm et al. [156], and Merchel et al. [115] used various adaptive n-interval forced choice methods, whereas Parsons and Griffin [139] and Morioka and Griffin [133] employed a 'yes-no' method.

Two frequency regions have been separated and fitted by first-order regressions. At lower frequencies, the perception threshold increases slightly at a rate of approximately 1 dB per octave. Above 150 Hz, the increase rises to

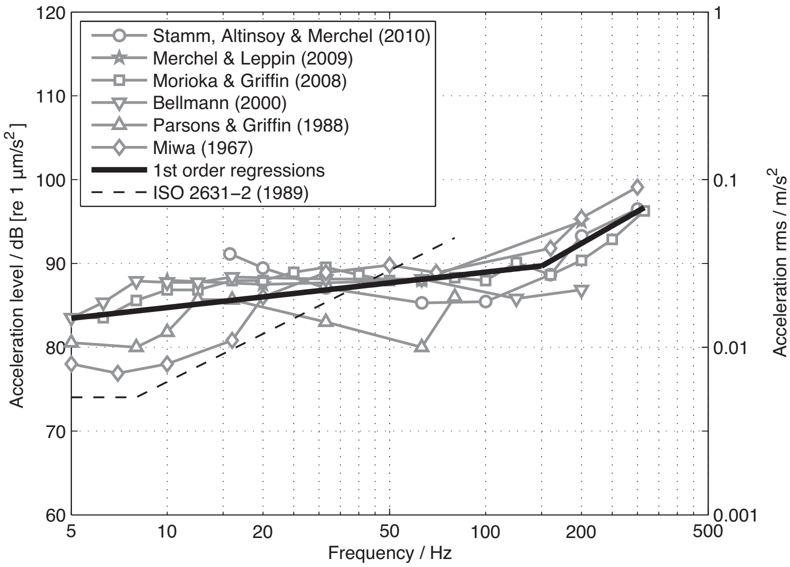


Fig. 2.8. Perception threshold for vertical sinusoidal WBVs from various laboratories [10, 115, 121, 133, 139, 156] in comparison to the threshold from ISO 2631:1989 [84]. A first-order regression was fitted to the data less than and greater than 150 Hz.

approximately 6 dB per octave. However, no prediction can be made outside of the shown frequency range. The regression curves must be interpreted carefully because the threshold curves vary between studies at lower frequencies and few measurement points are available at higher frequencies. In addition, large inter-individual differences have been reported [115].

The threshold curve from ISO 2631:1989 [84], which was removed from the revised ISO 2631:2003 [85], does not represent the data well. Thus, the fitted curves will be used as references below.

Little is known about curves of equal intensity perception for WBVs. Therefore, an experiment targeting this topic is described in Section 5.3.

Summary. *Both modalities exhibit frequency-dependent perception thresholds but with different trends. In addition, tactile perception is restricted to low frequencies. At 20 Hz, the usable amplitude range of both modalities is similar. However, with increasing frequency, the auditory dynamic range increases rapidly, whereas the tactile dynamic range remains fairly constant up to approximately 200 Hz.*

2.2.2 Age and Gender

Auditory

The threshold of hearing rises naturally with increasing age. This effect is referred to as *presbyacusis*, and it involves primarily frequencies greater than 3,000 Hz. Figure 2.9 presents data that depict the progression of hearing loss with age [155]. The data are averaged over men and woman; however, previous studies have found that presbyacusis starts more gradually in women but grows more quickly once it begins [19]. In addition, noise-induced hearing loss (sociocusis) is a common phenomenon today.

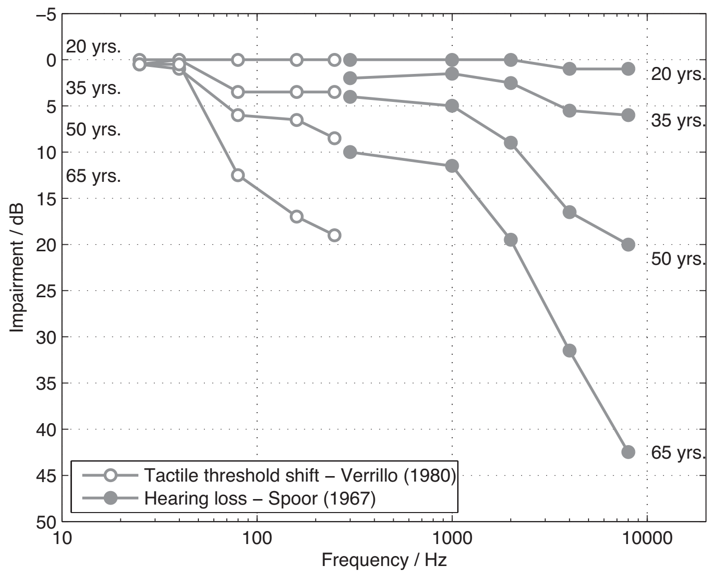


Fig. 2.9. Auditory and tactile threshold shift as a function of age. Auditory data depict presbyacusis (without the effects of severe occupational noise) [155]. Tactile data are achieved using a 2.9 cm^2 contactor at the thenar eminence [175] and plotted relative to the threshold at 20 years. The data points at 250 Hz are shifted slightly horizontally for better illustration.

Tactile

Similar to hearing, age has a considerable influence on tactile thresholds. The sensitivity for high frequencies decreases progressively with age [160, 180]. Figure 2.9 illustrates the shift in the tactile detection threshold for four age groups [174]. At higher frequencies, at which the Pacinian system is predominant, a strong loss of sensitivity was observed with increasing age. No effect was found for low frequencies.

In general, no gender differences were observed for tactile thresholds between men and women [109, 173]. Only Gescheider reported that women are slightly more sensitive to high-frequency vibrations at the thenar eminence a few days before menstruation [57].

Summary. *Both modalities show severe impairment of sensitivity with increasing age. This effect has a similar tendency: it is stronger toward the upper frequency limit of each modality. However, at approximately 250 Hz, the age-induced threshold shift appears stronger for the sense of touch than for hearing. This difference is particularly crucial in the context of this study because the tactile dynamic range is considerably smaller than the auditory dynamic range. A tactile threshold shift of 20 dB at 200 Hz almost halves the available amplitude range. In other words, vibrations that are strong for younger subjects might not be perceived at all by the elderly.*

2.2.3 Energy Integration

Another important characteristic of the auditory and tactile modalities that has an influence on the threshold is the ability to integrate energy. This ability is often discussed using the relationship between the duration and threshold of a stimulus.

Auditory

The auditory threshold of detection decreases with increasing duration, up to a stimulus duration of approximately 1 s. This relationship holds true for various types of stimuli over a broad frequency range [48]. Figure 2.10 presents data from Plomp and Bouman [143] and Florentine [41] for a stimulus frequency of 250 Hz. The curves follow the prediction made by the theory of temporal summation, which was formulated by Zwislocki in 1960 [197].

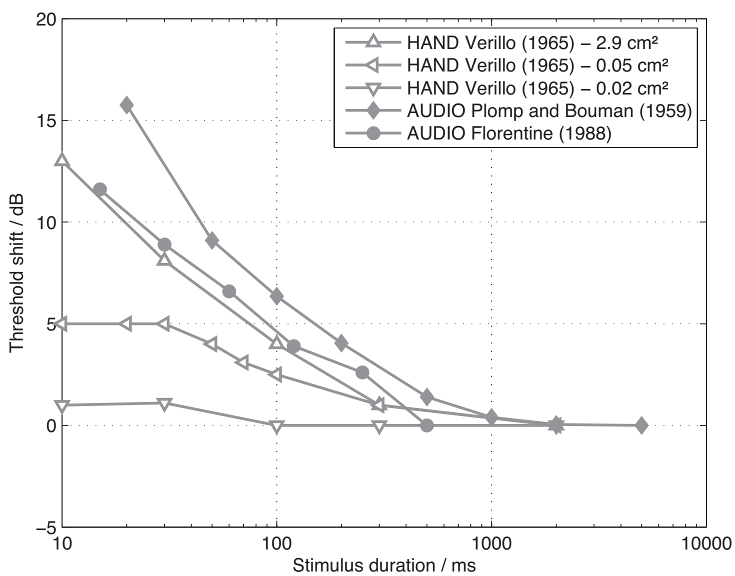


Fig. 2.10. Auditory and tactile threshold shift as a function of burst duration [41, 143, 170]. Data are plotted in dB relative to the detection threshold for the longest stimulus of each curve. The stimuli frequency was 250 Hz in all cases. The tactile stimuli were applied to the skin of the hand using different contactor sizes. The capital labels used in this and all following figures are abbreviations, e.g., AUDIO indicates data for the auditory sense, or HAND indicates data measured for the tactile modality at the hand.

Tactile

Temporal energy integration has also been observed in the tactile domain, but only in the Pacinian system [54, 62]. No temporal summation was observed for low frequencies, e.g., at 25 Hz [55]. Data from Verrillo [170] are plotted for comparison in Figure 2.10. Stimuli with a frequency of 250 Hz were delivered to the glabrous skin of the palm using a large contactor (2.9 cm^2). Verrillo measured almost a 3 dB reduction of threshold per doubling of duration up to a stimulus length of 300 ms, indicating a near complete integration of energy. Similar curves were observed at frequencies of 100 and 500 Hz, at which the Pacinian corpuscles are responsive to vibration. The data agree well with the curves observed in the auditory domain, indicating similar perceptual mechanisms.

Additional curves for smaller contactor sizes (0.05 cm^2 and 0.02 cm^2) can be observed in Figure 2.10 [170]. The dependence of duration on the threshold is reduced as the size of the stimulated area is reduced. Using smaller contact areas thus stimulates increasing numbers of non-Pacinian receptors (see Section 2.1.2). Consequently, the amount of temporal summation decreases.

In addition, absolute tactile sensitivity at higher frequencies depends strongly on the size of the stimulated area. Previous studies have shown that for frequencies between 80 and 320 Hz (Pacinian channel), the threshold decreases at 3 dB per doubling of the contact area at the thenar eminence of the hand [169, 171]. Similar results have been reported for the hairy skin of the forearm [172]. No effects were observed at lower frequencies [55].

Sound-induced vibrations can excite large areas of the skin, depending on the context and environment. It is assumed that sufficiently high frequencies ($>$ 40 Hz) and relatively large areas of the skin ($>$ 2.9 cm^2) are excited, resulting in predominant stimulation of the Pacinian channels.

Summary. *The auditory system is able to integrate energy over time for stimulus durations of up to approximately 1 s. A similar effect has been observed in the tactile system for sufficiently high frequencies and relatively large stimulation areas, indicating that PC receptors are capable of integrating energy over time. In addition, energy integration over space has been observed. However, non-Pacinian receptors appear to be unresponsive to changes in duration or stimulation area.*

Until now, only a single stimulus has been examined. However, in everyday life, two or more simultaneous stimuli are not unusual. If subjects are asked to judge the combined intensity of two tones, the result in audition is proportional to the overall energy if the frequencies lie within a critical band. However, if frequency components outside the critical bandwidth are added, the perceived intensity grows much stronger and the sensation magnitudes of the individual components can be totaled [36]. Interestingly, similar effects have been observed in the tactile domain. Evidence for energy integration within the Pacinian channel has been discussed above, and the addition of sensation magnitudes between mechanoreceptive channels has been reported [104, 178]. Thus, it has been suggested that the Pacinian channel is analogous to a critical band in the auditory system [102].

2.2.4 Masking

Multiple stimuli might interfere with one another if they are heard or felt in close temporal proximity. One such effect is the influence of one sound on the audibility of another sound, which is called *masking*. This phenomenon has been successfully used for audio data compression and might be useful for the reduction of the sound radiation of a vibration actuator in the context of sound-induced vibration reproduction, as is described in Section 7.2.2. This phenomenon is thus discussed below.

Auditory

Early experiments to investigate masking used sinusoids both as a masker and test signal [189]. However, beats occurred and complicated the results when both signals were close together in frequency. To avoid this problem, later studies used narrow band noise as a masker. The shifted threshold was determined for detecting a test tone at various frequencies in the presence of a masker with a fixed center frequency and amplitude. This masked threshold is often referred to as a *masked audiogram* or masking pattern. It is strongly correlated with the excitation pattern that the masker generates on the basilar membrane [26]. An exemplary masking pattern is presented in Figure 2.11 with data from Egan and Hake [28]. For the plotted curve, a 90 Hz wide band of masking noise is centered at 410 Hz with a level of 40 dB SPL. A narrow masking region can be observed. However, for higher sensation levels (SLs), which are not plotted here, the masking pattern spreads more toward the high-frequency side.

In general, auditory masking patterns are dependent on masker frequency, duration, and level. These patterns exhibit steep slopes toward lower frequencies and flatter slopes toward higher frequencies on a logarithmic frequency axis. However, masking patterns become increasingly symmetrical toward low sensation levels or low frequencies [28, 163], as illustrated in Figure 2.11. Interestingly, the maximum effects of low-frequency maskers (e.g., at 150 Hz)

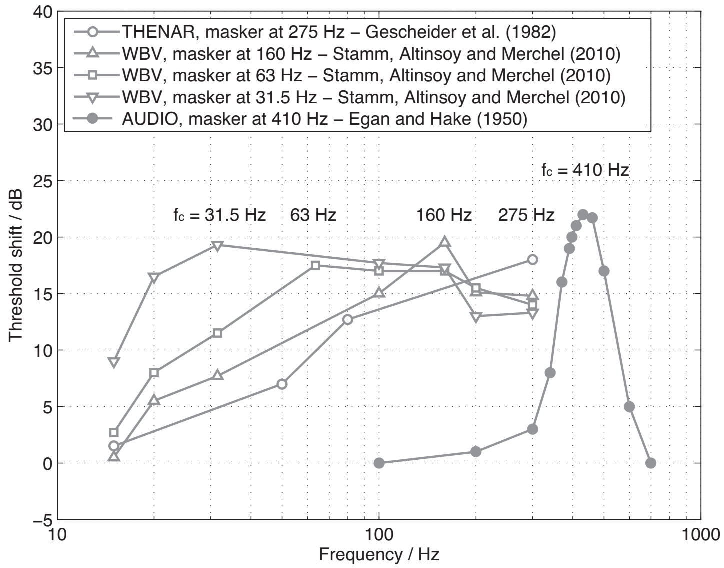


Fig. 2.11. Auditory and tactile masked thresholds relative to unmasked conditions as a function of frequency. The vibration maskers were narrow band noises centered at 31.5, 63, 160, and 275 Hz with a fixed level approximately 25 dB above the threshold. Data from Stamm, Altinsoy, and Merchel [156] are plotted for WBVs (25 Hz noise bandwidth) and from Gescheider et al. [59] for vibrations at the thenar eminence (100 Hz noise bandwidth). For comparison, an auditory masking pattern is plotted for a 90 Hz -wide band of masking noise that is centered at 410 Hz with a level of 40 dB SPL [28]. Test stimuli were simultaneously presented as sinusoids in all conditions.

appear to be slightly shifted toward higher frequencies [163], and the low-frequency masking pattern broadens significantly when plotted on a logarithmic scale [29, 30].

In the above studies, the masker and test signal were presented to the same ear or to both ears diotically. However, masking was observed in even dichotic conditions [32, 33].

Further, masking effects have been reported for experiments in which the masker and test signal are presented one after the other. This type of masking is referred to as *post-masking* (forward masking) if the test signal comes slightly behind the masker or as *pre-masking* (backward masking) if the test signal precedes the masker, as illustrated in Figure 2.12 using data from Elliott [32]. A 50 ms-long white noise masker at 90 dB SPL was used to mask a 7 ms-long test tone at 500 Hz. Post-masking is active until approximately 100 ms. Other studies have reported slightly longer post-masking intervals, e.g., Jesteadt et al. [90] used tones from 125 to 4,000 Hz and reported that post-masking was more common at very low frequencies than at high frequencies.

Pre-masking is believed to be much weaker than post-masking. Some studies have even shown that pre-masking diminishes or almost disappears if the subjects are highly trained [138].

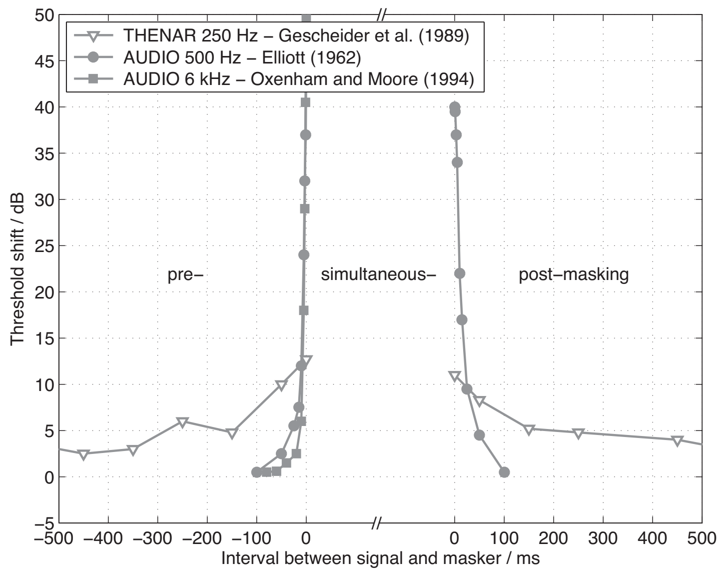


Fig. 2.12. Auditory and tactile pre- and post-masking as a function of the gap between the test signal and masker. Data from Gescheider et al. [60] are plotted using a 250 Hz vibration masker at the thenar eminence with a 20 dB sensation level. The test signal was also a 250 Hz vibration. Auditory data from Elliott [32] using a white noise masker at a 90 dB SPL are plotted for comparison. The test signal was a tone at 500 Hz. Additionally, pre-masking is plotted based on data from Oxenham and Moore [138] using a noise masker and a 6 kHz tone.

Tactile

Similar to audition, the detectability of a vibration might be reduced by another vibration. Again, this effect depends on the frequency, intensity, and timing of both stimuli. As in audition, masking increases as a function of increasing masker intensity and decreasing frequency separation. However, there is considerable evidence that the different mechanoreceptive channels do not mask each other [59, 102]. Tactile masking patterns from Stamm et al. [156] and Gescheider et al. [59] are plotted in Figure 2.11. Narrow band masking noise was simultaneously presented with sinusoidal test stimuli. Strong masking toward higher frequencies, which might be due to masking within the Pacinian channel, can be observed. For decreasing frequencies lower than the masker, the threshold of the Pacinian channel might exceed the threshold of another tactile channel, e.g., RA1, which takes over and gradually reduces the masking effect [60]. In this sense, the overlapping tactile channels could be regarded as similar to overlapping auditory bands, although with only a few fixed filters. This similarity would explain the strong asymmetry of tactile masking patterns plotted here.

Thresholds might be elevated even if two vibrations stimulate the body at different locations [56, 64].

Similar to in audition, masking is strongest for simultaneous stimulus presentation and decreases with increasing intervals between the test signal and masker [64, 103]. This relationship is illustrated in Figure 2.12. Tactile masking at the thenar eminence is plotted with data from Gescheider et al. [60] for a sinusoidal masker and for a test signal at 250 Hz. These researchers found that the rate of decay of post-masking is similar to that of pre-masking, independent of the masker type (sinusoidal or noise) and stimulated mechanoreceptor. Compared to audition, temporal masking appears to be much more extended for vibrations on the skin.

Other changes in sensation have also been reported for cases in which more than one stimulus is presented. For instance, a stimulus can cause a subsequent stimulus to appear more intense, with increasing intensity for a decreasing time interval between the two stimuli. This effect is referred to as *enhancement*, and it has been reported for short tone bursts in audition [198] and tactile perception [178].

Summary. *Both modalities illustrate the ability of one stimulus to mask (or enhance) another. In comparison, in the tactile modality, broader masking patterns are excited around the masker frequency, with strong masking toward higher frequencies. Furthermore, in the time domain, the tactile threshold rises over a longer period around the duration of a masker.*

2.2.5 Adaptation and Fatigue

In the previous section, masking, i.e., the ability of an intense stimulus to obscure a second, weaker test stimulus, was described. In this section, the ability of a temporally extended stimulus will be discussed to desensitize a sensory channel gradually. Such gradual desensitization might result in the decline of the apparent magnitude of a stimulus during presentation. Even some time after the stimulus has stopped, it might be more difficult to detect a test signal.

Auditory

In audition, *adaptation* and *fatigue* are often distinguished. Auditory adaptation refers to the decline in sensitivity within the first few minutes of stimulus presentation [126]. However, this effect appears to be restricted to low sensation levels or high frequencies [75, 186]. Thus, it might be of minor interest in the context of auditory-tactile music perception, so it will not be further discussed here. Auditory fatigue is often understood as a shift in threshold after excessive exposure to a fatiguing stimulus. This temporary threshold

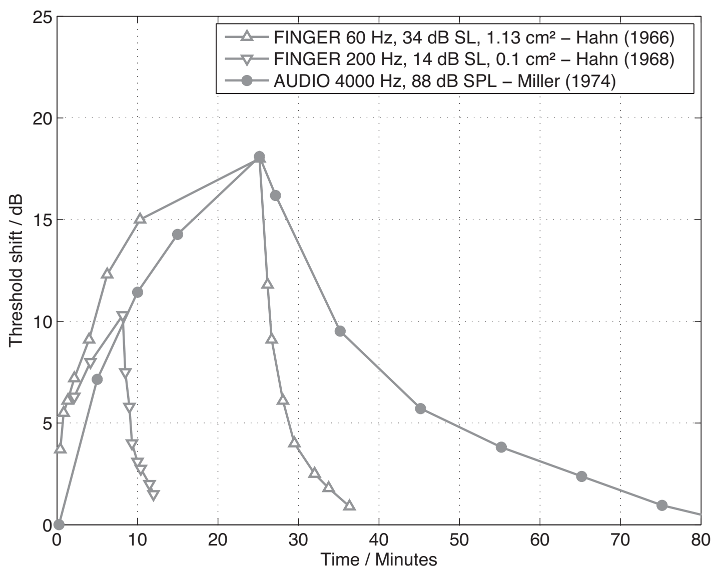


Fig. 2.13. Auditory and tactile temporary threshold shifts during and after exposure to long-lasting stimulation. Data from Hahn [69, 70] are plotted for vibratory stimulation of the Pacinian channel with different intensities and durations. An exemplary TTS for the auditory system based on Miller [120] is plotted for comparison.

shift (TTS) is well known from rock music [27] and will be summarized below. The TTS generally increases with increasing intensity and duration of the fatiguing stimulus. Similar to masking, larger TTSs have been observed with decreasing frequency separation between fatiguing and test stimulus. Interestingly, fatigue effects are less marked at low frequencies, possibly due to the middle ear reflex [126]. After cessation of the fatiguing stimulus, hearing recovers from the TTS approximately proportionally to the logarithm of the recovery time if the TTS is not excessively large (e.g., < 40 dB) and the exposure time is not overly long (e.g., < 1 day) [120]. Such an exemplary TTS curve is plotted in Figure 2.13 for 25 min of stimulation at 4 kHz, a frequency at which auditory fatigue is most effective.

Tactile

Similar to audition, the absolute perception threshold for vibration increases and recovers over time due to prolonged stimulation. In the vibro-tactile literature, this effect is often referred to as fatigue or adaptation. The TTS increases again with increasing intensity and duration of stimulation. For intense stimulation over longer periods, the recovery time can last up to several minutes. Compared to audition, much lower sensation levels are typically required for the effect to appear, and considerably steeper slopes have been reported [18, 51, 58, 188].

Two exemplary TTS curves are plotted in Figure 2.13 using data from Hahn [69, 70]. The upper curve was measured using a large contact area on the fingerpad vibrating with 60 Hz. Only a 34 dB sensation level was necessary to reach 17 dB TTS after 25 min of exposure. However, the TTS recovered more quickly compared to in audition. The curve with triangles facing down was measured using a small contact area on the fingerpad vibrating at 200 Hz and only a 14 dB sensation level. Again, steep rising and falling slopes were observed.

Similar to masking, it is widely believed that adaptation cannot occur between different tactile channels [70, 78].

Assuming that the described effects are also valid for WBVs at low levels, these effects might be interesting in the context of long-lasting music-induced vibrations, such as those generated by a church organ. However, continuous tones at low frequencies lasting more than, e.g., 30 s are very rare. Tones and breaks typically alternate, giving the tactile system time to recover. However, fatigue (and masking) might still be important in the context of automobile vibration reproduction systems, in which steady background vibrations can occur.

Summary. *Temporary threshold shifts due to prolonged stimulation occur in both modalities. In audition, high levels or long exposure times are necessary. In the tactile domain, even small sensation levels result in a TTS that grows and recovers quickly.*

2.3 Differential Sensitivity

In addition to absolute sensitivity, the smallest detectable changes of a stimulus are useful for a psychophysical comparison between the auditory and tactile modalities. Therefore, difference limen for intensity, frequency, duration, and location will be discussed below.

2.3.1 Intensity Discrimination

Auditory

In Figure 2.14, auditory just-noticeable differences in level (JNDLs) are plotted against frequency according to Florentine et al. [40] and Jesteadt et al. [89]. For high sensation levels (70 dB and 80 dB above the threshold), the auditory system is very sensitive to intensity changes, with a differential threshold of only 0.5 dB to 1 dB. This sensitivity is observed over a broad frequency range. However, for low sensation levels (30 dB and 50 dB), the JNDLs rise and some frequency dependence can be observed. Sensation level is slightly more important at high frequencies than at low frequencies, at which the curves tend to converge. Unfortunately, no data are available for frequencies less than 250 Hz.

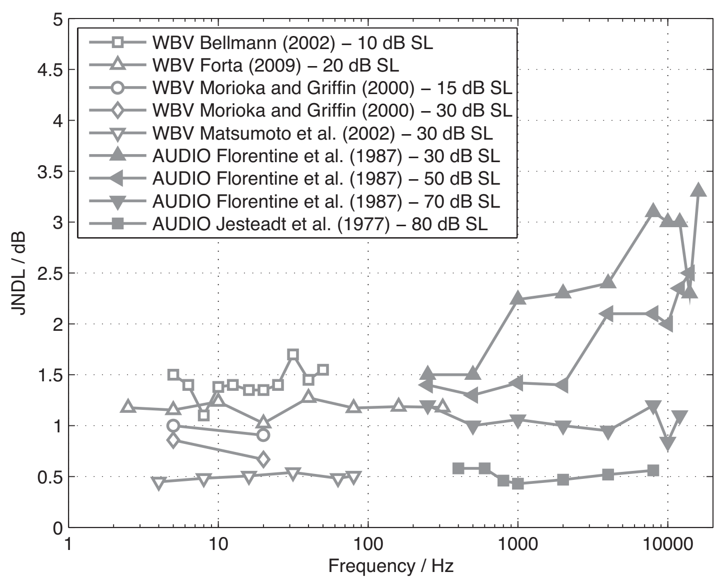


Fig. 2.14. Auditory [40, 89] and tactile [10, 44, 108, 134] JNDLs as a function of stimulus frequency.

If a single frequency is selected, the difference limen can be replotted as a function of sensation level. Figure 2.15 presents the differential threshold for a 1 kHz tone using data from various studies [40, 71, 89, 140]. The auditory JNDLs decrease significantly with increasing sensation level. This phenomenon is known as the 'near miss' to Weber's law, which would predict a constant JNDL in decibels, independent of sensation level.

Tactile

Tactile difference thresholds for level have also been studied extensively for various body sites. Different values between 0.4 dB and 2.3 dB have been reported in the literature [44, 147, 161]. In the context of this study, special attention is given to WBVs. Previous studies have shown that JNDLs for WBVs can be as small as 0.5 dB [108]. Other WBV studies [10, 44, 134] have obtained slightly higher values, which are depicted in Figure 2.14. None of the previous studies reported a dependence of the JNDL on stimulus frequency.

The absolute variations between results can be partially explained by the use of different experimental paradigms, resulting in different measurement points on the psychometric function. Additionally, the variations in absolute

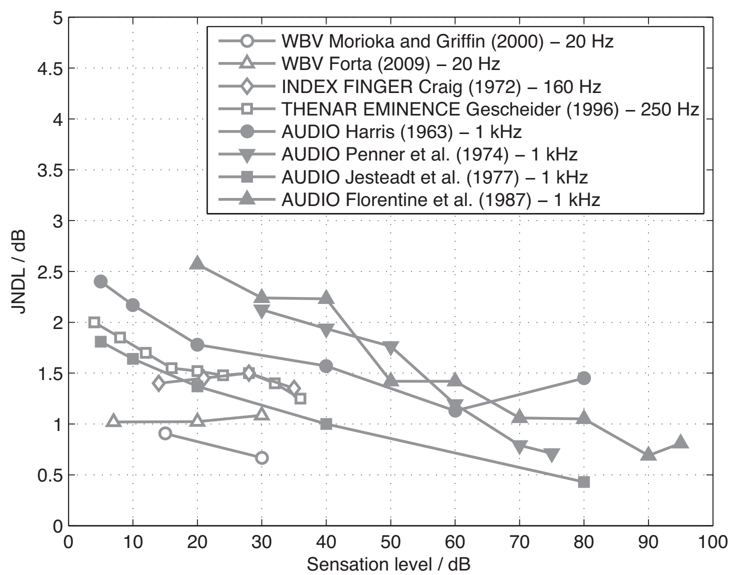


Fig. 2.15. JNDLs for 1 Hz tones as a function of sensation level [40, 71, 89, 140]. JNDLs for vibrations at various body sites are plotted using different frequencies for comparison [20, 44, 61, 134].

values between studies might be due to the applied sensation levels. Most studies have documented absolute acceleration levels, which were converted to sensation levels for comparability using the averaged threshold discussed in Section 2.1.2. The study with the lowest sensation levels measured the highest JNDLs (Bellmann [10]). The study with the strongest vibrations revealed the lowest difference thresholds (Matsumoto et al. [108]). This finding suggests a similar dependence of the JNDL on sensation level, as in audition. However, there are not sufficient data in the literature to test this hypothesis. Figure 2.15 presents different tactile studies that measured difference limen as a function of level. Only Gescheider [61] reported a significant decrement of the JNDL with increasing sensation level. Other studies did not observe an effect. However, a smaller dynamic range [20] and much lower vibration frequencies [44, 134] were tested. Thus, it is difficult to compare the results.

Summary. *The JNDLs for sound and vibration appear to be remarkably similar at low frequencies. A dependence on sensation level can be assumed in both modalities. Thus, intensity might be an important property for audio-driven vibration generation. However, the discrimination results summarized here do not prove that the perceived magnitude grows equally with increasing level in both modalities. However, the perceived magnitude might be essential for auditory-tactile music perception; thus, it will be evaluated further in Section 5.2.*

2.3.2 Frequency Discrimination

Auditory

One of the fundamental characteristics of the auditory system is its ability to discriminate between different frequencies. Just-noticeable differences in frequency (JNDFs) smaller than 1 Hz can be perceived at low frequencies. Figure 2.16 summarizes the data from various laboratories [34, 125, 127, 146, 193]. The plotted auditory JNDFs become larger with increasing frequency. The JNDFs are for tones with a minimum length of 200 ms. For shorter tones, the JNDFs increase rapidly [35]. Again, an influence of sensation level on the difference limen was observed, with a higher level resulting in smaller JNDFs [193].

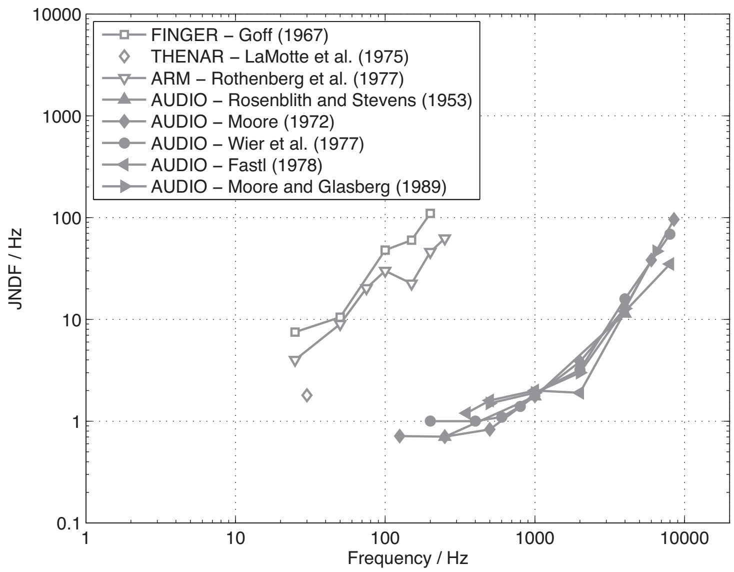


Fig. 2.16. Auditory [34, 125, 127, 146, 193] and tactile [65, 99, 147] thresholds for frequency discrimination of subsequent sinusoids (stimuli length > 200 ms) as a function of base frequency. The results from several studies are plotted for each modality. Auditory stimuli levels were in the range from 30 dB to 70 dB above threshold.

Tactile

The tactile ability to discriminate between vibration frequencies is quite limited when compared to the auditory system. However, data for tactile JNDF are lacking, perhaps because of the difficulty in eliminating concomitant cues, such as intensity differences, during experiments.

Most studies have focused on stimulation of the hand and forearm. The JNDFs obtained in these studies are plotted for comparison in Figure 2.16. Goff [65] investigated sinusoidal stimulation at the fingertip. Five frequencies (25, 50, 100, 150, and 200 Hz) were selected, and their magnitudes were adjusted to equal intensities (approximately 20 dB above the threshold). Goff found that the JNDF ranged from 8 to over 100 Hz, increasing with increasing reference frequency.

Rothenberg et al. [147] experimented with sinusoidal stimuli at the volar forearm. Frequencies between 25 and 250 Hz were evaluated. Their amplitudes were normalized to achieve a uniform subjective magnitude (approximately 14 dB above the threshold). The results revealed difference limen ranging from 4 to over 75 Hz.

The ability to detect changes in vibration frequency at the thenar eminence was measured by LaMotte and Mountcastle [99] using 30 Hz sinusoids. They found a JNDF of 1.8 Hz. This high sensitivity might be explained by the considerable training of their participants or by concomitant cues.

Summary. *The difference limen of tactile frequency discrimination at the hand and forearm are considerably higher than those values obtained in audition. Because of the sparseness of the data, an experiment will be described in Section 4.2, in which JNDFs for WBVs were measured. This topic is important in the context of this study because frequency information is one of the fundamental components of music, resulting in pitch perception. The question arises of whether it is possible and necessary to transform this information into music-driven vibrations. This question is subsequently investigated in Section 7.2.*

2.3.3 Temporal Discrimination

Another interesting aspect of both modalities is the ability to make temporal discriminations. Different stimuli and approaches have been used for investigations in the auditory domain, e.g., recognition of amplitude modulation [49] or the identification of an increase in duration [26]. However, only a few studies have considered the tactile domain. A lucid evaluation of temporal resolution is provided by the minimum detectible separation between two successive stimuli. This minimum detectible separation is referred to as the gap detection threshold, and it will be used for comparison purposes below.

Auditory

Numerous studies have investigated gap detection thresholds using different stimuli [38, 42, 43, 66, 72, 128, 129, 130, 152]. The minimum auditory temporal resolution was observed for clicks and for broad noise, and it is in the order of 2-3 ms. Exemplary data from Gescheider [53] and Plomp [142] are plotted in Figure 2.17. The gap detection threshold depends on sensation level and increases significantly toward lower intensities.

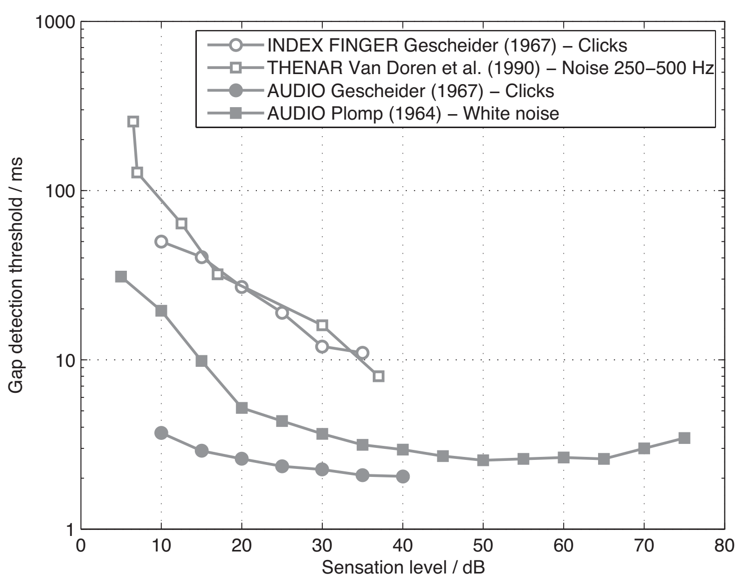


Fig. 2.17. Auditory and tactile thresholds for the detection of silent intervals in noises and between clicks as a function of sensation level. The results from several studies [53, 142, 166] are plotted for comparison.

This relationship is also true for sinusoidal excitation. Data from Moore et al. [131] are plotted in Figure 2.18. Sinusoidal gap detection thresholds are roughly constant at levels that are adequately audible but increase rapidly for levels close to the perception threshold. Minimum gap thresholds have been found at approximately 17 ms for the 100 Hz stimulus and at 6-9 ms for frequencies from 200 to 2,000 Hz. Slightly lower gap detection thresholds have been reported in other studies, such as 5 ms for 400 Hz by Shailer and Moore [153], which might be explained by different experimental procedures. No influence was observed for the embedding burst duration or temporal position of the gap [43, 72].

Tactile

Figures 2.17 and 2.18 compare tactile gap detection thresholds for noise, clicks, and sinusoids delivered to the hand from different publications [53, 63, 166, 166]. The minimum detectible gap between two tactile stimuli was found to be approximately 8 ms. Such thresholds were obtained for signals with medium to high sensation levels. For lower intensities, the minimum detectible gap increases, similar to our hearing. In contrast to the auditory modality, vibratory

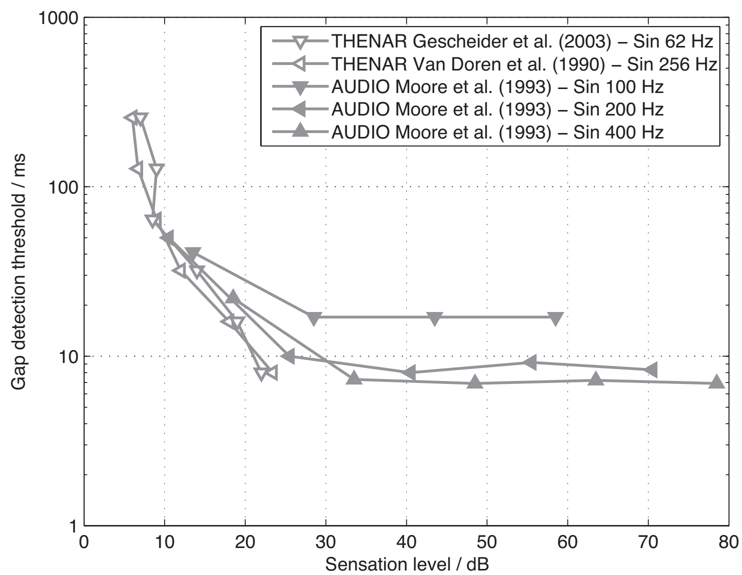


Fig. 2.18. Auditory and tactile thresholds for the detection of silent intervals in sinusoids as a function of sensation level. The results from several studies [63, 131, 166] are plotted for comparison.

gaps appear to be more difficult to detect between noise bursts than between sinusoidal bursts. Gap detection thresholds for noise and clicks were found to be 3-10 times higher for tactile perception than for hearing. In contrast, sinusoidal gap thresholds appear to be comparable between the modalities at low sensation levels. The reason for this behavior is not yet understood.

The ability to detect temporal gaps in vibration decreases marginally with age [166]. Similarly, a slight increase in auditory gap detection threshold with age has been reported [72, 130]. However, it can be assumed that aging does not result in a severe reduction in temporal resolution in either modality.

Summary. *Gap detection thresholds for sinusoidal stimuli are comparable in the tactile and auditory systems. However, thresholds for noises and clicks are not comparable between the two types of systems. The influence of the sensation level on auditory and tactile temporal resolving power is remarkably similar. Additionally, the gap detection thresholds are in the millisecond range, indicating good temporal resolution for both modalities. As will be described in Chapter 6, sounds with high sensation levels are necessary to excite just perceivable vibrations at the skin. Thus, temporal acuity for different intensities must be compared depending on the respective application.*

2.3.4 Location Discrimination

Localization is quite distinct between the auditory and tactile domains. In audition, only two input signals from the ears are available to estimate the position of an auditory event somewhere in space. In contrast, mechanoreceptors are spatially distributed all over the body, and tactile events are mostly perceived in proximity to the body.

Auditory

The localization ability of the auditory system can be partially described using the minimum angle at which two sources can be separated. This minimum audible angle (MAA) depends on the character of the sound and the position of the source relative to the listener. For impulsive sounds in front of the listener, MAAs of approximately 1° have been observed [96]. If the source moves toward one site or in the vertical direction, the minimum audible angle increases by up to several degrees. Additionally, the frequency content plays a dominant role. Distance perception is quite blurred, and familiarization with the sound plays an important role in estimating the distance of an auditory event [11].

Tactile

The spatial sensitivity of the tactile sensory system can be measured by two-point discrimination tasks, in which two spatially separated stimuli are presented either simultaneously or in rapid succession. The subjects must decide whether they feel one or two contact points. Tactile spatial acuity varies significantly across the body surface. The thresholds were observed to vary between approximately 1-2 mm and 4-5 mm depending on the location on the skin [190]. Regions with high receptor density, e.g., the fingers, have low spatial discrimination thresholds, whereas areas with low receptor density, e.g., the back, have low spatial acuity.

Effects similar to auditory phantom source localization or precedence have been observed when simulating two spatially separated areas [183], suggesting similar neural mechanisms for both modalities.

Some research has even attempted to reproduce the localization ability of our hearing system. If two spatially separated microphones are used to drive two vibration actuators mounted to the forearms [181] or fingertips [52], subjects can accurately localize sound sources after some training. Interestingly, many subjects reported that 'tactile sensations were projected out into space' to match the position of the corresponding sound source. This decoupling of receptor location and the perceptual event is known from vision and audition [161]. In the context of music-driven vibrations, this effect might explain the unconscious existence of vibrations, e.g., in a concert hall, as will be described in Chapter 6. The tactile receptors in the skin are stimulated; however, the localization of the perceptual event might be projected toward the source of the sound and vibration.

Summary. *It is difficult to compare the localization ability between the two modalities. Auditory events can be perceived everywhere around the listener; however, their resolution is quite limited. The spatial resolution of somatosensation is generally more detailed, but tactile events are restricted to the proximity of the skin. However, the feasibility of the projection of tactile events toward a sound source has been demonstrated. Sensory substitution systems for the hearing impaired use the good location discrimination of the tactile system to encode information, such as the frequency of a sound, to overcome shortcomings in tactile frequency perception.*

3

Development of the Experimental Setup

To investigate the perception of vibrations and sound further, it was necessary to develop a reproduction system that was capable of generating WBVs and audio signals separately. In Section 3.1, the audio hardware used is briefly introduced, and in Section 3.2, the design of a vibration chair is discussed.

3.1 Audio Reproduction

The audio signals in the following experiments were reproduced using Matlab software, and they were delivered through an external Hammerfall DSP Multiface II sound card. For the fundamental experiments, which will be described in Chapters 4 and 5, the sounds were amplified using a Phone-Amp G93 and were reproduced through a set of Sennheiser HDA 200 closed dynamic headphones. These headphones were selected to attenuate possible sound radiation from the shaker passively. The passive attenuation at 100 Hz was approximately 14 dB and increased toward higher frequencies.

During the music reproduction experiments in Chapter 7, a surround setup was used, according to ITU-R BS.775-1 [86], with five Genelec 8040A loudspeakers and a Genelec 7060B subwoofer. The setup was build in front of a silver screen for video projection. The system was equalized to a flat frequency response at the listener position.

3.2 WBV Reproduction

Pneumatic, hydraulic, or electrodynamic systems are typically used to generate WBVs. These systems differ in terms of their usable frequency ranges, displacement, and load. Pneumatic and hydraulic systems can handle large loads and produce large displacements, but they create some noise and are limited to low frequencies [5]. New hydraulic systems with wider bandwidths have recently been developed [8]. When investigating audio-driven vibrations,

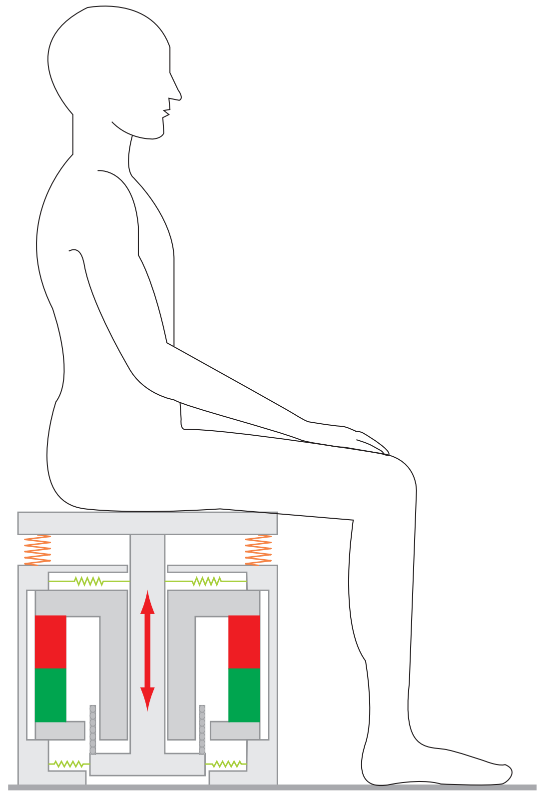


Fig. 3.1. Vibration chair with electrodynamic exciter.

frequencies from 10 to 200 Hz or more are important. In this frequency range, electrodynamic exciters are advantageous. Such systems have good frequency response and compact construction. Most consumer products are based on electrodynamic reproduction technology due to the low manufacturing costs and ease of installation.

This study was conducted using an electrodynamic shaker (self-made, based on an RFT Messelektronik Type 11076 with an Alesis RA 150 amplifier). WBVs were generated vertically, as shown in Figure 3.1.

The subjects were asked to sit on a flat, hard wooden seat (46 cm x 46 cm) with both feet flat on the ground. If necessary, plates were placed beneath the subjects feet. The transfer characteristic of the vibrating chair (relation between acceleration at the seat surface and input voltage) was strongly dependent on the individual person. This phenomenon is referred to as the body-related transfer function (BRTF). Differences of up to approximately 10 dB have been measured for different subjects [7]. Considering the just-noticeable difference in thresholds for vertical WBVs discussed in Section 2.3.1, which

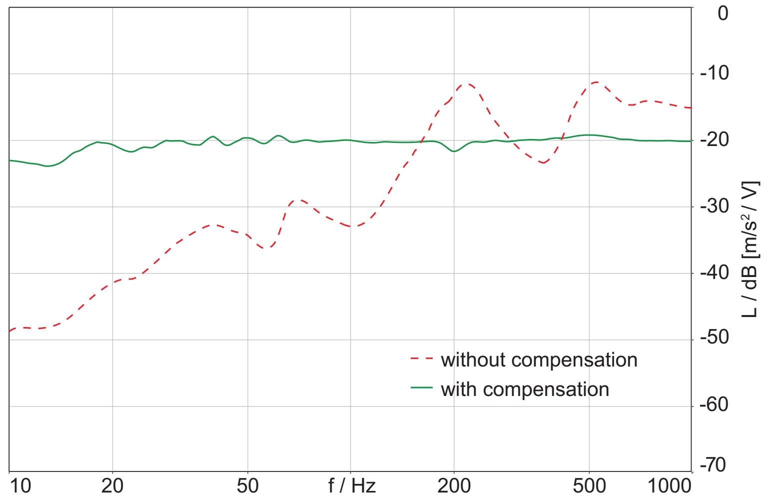


Fig. 3.2. Body-related transfer functions (FFT 65536, 1/24th octave intensity averaging) measured at the seat surface of the vibration chair, with and without compensation.

is approximately 1 dB, the individual BRTFs should be compensated for during perceptional investigations. The BRTF of each subject was individually monitored and equalized during all experiments. Subjects were instructed not to change their sitting posture after calibration until the end of the experiment. The transfer functions were measured using a vibration pad (B&K Type 4515B) and a Sinus Harmonie Quadro measuring board, and they were compensated for using inverse filters in Matlab. This procedure resulted in a flat frequency response over a broad frequency range (±2 dB from 10 to 1,000 Hz). An exemplary BRTF, with and without individual compensation, is shown in Figure 3.2.

However, although the frequency response in the vertical direction is interesting, the crosstalk to the horizontal vibration axis must also be considered. This crosstalk was minimized using hard, nonslotted disk springs to center the shaker. This configuration has the disadvantage that high power is needed to drive the shaker; however, more than a -10 dB difference in the acceleration levels between vibrations in the vertical and horizontal directions could be achieved over a broad frequency range. The resulting crosstalk between shaker axes is shown in Figure 3.3. The remaining crosstalk might not be crucial because a 10 dB increase in acceleration level corresponds to more than a twofold increase in perceived vibration intensity (refer to Section 5.2 for a detailed discussion). In addition, the perception threshold for horizontal seat vibrations is approximately 10 dB higher than for vertical WBVs [133].

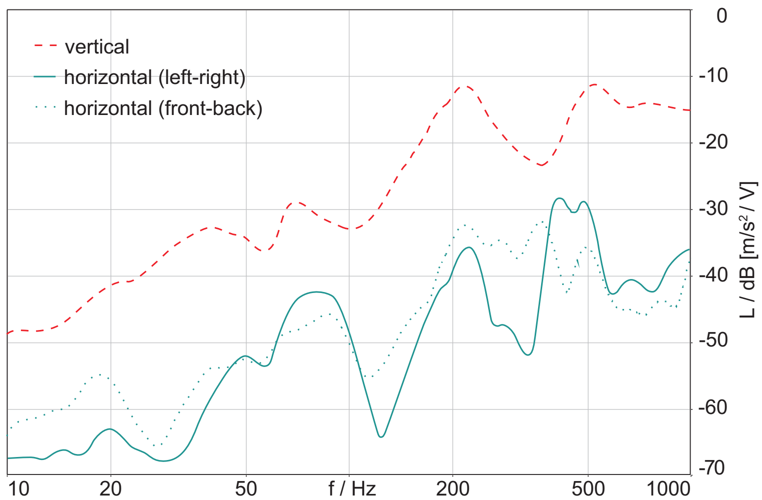


Fig. 3.3. Crosstalk between the vertical and horizontal vibration axes for body-related transfer functions (FFT 65536, 1/24th octave intensity averaging) of the vibration chair.

4

Frequency Perception

4.1 Introduction

It was found in Chapter 2 that frequency resolution is one of the major differences between the auditory and tactile senses. Thus, frequency resolution will be discussed in greater detail in this chapter. An experiment will be described that determined the ability to discriminate between different frequencies of WBVs.

In audition, stimuli frequency is related to pitch perception. The total number of perceptible pitch steps in the range of human hearing is approximately 1,400 [137]. The lowest frequency at which sound is perceived as a tone is approximately 16 Hz. For even lower frequencies, it is possible to follow the time structure of a signal. The perceived character of the sound changes, and pitch perception fades [124].

In the tactile domain, it is also possible to sort vibrations on a frequency-related scale. However, Békésy [182] already noted that there is not a high degree of similarity between pitch sensation in hearing and on the skin. As discussed in Section 2.3.2, the difference limen of tactile frequency discrimination at the hand and forearm are much higher compared to those in audition. There are, however, only limited data available for WBVs.

4.2 Frequency Discrimination of WBVs

Only one study, conducted by Bellmann [10], has addressed frequency discriminability of WBVs to date. Six subjects participated in his experiment. Four reference frequencies (5, 10, 20, and 40 Hz), with a fixed acceleration level of $L_{\mathrm{acc}} = 96\,\mathrm{dB}$, were tested. Bellmann found a linear increase in the JNDF with increasing frequency. The difference limen ranged from approximately 0.4 Hz (for a reference frequency f_{ref} of 5 Hz) to 12 Hz (for f_{ref} equals 40 Hz), corresponding to Weber fractions ($\frac{\mathrm{JNDF}}{f_{\mathrm{ref}}}$) of 8% to 30%, respectively. There are no known studies that have determined JNDFs for WBVs above

40 Hz, the frequency range that is most interesting in the context of music perception. Therefore, JNDFs were measured over a broader frequency range in this study.

4.2.1 Setup

WBVs were generated vertically using the electro-dynamic shaker described in Section 3.2. The participants were asked to sit on a flat, hard wooden seat with both feet on the ground. BRTFs were calibrated individually. The audio signals were delivered using closed dynamic headphones.

4.2.2 Subjects

Fifteen subjects (nine male and six female) voluntarily participated in the experiment. Most of the participants were students and were between 19 and 27 years old (mean, 23 years old). The participants weights ranged from 62 to 85 kg (mean, 71 kg). The subjects indicated that they had no hearing impairments or spinal damage.

4.2.3 Stimuli and Experimental Design

To accurately measure the differential frequency thresholds of vibratory stimuli, concomitant changes in the subjective intensity should be eliminated. Therefore, equal-vibration perception curves were measured for WBVs in a separate experiment (see Section 5.3). Using these data, the stimulus amplitudes in the current experiment were carefully normalized to achieve a uniformly perceived magnitude (approximately 20 dB above the threshold).

A preliminary experiment indicated that above 90 Hz, the perceived location of the vibration started to move along the thigh depending on various factors, such as clothing and posture. These localization cues for stimulus discrimination complicated the measurements of the differential frequency thresholds. Thus, reference frequencies (f_{ref}) between 20 and 90 Hz at 10 Hz intervals were chosen.

An adaptive, two-interval, forced-choice, one-up/one-down psychophysical procedure was used. The subjects were presented with two consecutive randomized stimuli and asked to decide whether the frequencies were identical or not. If no difference was observed, the frequency spacing was increased or otherwise decreased. Both stimuli had a length of 1 s, with a 0.5 s break between them. The test stimulus started at a frequency (f_{test}) of 80 Hz above the frequency of the reference stimulus (f_{ref}). Thus, this experiment measured only positive JNDFs ($f_{\text{test}} > f_{\text{ref}}$). The initial step size of 20 Hz was halved after each upper reversal (up to down) to a final step size of 2.5 Hz. The measurement phase started after the final step size was reached, and this phase was terminated after five upper reversals. The individual difference threshold was obtained by calculating the mean of the values during the measurement phase. Each frequency was measured twice for each subject. The entire experiment lasted approximately 30-40 min per participant, including an initial familiarization phase.

4.2.4 Results and Discussion

The measured individual difference limens in frequency are plotted in Figure 4.1 as the mean values and their intra-individual standard deviations. The intra-individual standard deviations remained almost constant, whereas the inter-individual variations increased with increasing frequency (e.g., approximately 2 Hz at $f_{\text{ref}} = 20$ and 27 Hz at $f_{\text{ref}} = 90$ Hz). These relationships can be seen in Figure 4.2, which presents the data averaged across all of the subjects with the corresponding inter-individual standard deviations. The mean JNDF increased with increasing frequency, from approximately 7 Hz at $f_{\text{ref}} = 20$ to 66 Hz at $f_{\text{ref}} = 90$ Hz. The Pearson product-moment correlation between the logarithmized JNDFs and the logarithmized reference frequency was statistically significant, with $r^2 = 0.94$ ($p<0.01$). Therefore, a linear regression curve was calculated using a least-squares approach in the double-logarithmic

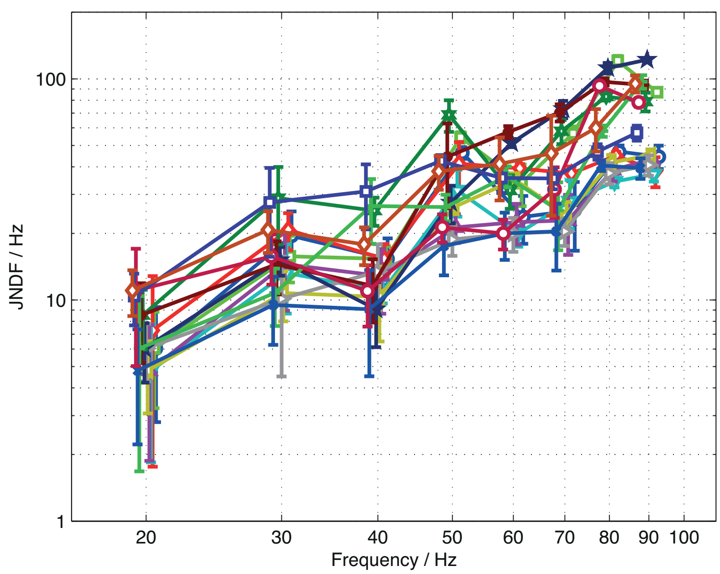


Fig. 4.1. Mean of the individual difference threshold ± intra-individual standard deviations for all 15 subjects. To illustrate the results better, the frequency of each data point was shifted slightly.

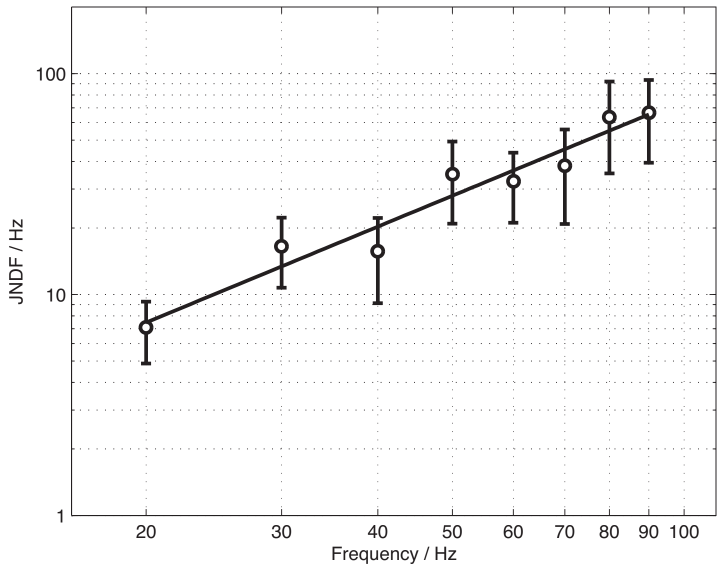


Fig. 4.2. Averaged frequency difference thresholds ± inter-individual standard deviations for all 15 subjects. Additionally, a first-order regression line was fitted to the data.

domain. This curve corresponds to a power function in the linear domain: $\text{JNDF} = 0.1 * f_{\text{ref}}^{1.44}$. The regression curve is plotted in Figure 4.2.

Because the difference limen tends to vary directly with frequency, it is common to plot the results as Weber fractions ($\frac{\Delta f}{f_{\text{ref}}}$). Weber found that the perceivable difference in the stimulus intensity is a constant fraction of the reference intensity for medium-intensity stimuli [187].

The solid line in Figure 4.3 represents the data plotted in Figure 4.2. The Weber fraction increases from approximately 35% to 80% at higher frequencies. The values at 20 and 40 Hz are comparable to the WBV results determined by Bellmann, who used a similar psychophysical procedure [10]. However, his values were considerably smaller at lower frequencies. In his study, improved discrimination could at least partially be explained by the absence of compensation for concomitant changes in the subjective intensity.

When comparing the results of the current study to the frequency difference limens reported in the literature for various locations on the arm and hand, inferior frequency discrimination was observed. This finding could be partially explained by the fact that only positive JNDFs ($f_{\text{test}} > f_{\text{ref}}$) were measured here. The frequency difference threshold is typically considered to be half of the difference limen for positive and negative JNDFs. Another possible reason for the superior discrimination results in the literature might be

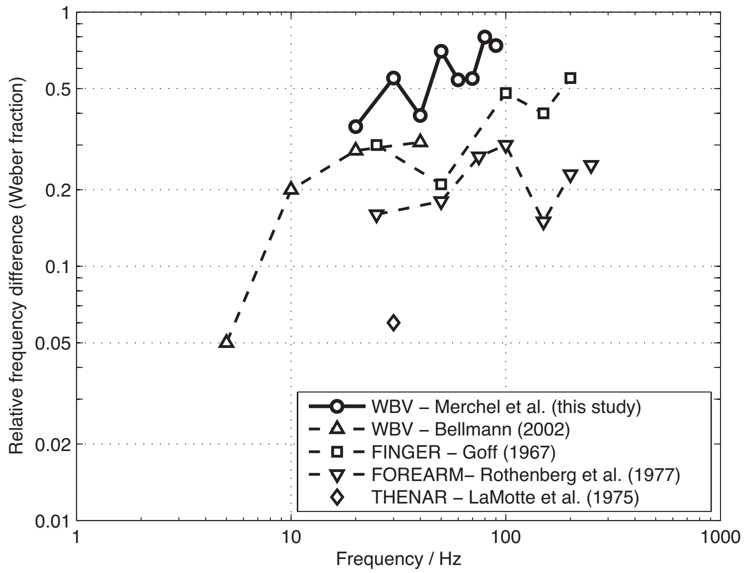


Fig. 4.3. The relative frequency difference ($\frac{\Delta f}{f_{\text{ref}}}$; Weber fraction for frequency) for sinusoidal vibrations at various body sites is shown as a function of the reference frequency. The dashed lines and the diamond indicate results from other laboratories.

the higher tactile sensitivity of the hand and/or forearm. Interestingly, the frequency discrimination observed by Rothenberg et al. at the forearm [147] was better than that observed by Goff at the finger, which is more sensitive [65]. The even more superior result by LaMotte et al. at 30 Hz [99] could be due to the use of highly trained subjects.

4.2.5 Summary

In this section, the ability to discriminate among various vertical, sinusoidal frequencies is discussed for WBVs. JNDFs were measured while carefully eliminating concomitant changes in the subjective intensity or stimulation location. The results can be summarized as follows.

- The JNDFs increased with frequency, from approximately 7 Hz at $f_{\text{ref}} = 20$ to 66 Hz at $f_{\text{ref}} = 90$ Hz.
- This increase corresponds to Weber fractions ($\frac{\Delta f}{f_{\text{ref}}}$) between approximately 35% and 80%.
- Only positive JNDFs were measured ($f_{\text{test}} > f_{\text{ref}}$).

It is known from audition that the threshold required to discriminate between two slightly different frequencies depends on the stimulus level to some extent. Increased sound pressure levels result in lower JNDFs [193]. In this study, moderate acceleration levels (20 dB above the threshold) were used. In future studies, the dynamic ranges of the sinusoidal WBVs could vary. In addition, negative JNDFs ($f_{\text{test}} < f_{\text{ref}}$) or discrimination thresholds that are dependent on signal length could be measured. However, the above data appear sufficient to draw conclusions in the context of music-induced vibration perception.

When comparing the present results with our hearing system, much higher JNDFs were observed for tactile perception. Weber fractions of less than 1% are common for auditory frequency discrimination in a broad frequency range (see Section 2.3.2). For instance, at approximately $f_{\text{ref}} = 100$ Hz, JNDFs less than 1 Hz have been found, which is more than one order of magnitude lower than the tactile JNDFs measured in this study. Thus, only few different tactile sensations can be evoked by varying the vibration frequency.

In summary, tactile frequency perception is weak, and thus, pitch might be of minor importance in the process of vibration generation from music. For instance, it might be possible to strongly compress the frequency range of a vibration signal without influencing the overall vibration perception. This procedure would allow for the use of simple and inexpensive narrow-band inertial vibration actuators and will be evaluated further in Sections 7.2.3 and 7.2.4. This hypothesis is supported by a study from Abercrombie and Braasch [3], who found that variations in frequency content could barely explain the perceived differences between music-induced stage vibrations. They presented listeners with various WBV signals and asked them to rate the degree of perceived difference in a paired comparison test. Multidimensional scaling was used to explore the underlying perceptual dimensions. Interestingly, sensation level strongly dominated the perceived differences between vibration signals, whereas frequency content or audio-tactile time delays had negligible influence. Thus, the next chapter discusses auditory and tactile intensity perception in further detail.

5

Intensity Perception

5.1 Introduction

In Chapter 2, the auditory and tactile absolute and differential sensitivities were compared in terms of stimulus intensity. However, to characterize a sensory channel, not only are thresholds and difference limens important but also the perceived magnitude of a stimulus above the threshold.

The examination of auditory intensity perception has a long history, and comprehensive knowledge exists. However, tactile intensity perception has not been studied as thoroughly. A short literature review will provide an overview of the current state of research, with a focus on perceived vibration magnitude. To broaden our knowledge, tactile and auditory-tactile intensity perception will be investigated in the following sections. The growth of perceived intensity of WBVs with increasing vibration level will be compared to auditory loudness in Section 5.2. Therefore, a magnitude estimation experiment will be performed. A new descriptor (perceived vibration magnitude M in vip) is defined to represent human vibration intensity perception, comparable to auditory loudness N in sones [158], to be able to quantify the tactile sensation magnitude.

Using the magnitude estimation data in Section 5.2, curves of equal vibration intensity can be determined. These curves can also be measured directly by matching the intensities of various test vibrations to reference vibrations at different levels. These measurements will be depicted for comparison in Section 5.3.

If the tactile and auditory senses are stimulated simultaneously, e.g., in a musical context, it might be necessary to match the sensation magnitudes in both modalities to form a coherent and immersive perception. To provide the necessary groundwork for this procedure, Section 5.4 addresses the measurement of cross-modal intensity matching using sinusoids.

Additionally, cross-modal interactions might occur. Thus, Section 5.5 investigates whether WBVs have an influence on loudness perception.

Auditory

The perceived magnitude in audition is referred to as loudness. The unit of loudness is the sone, which was defined by Stevens [158]. The relationship between sensation and physical stimuli can be described by Stevens' power law [159], which states that the sensation magnitude Ψ grows as a power function of the physical stimulus magnitude ϕ:

$$\Psi = k\phi^n. \tag{5.1}$$

The constant k changes depending on the units of measurement. More interestingly, the value of the Stevens' exponent n is characteristic for each sensory modality. The power law can be extended using an additive constant ϕ_0, which represents the perception threshold [31]:

$$\Psi = k(\phi - \phi_0)^n. \tag{5.2}$$

If plotted with logarithmic scales on both axes, the power function becomes a straight line. Therefore, data fitting can be undertaken easily using linear regression by reformulating Equation 5.2 in terms of logarithms:

$$\log \Psi = n \log(\phi - \phi_0) + \log k. \tag{5.3}$$

The exponent n now represents simply the slope of the line. For loudness perception, Stevens' exponent depends on frequency. At 1 kHz, this exponent was observed to be approximately 0.6, which corresponds to an increase of 10 dB in sound pressure level per doubling of loudness. However, for lower frequencies or low sensation levels, the perceived magnitude grows more rapidly. This relationship is illustrated for a tone with 250 Hz in Figure 5.1 using data from Hellman and Zwislocki [74].

Tactile

Stevens' exponent for perceived vibration magnitude depends on the stimulated body site. An example for stimulation at the finger and volar arm is plotted for 250 Hz in Figure 5.1, according to Verrillo and Chamberlain [177]. A 0.28 cm^2 contactor was used in both cases. Stevens' exponent is higher for vibrations at the arm compared to the finger. It was hypothesized that this growth would be slightly more rapid for body sites with a lower density of neural innervation. This inverse dependence has also been reported for other regions of the body, such as the forearm compared to the thigh [17]. However, it could also be supposed that the rate of growth is inversely correlated with 'the total number of sensory neural units stimulated rather than being related simply to the density of neural innervation' [177]. This hypothesis is

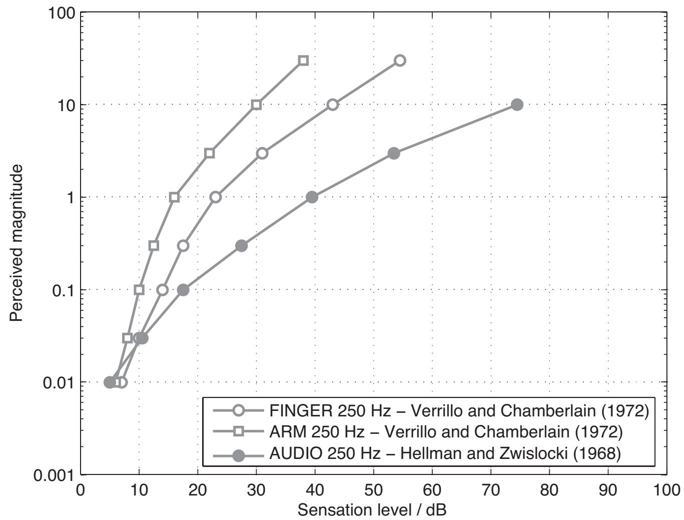


Fig. 5.1. The growth of perceived magnitude as a function of sensation level for acoustical and vibratory stimuli at 250 Hz. The acoustical stimulus was reproduced diotically via earphones [74]. Vibrations were reproduced at the finger and forearm using a 0.28 cm^2 contactor [177]. All of the studies plotted here used the method of numerical magnitude balance.

interesting in the context of music perception because sound might stimulate the whole body surface or a large portion thereof, e.g., via the seat in a concert hall.

WBVs have been investigated mainly in the contexts of discomfort, annoyance, and health risk estimation. Thus, most examinations have considered high vibration magnitudes and low frequencies. For clarity, this comparison will concentrate solely on seat vibrations in the vertical direction. Controversial data exist regarding the *perceived magnitude* of sinusoidal WBVs for subjects seated on rigid seats. Widely varying Stevens' exponents have been reported, ranging from 0.46 to 1.75 [45, 76, 80, 91, 122]. The variation in the data might be due to varying experimental procedures and the different ranges of vibration intensities investigated. Miwa [122] reported an exponent of 0.46 for vertical vibrations between acceleration levels of 107 dB and 147 dB at 5, 20, and 60 Hz. He used the corrected ratio method, which consists of two experiments based on fractionation and equisection judgments. Jones and Saunders [91] investigated the subjective response to vertical vibrations between 5 and 40 Hz. They used a magnitude estimation procedure and obtained an exponent of 0.93. Using cross-modality matching techniques, Hempstock and Saunders [76] investigated the subjective growth of whole-body vertical sinusoidal vibrations between acceleration levels of 114 and 136 dB. Frequencies between 5 and 80 Hz were used. They found exponents between 0.49 and 1.42, that were greatly influenced by the applied method. Using the method of magnitude production and magnitude estimation, Fothergill and Griffin [45] obtained an exponent as high as 1.75 for a 10 Hz vibration and acceleration levels between 110 and 125 dB. Despite their contradictory results, none of the studies directly measured the perceived vibration intensity close to the threshold, i.e., the range that is most interesting for music-induced vibrations. Unfortunately, there are no data available for frequencies greater than 80 Hz.

Some studies exist, which did not investigate the perceived intensity directly. For instance, Howarth and Griffin [80] examined the *annoyance* for frequencies between 4 and 63 Hz using magnitude estimation. They reported a mean exponent of 1.2 for acceleration levels between 92 and 112 dB. Morioka and Griffin [132] asked their test participants to estimate the perceived *discomfort* of seat vibrations over a broad dynamic range. They measured relatively low Stevens' exponents and hypothesized that there would be a frequency dependence. A decrease of the exponent with increasing frequency from 16 Hz ($n \approx 0.8$) to 100 Hz ($n \approx 0.4$) was reported. However, it is questionable whether (dis-)comfort or annoyance can be compared with perceived vibration magnitude.

Therefore, the growth of perceived intensity for vertical WBVs will be investigated in Section 5.2 using a broad range of frequencies (10 to 200 Hz) and amplitudes from hardly perceivable to annoying.

Before this investigation is presented, the summary of the current state of research will be completed by discussing the influences of stimuli duration, multiple stimulation, gender, and age on the perceived vibration magnitude. The perceived intensity of a vibration also depends on stimulus duration. Gescheider and Joelson [55] reported growth of sensation magnitude during the first second, most likely due to temporal summation in the Pacinian channel (refer to Section 2.2.3). For longer exposures, the subjective magnitude decreased due to adaptation [51] (refer to Section 2.2.5).

Until now, only experiments with single sinusoidal vibrations have been discussed. The combined perceived intensity of vibratory stimulus pairs was investigated by Verrillo and Gescheider [178]. Summation of sensation magnitude was observed for stimuli that activate different tactile channels. Energy summation was observed for components close in frequency. This effect is similar to the characteristics found in audition: the sensation magnitude (loudness) is summed if two signals have a frequency spacing greater than a critical band. Intensity summation occurs for acoustical signals within a critical band.

Vibratory sensation magnitude grows slightly more rapidly for women than for men [173]. This difference was found to result in a stronger vibration sensation for women. No similar differences have been observed in audition.

The growth of perceived magnitude increases slightly at higher frequencies with age (Pacinian system) [180]. This relationship might be due to the decreasing dynamic range with age. A similar effect (loudness recruitment) has been established from audition for hearing-impaired people (see Marozeau and Florentine [105] for a review). However because the amounts of hearing loss and tactile threshold shift are dissimilar between modalities (depending on age and frequency, as discussed in Section 2.2.2), it might be expected that the concomitant steepening of the curve of perceived intensity would also be dissimilar.

5.2 Perceived Magnitude of WBVs

In this section, the perceived vibration magnitude of WBVs will be measured as a function of sensation level using frequencies from 10 to 200 Hz and a large amplitude range.

If the growth of perceived magnitude for vertical WBVs deviates from audition, these data can be used to understand and compensate for intensity mismatch effects in auditory-tactile music perception (see Section 7.3.1).

5.2.1 Setup

WBVs were generated vertically using the setup described in Section 3.2. Pink noise was presented through a set of Sennheiser HDA 200 closed dynamic headphones at 74 dB(A) to acoustically mask the noises emitted by the chair.

5.2.2 Subjects

Ten subjects (eight male and two female) voluntarily participated in the study. Most of the participants were students between 20 and 27 years old (mean, 23 years). The participants weighed between 62 kg and 83 kg (mean, 73 kg) and stated that they had no hearing or spinal damage.

The subjects were instructed to sit upright in a comfortable posture with their hands on their thighs and both feet flat on the ground. Additional plates were used to adjust the height of the feet until the thighs were approximately horizontal and level with the seat.

5.2.3 Stimuli and Experimental Design

The subjects were asked to sit upright on the vibration chair and to remain in the same position throughout the experiment. The test participants were asked to judge the intensity of the WBV at each of the selected frequencies (10, 20, 50, 100, 150, and 200 Hz). The vibrations varied in acceleration level from 90 to 130 dB in 5 dB steps. The range of stimulus magnitudes was extended to 140 dB for 100, 150, and 200 Hz to account for the rising perception threshold at higher frequencies. All of the vibration signals were faded in and out using ramps of 50 ms with the shape of half a Hann window. The presentation order of the frequencies and magnitudes was completely randomized.

The method of magnitude estimation was employed to determine intensity judgments. Stimulus pairs, consisting of a reference vibration and test vibration, which had a duration of 1 s each with a 1 s interval between them, were presented. The reference vibration was fixed at 20 Hz with an acceleration level of 110 dB. The task of the subject was to assign a number representing the perceived intensity of the test vibration relative to the intensity of the reference vibration, assuming the intensity of the reference vibration corresponded to 100. If the subjects did not perceive the test stimulus, they were instructed to type 0. These data were removed before further analysis.

The reference and test stimuli were marked visually using the experimental interface controlled by Matlab. Each test participant judged all stimuli three times in one session, lasting approximately 30 min. Before starting the experiment, the subjects were familiarized with the magnitude estimation task by judging the lengths of lines drawn on paper.

5.2.4 Results and Discussion

The mean perceived intensities were calculated by averaging the assigned numbers for each subject over the three repetitions. An example is plotted in Figure 5.2 for one exemplary subject and a vibration frequency of 20 Hz. A curvilinear relationship was observed for most test participants. For low sensation levels, the data followed Stevens' power law with an additive constant (Eq. 5.2). However, some deviation at acceleration levels greater than 120 dB were observed. This finding supports the hypothesis that there would be a tendency for subjects to apply a nonlinear scale for mapping stimulus intensities to numbers [144]. The topic has been debated for many decades, and there is

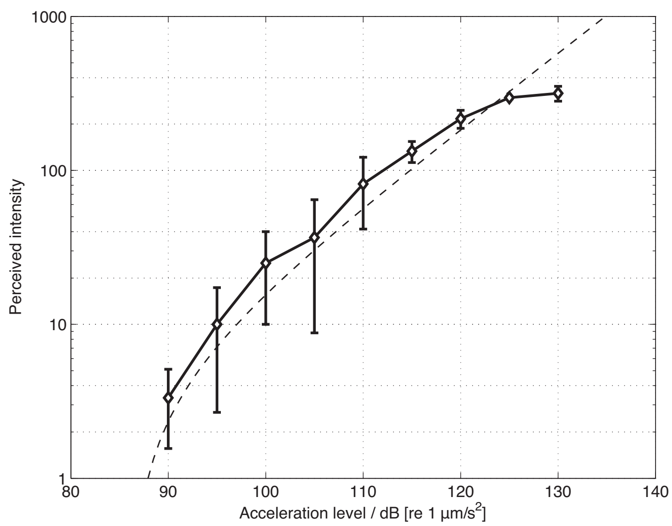


Fig. 5.2. Example of the growth of sensation for a single test participant and 20 Hz vertical WBVs. Plotted are mean values ± one standard deviation calculated for the three intra-individual repetitions. The data were fitted using Equation 5.3. The acceleration level of the reference vibration was 110 dB.

still disagreement about the adequate method for measuring suprathreshold sensation magnitude (for a review see Hellbrück and Ellermeier [73]). Stevens' power law with an additive constant was used in this study to fit the individual data to determine the relationship between the sensation magnitude and corresponding acceleration levels. The constant ϕ_0 was taken from the mean perception threshold for vertical WBVs, as discussed earlier in Section 2.2.1. All individual regressions fit the data well with coefficients of determination, r^2, of 0.9 or greater.

The mean rates of growth of sensation were determined by averaging the individually fitted exponents over all test participants. The resulting Stevens' exponents n are plotted in Figure 5.3. Mean values between 0.75 and 0.97 were obtained but with large inter-individual variances. No significant variation with frequency was measured. This finding is supported by other studies, which found little evidence of a frequency dependence of Stevens' exponent for vertical seat vibrations [80, 91, 122].

As discussed earlier, the exponents found in those studies differed depending on numerous factors. For instance, Miwa [122] obtained lower Stevens' exponents compared to the current study using the method of magnitude production. Interestingly, magnitude production typically leads to higher values than magnitude estimation [45]. However, the applied sensation levels were rather high in Miwas study. Thus, the shallower section of the curve was mea-

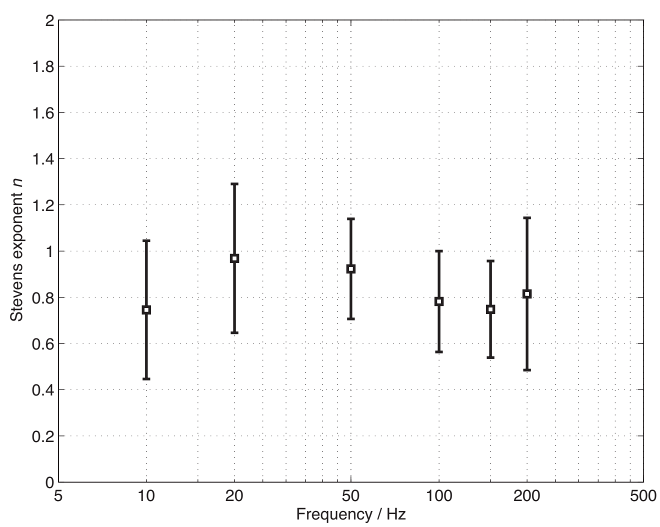


Fig. 5.3. The rate of growth of sensation expressed as Stevens' exponent as a function of vibration frequency. The means ± one standard deviation are plotted for 10 subjects. A mean Stevens' exponent of 0.83 was obtained.

sured, which could explain the lower exponents. In contrast, higher Stevens' exponents have been measured by Howarth and Griffin [80], among others. They applied only low acceleration levels and thus measured the steeper section of the curve.

Based on his data, Miwa [122] proposed a scale of *vibration greatness* to measure vibration sensation magnitude. He suggested that the unit value should correspond to a 120 dB acceleration level. Further, he provided two separate functions (based on Stevens' power law *without* an additive constant), which approximated the vibration greatness for acceleration levels less than or greater than 120 dB. However, his data seemed to underestimate the growth of sensation as reported above, and the utilization of an additional constant, representing the threshold, appears to be beneficial. Therefore, Stevens' power law with an additive constant is more suitable for approximation. In addition, according to his own data [123], a reference value of 120 dB is very close to unpleasant, indicating that typical vibrations must be described using fractions of vibration greatness, which appear to be cumbersome.

Therefore, an alternative perceptually valid scale for vibration sensation magnitude is proposed below. To quantify the tactile sensation magnitude, a new descriptor is defined to represent human vibration intensity perception. This new descriptor is called perceived vibration magnitude *M*. The measurement unit of the perceived vibration magnitude M is called Vip with the unit symbol vip, which stands for '*v*ibration *i*ntensity *p*ercept'. The perceived vibration magnitude of 1 vip is proposed to be equivalent to the perceived vibration intensity of a 20 Hz vibration, with a sensation level of 15 dB. The value of a 15 dB sensation level is reasonable because it corresponds approximately to one third of the tactile dynamic range. For vertical WBVs, this value relates approximately to a 100 dB acceleration level. The selected reference frequency of 20 Hz is relatively low in the context of music-generated vibrations; however, much lower frequencies are useful for automobile applications or in the field of occupational health. Therefore, 20 Hz is a good compromise.

Stevens' power law with an additive constant can now be formulated as follows:

$$M = k(a - a_0)^n, \tag{5.4}$$

where M represents the vibration magnitude in vip, a is the acceleration, and a_0 is the acceleration at the threshold. Using the acceleration level L_{acc} and the acceleration level at threshold L_0, Eq. 5.4 becomes

$$M = k(10^{\frac{L_{\mathrm{acc}}}{20}} - 10^{\frac{L_0}{20}})^n. \tag{5.5}$$

The measured exponent n in this study is a good starting point for modeling the perceived magnitude of vertical seat vibrations. The acceleration levels at threshold L_0 for WBV were discussed in Section 2.2.1. Finally, the constant k can be determined so that a vibration magnitude of 1 vip corresponds to a 15 dB sensation level at 20 Hz, as defined above. This calculation results in $k = 13 \cdot 10^{-6}$ for the above data.

To warrant a standardized model for the perceived vibration magnitude, the results from different laboratories should have reasonably good agreement and might then be averaged and approximated. Unfortunately, sufficient data are not available. As discussed above, Stevens exponents determined in different studies have varied. It might also be reasonable to exclude studies that investigated annoyance or discomfort. In addition, it might be beneficial to consider different common situations in which vibrations are perceived. These situations include hand and finger vibrations, as well as vibrations for standing, sitting, and lying subjects in vertical and horizontal directions. However, whether such a universal model is possible remains unresolved because the growth of perception appears to be dependent on the vibration location or number of receptors stimulated (see Figure 5.1).

Using the above definition, the mean perceived vibration magnitudes obtained in this study at 20 Hz are plotted versus sensation levels between 10 and 40 dB in Figure 5.4 by averaging over all of the test participants. The inter-individual standard deviation was low at a 24 dB sensation level, which corresponds to the reference stimulus at a 110 dB acceleration level. As expected, the variances increase with increasing distance from the reference amplitude.

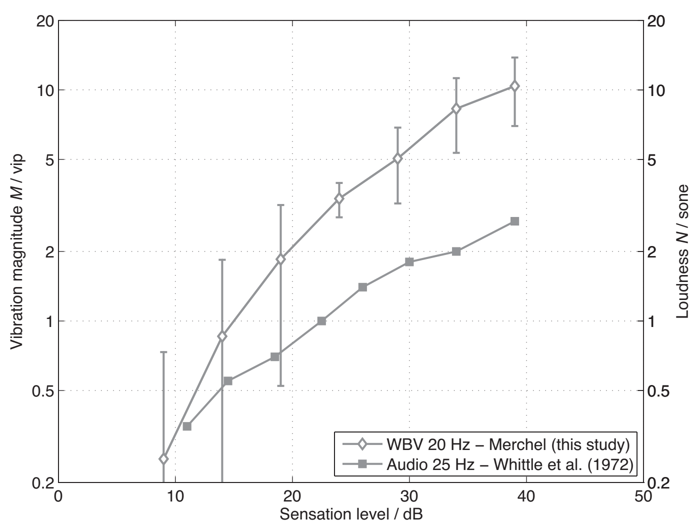


Fig. 5.4. The growth of perceived magnitude as a function of sensation level for acoustical and vibratory stimuli of approximately 20 Hz. The acoustical stimulus was reproduced using a pressure cabinet, and the method of magnitude estimation was applied [192]. The same method was used to evaluate vertical WBVs in this study. Plotted are the mean values averaged over all of the test participants ± one inter-individual standard deviation.

For comparison, auditory loudness is plotted for a 25 Hz tone in Figure 5.4 using data from Whittle et al. [192]. A pressure cabinet was used for sound reproduction, and the method of magnitude estimation was applied. The growth of vibration magnitude is clearly steeper compared to loudness, particularly at low sensation levels. This difference increases toward higher frequencies because Stevens' exponent decreases toward higher frequencies in audition, whereas Stevens' exponent remains constant in the tactile domain. These results are very interesting in the context of auditory-tactile music perception because the inconsistent growths of perceived magnitude might lead to an intensity mismatch between modalities.

5.2.5 Summary

In this section, the growth of perceived intensity was investigated for the vertical WBVs of seated subjects. The following results were obtained.

- The relationship between sensation magnitude and sensation level was determined for WBVs. A mean Stevens' exponent of 0.83 was obtained.
- The perceived vibration magnitude M was defined to represent human vibration intensity perception, comparable to auditory loudness N. The measurement unit of the perceived vibration magnitude M is called Vip, with the unit symbol vip. The perceived vibration magnitude of 1 vip is proposed to be equivalent to the perceived vibration intensity of a 20 Hz vibration with a sensation level of 15 dB.
- Compared to audition, the increase in perceived magnitude is steeper, particularly at low sensation levels.

Because weak vibrations close to the perception threshold are typically coupled with relatively loud sounds in real-life situations as discussed in Chapter 6, the steeper increase in perceived magnitude is particularly important in the context of this study. The perceived vibration magnitude might rise rapidly from imperceptible to strong if vibrations are generated from an audio signal with a large dynamic range. Therefore, it might be advantageous to apply dynamic compression during the generation process. This effect will be evaluated further in Section 7.3.1.

5.3 Equal Intensity Contours of WBVs

The magnitude estimation data in the last section can also be plotted differently. Equivalent intensity contours can be constructed using the Stevens' exponents shown in Figure 5.3. For vibration magnitudes M equal to 0.25, 0.5, 1, 2, 4, 8, and 16 vip, the corresponding acceleration levels were calculated for all frequencies. The resulting contours are shown in Figure 5.5. The sensitivity to acceleration decreases toward higher frequencies. There is only a gradual decrease in sensitivity as the frequency increases between 10 and 100 Hz, although a more rapid reduction toward higher frequencies is observed.

These curves can also be obtained directly by matching the intensity of various test vibrations to reference vibrations at different levels. Two experiments were conducted to determine the curves of equal-intensity perception of sinusoidal vertical WBVs for seated subjects. The test participants were asked to match the intensity of a test stimulus to a reference vibration using a method of adjustment. In the first experiment, two different reference frequencies (20 and 100 Hz) were used with the same level. One reference frequency was selected and used in the second experiment to determine equal intensity curves over a wider dynamic range.

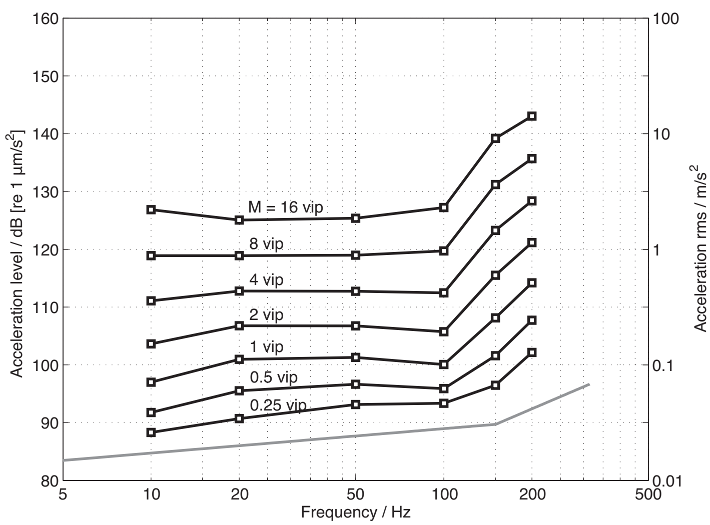


Fig. 5.5. Contours with equivalent vibration intensity perception of vertical WBVs, for perceived vibration magnitudes M from 0.25 to 16 vip. The contours were calculated from magnitude estimation data.

5.3.1 Setup

The setup was the same as in Section 5.2 using the individually calibrated vibration seat. As in the previous experiment, test subjects were exposed to masking pink noise at 74 dB(A) via a pair of headphones.

The participants were able to control the amplitude of the test vibration using a rotary knob that was infinitely adjustable and that did not possess any visual indicators (Griffin Technology, PowerMate).

5.3.2 Subjects

Ten subjects (six male and four female) voluntarily participated in both experiments. Most of the participants were students between 19 and 27 years old (mean, 23 years). The participants weighed between 62 kg and 85 kg (mean, 71 kg) and stated that they had no hearing or spinal damage.

The subjects were instructed to sit upright in a comfortable posture, with their hands on their thighs and both feet flat on the ground. Additional plates were used to adjust the height of the feet until the thighs were approximately horizontal and level with the seat.

5.3.3 Stimuli and Experimental Design

Seven test frequencies over a wide frequency range (10, 20, 50, 100, 150, 200, and 250 Hz) were selected for the experiments. In the first experiment, two frequencies were selected as reference frequencies (20 and 100 Hz) with a fixed acceleration level of 110 dB. The reference vibration was presented for 1 s, followed by a 0.5 s break. Afterwards, one of the test frequencies was reproduced for 1 s. The order of the test frequencies was completely randomized. All of the vibration signals were faded in and out using ramps of 50 ms with the shape of half a Hann window. Reference and test stimuli were marked visually using the experimental interface, controlled by Matlab. The task of the subject was to adjust the perceived intensity of the test vibration adaptively to the perceived intensity of the reference vibration using the rotary knob. The test frequency was adjustable with a minimum step size of 0.25 dB. The initial acceleration level of the test WBV was 90 dB $\pm$ 5 dB (a random offset was used for each trial). This sequence was automatically repeated until the participant was satisfied with his/her match. The subject was free to take as much time as necessary to make the proper adjustments. The total duration of the experiment varied between participants and took between 25 and 35 min to match both reference vibrations with all seven test vibrations (including a familiarization phase at the beginning). For the 100 Hz reference condition, each match was repeated three times for all of the participants to check intra-individual repeatability. After analysis of the first experiment, a second experiment was conducted in a separate session.

In the second experiment, a reference frequency of 20 Hz was used with various acceleration levels, from relatively low to moderately strong values (100, 105, 110, 115, and 120 dB). The experimental design was identical to the first experiment. The total duration of the experiment varied between 35 and 45 min depending on the individual subject.

5.3.4 Results and Discussion

Experiment 1 - Different Reference Frequencies

The equal-intensity contours are plotted in Figure 5.6, with mean and inter-individual standard deviations. The mean values increased depending on the frequency from 10 to 250 Hz. The inter-individual standard deviations increased with increasing distance from the reference frequencies. However, the averaged contours agreed reasonably well. Thus, the 20 Hz stimulus was used as a reference in the following experiment.

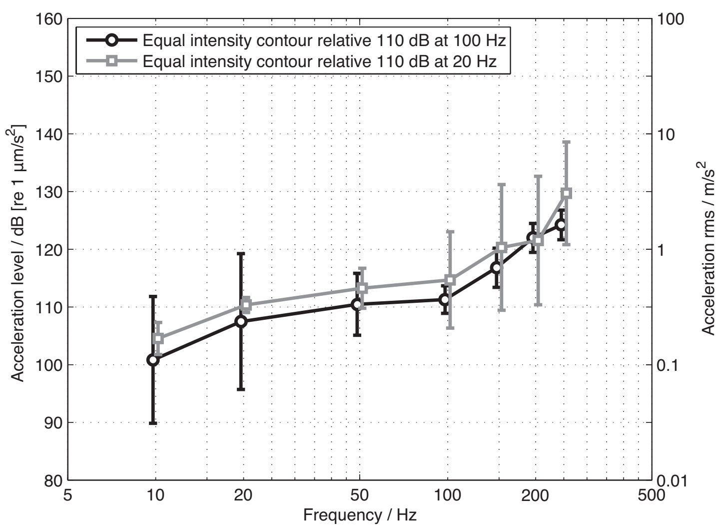


Fig. 5.6. Comparison between averaged equal intensity contours for vertical sinusoidal WBVs using reference frequencies of 20 and 100 Hz with an acceleration level of 110 dB. The plot presents mean values ± one standard deviation for 10 subjects.

Experiment 2 - Different Magnitudes

Figure 5.7 presents the averaged equal-intensity contours for vertical sinusoidal WBVs using a reference frequency of 20 Hz with acceleration levels of 100, 105, 110, 115, and 120 dB. The contours are almost parallel at lower frequencies and exhibit some deviations at higher frequencies. This finding can be explained by the increasing distance from the reference stimulus. Two subjects in particular presented inconsistencies for higher test frequencies. The inter-individual standard deviations increased with increasing frequency and exhibited no dependence on vibration magnitude. The deviations were comparable to the results shown in Figure 5.6 and thus will not be plotted here. The averaged contours increased depending on the frequency from 10 to 250 Hz. The slope was shallow up to 100 Hz and steepened for higher frequencies up to 250 Hz. The curves increased above 100 Hz at a rate of approximately 9 dB per octave.

The data from Section 5.2 were added to Figure 5.7 for comparison. Good agreement between the datasets can be observed. Both measurements exhibit a pronounced bend at approximately 100 Hz. However, a slightly stronger increase in acceleration with increasing frequency was observed in the current experiment.

Few studies have investigated equal perception contours for low magnitudes of vibration and/or a broad frequency range. Bellmann [10] measured an equal intensity contour between 5 and 80 Hz using an adaptive two-interval

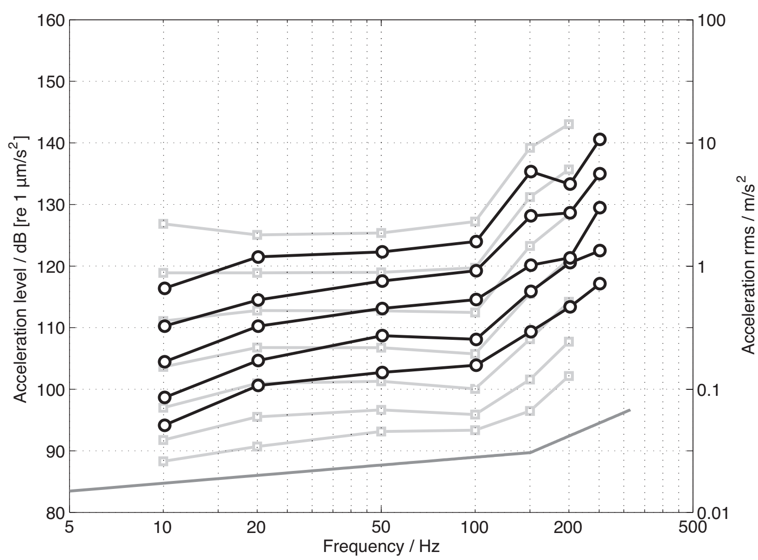


Fig. 5.7. Comparison of equal intensity contours measured in this experiment (black) and contours derived from magnitude estimation data in Section 5.2 (gray).

forced-choice procedure. A reference frequency of 20 Hz at 100 dB was used, which corresponds to the weakest reference stimulus in the present study. The resulting contour is plotted in Figure 5.8 for comparison, indicating reasonable agreement.

Morioka and Griffin [132] used magnitude estimation to determine judgments of *discomfort* caused by vertical vibrations over a wide frequency and dynamic range. Equal (dis-)comfort contours were then calculated from the data. The resulting curves are plotted in Figure 5.8 for comparison. The bottom and top contours in the plot do not predominantly represent measured vibrations but are determined by extrapolation. The contours exhibit a similar tendency as in the present study but with no pronounced break at 100 Hz. However, a dependence of the magnitude on the shape of the equal comfort contours was reported. This result differs somewhat from the present findings but might be explained by the different experimental question (discomfort versus intensity).

Earlier data can be found in Howarth and Griffin [80], who used an acoustical reference stimulus (one-third-octave band noise centered at 1 kHz) to measure *annoyance* caused by vibration using a method of magnitude estimation. Only frequencies up to 63 Hz were measured using a small dynamic range. Interestingly, the resulting annoyance contours were similar to the equal intensity contours from the present study, as illustrated in Figure 5.8.

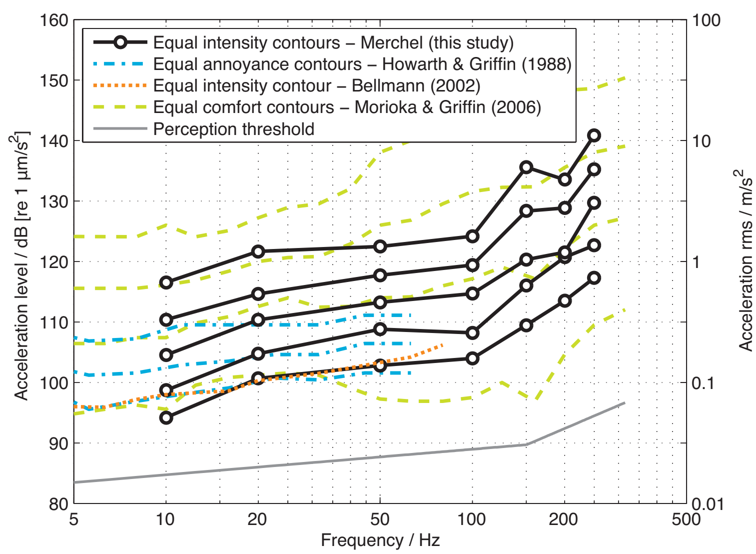


Fig. 5.8. Comparison of equal perception contours for sinusoidal vertical WBVs from different laboratories.

5.3.5 Summary

In this section, contours of equally perceived intensity were determined for vertical sinusoidal WBVs. The shapes of the equal intensity contours in the present experiment agreed reasonably well with the contours from the experiment in Section 5.2 and with other studies despite the use of different methodologies and experimental questions. The data from Morioka and Griffin [132] exhibit a level dependence in their equal comfort contours, indicating that the rate of growth of perceived vibration comfort varied with frequency. However, the results for vibration intensity discussed above do not support this hypothesis, at least at frequencies of less than 250 Hz.

In contrast, auditory intensity perception is strongly dependent on frequency, as illustrated in Figure 2.7a. If the tactile and auditory senses are stimulated simultaneously, e.g., in a musical context, it might be necessary to match the sensation magnitudes in both modalities to form a coherent and immersive perception. To provide the necessary groundwork, a cross-modal matching experiment will be described in the following section.

5.4 Auditory-Tactile Intensity Matching

The equivalence of sound and WBVs has rarely been investigated in terms of sensation magnitude. Previous studies have focused on the level of participant annoyance and comfort. The stimuli used in these studies have varied from sinusoidal signals (e.g., a 1,000 Hz tone matched with a 10 Hz vibration [39]) to railway noise [79]. The combined effects of noise and vibration on participant performance have also been investigated [149]. Moreover, Kaufmann et al. [94] matched the narrow-band white noise of vibration (center frequency 31.5 Hz) and sound (center frequency 100 Hz) at three different levels of sound pressure (70, 75, and 80 dB). The results of their study indicated that a 5 dB increase in sound pressure level led to a 2 dB increase in matched acceleration level. However, only stimuli with non-identical center frequencies were investigated using a narrow dynamic range.

In this section, the point of subjective equality (PSE) between the intensity of sound and vibration will be determined using three sinusoidal stimuli (50, 100, and 200 Hz) at various intensities.

Two experiments were conducted to determine the PSE for pure tones and sinusoidal WBVs. In the first experiment, tones and vibrations with identical frequencies were employed, whereas WBVs and tones with differing frequencies were investigated in the second experiment.

5.4.1 Setup

WBVs were generated vertically using the electro-dynamic shaker described in Section 3.2. The subject was asked to sit on a flat, hard wooden seat with both feet on the ground. The BRTFs were calibrated individually. The audio signals were delivered as described in Section 3.1 using closed dynamic headphones. The participant was able to control the amplitude of the vibration using a rotary knob that was infinitely adjustable and that did not possess any indicators, such as an on-or-off mark (Griffin Technology, PowerMate).

5.4.2 Subjects

Ten subjects voluntarily participated in both experiments (eight male and two female). Most of the participants were students between 20 and 29 years old (mean, 23 years). The participants weighed between 58 and 95 kg (mean, 77 kg) and stated that they did not have any hearing or spinal damage.

5.4.3 Stimuli and Experimental Design

Three sinusoidal frequencies were selected for this study (50, 100, and 200 Hz). Tonal and vibrational signals were simultaneously emitted for 1 s and were faded in and out using ramps of 50 ms with the shape of half a Hann window. Tones with a fixed loudness of 1, 4, 16, and 64 sone were used as references. Figure 5.9 presents the selected reference tones and corresponding isophones (ISO 226:2003 [83]). Isophones were selected as a reference for cross-modality matches to compare the results to equal-vip contour plots from previous experiments.

The task of the participant was to match the amplitude of a vibration to the loudness of a tone. The subjects were able to adjust the intensity of the vibration adaptively using a rotary control knob with a minimum step size of 0.25 dB. The initial acceleration level of the WBV was 90 dB $\pm$ 5 dB. A random offset within this range was used for each trial. A low initial acceleration was necessary because the dynamic range of the perception of vibration was small and a high level of vibration could cause discomfort (see Section 2.2.1). However, during the training period, the subjects were encouraged to test the entire dynamic range of the reproduction system using the manual amplitude adjustment.

The test stimulus was followed by a 1 s break, and this sequence was repeated until the subject was satisfied with the intensity match. The subject

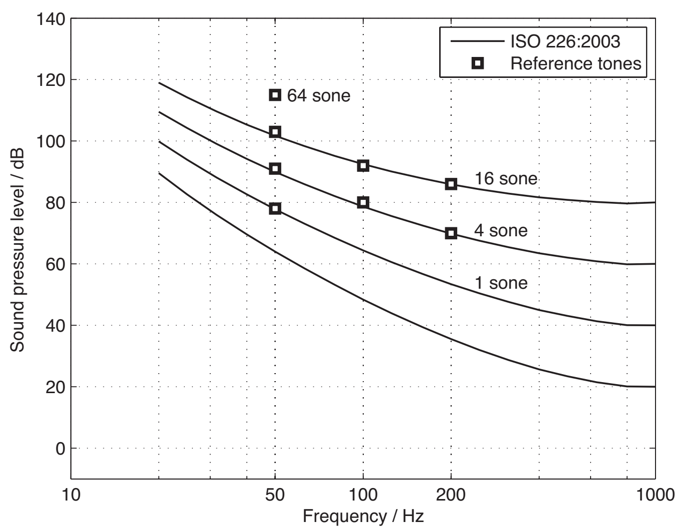


Fig. 5.9. The reference stimuli were selected to fit equal-loudness contours based on ISO 226:2003 [83].

was free to take as much time as necessary to make the proper adjustment. Each condition was repeated five times for each participant. Before the test began, a 5 min training period was conducted to familiarize the participant with the stimuli and test procedure. The total duration of the experiments varied between 20 and 40 min depending on the individual participant.

In the first experiment, the frequencies of the tone and vibration were identical. Three representative frequencies were selected from the overlapping auditory-tactile frequency range: 50, 100, and 200 Hz. Each frequency was tested at loudnesses of 4 and 16 sones. To evaluate the effect of loudness in more detail, additional tones with loudnesses of 1 and 64 sones, at a frequency of 50 Hz, were selected.

Sound and vibrations have been shown to integrate well into one multimodal percept if the frequency of the tone is a harmonic of the vibration frequency [6]. To use this effect for the generation of a vibrational signal from an audio recording (see Section 7.2.3), the relationship of cross-modal intensity was investigated for non-identical frequencies in the second experiment. In this experiment, the frequency of vibration was fixed at 50 Hz and the frequency of the tone varied between 50, 100, and 200 Hz at a constant loudness of 4 sones.

5.4.4 Results and Discussion

Experiment 1 - Identical Frequencies

Figure 5.10 presents the level of acceleration that was matched to the 4 sone (lower curve) and 16 sone (upper curve) tones (averaged over all of the subjects and repetitions). As expected, the amplitude of the matched vibrations was higher for the 16 sone tone than for the 4 sone tone. The curves were parallel at a distance of approximately 6 dB. In addition, the 50 and 100 Hz tones were matched at the same level of acceleration. However, at 200 Hz, the matched acceleration increased by approximately 10 dB. The loudness of the reference tones was equal; thus, these curves paralleled the equal-vip contours. This result can be seen in Figure 5.11, in which the results from cross-modal matching are compared with the equal-intensity contours determined in Section 5.3.

In this study, inter-individual differences were much larger than intra-individual deviations. This effect can be readily observed in the results shown in Figure 5.12, which displays the results of the 4 sone reference tones. The plots illustrate good intra-individual repeatability. However, large deviations between subjects were observed. This finding might be partially explained by varying perception thresholds. In addition, the individual experience with combined vibro-acoustical stimuli might have influenced the cross-modal matching.

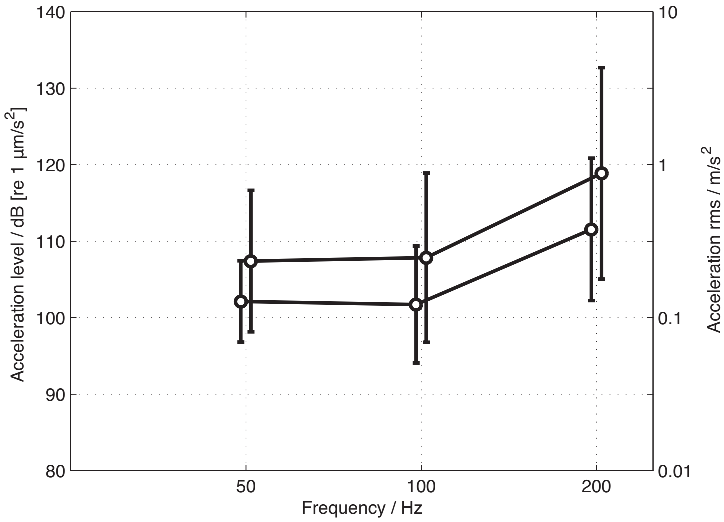


Fig. 5.10. Mean ± one standard deviation of the inter-individual cross-modality matching results, based on reference tones with a loudness of **4 sones** (lower curve) and **16 sones** (upper curve). The frequency of each data point was shifted slightly to illustrate the results more clearly.

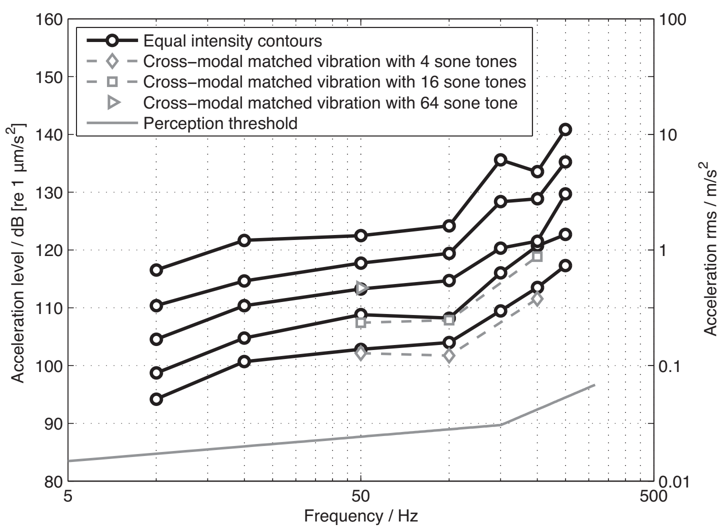


Fig. 5.11. Comparison of equal-intensity contours for sinusoidal vertical WBVs (Section 5.3) with cross-modally matched vibrations with tones from equal-loudness contours (this section).

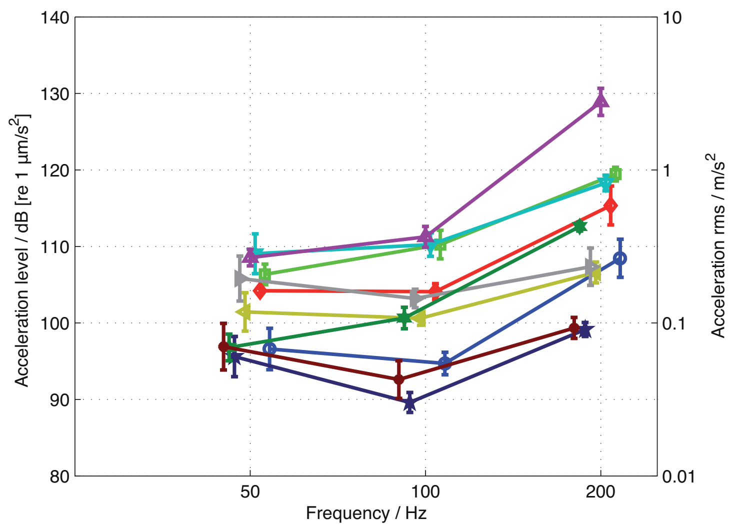


Fig. 5.12. Mean ± the standard deviation of individual results of cross-modality matching based on reference tones with a loudness of **4 sones**. To illustrate the results clearly, the frequency of each data point was shifted slightly.

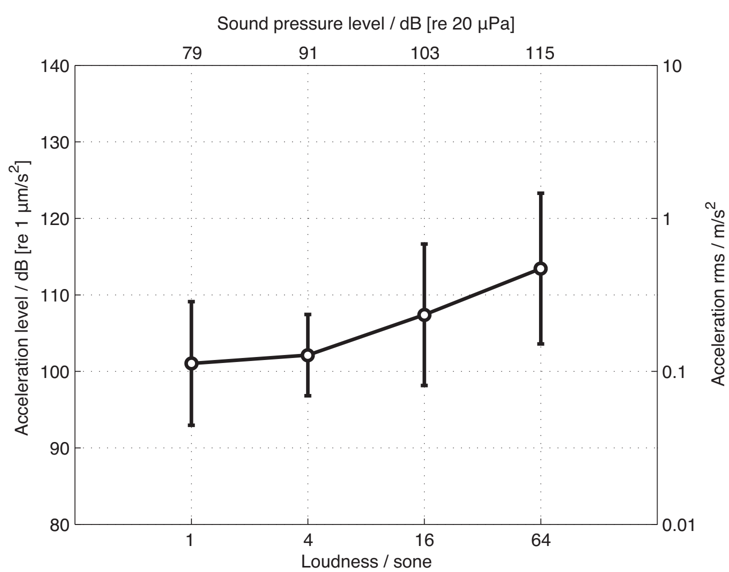


Fig. 5.13. Mean and standard deviation of the inter-individual results of cross-modality matching with a **50 Hz** vibration and **50 Hz** reference tone at a loudness of 1, 4, 16, and 64 sones. At 50 Hz, a quadruplication of loudness corresponded to an increase of approximately 12 dB in sound pressure level (see Figure 5.9).

Figure 5.13 presents the acceleration that was matched to a 50 Hz tone with a loudness of 1, 4, 16, and 64 sones. No significant difference was observed between the matched acceleration of a tone with a loudness of 1 and 4 sones. At higher sound pressure levels, the matched acceleration increased by approximately 5-6 dB, with a quadruplication of loudness. For a tone with a frequency of 50 Hz, this corresponded to an approximately 12 dB increase in sound pressure level (see Figure 5.9). Interestingly, the results from the current experiment suggest an even steeper growth of perceived vibration magnitude compared to Section 5.2. In other words, only a small increase in acceleration level is necessary at low sensation levels to increase the perceived vibration magnitude strongly. These results will be reviewed again in Section 7.3.1 for the discussion of the amount of compression that might be advantageous for music-driven vibration generation.

Experiment 2 - Different Frequencies

In the second study, different tones (50, 100, and 200 Hz) of equal loudness (4 sones) were matched to a single vibration (50 Hz). As expected, the amplitude of the matched vibration was equal under all three conditions, as shown in Figure 5.14. Again, large inter-individual variances were observed. The absolute value of the matched acceleration was similar to previously reported

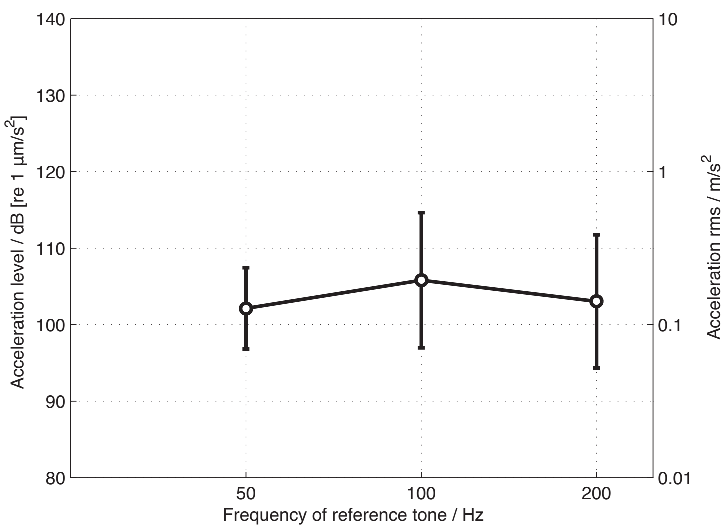


Fig. 5.14. Mean acceleration levels ± one standard deviation for the cross-modality matched vibration at **50 Hz** to reference tones at 50, 100, and 200 Hz with a loudness of 4 sones.

results. Kaufmann et al. [94] observed a matched acceleration of approximately 107 dB using narrow-band white noise at 4 sones (vibration: center frequency = 31.5 Hz; sound: center frequency = 100 Hz, and a sound pressure level of 80 dB, which corresponded to a loudness of approximately 4 sones). In comparison, the results from this study indicated that a vibration at 50 Hz with an acceleration level of 106 dB matched a 100 Hz reference tone with 4 sones.

The method of adjustment used in this study was fast and reliable. However, the initial acceleration levels were consistently lower than the matched levels; thus, the results might have underestimated the PSE of intensity.

5.4.5 Summary

In this section, PSEs were determined for the loudness of pure tones and the perceived magnitude of sinusoidal WBVs, and the following results were obtained.

- The matched acceleration level of a tone with a specific loudness (4 or 16 sones) was similar at 50 and 100 Hz; however, the matched acceleration increased significantly at 200 Hz. This finding correlates well with data representing equal-vibration-intensity contours determined in previous sections. The difference between isophones and equal-vip contours could be compensated for during the process of generating vibrations from music to cross-modally match intensities (see Section 7.3.2).
- Tones of equal loudness were matched to the same vibration magnitude even if the acoustic frequency was different from the frequency of vibration.
- A quadruplication of loudness resulted in a 5-6 dB increase in the matched acceleration level for loudness values greater than 4 sones. This finding confirms a steeper slope for vibration-intensity perception than for loudness. If vibrations are generated from sound, this perceptual difference could be compensated for using compression to form coherent multimodal perception. The resulting compression algorithms are evaluated in Section 7.3.1.
- Small intra-individual and very large inter-individual variations were observed. These results could lead to large differences in preferred vibration levels for tactile music perception, which are discussed in Section 7.4.1.

5.5 Influence of WBVs on Loudness Perception

Until now, hearing and tactile perception have been investigated as if they were isolated processes; however, perception is a multi-sensory phenomenon that involves the interaction of individual sensory modalities. Understanding this interaction is important for not only the vibro-acoustical reproduction of music but also the design of virtual reality experiences, consumer product design, and the understanding of sound and vibration perception in the automotive industry. Discussions about auditory-tactile perception first appeared in as early as 1985, focusing on the interaction between vibration and sound in automobiles [162]. Although the mechanisms of single modalities are better understood today, the interactions between modalities have not been studied as extensively. In this section, one of these interactions will be discussed, namely, the influence of WBVs on loudness perception. The discussion investigates the hypothesis that acoustic signals would be rated as being louder in the simultaneous presence of WBVs.

Different neurological studies have revealed integration and interaction effects between auditory and tactile signals, both in the auditory cortex of monkeys [95] and in the cerebral cortex of human beings [47]. In addition, tactile stimuli alone have been observed to activate the auditory cortex [15, 46]. These studies indicated an early interaction of the two modalities in the brain. However, the measurement of action potentials does not provide information regarding the influence of this interaction on a specific perceptual dimension, such as loudness.

Several psychophysical studies have investigated audio-tactile loudness perception. The direct influence of hand vibration on loudness perception was demonstrated by Schürmann et al. [151]. Their results demonstrated that audio-tactile stimuli were perceived as 12.4% louder than purely auditory stimuli, corresponding to a sound pressure level difference of 1.1 dB. Other studies did not find a significant influence of vibration on loudness perception. Specifically, Bellmann [9] observed only a slight difference while measuring a 60-phon loudness contour, with and without simultaneous WBV. Using sinusoidal and broadband signals (automotive noises), Lange [100] also did not find an influence of WBV on loudness perception. The conflicting results across the aforementioned studies, which differed from each other in terms of both the methods and stimuli, motivated the further investigation of the influence of vibration on loudness perception in the current study. Additional variables, including vibration frequency and vibration amplitude, are also considered.

5.5.1 Setup

As in the previous experiments, WBVs were generated vertically using the electro-dynamic shaker, as described in Section 3.2. The subject was asked to sit on a flat, hard wooden seat with both feet on the ground. The BRTFs were calibrated individually. The audio signals were delivered as described in Section 3.1 using closed dynamic headphones. Masking pink noise was reproduced throughout the experiment at 74 dB(A). The participant was able to control the amplitude of the test stimulus using a rotary knob that was infinitely adjustable and that did not possess any indicators, such as on-or-off markings (Griffin Technology, PowerMate). During the experiment, the control knob was used in the same manner as the volume control of a stereo system.

5.5.2 Subjects

Twenty-eight subjects voluntarily participated in the experiment (16 male and 12 female). Most of the participants were students. The participants were between 19 and 52 years old (mean, 27.3 years), and they had no hearing or spinal damage and had not been previously involved in psychophysical experiments.

5.5.3 Stimuli and Experimental Design

Sinusoidal stimuli were selected to investigate the influence of vibration frequency on loudness perception. The use of sinusoidal stimuli is ecologically valid because acoustical and vibratory sinusoidal components are often perceived simultaneously, such as when driving a car or listening to a concert. Frequencies of 10, 20, 63, and 200 Hz, which can be perceived by both the auditory and tactile senses, were chosen. To analyze the influence of vibration level, vibration magnitudes of up to 1 vip were generated at 4, 8, and 12 dB above the individual tactile perception threshold for each subject. The reference tone was chosen with a constant sensation level of 10 dB with reference to the individual detection threshold in 74 dB(A) pink noise.

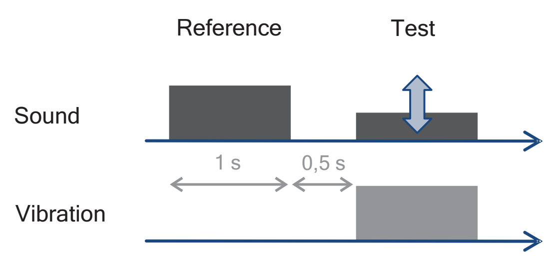


Fig. 5.15. Experimental procedure: the reference and test tones were presented in alternation until a participant was satisfied with her/his loudness matching. The amplitude of the acoustical reference stimulus was fixed at a level of 10 dB above the threshold. The amplitude of the acoustical test stimulus could be adjusted using a rotary knob. Throughout each trial, the vibration amplitude remained constant at a 4, 8, or 12 dB sensation level.

The task was to compare a reference tone without vibration to a test tone reproduced simultaneously with a vibration. During this comparison, the level of the test tone could be adjusted using the rotary knob. The participants were instructed to focus solely on the loudness of the sound. A trial was complete when a participant judged both of the tones as equally loud. A diagram of a sample trial is shown in Figure 5.15. The reference tone without vibration and the test stimulus with vibration were played back in turns. Each tone was 1 s long and was separated from adjacent tones by 0.5 s breaks. The loudness-matching trial was repeated 10 times for each participant. Additionally, purely acoustical reference loudness matching was conducted by having the participants compare a test tone without vibration to a reference tone without vibration.

5.5.4 Results and Discussion

The results are presented in Figure 5.16. As expected, a level difference of 0 dB was observed between the reference tone and adjustable test tone when no vibration was reproduced. When the test tone was accompanied by vibration, its adjusted level fell by an average of 1 dB compared to the reference tone (analysis of variance (ANOVA), $p<0.05$). In other words, the simultaneous

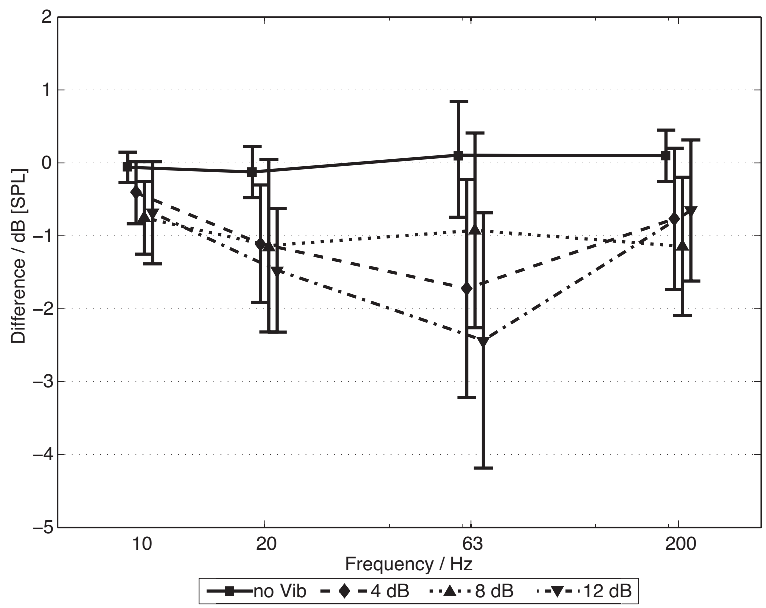


Fig. 5.16. Results of the loudness-matching experiment, indicating the mean values and standard deviations. The difference between the adjusted level of the reference tone and the level of the test tone is presented for all vibration conditions.

presence of vibration had a significant influence on loudness perception. Interestingly, as shown in a series of pairwise comparisons, there were no significant differences across vibration levels or frequencies. However, greater standard deviations were observed at 63 Hz, although the reasons for this effect were unclear.

At first glance, a 1 dB difference in sound pressure level can appear relatively small. However, the dynamic range of the auditory system decreases with decreasing frequency. The increased loudness perception cannot be explained by bone-conducted sound from the vibration chair. If this were the case, the value of the acceleration level would influence loudness matching. As this was not the case with our data, an integration effect provides a more likely explanation. In addition, the reproduced acceleration levels were quite small.

The results of this study essentially deviate from the results of Bellmann [9] and Lange [100], who did not find that WBV had an influence on loudness perception. However, the observed effect in this study has the same order of magnitude as that reported by Schürmann et al. [151]. Several factors might explain the differences between our results and those in the literature.

- The participants in the Lange study [100] were more experienced than those in this study. This difference might have influenced the study findings due to the tendency of experienced participants to judge loudness more critically than inexperienced participants.
- In this study, the acceleration level was adjusted according to the perception thresholds of the individual participants to ensure clearly perceptible vibrations.
- The participants adjusted loudness using an infinite rotary knob with no markings, and thus, they were forced to rely solely on the rendered signals.

5.5.5 Summary

The following results were obtained

- Even under the critical conditions of this study, an influence of WBVs on loudness was observed: an auditory-tactile loudness illusion.
- Interestingly, there was no systematic influence of frequency or acceleration level on this effect. The influence of higher acceleration levels on loudness perception would be an interesting direction for future research.

This auditory-tactile loudness illusion is believed to have an even stronger effect when the subjects are not instructed to focus solely on loudness. The perceived overall intensity of the low frequencies in music might increase if vibrations are reproduced using a shaker. This could result in a misbalanced ratio between bass and treble. Therefore, it might be necessary to compensate for this effect by adjusting the audio equalization in a vibrationally enhanced reproduction setup. The preferred audio balance between high and low frequencies in such a scenario is discussed further in Section 7.4.2.

6

Measurements of Music-Driven Sound and Vibration

To understand the perception of sound-induced vibrations (in a playback scenario or in a concert hall), it is important to know the perceptual abilities of our senses, as reviewed in Chapter 2 and investigated in Chapters 4 and 5. Our experience also plays an important role. This holds particularly true for the judgment of quality, and it will be discussed in detail in Chapter 7. Therefore, this chapter considers sound and vibration when we experience music in different real-life situations. A typical example is the vibration perceived during an organ performance in a church or while listening to a concert. The following questions come to mind: Can vibrations be perceived, even in a conventional classical concert hall during a concert performance? Do the acceleration levels exceed the perception threshold? Which dynamic range can be expected? Is there a linear relationship between the sound pressure level and acceleration level for WBV? How does the vibration intensity vary depending on receiver location? To answer these questions, comprehensive sound and vibration measurements were obtained in two exemplary locations: a classical concert hall and a church.

Only a few studies have been published that have involved vibration measurements in a musical context. Daub [21] measured sound and vibrations in two different venues using musical instruments as sound generators. Thus, it was difficult to separate the contribution of the sound source from the transfer characteristics of the room.

A study by Simon et al. [154] reported audio-induced vibrations in a car generated by a music sequence, which was played back by an automotive audio system. They found that high vibration levels between 50 and 75 Hz were excited in the seat and floor. However, whether these characteristics resulted from the spectral content of the source material remains unclear. Simon et al. also reported a relatively linear relationship between sound and vibration in this frequency range (a 4.5 dB increase in the bass level resulted in an approximately 4.5 dB higher acceleration level).

Abercrombie and Braasch [2] measured structural impulse responses on different stage floors, which were excited using a sledgehammer. They found that the acceleration level decreased and the propagation time increased with increasing distance from the source (maximum of 4 m measured). Both the acceleration level and propagation time were strongly dependent on the stage construction. Unfortunately, it was not possible to separate air- from structure-borne vibrations in their measurements or to predict the resulting vibrations in the auditorium.

For this reason, vibro-acoustical measurements in the auditorium of the Semperoper Dresden and Lutherkirche Radebeul were conducted.

6.1 Concert Hall

This section discusses the relationship between sound and vibration for frequencies of up to 1 kHz in the Semperoper Dresden (Figure 6.1). Therefore, the transfer function between sound pressure and acceleration was measured at different listening positions.

6.1.1 Setup

A dodecahedral loudspeaker (Outline, Globe Source with Subwoofer) was used as the sound source. It was placed on the lifted orchestra pit 4 m from the edge of the stage and 1.5 m to the side of the middle axis. Six receiver positions were selected: three in the parquet (R1-R3), one in the loge (R4), and two in

Fig. 6.1. Auditorium of the Semperoper Dresden with a lifted orchestra pit.

the balconies (R5 and R6). These positions are illustrated in Figure A.1 in Appendix A. To measure room impulse responses, a measurement microphone at ear height (B&K, 4188), a spherical microphone array, and a Kemar dummy head with blocked ear canals were used. The aforementioned recordings were used to reproduce the opera house virtually in the lab, e.g., using wave field synthesis or binaural reproduction, to conduct perceptual experiments. In this study, only the data from the measurement microphone will be discussed.

The measurements could only be obtained in the empty concert hall at night. However, compared with a situation in which two-thirds of the seats are filled, the reverberation time below 2 kHz was prolonged by only 0.5 s, which resulted in an approximately 3 dB increase in the sound pressure level averaged over receiver positions R1-R6 [98]. The stage is typically narrowed for orchestral concerts, using moveable wall elements to build a concert dome. Because this dome was not available, the measurements were obtained with the safety curtain closed. The reverberation times in both situations are again very similar [98].

White noise was used as a measurement signal and was reproduced via the loudspeaker. However, the sound source exhibited some coloration. The magnitude spectra of the resulting sound pressures at all receiver positions at ear height are shown in Figure 6.2. A homogeneous energy distribution over all receiver positions was observed. However, a strong increase toward lower frequencies was also observed, which resulted mainly from the characteristics of the sound source. This characteristic did not distort subsequent results because the transfer function between the sound pressure at ear height and the

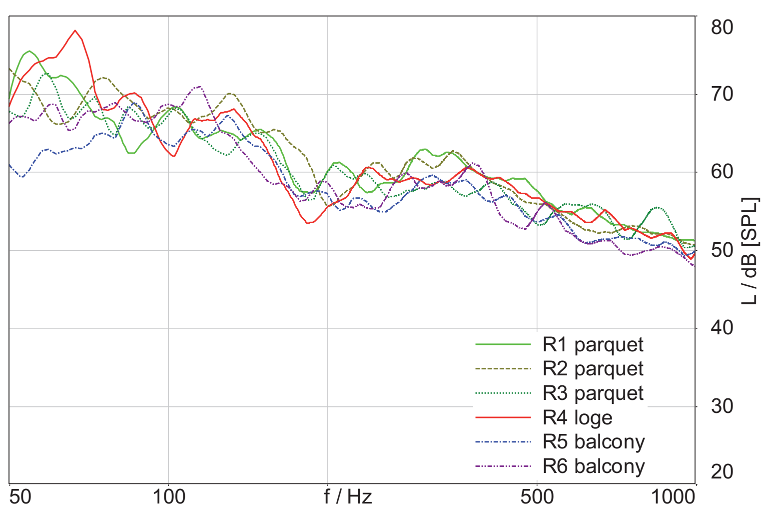


Fig. 6.2. Magnitude spectrum (FFT 65536, 1/24th octave intensity averaging) of sound pressure at ear height for all receiver positions.

acceleration at different surfaces was calculated. The influence of the overall sound pressure level on this transfer function will be discussed later.

To measure the vibrations at different surfaces, accelerometers (Kistler, 8636C10) were mounted to small metal plates 10 cm in diameter and were placed on the ground, seat, and armrest (see Figure 6.3). The measurement position was then loaded with a person (80 kg). All of the other seats were unoccupied. The influence of a larger audience on the measured vibrations cannot be easily assessed. However, the presence of a second person in an adjacent seat did not specifically change the results of the test measurements.

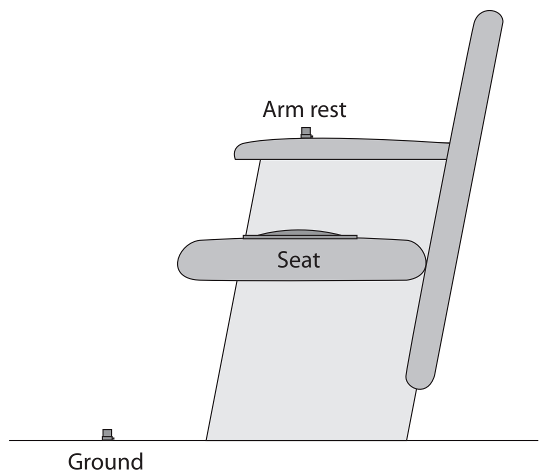


Fig. 6.3. Measurement setup with accelerometers on the ground, seat, and arm rest.

6.1.2 Results

Figure 6.4 illustrates the transfer function between the acceleration on the ground and the sound pressure at ear height for the same measurement position. This function corresponds to the difference between the acceleration level and sound pressure level $L_{\mathrm{acc}} - L_{\mathrm{SPL}}$. A horizontal line at 0 dB represents equal levels for sound at ear height and vibrations on the ground. The overall sound pressure level at each measurement position was approximately 90 dB. Higher acceleration levels were measured in many cases, particularly in the parquet (R1-R3) for frequencies less than 200 Hz. Interestingly, this location-dependent difference disappeared at the seat surface plotted in Figure 6.5. The frequency response was relatively homogeneous over a broad frequency range, with a slight decrease toward lower frequencies. Only a few positions exhibited isolated resonances (e.g., 100 Hz at receiver position R4).

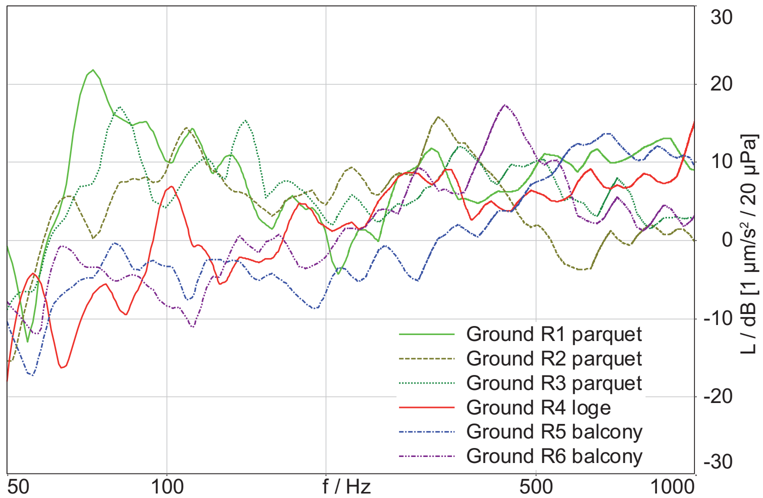


Fig. 6.4. Transfer functions (FFT 65536, 1/24th octave intensity averaging) between acceleration on the ground and the sound pressure at ear height at all receiver positions.

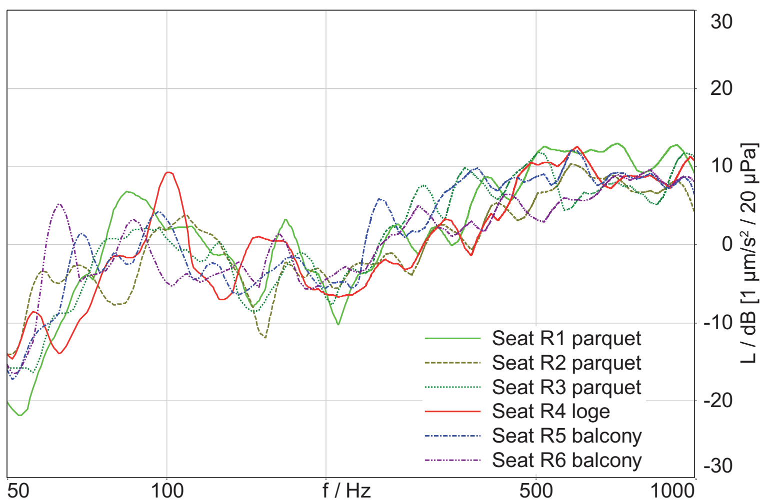


Fig. 6.5. Transfer functions (FFT 65536, 1/24th octave intensity averaging) between acceleration on the seat and the sound pressure at ear height at all receiver positions.

No distinct dependence of the overall level on the distance between the receiver and source was observed, e.g., there were no considerable differences between the accelerations at positions R1 and R3. This finding indicates that the vibrations were not transmitted via the ground from the loudspeaker to the seat. Therefore, it is hypothesized that the airborne sound excites the surface near the listener. To test this hypothesis, the loudspeaker was decoupled from the stage floor using a sheet of foam (55 x 45 x 16 cm). The vertical resonance frequency of this system was measured to be approximately 8 Hz, resulting in effective vibration isolation in the interesting frequency range above 50 Hz. The exemplary transfer functions at position R4 (seat) in Figure 6.6 exhibit no considerable differences with or without the isolated loudspeaker. This finding supports the hypothesis that the vibrations are excited via airborne vibrations in the auditorium. Airborne transmission could also explain the lower levels on the ground for positions R4-R6 due to the smaller floor areas in the balconies.

If the sound pressure level rises, the excited vibrations should increase accordingly. However, whether there is a linear relationship between both levels remains unclear. Therefore, a few measurements were obtained at different sound intensities. Figure 6.7 presents two exemplary transfer functions at position R4 (seat) with a 30 dB difference in the sound pressure level. Both curves are almost identical. This finding proves a linear relationship between the two physical variables, which is a prerequisite for meaningful transfer functions.

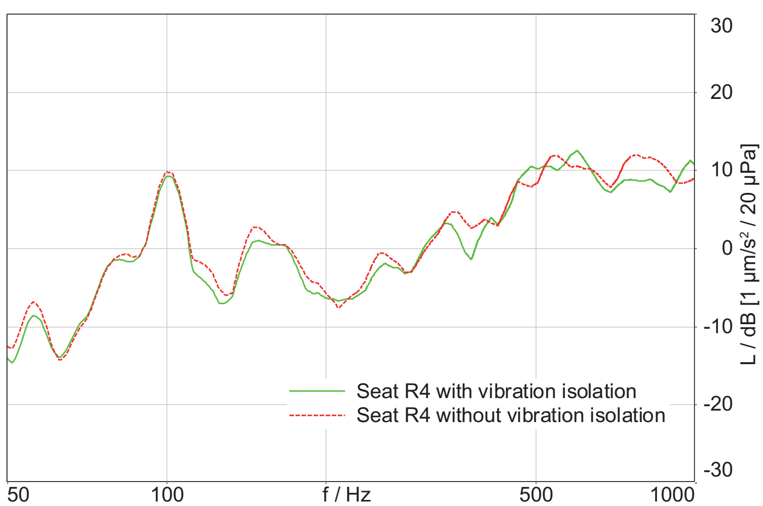


Fig. 6.6. Transfer functions (FFT 65536, 1/24th octave intensity averaging) between acceleration on the seat and the sound pressure at ear height at the receiver position R4 with and without vibration isolation of the sound source.

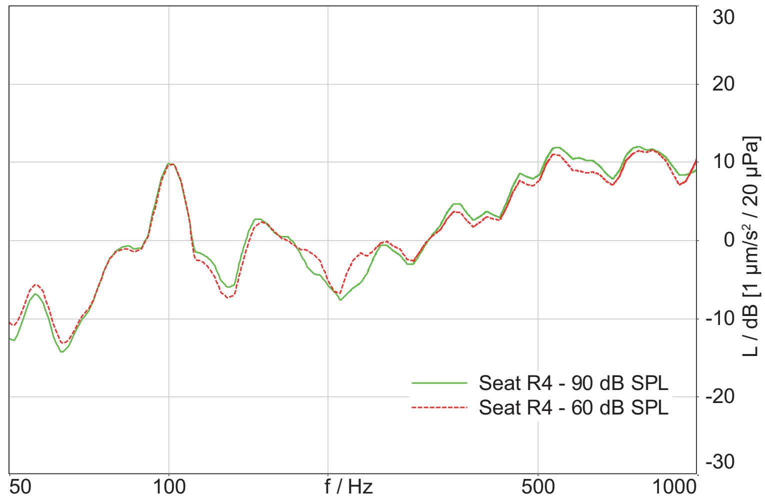


Fig. 6.7. Transfer functions (FFT 65536, 1/24th octave intensity averaging) between acceleration on the seat and the sound pressure at ear height at the receiver position R4 for different sound pressure levels.

6.1.3 Discussion

Typical sound pressure levels in a concert hall for fully orchestrated passages are between 80 and 90 dB (forte) depending on the instrumentation and room [119]. Fortissimo can reach average sound pressure levels approximately 10 dB higher [191]. For example, there have been measurements in the Semperoper with L_{AI} from 96 to 98 dB for themes from Wagner's Lohengrin [98]. The peak level at low frequencies can reach even higher values. Considering the perception threshold for sinusoidal seat vibrations, which is approximately $L_{\mathrm{acc}} \approx 90$ dB for frequencies less than 150 Hz (see Section 2.2.1), the above measurements indicate perceivable vibrations for the forte and fortissimo parts of orchestral music. However, these vibrations might not be perceived separately because of the integration of the tactile sense with the other sensory modalities into one multimodal concert event. During subsequent concert visits to the Semperoper, the author paid special attention to music-induced vibrations and could clearly identify them during a classical concert and a jazz concert. Interestingly, other concert visitors, who had been unaware of music-induced vibrations in the concert hall before, confirmed these findings.

The measurements suggest that differences between positions and the influence of local resonances should be clearly perceivable because they are considerably larger than the perceivable difference in the acceleration level, which is below 1.5 dB (see Section 2.3.1). In addition, the growth of perceived vibration magnitude with increasing acceleration level is rather steep, particularly at low sensation levels, as discussed before, e.g., in Section 5.2.

This growth results in strongly perceived vibration intensity differences for even small changes in the acceleration level. These differences are expected to increase further between different venues. Therefore, a second measurement series was obtained in a church.

6.2 Church

This section discusses the relationship between sound and vibration in the Lutherkirche in Radebeul, a typical church build in 1892 with massive bearing walls and a stone floor. An outline of the church is provided in Figure A.2 in Appendix A. An organ loft is located in the back of the church. Wooden pews are located in the nave and on the wooden galleries. Again, transfer functions between sound pressure and acceleration were measured at different exemplary listening positions.

6.2.1 Setup

The same method and equipment were used as in the concert hall measurements described above. The dodecahedral source S1 was placed in the organ loft to simulate organ stimulation. Various measurement positions were selected, but only two exemplary receiver positions (R1 in the nave and R2 in the gallery) are discussed here.

Again, different microphone setups were used to record various room impulse responses. Additional vibration impulse responses were measured on the ground, footrest, seat, and back rest of the wooden pews using accelerometers (Kistler, 8636C10). This process is illustrated in Figure 6.8. The measurement position was then loaded with the same person (80 kg) as before.

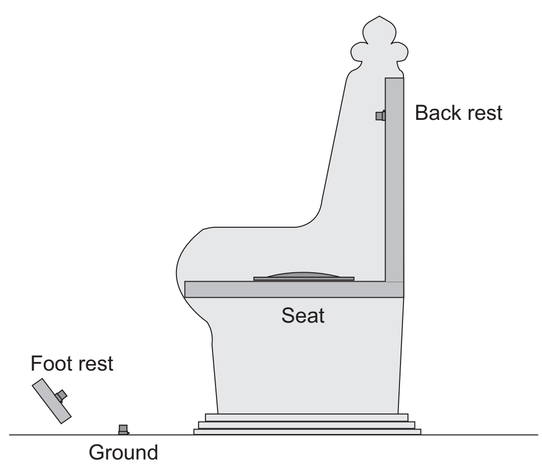


Fig. 6.8. Measurement setup with accelerometers on the ground, foot rest, seat, and back rest.

6.2.2 Results

Figure 6.9 plots the transfer function between acceleration on the ground and the sound pressure at ear height for the same position. The acceleration on the ground, which was excited by the sound, differed significantly between positions due to the massive stone floor in the nave and the wooden construction of the gallery. Again, a broad vibration spectrum was excited with a slight roll-off toward lower frequencies.

This frequency spectrum differed completely for measurements at the footrest, which consisted of a long wooden board mounted only at its ends. Figure 6.10 illustrates the strong resonances for footrest vibrations at both receiver positions. This resonance pattern was also dependent on the position of the accelerometer along the board, which is not plotted here. The acceleration level varied considerably with frequency; however, the overall level was similar under both conditions.

Comparable levels at both positions were also measured at the seat (see Figure 6.11) and backrest. Compared with the concert hall, the overall acceleration level at the seat was significantly higher. This finding might have been due to the missing seat upholstery and the large continuous surfaces, which could be excited more intensely by the sound.

Similar to the concert hall, no change in the transfer function was observed when the subwoofer was decoupled from the ground, supporting the hypothesis of the dominance of airborne vibrations in the auditorium. There was also no

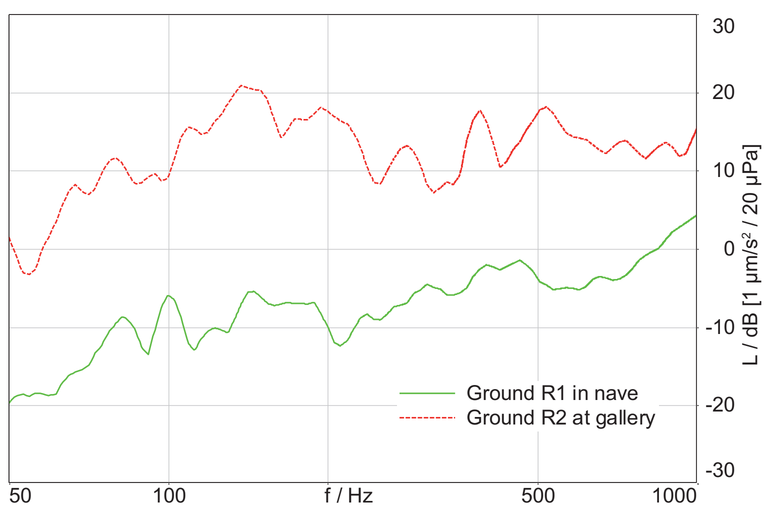


Fig. 6.9. Transfer functions (FFT 65536, 1/24th octave intensity averaging) between acceleration on the ground and the sound pressure at ear height in the nave and in the gallery.

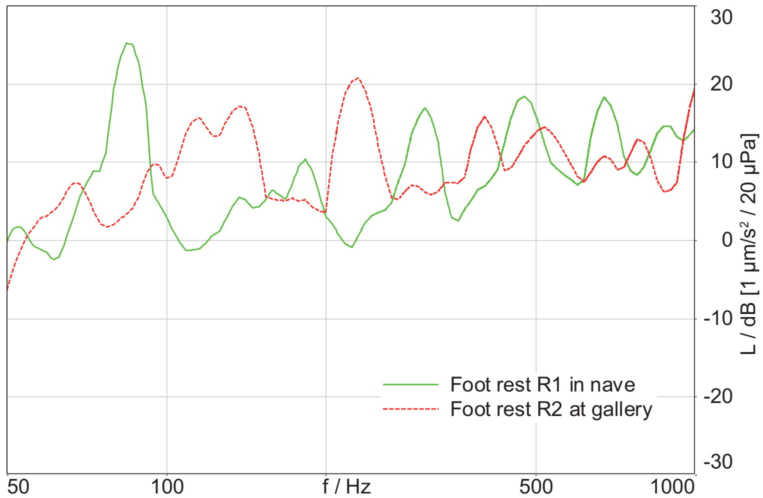


Fig. 6.10. Transfer functions (FFT 65536, 1/24th octave intensity averaging) between acceleration at the feet and sound pressure at ear height in the nave and in the gallery.

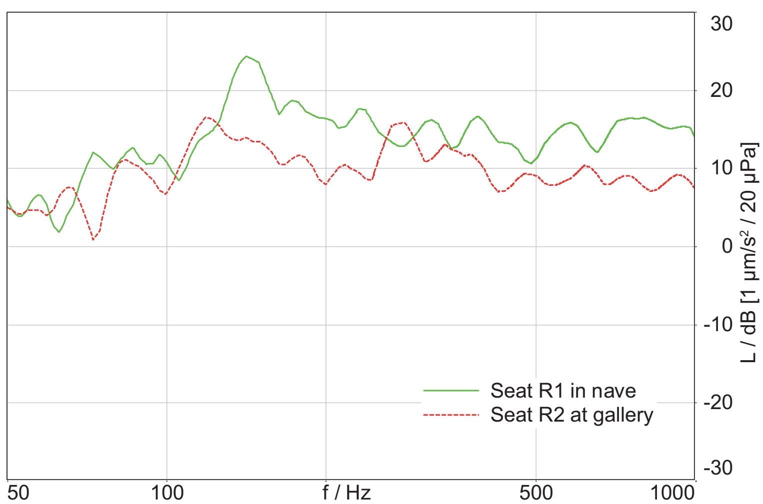


Fig. 6.11. Transfer functions (FFT 65536, 1/24th octave intensity averaging) between acceleration on the seat and the sound pressure at ear height in the nave and in the gallery.

effect on the overall sound pressure level, confirming the linear relationship between sound pressure and acceleration discussed in Section 6.1.2.

6.2.3 Discussion

Organ-generated sound pressure levels in a church can be quite high. A sample sequence with a significant low-frequency content (Max Reger, Introduktion d-Moll), performed by organ player Gottfried Trepte in the Lutherkirche Radebeul, reached sound pressure levels of 90 dB(A) at both receiver positions. The resulting seat vibrations at R2 (gallery) are plotted in Appendix B. Clear-to-strong vibrations (100 dB) were excited most of the time over a broad frequency range. However, lower frequencies dominated the vibration experience of the author, which might have been due to the strong tactile masking effects toward higher frequencies discussed in Section 2.2.4. This finding led to the idea of reproducing only the fundamental frequency of a complex vibration in a reproduction scenario, which will be studied further in Section 7.2.2.

6.3 Summary

The described measurements demonstrated that sound can excite perceivable surface vibrations. Furthermore, the results indicated that our experience with audio-induced vibration can vary heavily. Even within one venue, the vibration intensities and frequency spectra are strongly dependent on listener position. However, the measured sound-induced vibrations are only exemplary in nature. The measurements provided no insights into the perceived quality of such music-induced vibrations. No ideal sound-to-vibration transfer function could be deduced. If an ideal sound-to-vibration transfer function exists, it might be possible to improve the concert experience by modifying vibrations through architectural changes or artificial generation. The latter case is particularly interesting for audio-reproduction systems, but it could improve the music experience even for classical concert halls. Comprehensive listening tests are necessary to identify which vibrations are favorable. Therefore, a variety of different vibration generation approaches will be selected and evaluated in Chapter 7.

7

Quality of Auditory-Tactile Music Reproduction

7.1 Introduction

The quality of a concert experience depends on many factors. Images, sound, and vibrations are perceived through our senses. The resulting multimodal perception is reflected in our brain, and qualities and features are recognized and designated. These features are then compared with our expectations. If the expected features are met or exceeded, good quality is assigned to the concert experience. This process of quality evaluation is illustrated schematically in Figure 7.1 using the following definitions provided by Jekosch [88]:

- **Quality:** The result of judgment of the perceived composition of an entity with regard to its desired composition
- **Perceived composition:** The totality of features of an entity; a signal for the identity of the entity visible to the perceiver
- **Entity:** A material or immaterial object under observation [25]
- **Desired composition:** The totality of features of individual expectations and/or relevant demands and/or social requirements
- **Feature:** A recognizable and nameable characteristic of an entity

The immaterial object under observation in this chapter is the concert experience. The relevant physical components are assumed to be images, sound, and vibrations. As described in Section 6, measurements in a concert hall and church confirmed the existence of WBVs during music performances. If a kettledrum is struck or the organ plays a tone, the ground and chair vibrate. The vibratory intensity and frequency spectra are dependent on various factors, such as room modes and the construction parameters of the floor. Nevertheless, in many cases, the concert listener might not recognize the vibrations as a separate feature because tactile perception is integrated with the other senses (e.g., vision and hearing) into one multimodal perception. Even if the listener is unaware of vibrations, they can have an influence on recognizable features of the concert experience, e.g., the listener's presence or envelopment,

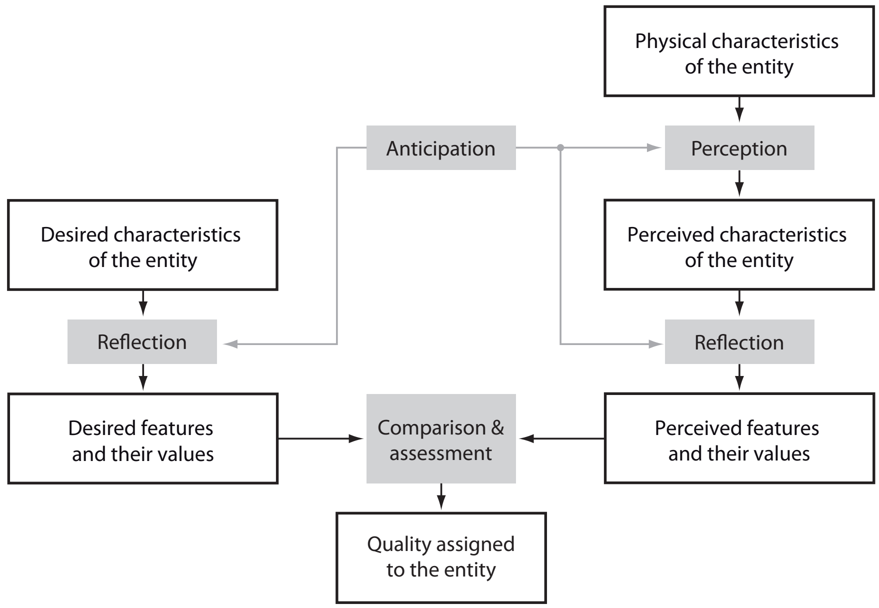


Fig. 7.1. Schematic representation of the quality event according to Jekosch [88].

which are parameters that are of vital importance in determining the quality of concert halls [16]. In an experiment, test participants could be guided to compare the totality of those features with their expectations. This comparative process would then result in a quality assessment, either by totaling the weighted individual assessments or by separately evaluating the overall quality of the concert experience. However, guiding the participant in this manner, it is difficult to ensure that all of the relevant features are considered. However, which features are actually relevant? Is it even possible to describe the typical features generally, irrespective of the individual observer?

To avoid the aforementioned problems of a controlled assessment method, the test participants in the following experiments were asked to evaluate the *overall quality of the concert experience* only. This method of *random quality assessment* can be regarded as an 'intrinsically initiated process of comparing the totality of the perceived features [...] in terms of its suitability to fulfill all the individual's [...] expectations of features' [88].

The main hypothesis to be evaluated in this study was that WBVs would be important for the perception of music. If the vibratory component were missing, there might be a loss of perceived quality in the concert experience. From another perspective, the perceived quality of a concert hall or conventional audio reproduction system might be improved by adding vibrations. These vibrations can be excited directly via the air or via the surfaces that are in contact with the listener. This study focuses on WBVs for a seated person, such as the vibrations that can be perceived in a classical concert hall or church.

Unfortunately, there is no vibration channel in conventional music recordings. Therefore, it would be advantageous if a vibration signal could be generated using the information stored in existing audio channels. This approach might be reasonable because the correlation between sound and vibration is naturally strong in everyday situations, as measured in Section 6.

Two pilot experiments were conducted and described by Merchel et al. [111, 112], who investigated the influence of WBVs on the overall quality of the reproduction of concert DVDs. Low-pass-filtered audio signals were used for vibration generation. Subjects were observed to prefer vibrations in many cases. In particular, if the vibration seat was turned off, the subjects reported that something was missing. However, different complaints were noted during the execution of the pilot experiments. It was stated that the high-frequency vibrations were sometimes prickling and therefore unpleasant. Several subjects reported that some vibrations were too strong and that others were too weak or completely missing. It was also noted that the sound generated by the vibration chair at higher frequencies was disturbing.

In the aforementioned experiments, a precisely calibrated vibration actuator was applied that was capable of reproducing frequencies from 10 to 200 Hz and higher. In practical applications, smaller and less expensive vibration actuators would be beneficial. However, these shakers are typically limited to a small frequency range around a resonance frequency.

This work aimed to broaden our understanding of the coupled perception of music and vibration by addressing the following questions:

- Can vibration-generation algorithms be obtained that result in an improved *overall quality of the concert experience* compared with reproduction without vibration?
- Which algorithms are beneficial in terms of *silent and simple* vibration reproduction?

In this chapter, algorithms are developed and evaluated to improve music-driven vibration generation considering the above questions and complaints. Four music sequences are selected in Section 7.1.1 to test the different approaches. The experimental setup, subjects, and experimental procedure are described.

Different vibration-generation approaches that primarily modify the frequency content are discussed in Section 7.2. Approaches related to intensity are evaluated subsequently in Section 7.3. Section 7.4 is concerned with the preferred levels of sound and vibration. Finally, temporal processing in the vibration generation process is studied in Section 7.5.

7.1.1 Stimuli

To represent typical concert situations for both classical and modern music, four sequences were selected from music DVDs [12, 97, 195, 196] that included low-frequency content. A stimulus length of approximately 1.5 min was chosen to ensure that the participants had sufficient time to become familiar with a stimulus before providing their quality judgment. The following sequences were selected:

- Bach, Toccata in D minor (*church organ*)
- Verdi, Messa Da Requiem, Dies Irae (*kettledrum, contrabass*)
- Dvořák, Slavonic Dance No. 2 in E minor, op. 72 (*contrabass*)
- Blue Man Group, The Complex, Sing Along (*bass, percussion, kick drum*)

The first piece, Toccata in D minor, is a well-known organ work that is referred to as BACH. A spectrogram of the first 60 s is plotted in Figure 7.2. A rising and falling succession of notes covering a broad frequency range can be seen in the figure. Additionally, steady-state tones with a rich overtone spectrum dominate the composition. Strong vibrations would be expected in the church measured in Section 6.2 for this piece of music.

The second sequence, Dies Irae, is abbreviated as VERDI. It is a terrifying composition for double choir and orchestra. A spectrogram is plotted in Figure 7.3. Impulsive fortissimo sections with a concert bass drum, kettledrum, and tutti orchestra alternate quickly with sections that are dominated by the choir, bowed instruments, and brass winds. The sequence is characterized by strong transients.

The third stimulus, Slavonic Dance No. 2 in E minor, is referred to as DVORAK. Again, a spectrogram is plotted in Figure 7.4. It is a calm orchestral piece, dominated by bowed and plucked strings. Contrabasses and cellos continuously generate low frequencies at a low level.

An organ work and two orchestral compositions, with transient and steady-state content at low frequencies, have been described thus far. The fourth and final sequence, Sing Along, is a typical pop music example. It is performed by the Blue Man Group, which is further shortened to BMG. The sequence is characterized by the heavy use of drums and percussion. These instruments generate transient content at low frequencies, which can be seen in the corresponding spectrogram in Figure 7.5. Additionally, a bass line can be easily identified.

To generate a vibration signal from these sequences, the sum was calculated of the LFE channel and the three respective frontal channels. No low-frequency content was contained in the surround sound channels in any situation. Pure Data (Pd), a real-time programming environment for audio processing, was used. The detailed processing is described in the respective subsections.

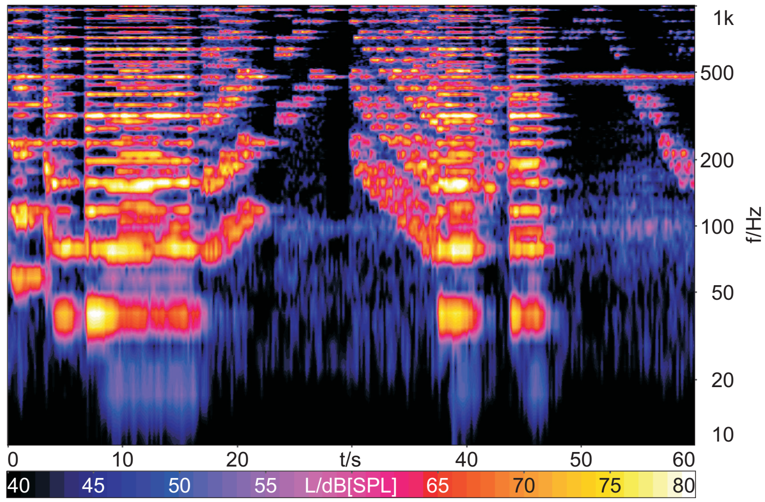


Fig. 7.2. Spectrogram of the mono sums for 60 s from the BACH sequence. The short-time Fourier transforms (STFTs) were calculated with 8,192 samples using 50% overlapping Hann windows.

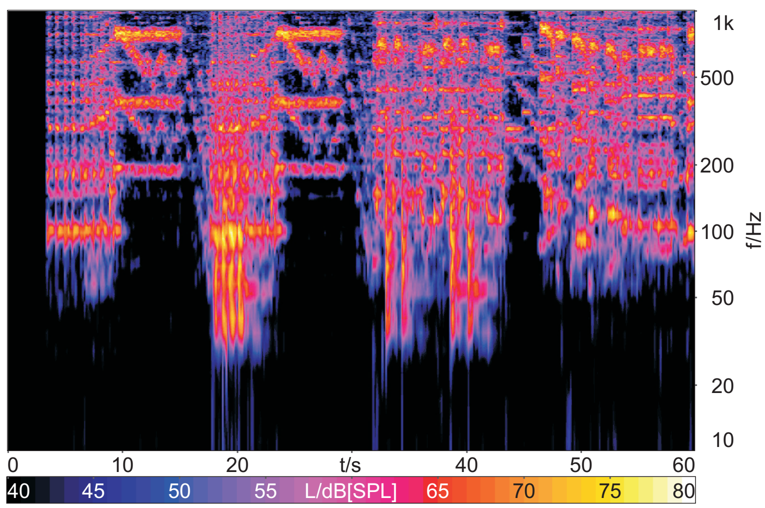


Fig. 7.3. Spectrogram (STFTs: 8,192 samples, 50% overlapping Hann windows) of the mono sums for 60 s from the VERDI sequence.

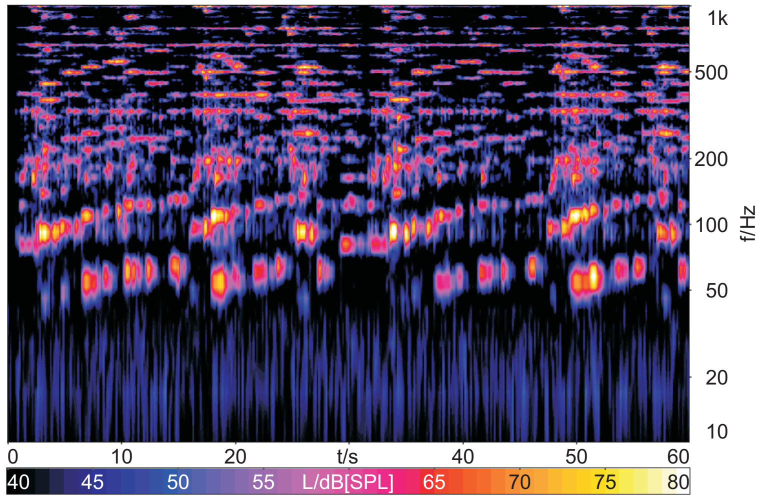


Fig. 7.4. Spectrogram (STFTs: 8,192 samples, 50% overlapping Hann windows) of the mono sums for 60 s from the DVORAK sequence.

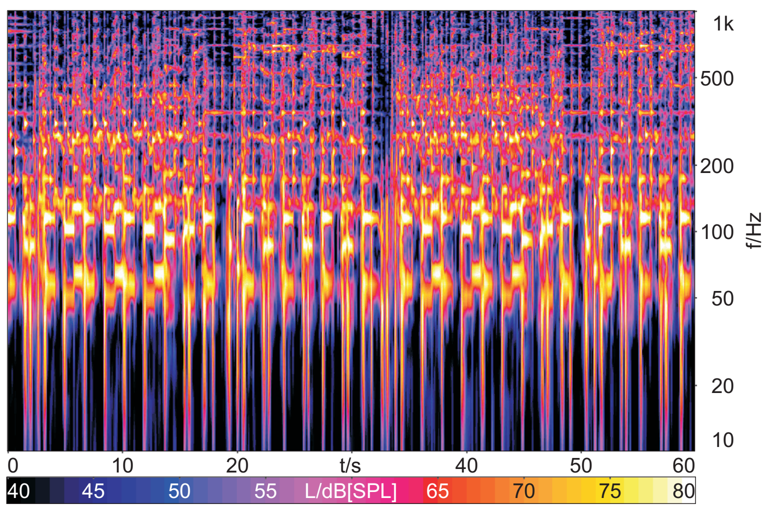


Fig. 7.5. Spectrogram (STFTs: 8,192 samples, 50% overlapping Hann windows) of the mono sums for 60 s from the BMG sequence.

7.1.2 Synchronization

For a good multisensory concert experience, it is recommended that input from all sensory systems should be integrated into one unified perception. Therefore, the delay between different sensory inputs is an important factor. Many published studies have focused on the perception of synchrony between modalities, mostly related to audio-visual delay (e.g., [165, 167]). Few studies have focused on the temporal aspects of acoustical and vibratory stimuli. These studies have differed in the types of reproduced vibration (vibrations at the hand, forearm, or WBV), types of stimuli (sinusoidal bursts, pulses, noise, instrumental tones, or instrumental sequences), and experimental procedures (time-order judgments or the detection of asynchrony). However, some general conclusions can be drawn.

It was reported that audio delays were more difficult to detect than audio advances. Hirsh and Sherrick [77] found that a sound must be delayed 25 ms against hand-transmitted sinusoidal bursts to detect that the vibration preceded the sound. However, vibrations had to be delayed only 12 ms to detect asynchrony. A similar asymmetry was observed by Altinsoy [4] using broadband noise bursts reproduced via headphones and broadband vibration bursts at the fingertip. Altinsoy found that stimuli with audio delays of approximately 50 ms to -25 ms were judged to be synchronous. He also reported that the point of subjective simultaneity (PSS) shifted toward an audio delay of approximately 7 ms. Detection thresholds for auditory-tactile asynchrony appear to also depend on the type of stimulus. In an experiment reproducing broadband noise and sinusoidal WBVs, audio delays from 63 ms to -47 ms were found to be synchronous [5]. Using the same setup, it was reported that audio delays from 79 ms to -58 ms were judged to be synchronous regarding sound and WBVs from a car passing a bump [5].

For musical tones, the point of subjective simultaneity appears to vary considerably for instruments with different attack or decay times. For example, PSS values as high as -135 ms for pipe organ or -29 ms for bowed cello have been reported [22, 185]. In contrast, PSS values as low as -2 ms for kick drum or -7 ms for piano tones were obtained [185]. Similarly, low PSS values were obtained using impact events reproduced via a vibration platform [106].

Thus, auditory-tactile asynchrony detection appears to depend on the reproduced signal. Impulsive content is clearly more prone to delay between modalities. Because music often contains transients, the delay between sound and vibration in this study was set to 0 ms. However, for a real-time implementation of audio-generated vibration reproduction, a slight delay appears to be tolerable or even advantageous in some cases. Additionally, the existence of perceptual adaptation mechanisms, which can widen the temporal window for auditory-tactile integration after prolonged exposure to asynchronous stimuli, has been demonstrated [136].

7.1.3 Setup

A surround setup was used, according to ITU-R BS.775-1 [86], with five Genelec 8040A loudspeakers and a Genelec 7060B subwoofer. The system was equalized to a flat frequency response at the listener position. To place the subject in a standard multimedia reproduction context, an accompanying picture from the DVD was projected onto a silver screen. The video sequence showed the stage, conductor, or individual instrumentalists while playing.

In addition, vertical WBVs were reproduced using an electrodynamic vibration seat. The construction and specifications of this seat are provided, together with the description of the audio setup, in Chapter 3. The BRTFs were again calibrated for each subject individually before each experimental run using inverse filters in Matlab.

There was also crosstalk between the different systems. For example, the subwoofer excited some seat vibrations. These vibrations were measured to be less than $0.5\,\text{mm/s}^2$. The peak vibrations, reproduced with the shaker, reached 250 to $600\,\text{mm/s}^2$, which is a factor of approximately 1,000 greater than the sound-induced vibrations. Thus, this type of crosstalk is not critical.

The vibration seat itself generated some sound, particularly at higher frequencies, depending on the vibration-generation algorithm and acceleration level. This sound was not critical in most cases because the loudspeaker-generated sound was much louder. However, during quiet passages, sound from the shaker could be heard sometimes. Therefore, attempts were undertaken to reduce the sound generated by the vibration reproduction system in the subsequent experiments. The loudness and vibration intensity of the four sequences were equalized in a pilot experiment using three subjects. The resulting equivalent continuous sound pressure level was measured; for instance, a value of approximately 75 dB was obtained for the BACH sequence.

7.1.4 Subjects

Twenty subjects voluntarily participated in this experiment (14 male and six female). Most of them were students between 20 and 55 years old (mean, 24 years) and between 58 and 115 kg (mean, 75 kg). All of the participants stated that they had no known hearing or spine damage. The average number of self-reported concert visits per year was nine and ranged from one to approximately 100. Two subjects were members of bands. The preferred music styles varied, ranging from rock and pop to classical and jazz. Fifteen subjects had not been involved in music-related experiments before, whereas five had already participated in two similar pilot experiments [111, 112].

7.1.5 Experimental Design

The concert recordings were played back to each participant using the setup described above. Additional vibrations were reproduced using the vibration chair. The vibration intensities were initially adjusted so that the peak acceleration levels reached approximately 100 dB and were thus clearly perceptible. However, the vibration level could be varied easily if such a reproduction system were to be implemented at home. Additionally, the perception thresholds (see Section 2.2) and level preferences (see Section 5.4 and 7.4.1) varied between subjects. Therefore, each subject was asked to adjust the vibration amplitude individually to the preferred level. This adjustment was typically performed within the first 5-10 s of a sequence. Subsequently, the subject had to judge the *overall quality of the concert experience* using a quasi-continuous scale. Verbal anchor points ranging from bad to excellent were added, similar to the method described in ITU-T P.800 [87]. Figure 7.6 presents the rating scale that was used.

Fig. 7.6. Rating scale for evaluation of the overall quality of the concert experience.

To prevent dissatisfaction, the subject could interrupt the current stimulus as soon as she/he was confident with her/his judgment. The required time varied between subjects, from 30 s to typically no more than 60 s. After rating the overall quality, the subjects were encouraged to formulate reasons for their judgments briefly.

Each subject was asked to listen to 84 completely randomized stimuli, 21 for each music sequence. The stimuli were divided into blocks of eight stimuli. After each block, the subject had the opportunity to relax before continuing with the experiment. Typically, it took approximately 35 min to complete three to four blocks. After 45 min at most, the experimental session was interrupted and was continued on the next day (and the next, if necessary). Thus, two to three sessions were required for each subject to complete the experiment.

Before starting the experiment, the subjects had to undergo training with three stimuli to become familiar with the task and stimulus variations. The stimuli used were the first 1.5 min from BMG using three very different vibration-generation approaches. This training was repeated before each subsession.

Matlab was used to control the entire experimental procedure (multimodal playback, randomization of stimuli, measurement and calibration of individual BRTFs, guided user interface, and data collection).

7.2 Frequency Approaches

Different approaches that modified the frequency content of the audio signal to generate suitable vibrations were evaluated and will be discussed below.

7.2.1 Low Pass

The simplest approach would be to route the sound (sum of the three frontal channels and LFE channel) directly to the vibration actuator. With some deviations, this process would correspond to the approximately linear transfer functions between sound pressure and acceleration measured in Chapter 6. However, participants typically chose higher acceleration levels in the laboratory, which resulted in significant sound generation from the actuator. To reduce high-frequency noise from the actuator, the signal was low-pass-filtered using two cutoff frequencies (100 and 200 Hz), as illustrated in Figure 7.7. However, the sound generated by the vibration system could not be completely suppressed. The low-pass filter was a steep 10th-order Butterworth filter. The resulting multimodal sequences were reproduced and evaluated in the manner described above.

For the statistical analysis, the individual quality ratings were interpreted as numbers on a linear scale from zero to 100, with zero corresponding to ‘bad’ and 100 to ‘excellent’. The data were checked for a sufficiently normal distribution with the Kolmogorov-Smirnov test (KS test). A two-factor repeated-measures ANOVA was performed using the SPSS statistical software, which also checks for the homogeneity of variances. The two factors were the played music *sequence* and the applied *treatment*. Averaged results (20 subjects) for the overall quality evaluation are plotted in Figure 7.8 as the mean and 95% confidence intervals. The quality ratings for the concert reproduction *without vibration* are shown on the left.

Reproduction with vibration was judged to be better than reproduction without vibration. Post-hoc pairwise comparisons confirmed that both low-pass treatments were judged to be better than the reference condition at a highly significant level ($p<1\%$), both with an average difference of 27 scale units, using Bonferroni correction for multiple testing. This finding corresponds to approximately one unit on the five-point scale shown in Figure 7.6. The effect appeared to be strongest for the BMG pop music sequence; however, no significant effects for differences between sequences or interactions between sequences and treatments was observed.

Using the 100 Hz cutoff frequency, the participants indicated that some vibrations were missing, which was expected. Using the 200 Hz cutoff frequency, the overall vibration magnitude appeared to match the expectations better; however, the subjects occasionally reported tingling sensations on the buttocks or thighs, which only a portion of the participants liked. This finding could explain the slightly larger confidence intervals for this treatment.

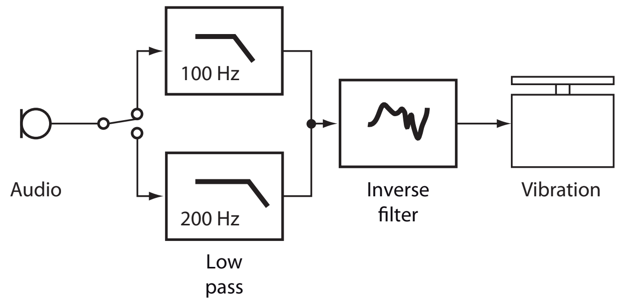


Fig. 7.7. Signal processing to generate vibration signals from the audio sum. The signal was filtered with a variable low-pass filter, and the BRTF of the vibration chair was compensated for individually.

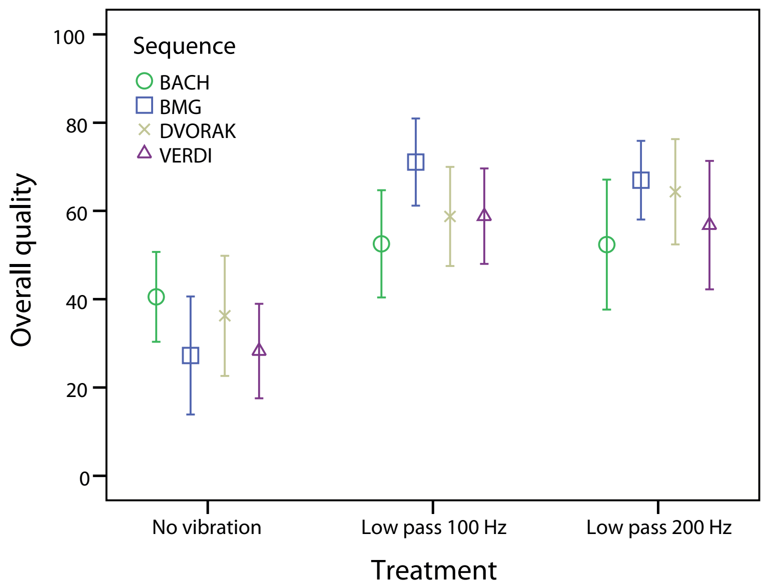


Fig. 7.8. Mean overall quality evaluation for reproduction using different vibration-generation approaches, plotted with 95 % confidence intervals. The two low-pass conditions were both judged to be significantly better than reproduction without vibrations.

The positive effect of reproducing vibrations generated by simple low-pass filtering and the negligible difference between the low-pass frequencies of 100 and 200 Hz are in agreement with earlier results [112].

7.2.2 Fundamental Frequency

In the previous section, low-pass-filtered vibrations were found to be effective for multimodal concert reproduction. However, for the low-pass 200 Hz condition, some sound was generated by the vibration system. This fact is particularly critical if the audio signal is reproduced for one person via headphones. A second person in the room would be quite disturbed by only hearing the sound generated by the shaker. Therefore, an attempt was undertaken to further reduce the sound generated by the vibration system. This goal could be accomplished, e.g., by insulating the vibrating surfaces as much as possible. Because good insulation is difficult to achieve in this case, it would be advantageous to reduce the sound generated by the shaker by modifying the vibration signal. One effective approach would be to reduce the vibration signal to the lowest frequency contained in the signal. This approach will be evaluated in this section. Another approach, which will be discussed in the subsequent section, would be to shift the frequencies down.

A typical tone generated by an instrument consists of a strong fundamental frequency and several harmonics, with multiples of the fundamental frequency. If different frequencies are presented simultaneously, strong masking effects toward higher frequencies can been observed in the tactile domain (see Section 2.2.4). It can be assumed that the fundamental component considerably masks higher frequencies. Therefore, it might be possible to remove the harmonics completely in the vibration generation process without noticeable effects. This approach is illustrated in Figure 7.9. The fundamentals of the summed audio signal were tracked using the Fiddle algorithm in Pd, which detects peaks in the frequency spectrum. A first-order low-pass filter was adaptively adjusted to the lowest frequency. If no fundamentals were detected, the low-pass filter was set to 100 Hz to preserve broadband impulsive events.

The results from the evaluation of the resulting concert reproduction are plotted in Figure 7.10. The statistical analysis was executed in the same manner as in the previous section. Again, the overall quality of the concert experience improved when vibrations were added (very significant, $p<1\%$). At the same time, the generation of high-frequency components could be reduced, except for conditions in which the fundamental frequency was almost 200 Hz, e.g., in the VERDI sequence (see Figure 7.3). For VERDI and DVORAK, some subjects again reported tingling sensations. For BMG and DVORAK, the subjects reported that it was difficult to adjust the vibration magnitude because the vibration intensity varied unexpectedly.

The average difference in perceived quality with and without vibrations was 26 scale units. Interestingly, the differences between sequences increased. The strongest effect was observed for the BMG sequence compared with the other sequences (significant interaction between treatment and sequence, $p<5\%$). The spectrogram in Figure 7.5 reveals that for the BMG sequence, the fundamentals always lay below 100 Hz and the first harmonic almost always lays above 100 Hz. Therefore, the fundamental filtering, as implemented

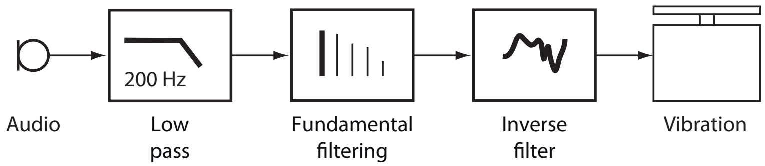


Fig. 7.9. Signal processing to generate vibration signals from the audio sum. The fundamentals below 200 Hz were tracked, and an adaptive low-pass filter was adjusted to this frequency to suppress all harmonics. If no fundamentals were detected, the low-pass filter was set to 100 Hz.

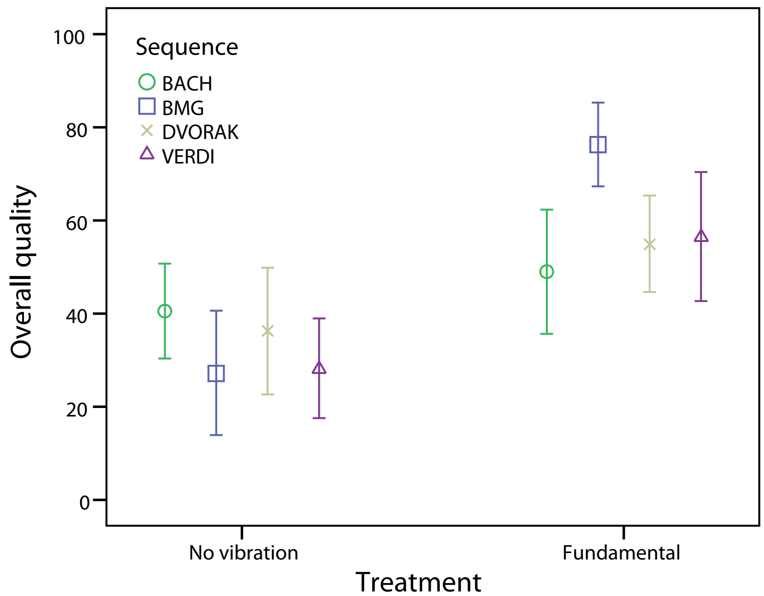


Fig. 7.10. Mean overall quality evaluation for reproduction using different vibration generation approaches plotted with 95 % confidence intervals.

here, almost corresponded to the low-pass-filtering condition, with a cutoff at 100 Hz. As expected, the resulting overall quality was judged to be similar in both cases (no significant difference; compare with Figure 7.8).

In addition, Figure 7.5 reveals that the first harmonic of the electric bass is slightly stronger than the fundamental. However, the intensity balance between fundamentals and harmonics is constant over time, resulting in a good match between sound and vibration. This relationship is not the case for the BACH sequence, plotted in Figure 7.2. The intensity of the lowest-frequency component is high within the first 10 s and then suddenly weakens, whereas the intensities of higher frequencies increase simultaneously. If only the lowest frequency is reproduced as a vibration, this change in balance between frequencies might result in a mismatch between auditory and tactile perception, which would explain the poor quality ratings for the BACH sequence using the fundamental frequency approach.

With increasing loudness, the tone color of many instruments is characterized by strong harmonics in the frequency spectrum [119]. However, the fundamental does not necessarily need to be the most intense component or can be completely missing. However, the auditory system still integrates all harmonics into one tone, in which all partials contribute to the overall intensity. In addition, different simultaneous tones can be played with different intensities depending on the composition. Therefore, a more complex approach could be beneficial. The lowest pitch could be estimated and used to generate the vibration. However, the intensity of the vibration should still depend on the overall loudness within a specific frequency range. In this manner, a good match between both modalities might be achieved. However, the processing is complex and could require greater computing capacity. Better matching of the intensities appears to be a crucial factor and will be further evaluated in Section 7.3.

7.2.3 Octave Shift

Another approach would be to shift down the frequency of the vibration signal. In this manner, the high-frequency sound generation could be further reduced and the tingling sensations could be eliminated.

The frequency resolution of the tactile sense is considerably worse than that of audition (see Section 2.3.2 and Chapter 4). Therefore, it might be acceptable to strongly compress the signal in the frequency range. However, the resulting vibrations should still perceptually integrate with the sound. Earlier experiments have been conducted to test whether subjects can match the frequencies of sinusoidal tones and vibrations presented through a vibration seat [6]. The results are summarized in Figure 7.11. The subjects were able to match the frequencies of both modalities with some tolerance. In most cases, the subjects also judged the lower octave of the auditory frequency to be suitable as a vibration frequency. Therefore, the decision was made to shift all the frequencies in this study down one octave relative to their original frequencies. This shift corresponds to compression in the frequency range, as illustrated in Figure 7.12, with stronger compression toward higher frequencies. A 50 Hz component would become 25 Hz, and a 100 Hz partial would become a 50 Hz partial. Again, Pd was used for pitch-shifting using a granular synthesis approach. The signal was cut into grains of 1,000 samples, which were slowed by

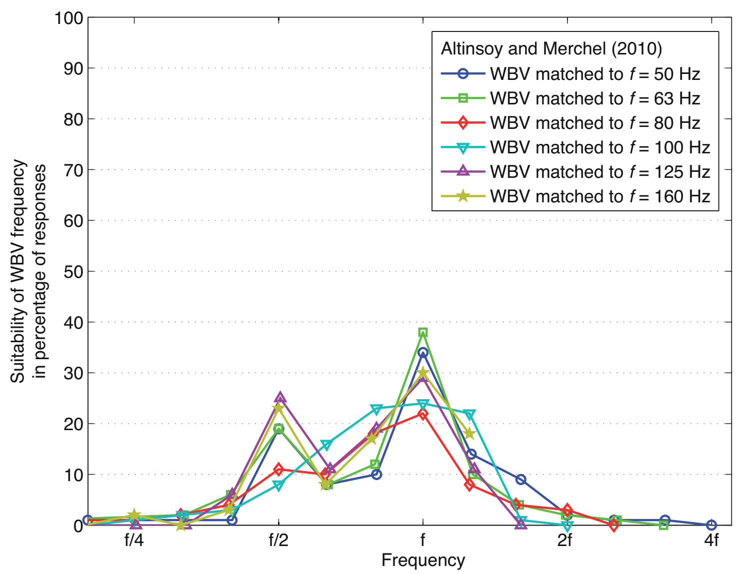


Fig. 7.11. Distribution of cross-modality frequency-matched WBVs to acoustical tones with various frequencies f according to Altinsoy and Merchel [6].

half and summed again using overlapping Hann windows. Using this method, some high-frequency artifacts occurred, which were subsequently filtered out using an additional low-pass filter at 100 Hz. The resulting low-pass-shifted vibration signals were evaluated as described above. The results are plotted in Figure 7.13. Again, the statistical analysis was performed using ANOVA after testing the preconditions.

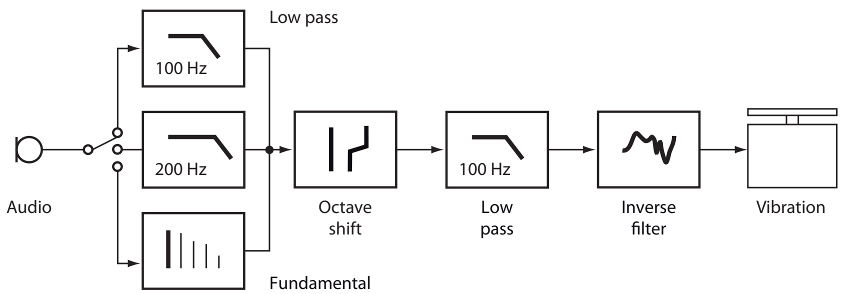


Fig. 7.12. Signal processing to generate vibration signals from the audio sum. Compression was applied in the frequency range by shifting all of the frequencies down one octave using granular synthesis. To suppress high-frequency artifacts, a 100 Hz low-pass filter was subsequently inserted.

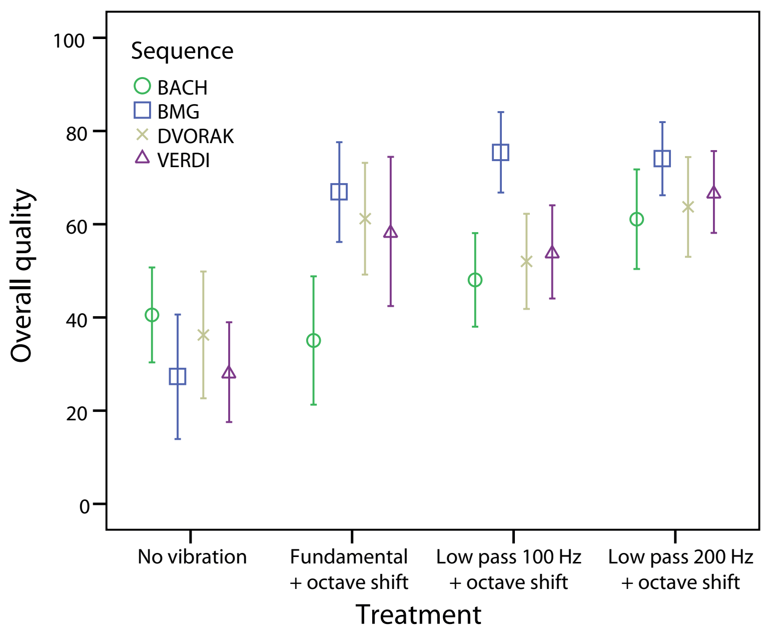


Fig. 7.13. Mean overall quality evaluation for reproduction, using different vibration-generation approaches, plotted with 95 % confidence intervals.

For the BACH sequence, shifting the lowest fundamental even farther down resulted in poor quality ratings. The occasionally weak fundamental vibrations in this sequence (as discussed in the last section) still did not cross-modally match the loudness of the sound. Therefore, no improvement in perceived quality could be expected. If increasing numbers of partials are added using low-pass filters with increasing bandwidths, the perceived quality increases for BACH, most likely due to a better intensity match between modalities.

The quality scores for the BMG sequence depend much less on the initial filtering. As discussed before, the difference between the 'fundamental' condition and the 'low-pass 100 Hz' condition are small. By octave-shifting the signals, the character of the vibration changed. Some subjects described the vibrations as 'wavy' or 'bumpy' rather than as 'humming' perceptions, as they had previously. However, many subjects liked the varied vibration character, and the averaged quality ratings did not change significantly compared with Figure 7.8 and 7.10. Adding more signal content above 100 Hz resulted in no further improvement for BMG for the reasons already discussed in the previous section.

The results were different for the DVORAK and VERDI sequences. In Section 7.2.1, no preference for one of the two low-pass conditions was observed. However, if these sequences are shifted in frequency, Figure 7.13 reveals an increase in quality for the 200 Hz low-pass treatment. This could be explained by considering the periods during which the lowest-frequency component is greater than 100 Hz (e.g., VERDI second 10 to 17). By octave-shifting these components while retaining their acceleration levels, they become perceptually more intense due to the decreasing equal-intensity contours for WBVs (refer to Section 5.3). In addition, the vibrations were reported to cause less tingling. The same result held true for octave-shifting the fundamental.

The dependence of the quality scores on the music sequence and the filtering approach was confirmed statistically by the very significant ($p<1\%$) effects for the factor sequence, the factor treatment, and the interaction of both. On average, all of the treatment conditions were judged to be better than without vibrations on a very significant level ($p<1\%$). No statistically significant differences between the 'fundamental' and the 'low-pass 100 Hz' conditions were observed. However, the 'low-pass 200 Hz' condition was judged to be slightly but significantly, better ($p<5\%$) than the 'fundamental' (averaged difference = 11) and the 'low-pass 100 Hz' (averaged difference = 9) treatments with octave-shifting. As explained above, these main effects must be interpreted in the context of the differences between sequences.

It can be concluded that octave-shifted vibrations appeared to be integrable with the sound in many cases. The best quality scores were achieved, independent of the sequence used, by applying a higher low-pass frequency, e.g., 200 Hz.

7.2.4 Substitute Signals

It was hypothesized in the previous section that the variance in the vibration character that resulted from the frequency shift would not negatively influence the quality scores. Thus, it might be possible to compress the frequency range even more. This approach was evaluated using several substitute signals, and it is discussed in the following section. Figure 7.14 presents the signal-processing chain. A signal generator was implemented in Pd to produce continuous sinusoidal tones at 20, 40, 80, and 160 Hz. The frequencies were selected to span a broad frequency range and to be clearly distinguishable considering the JNDFs for WBVs measured in Chapter 4. Additionally, a condition was included using white Gaussian noise (WGN), low-pass-filtered at 100 Hz. These substitute signals were further multiplied with the envelope of the original low-pass-filtered signal to retain the timing information. An envelope follower was implemented, which calculated the RMS amplitude of the input signal using successive analysis windows. Hann windows were applied, and the window size was set to 1,024 samples, which corresponded to approximately 21 ms, to avoid smearing the impulsive signal content. The period for successive analysis was half of the window size.

The reduction of vibration bandwidth, using the substitute signal approach, is illustrated in Appendix D. For comparison, the BRTF of a low-cost inertial shaker is appended as well.

The quality scores, using substitute vibration signals, are presented in Figure 7.15. Again, an ANOVA was applied for the statistical analysis. All of the substitute vibrations, except for the 20 Hz condition, were judged to be better than reproduction without vibration at a highly significant level ($p < 1\%$). The average differences, compared with the no-vibration condition, were between 29 scale units for the 40 Hz vibration and 18 scale units for WGN and the 160 Hz vibration. There was no significant difference between the 20 Hz vibration and the no-vibration condition. The subjects indicated that the 20 Hz vibration was too low in frequency and did not fit with the audio content. In contrast, 40 and 80 Hz appeared to fit well. No complaints about a mismatch between sound and vibration were noted. The resulting overall quality was judged to be comparable to the low-pass conditions in Figure 7.8 (no significant difference).

Interestingly, even the 160 Hz vibration resulted in fair quality ratings. However, a trend toward worse judgments, compared with the 80 Hz condition, was observed ($p \approx 11\%$). A much stronger effect was expected because this vibration frequency is relatively high, and tingling effects can occur. There was some disagreement between subjects, which can be observed in the larger confidence intervals for this condition.

Even more interesting, the reproduction of WGN resulted in fair quality ratings. However, this condition was still judged to be slightly worse than the 40 and 80 Hz vibrations (average difference = 11, $p < 5\%$). The effect was strongest for the BACH sequence, which resulted in poor quality judgments

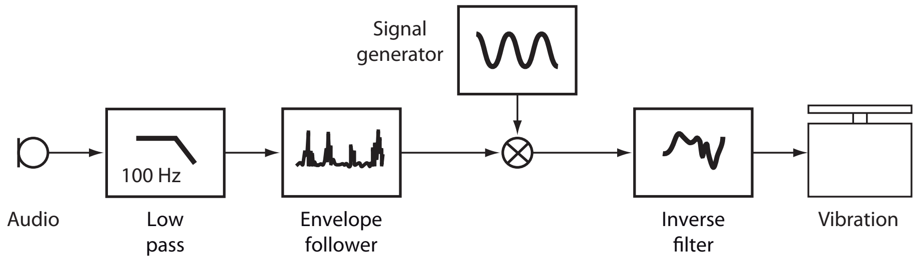


Fig. 7.14. Signal processing to generate vibration signals from the audio sum. The envelope of the low-pass-filtered signal was extracted and multiplied with substitute signals, such as sinusoids at 20, 40, 80, and 160 Hz or white noise.

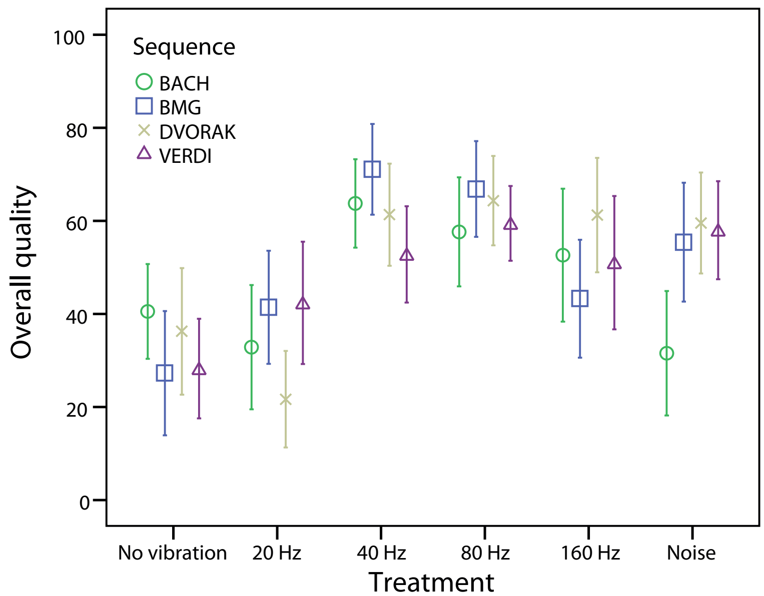


Fig. 7.15. Mean overall quality evaluation for reproduction, using different substitute vibration-generation approaches, plotted with 95 % confidence intervals.

(very significant interaction between sequence and treatment, $p < 1\%$). The BACH sequence contained long tones that lasted for several seconds, which did not fit with the 'rattling' vibrations excited by the noise. In contrast, in the BMG, DVORAK, and VERDI sequences, impulses, and short tones resulted in brief vibration bursts of white noise, which felt less like 'rattling'. Nevertheless, the character of the bursts was different from sinusoidal excitation. In the BMG sequence in particular, the amplitude of the transient vibrations generated by the bass drum varied depending on the random section of the noise. This finding is most likely one of the reasons why the quality judgment for BMG in the noise condition tended to be worse compared, e.g., with the approach using a 40 Hz vibration.

Even simple vibration signals can result in good reproduction quality. For the tested sequences, amplitude-modulated sinusoids at 40 and 80 Hz worked well.

7.3 Intensity Approaches

In the previous experiments, the overall vibration intensity was adjusted individually by each test participant. However, the intensity differences between consecutive vibration components or between vibration components at different frequencies were held constant. It was reported that the expected vibrations were missing in some cases, possibly due to the differing frequency-dependent thresholds and differences in growth of sensations for the auditory and tactile modalities, as discussed in Section 2.2.1 and Chapter 5. Therefore, an attempt was undertaken to better adapt the signals to the different dynamic ranges. The influence of dynamic compression is evaluated in Section 7.3.1, and the differences between thresholds is compensated for in Section 7.3.2.

7.3.1 Compression of Dynamic Range

To better match cross-modally the auditory and tactile growth of sensation with increasing intensity, the music signal was compressed in the vibration generation process, as illustrated in Figure 7.16. As one moved toward lower frequencies, the auditory dynamic range decreased gradually and the growth of sensation with increasing intensity rose more quickly (see Section 2.2.1). In the tactile modality, the dynamic range is generally smaller compared with audition; however, no strong dependence on frequency between 10 and 200 Hz was found in this study (see Section 5.2). Accordingly, there was not much variation between frequencies in the growth of sensation of WBVs with increasing intensity. Therefore, less compression appears necessary toward lower frequencies. However, a frequency-independent compression algorithm was implemented for simplicity.

The amount of compression needed for ideal intensity matching between both modalities was predicted using cross-modal matching data, determined in Section 5.4. For moderate sinusoidal signals at 50, 100, and 200 Hz, a 12 dB increase in sound pressure level matched well with an approximately 6 dB increase in acceleration level, which corresponds to a compression ratio of two. Further, the curve of sensation growth versus sensation level flattens toward higher sensation levels in the auditory [74] and tactile domains [132] (see Section 5.2, Figure 5.4). This finding might be important because loud music typically excites weak vibrations. The effect can be accounted for using higher compression ratios. Therefore, altogether, three compression ratios (two, four, and eight) were selected for testing. Attack and release periods of 5 ms were chosen to follow the source signals quickly.

Statistical analysis was applied as described above using a repeated-measures ANOVA and post hoc pairwise comparisons with Bonferroni correction. The quality scores for the concert experience, using the three compression ratios, are plotted in Figure 7.17. Again, the no-vibration condition was used as a reference. Compressing the audio signal using a ratio of two

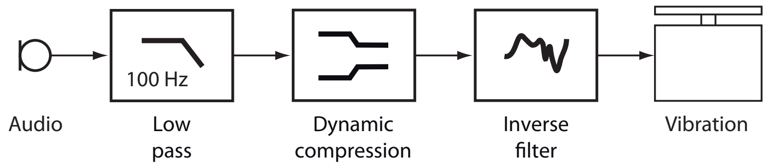


Fig. 7.16. Signal processing to generate vibration signals from the audio sum. The low-pass-filtered signal was compressed using different compression factors.

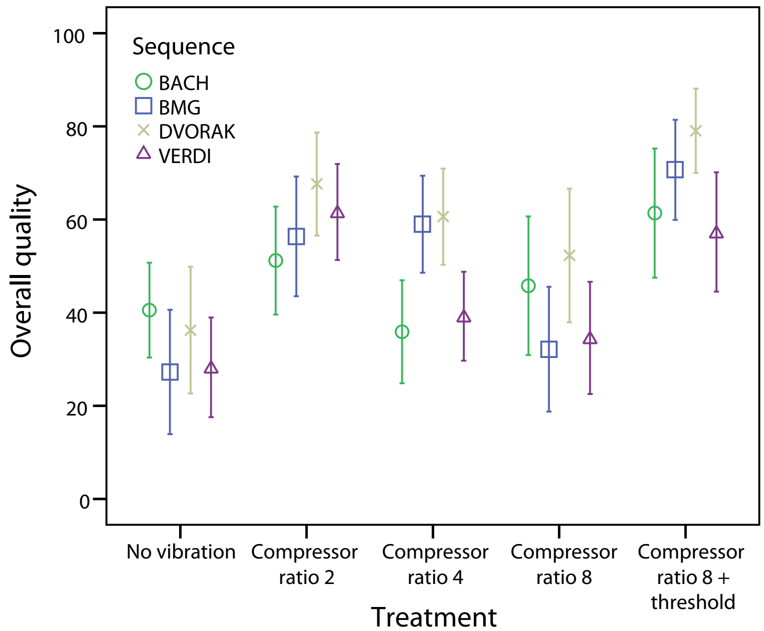


Fig. 7.17. Mean overall quality evaluation for reproduction, using different vibration generation approaches, plotted with 95 % confidence intervals.

resulted in significantly improved quality perception compared with the no-vibration condition (average difference = 26, $p < 1\%$). Although, the ratings were not statistically better than the 100 Hz low-pass condition in Section 7.2.1, some test participants reported that the initial level adjustment was easier, particularly for the DVORAK sequence. This finding is plausible because the DVORAK sequence covers quite a large dynamic range at low frequencies, which might have resulted in missing vibration components if the average amplitude was adjusted too low or in mechanical stimulation that was too strong if the average amplitude was adjusted too high. Therefore, compressing the dynamic range could have made it easier to select an appropriate vibration level.

Increasing the compression ratio further to four or eight reduced the averaged quality scores (average difference between a ratio of two and four = 11, $p < 5\%$; average difference between a ratio of two and eight = 18, $p < 1\%$). The reason for this decrease in quality appeared to be the noise floor of the audio signal, which was also amplified by the compression algorithm. This vibration noise was primarily noticeable and disturbing during the passages of music with little or no low-frequency content. In particular, for BACH and VERDI, such passages can be found. This fact would explain the bad ratings for these sequences already with a compression ratio of four. To check this hypothesis, the compression algorithm with a ratio of eight was tested again using a threshold. Loud sounds above this threshold were compressed, whereas quieter sounds remained unaffected. The threshold was adjusted for each sequence so that no vibrations were perceivable during passages with little frequency content below 100 Hz. The resulting perceptual scores are plotted on the right side in Figure 7.17. The quality was judged to be significantly better compared with the no-vibration condition (average difference = 34, $p < 1\%$) and with the compression ratios of four and eight without a threshold (average difference = 18 and 26, respectively, $p < 1\%$). However, there was no significant difference compared with the compression ratio of two or the 100 Hz low-pass condition in Section 7.2.1.

These findings indicate that even strong compression might be applied to music-induced vibrations without impairing the perceived quality of the concert experience. In contrast, compression appears to reduce the impression of missing vibrations and thus makes it easier to adjust the vibration level. However, a suitable threshold must be selected for strong compression ratios. This threshold appears possible if the source signal is uncompressed, which is typically the case for classical recordings. In contrast, modern music or movie soundtracks are occasionally already highly compressed with unknown compression parameters, which could be problematic in terms of vibration generation.

7.3.2 Cross-Modal Intensity Matching

As discussed earlier, the thresholds and equal-intensity contours differ for the perception of sound and WBVs. If a sweep is played from 20 to 200 Hz, with a constant sound pressure level via loudspeakers, the resulting perceived intensity becomes stronger with increasing frequency. If the sweep is reproduced with a constant acceleration level via a vibration chair, the perceived vibration intensity remains almost constant up to approximately 100 Hz and then decreases toward higher frequencies. Problems with the intensity mismatch between modalities were already discussed in Section 7.2. Therefore, an attempt is made in this section to match the intensities between both modalities. This attempt is implemented by frequency weighting the low-pass-filtered signal in the processing chain, as illustrated in Figure 7.18.

In the auditory domain, the isophones are not completely parallel depending on the intensity of the sound. Because sound pressure levels clearly above the threshold are typically associated with vibrations, it appears reasonable to use B- or C-weighting for compensation, representing medium- or high-level sounds, respectively [81]. B-weighting was removed from the current measurement standards; however, it is applied here because it represents curves of medium equal-loudness levels at 50-70 phon. The B-weighted signal was then further processed to account for the frequency-dependent sensitivity to WBVs. In the tactile domain, equal-intensity contours appear to be rather parallel (see Section 5.3). No filtering was necessary below 100 Hz because the contours are rather flat in terms of acceleration level. Above this frequency, the increasing contours were approximated by amplifying the signal, starting with 0 dB and increasing by 6 dB per octave toward higher frequencies. In the signal-processing chain for vibration generation, both weightings totaled an overall amplification of higher frequencies.

The quality judgments for these conditions are plotted in Figure 7.19. Reproductions with weighted vibrations were perceived to be better than those without vibrations. This effect was strongest for weighted low-pass-filtered vibrations less than 100 Hz (average difference = 18, $p < 1\%$). Increasing the low-pass frequency to 200 Hz degraded the resulting quality scores toward the no-vibration condition; however, statistically, only a trend was observed (average difference = 11, $p \approx 9\%$). This perception of a poor-quality experience can be explained by the tingling sensations that were reported by many test participants because of the amplified high frequencies. The goal to better match the intensities between both modalities could not be achieved because of the disturbing tingling effect. Only two subjects stated that the vibration intensity better matched the audio signal. Because the low-pass conditions without frequency weighting, which are plotted in Figure 7.8, were judged to be significantly better, the frequency weighting implemented here cannot be recommended.

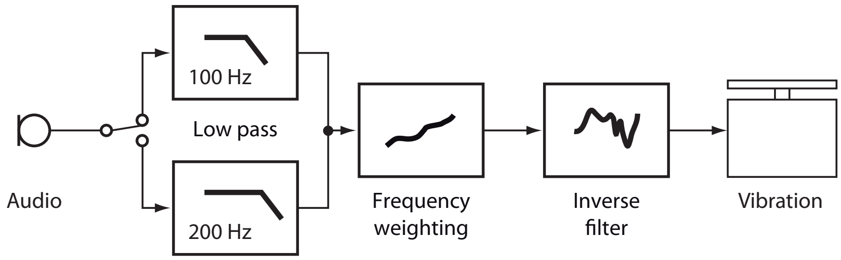


Fig. 7.18. Signal processing to generate vibration signals from the audio sum. The low-pass-filtered signal was frequency-weighted to match the intensities between both modalities.

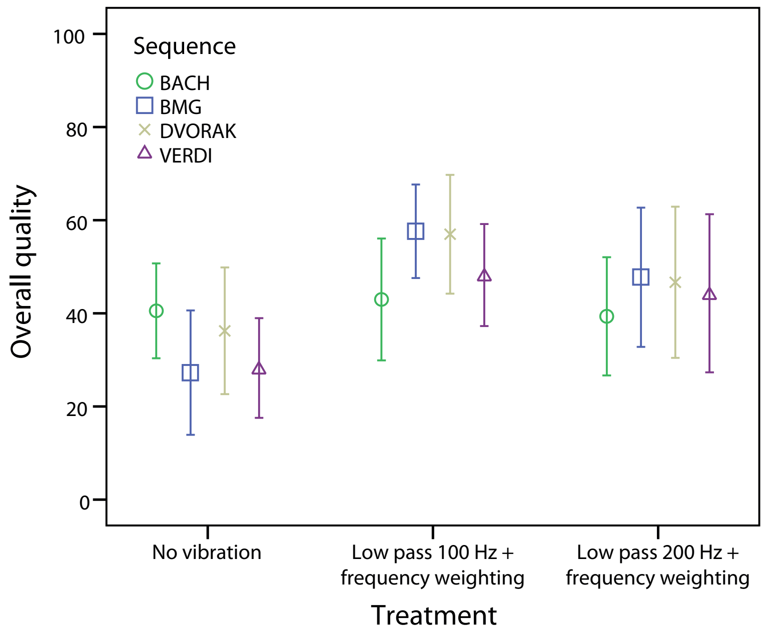


Fig. 7.19. Mean overall quality evaluation for reproduction, using different vibration generation approaches, plotted with 95 % confidence intervals.

7.4 Level Adjustments

7.4.1 Preferred Vibration Magnitude

In the previous experiments, the vibration intensity was adjusted individually by the test participant for each stimulus. The adjustment process was typically finished within the first 5 to 10 seconds. To check whether there was general agreement concerning an ideal vibration intensity, the preferred vibration amplitudes were logged during the above experiments. For the condition with low-passed vibrations at 100 Hz the mean sound pressure levels and acceleration levels are shown in Table 7.1. All of the values were calculated from the first 10 s of each sequence. The first 10 s were selected because this period is assumed to be the relevant period for level adjustment.

Table 7.1. Sound pressure levels and preferred acceleration levels for the four music sequences using the low-pass approach with a 100 Hz cutoff frequency. Additionally, the cross-modally matched levels of two sinusoids at 50 and 100 Hz are shown (see Section 5.4). The measured levels in the Semperoper and church are added for references (see Chapter 6).

		Sound		Vibration	
		$L_{F,max,LP100}$	L_{eq}	$L_{acc,F,max}$	L_{acc} (STD)
Sequence	BACH	76 dB	75 dB	104 dB	96 (5 dB)
	VERDI	77 dB	74 dB	101 dB	97 (5 dB)
	DVORAK	73 dB	71 dB	100 dB	90 (6 dB)
	BMG	88 dB	81 dB	105 dB	99 (6 dB)
Sinusoid at	50 Hz		91 dB		102 (5 dB)
	100 Hz		92 dB		108 (11 dB)
Measurement in	Semperoper		90 dB		≈ 90 dB
	Church		90 dB		≈ 100 dB

The equivalent continuous sound pressure level L_{eq} was given for all conditions. Additionally, the audio signal was low-passed at 100 Hz, and the highest RMS level was calculated using fast time-weighting ($L_{F,max,LP100}$, maximum sound pressure level). $L_{F,max,LP100}$ was slightly higher than L_{eq} for all of the music sequences, corresponding to L_{eq} for the sinusoids and measurements.

The preferred acceleration levels L_{acc} are given for the four music sequences. The standard deviations indicate variations between subjects, which might have been due to differing perception thresholds (as discussed in Section 2.2.1) or different preferences regarding vibration intensity. Similarly, large standard deviations were observed in the cross-modal intensity matching experiment in Section 5.4. Averaged over all of the sequences, L_{acc} was adjusted to approximately 20 dB greater than L_{eq}. However, it is difficult to compare both values directly. As shown in Appendix C, the vibration amplitude varied over time depending on the reproduced sequence. It was hypothesized that the maximum acceleration level would play a more dominant role for vibration adjustment. Therefore, the maximum acceleration level $L_{\mathrm{acc,F,max}}$, within the first 10 s of each sequence, is also provided in Table 7.1. A short time constant of 125 ms was used to calculate the level corresponding to fast time-weighting in acoustics. This constant was chosen because both modalities exhibit strong similarities in their temporal processing, as discussed in Chapter 2. The maximum acceleration level $L_{\mathrm{acc,F,max}}$ was adjusted to approximately 17-28 dB above the maximum low-passed sound pressure level $L_{\mathrm{F,max,LP100}}$. For this comparison, it was assumed that the vibrations were matched to low-frequency sounds only. However, musical tones consist of multiple frequency components (harmonics), which contribute to overall loudness. Using $L_{\mathrm{F,max,LP100}}$ as a reference might therefore underestimate the loudness with which the vibration is compared. This relationship could also explain the smaller level differences between sound and vibration determined in the cross-modality matching experiment using sinusoids in Section 5.4. Some results are provided in Table 7.1 for comparison. For a tone at 50 Hz, a 91 dB SPL resulted in a 102 dB acceleration level (11 dB difference), whereas a 92 dB SPL resulted in a 108 dB acceleration level (16 dB difference) for a tone at 100 Hz. In contrast to the perceived vibration magnitude, auditory intensity perception at frequencies less than 100 Hz is heavily dependent on frequency.

In comparison, the measured level differences in actual concert rooms appear to be much smaller. Even in the church, only an approximately 10 dB difference between acceleration level at the seat surface and the sound pressure level at ear height was reported in Chapter 6 for frequencies of approximately 100 Hz. In the Semperoper, the acceleration level and sound pressure level at less than 100 Hz were almost identical.

In summary, for the sequences under investigation, relatively high vibration levels were preferred in the concert reproduction situation. However, there was considerable variation in the preferred acceleration level depending on sequence and subject. This finding supports the decision for individual intensity adjustment in the previous experiments.

7.4.2 Preferred Audio Equalization

In the previous section, the preferred acceleration levels were compared with the reproduced sound pressure levels. However, the preferred bass level also might change under the influence of additional vibrations. In Section 5.5, vibrations were found to be able to influence the perceived loudness of a sound at low frequencies. Tones were perceived to be louder when vibrations were reproduced simultaneously. It could be expected that the preferred low-frequency equalization of a music signal would also be affected. Therefore, the listeners preference for low-frequency audio equalization was measured under the influence of WBVs. For instance, a preferred reduction of bass levels would be advantageous in automobile applications, in which pedestrians outside the car typically hear only the low-frequency components of music reproduced inside the car. This effect is unwanted, particularly for upper-class automobiles. It would also be beneficial to reduce the sound pressure level in discotheques to prevent hearing damage by reproducing additional vibrations. However, the direction of this effect is not predictable. On the one hand, the subjects could perceive the low-frequency content louder if vibrations are present and therefore reduce the bass amplitude to compensate for this effect. On the other hand, the listeners might expect a higher level of bass because of the vibrations. In this case, the test participants could prefer amplified low-frequency content in the audio signal to better match the vibrations.

The same setup and music sequences as in the previous sections were used for this experiment. Again, a stimulus length of approximately 1.5 min was observed to be useful to ensure that the participants had sufficient time for their evaluations. However, longer parts with little low-frequency content were removed from the sequences because they were found not to be useful for the evaluation of audio equalization, resulting in listener dissatisfaction. The task of the subjects was to adjust the intensity of the bass to their preferred level using an infinite rotary knob with no visible marks. The adjustable bass signal was derived from the mono sum of the 5.1 audio signal using a steep cross-over filter at 100 Hz (Butterworth, eighth order) and was reproduced via a subwoofer. Intensity increments were possible in 1 dB steps over a 25 dB range. The initial bass level in each trial was randomly varied over the complete dynamic range.

The vibration signals were generated from the audio sequences with the 100 Hz low-pass approach introduced in Section 7.2.1. The listeners were asked to adjust the bass equalization for four different WBV levels. The acceleration levels referred to the individual perception threshold to ensure that the perceived vibration magnitude was comparable between subjects. The thresholds of vibration perception were measured at 25 and 50 Hz. The results were comparable to the results of the studies summarized in Section 2.2.1. Because the threshold is rather flat below 100 Hz, an averaged perception threshold was calculated from the two measurements for each participant. The four WBV levels were then adjusted relative to the individual thresholds and were further labeled as follows: none; low (0 dB); medium (6 dB); and high (12 dB). The low-vibration setting was just perceptible, with an averaged acceleration level L_{acc} equal to the perception threshold. This level corresponds to a peak acceleration level $L_{\text{acc,F,max}}$ between 4 and 10 dB greater than the threshold depending on the music sequence. The high-vibration condition produced approximately the equivalent level of the preferred vibration magnitude determined in Section 7.4.1. In all cases, the BRTF of the reproduction system was individually compensated, as discussed in Section 3.2.

To test the reliability of each subject, there were two repetitions of each sequence. This process resulted in a total of 32 trials (four sequences x four vibration levels x two repetitions). The listeners were allowed to stop a trial as soon as they were satisfied with their adjustment. Before the experimental session, the listeners were familiarized with the procedure and all of the music sequences. The overall test time varied between subjects and took up to 1.5 h. All of the stimuli were presented in completely random order.

A total of 20 subjects participated in the experiment (16 male and four female). Most of them were students between 19 and 49 years old (mean, 25 years) and between 50 and 115 kg (mean, 73 kg). All participants stated that they had no known hearing or spine damage. Most of the test subjects had never participated in a listening experiment before. In general, the listeners can be regarded as naive, with no background in audio engineering or sound evaluation.

The results were analyzed with a repeated-measures ANOVA. The preferred bass equalization in the no vibration condition (reference) was dependent on the individual subject and music sequence. In general, the listeners preferred higher bass levels compared to a flat frequency response. The standard deviation for the two repetitions was approximately 4 dB, which appears quite small for naive listeners, considering that the repetitions were randomized within the entire experimental session, including various conditions with WBVs. To better illustrate the data, Figure 7.20 presents the mean preferred bass equalization levels (for different levels of WBVs) relative to the preferred bass levels without vibrations. WBVs had no significant effects on the pre-

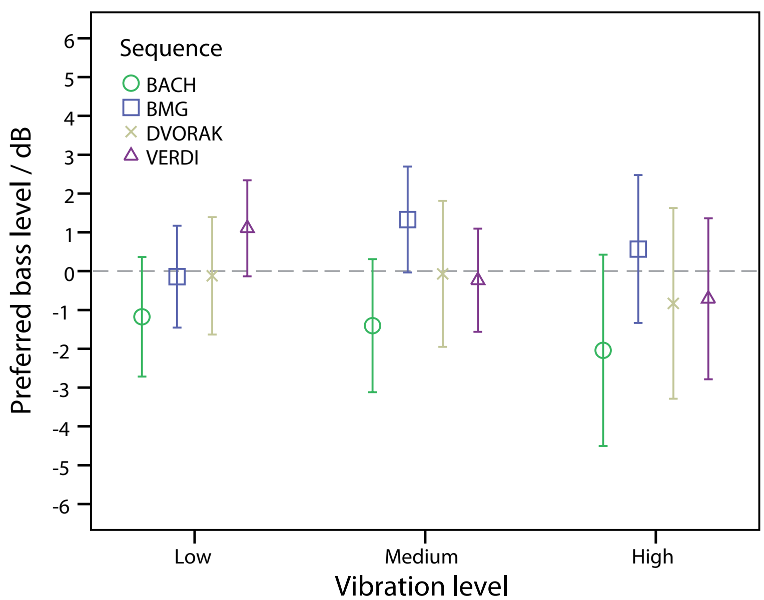


Fig. 7.20. Mean preferred bass equalization levels with 95 % confidence intervals for different levels of WBVs. The bass level is plotted relative to the individual preferred bass level without vibrations.

ferred bass equalization. This finding was true for all vibration levels. The results support neither the hypothesis that subjects would correct for increased loudness due to vibrations nor the hypothesis that subjects would increase the bass level because they expected more low-frequency content due to vibrations. Furthermore, no preference groups were observed.

One could argue that the cross-modal loudness illusion found in Section 5.5 was relatively small (approximately 1 dB) and was independent of vibration level. Therefore, only a negligible effect could have been expected in this experiment. However, in the loudness illusion experiment in Section 5.5, very short 1 s sinusoids were compared with each other in an A-B design. The test participants were aware of the added vibrations and were told to focus solely on the loudness of the two tones. Therefore, the experiment might have underestimated the increased loudness in a more realistic scenario. Interestingly, the subjects reported perceiving the bass more intensely with vibrations present during the current experiment. This finding supports the hypothesis of increased loudness because of the vibrations. However, the more intense bass might fit well with the expected bass amplification if vibrations are present. This theory could explain the consistent bass equalization preferences in the results.

The results indicated that the presence of WBVs did not affect the preferred bass equalization of naive listeners. In contrast, two other studies reported that vibrations associated with low-frequency reproduction of music in an automotive audio system produced a decrease in the level of preferred bass equalization. Data from Simon et al. [154] and Martens et al. [107] are plotted in Figure 7.21 for comparison.

Both studies investigated the same automotive audio system via binaural reproduction using headphones. Simon et al. [154] generated WBVs using a platform in a room. An adaptive staircase tracking procedure was applied for bass adjustment. Short audio segments were selected from four musical sequences. Martens et al. [107] simulated vibrations by mounting a low-frequency audio transducer under the seat of a car. The test procedure was similar to that applied in this study. Two looped audio sequences were chosen with lengths of 20-30 s. Both studies used trained listeners with backgrounds in audio evaluation. The vibration generation approaches appeared to be similar to the approach applied in this study; however, not all of the details are provided in the papers.

The data from both studies agree with a 1.5-2 dB decrease in the preferred bass level for a 4 dB increase in vibration level. At low vibration levels near the threshold (0 dB), no apparent effects were observed in any of the studies. The differences from the current results at higher relative vibration levels cannot be explained definitively. The studies differed in reproduced musical genres, with mainly classical programs in this study and with no orchestral or sacral music in the others. However, no significant differences were observed between musical genres in this study. Furthermore, the number and selection of subjects might have played a role. Audio experts could have different pref-

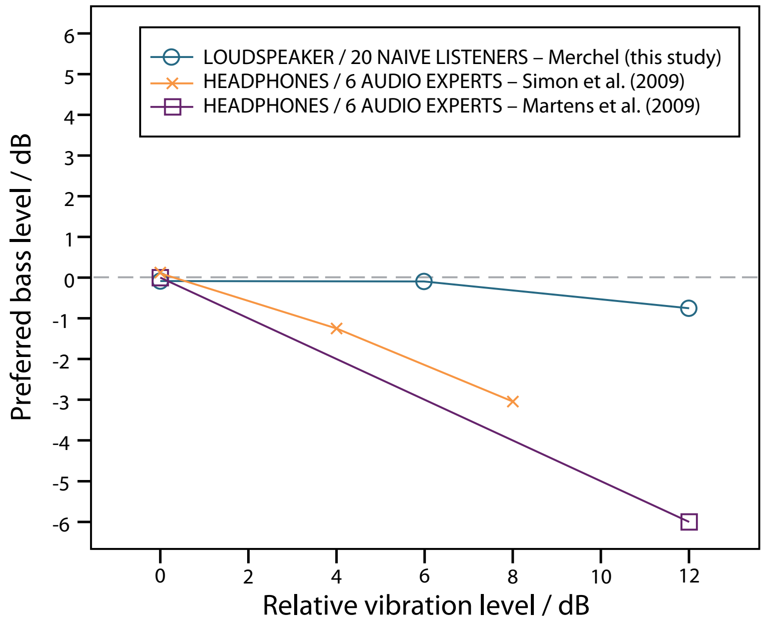


Fig. 7.21. Comparison of the relative preferred bass equalization levels from this study and from two other studies [107, 154] for different WBV levels.

erences than naive listeners. Additionally, reproduction via headphones might be important. The 'missing 6 dB' effect with headphones, which was partially related to the absence of tactile stimulation [148], might have been compensated for through the reproduction of vibrations. Martens et al. reported that the inclusion of WBVs at low levels improved the similarity between binaural reproduction and the in situ experience in the car. However, at higher vibration levels, the audio experts might have become annoyed and 'reduced the level of bass equalization, perhaps with the false hope that it might also reduce the vibration' [107]. This result would have biased the effects, resulting in lower preferred bass levels. In any case, the effects of WBVs on bass equalization preference cannot be conclusively reported.

7.5 Time Approaches

7.5.1 Envelope Processing

During the experiments in Sections 7.2 and 7.3, it was sometimes indicated by the test participants that the vibrations felt like tingling. This effect could be reduced by removing higher frequencies or by shifting the frequencies down in range. However, this processing also weakened the perceived tactile intensity, e.g., of broadband transients. The question arises of what perceptual relevance transients have for the perceived quality of music perception compared to steady-state vibrations. This question could be evaluated by isolating transients and sinusoids in the vibration signal. Therefore, an algorithm was implemented in Pd to identify broadband transients in the low-passed signal (see Figure 7.22). If an attack was detected simultaneously in multiple frequency bands, the amplitude envelope (calculation as in Section 7.2.4) of the low-passed signal was passed through for 100 ms or was otherwise muted. By multiplying this envelope or its inverse with the low-passed signal, the transients and steady-state tones can be separated.

The BMG sequence (see Figure 7.5) exhibits strong transients (kick-drum and percussions) and distinct steady-state tones (bass guitar). Thus, this sequence was selected for this experiment, and the above processing was applied to generate vibrations. Additionally, a further approach was selected, which reduced tingling sensations for steady-state tones and simultaneously kept the transients unaffected. This process was undertaken by fading out the steady components at 10 dB/s using a multi-band compressor with long attack and short decay.

The three different vibration conditions were evaluated as described above. Again, the results were analyzed using a repeated-measures ANOVA and post-hoc pairwise comparisons with Bonferroni correction. The overall quality ratings of the concert reproductions are plotted in Figure 7.23. The reproduction of only transients was rated similar to the no-vibration condition (no statistically significant improvement). The test participants frequently reported that the vibrations were missing. This was not the case for vibrations generated from steady-state tones only. Steady-state vibrations were rated very significantly better than reproduction without vibrations (average difference = 27, $p < 1\%$) and significantly better than transients only (average difference = 18, $p < 5\%$). The rating was similar to the quality ratings using only a low-pass filter (see Section 7.2.1). These results suggest that steady-state components are more important than transients for the reproduction of music-induced vibrations from the BMG sequence. However, conclusions should be drawn carefully because the transients in this sequence aligned well with the beginning of the tones. Reproducing vibrations generated from steady tones still resulted in some transient content because of the sharp fade-in of the sinusoidal vibrations.

Adding the transients again and fading out the tonal components resulted in surprisingly good quality scores (fade out) compared to the no-vibration

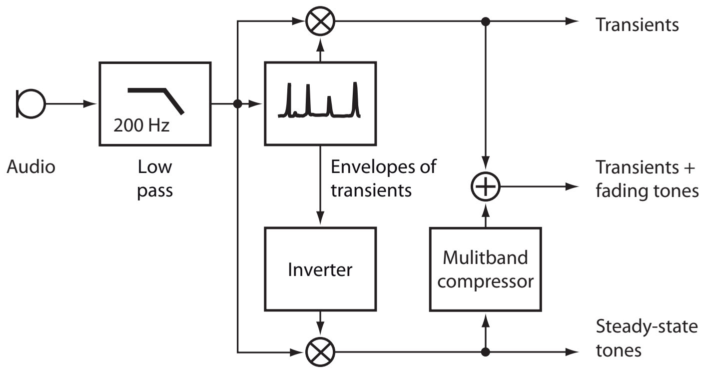


Fig. 7.22. Signal processing to generate vibration signals from audio sum. Transients were detected in the audio signal and were separated from steady-state tones by multiplying them by their (inverted) envelope. Steady tones were additionally faded out with 10 dB/s using a multi-band compressor with long attack and short release.

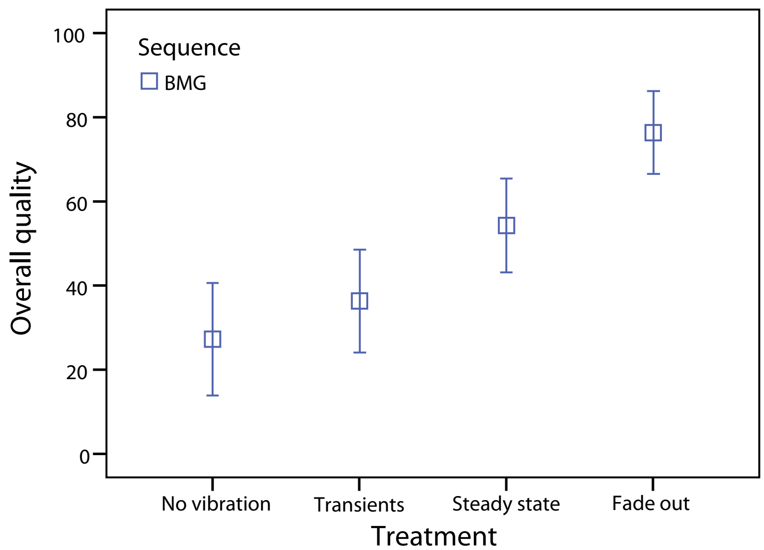


Fig. 7.23. Mean overall quality evaluation for reproduction, using different vibration-generation approaches, plotted with 95 % confidence intervals.

condition (average difference = 49, $p < 1\%$). It appears that the slow fading of steady-state vibrations did not impair the perceived quality here; in contrast, the best results were achieved. The reason for this finding was likely the reduction of the tingling sensations associated with higher vibration frequencies.

The last approach appears promising and should be further evaluated in a subsequent study using additional sequences. As described above, the processing is computationally intensive. However, it might be possible to simplify the system by reducing it to a single multi-band compressor. Using a long attack and short release period, transients could be maintained and continuous narrow-band content could be slowly faded out. Attack and release parameters could also be varied depending on frequency, such as using steeper fading slopes for higher frequencies. Additionally, the combination with one or more of the previously described approaches, such as octave shifting and compression of dynamic or frequency range, would be interesting.

7.6 Summary

Various audio-induced vibration-generation approaches were developed using the fundamental knowledge of auditory and tactile perception discussed earlier. The perceived concert quality, using combined sound and vibration reproduction, was evaluated. Seat vibrations can have a considerably positive effect on the experience of music. Because the test participants evaluated all of the approaches in Sections 7.2 and 7.3 mixed with each other in completely randomized order, the resulting mean overall quality values can be directly compared. The quality scores for concert experiences using some of the vibration-generation approaches are summarized in Figure 7.24 (all were judged to be significantly better than without vibrations, $p < 1\%$).

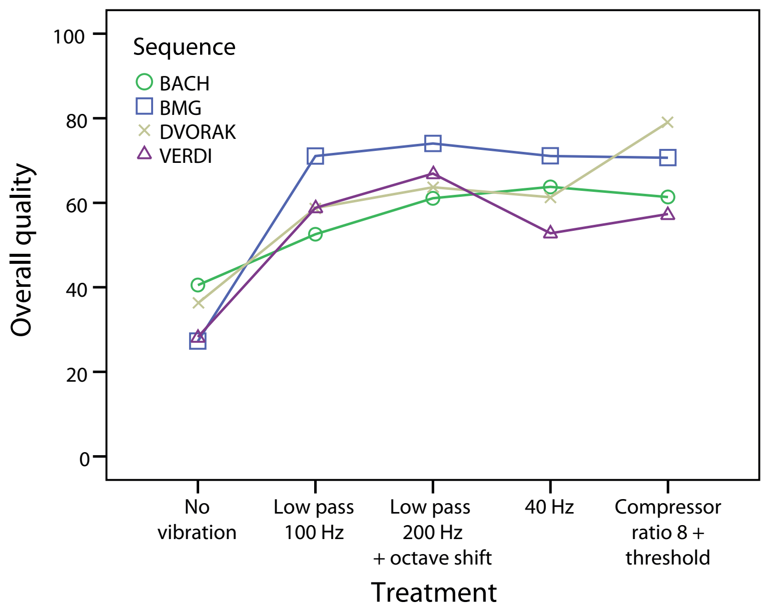


Fig. 7.24. Mean overall quality evaluation for music reproduction using selected vibration-generation approaches. For better illustration, individual data points have been connected with lines.

The low-pass filter approach was the most similar to the vibrations potentially perceived in real concert halls, resulting in good quality ratings. The approach is not computationally intensive and can be recommended for reproduction systems with limited processing power. Because the differences with a low-pass filter between 100 and 200 Hz were small, the lower cutoff frequency is recommended to avoid unwanted sound generation. By adding further processing, the sound generated by the vibration system can be further reduced while preserving the good quality scores for concert reproduction. One successful approach involves compression in the frequency range, e.g., using octave shifting. Surprisingly, even strong frequency compression to an amplitude-modulated sinusoid, e.g., 40 Hz, appears applicable. This process allows for much simpler and less expensive vibration-reproduction systems, such as in home cinema scenarios. However, some signal processing power is necessary, for instance, to extract the envelope of the original signal. Furthermore, it appears useful to apply some dynamic compression, which makes it easier to adjust the vibration level. However, the effect appears to depend on the musical material, and it did not result in an improved overall quality rating. In this study, source signals with a high dynamic range were used as a starting point. Further evaluation using audio data that are already compressed with unknown parameters is necessary.

The preferred acceleration levels were found to be significantly greater than the corresponding sound-pressure levels. Differences of 20-30 dB were common in the experiments, exceeding even the values measured in the church (approximately 10 dB). However, there was considerable variation between test participants. It is therefore recommended that users be given the ability to make individual adjustments.

In summary, the test participants appeared to be relatively tolerant of a wide range of music and seat vibration combinations. Perhaps our real-life experience with the simultaneous perception of auditory and tactile events is varied, and thus, our expectations are not strictly determined. For example, the intensity of audio-related vibrations might vary greatly between different concert venues. Additionally, the tactile sense is rather limited in some aspects compared with audition. In particular, the frequency resolution and the perception of pitch are strongly restricted, which allows for the modification of frequency content over a wide range. The effects of additional vibration reproduction depended on the selected music sequence to some extent. For example, the BMG rock music sequence was judged to be significantly better in most of the cases, including vibrations, than the classical compositions (see Figure 7.24). This finding is plausible because we expect strong audio-induced vibrations at rock concerts. However, adding vibrations clearly increased the perceived quality, even for classical music.

During the experiments, the test participants occasionally indicated that the vibrations felt like tingling sensations. One approach to reducing the tingling sensations would be to fade out continuous vibrations with a long attack and short release using a compressor. This type of temporal processing appears promising and should be evaluated further.

Another approach would be to code auditory pitch information into a different tactile dimension. For example, it would be possible to transform the pitch of a melody into the location of vibration along the forearm or back using multiple vibration actuators. This frequency-to-place transformation approach is typically applied in the context of tactile hearing aids, in which the tactile channel is used to replace the corrupt auditory perception

[145]. However, in such sensory substitution systems, the transformation code must be learned. This study demonstrated that it might not be necessary to code all available auditory information into the tactile channel to improve the perceived quality of music. However, there is still creative potential using this approach, which has been applied in several installations, [23, 24, 68, 93, 110]. This potential should be investigated further in subsequent studies.

The results presented in this chapter demonstrated that there is a general connection between vibrations and the perception of music. However, in this study, only seat vibrations were reproduced in a surround setup. Interestingly, none of the participants complained about an implausible concert experience. Still, one could question whether the 5.1 reproduction situation can be compared with a live situation in a concert hall or church. Because the test participants generally preferred higher acceleration levels, such as those that are typically excited in a concert venue, it is hypothesized that real halls could benefit from amplifying the vibrations in the auditorium. This effect could be achieved passively, e.g., by manipulating the mechanical floor construction, or actively, using electrodynamic exciters as in the above experiments. It would be very interesting to investigate the effect in an actual concert situation.

8

Summary and Outlook

The goal of this thesis was to improve our understanding of the coupled perception of sound and vibration using the example of auditory-tactile music perception. Therefore, the capabilities and limitations of both modalities were compared first. However, particularly for the perception of vibrations at low levels, only limited results have been reported to date. Therefore, the frequency discrimination and intensity perception of vertical WBVs was investigated in several experiments. Additionally, cross-modal effects were considered. In the second part of this work, this gained knowledge was applied to develop and evaluate various approaches to generating vibrations from music sequences. Vibrations were found to play a significant role in the perception of music.

An experimental setup was developed to conduct these experiments. WBVs were reproduced over a broad frequency and amplitude range via an electro-dynamic shaker. The corresponding BRTFs were individually equalized in all of the investigations using inverse filters.

In the following, the main results and conclusions that were gained with each sub-step will be recapitulated.

1. *Comparison of the sensory modalities*
   The anatomical and perceptual similarities and differences between the two modalities were compared to provide the groundwork for understanding auditory-tactile music perception. For the perception of music-induced vibrations, the Pacinian channel appears particularly relevant. Despite the smaller frequency range, this channel exhibits remarkable similarities to the auditory system. Similar to hearing, the absolute perception threshold depends on stimulus duration, frequency, and age, among other factors. Both modalities are able to integrate energy over time for short stimuli. The threshold rises temporarily after prolonged stimulation. Both modalities exhibit the ability of one stimulus to mask another. In comparison, in the tactile modality, broader masking patterns are excited around the

masker frequency, with strong masking toward higher frequencies. If the temporal discrimination abilities are compared in terms of gap-detection thresholds, good temporal resolution was found for both modalities in the millisecond range.
In the context of this study, two differences appeared to be most important. First, the tactile dynamic range is considerably smaller than the auditory dynamic range. However, the most evident difference is the reduced ability to distinguish between frequencies with the tactile modality.

2. *Determination of the tactile frequency-discrimination ability for WBVs*
JNDFs were measured for WBVs while carefully eliminating concomitant changes in the subjective intensity or stimulation location. For reference frequencies between 20 Hz and 90 Hz, difference limen between 7 Hz and 66 Hz were measured, which increased with increasing frequency. In comparison, the auditory JNDFs were approximately 1 Hz for similarly low frequencies. It was concluded that tactile frequency perception is weak, and thus, pitch might be of minor importance in the process of vibration generation from music.

3. *Investigation of the perceived vibration magnitude for WBVs*
The perceived vibration magnitude of WBVs was measured as a function of sensation level for frequencies between 10 and 200 Hz, and a large amplitude range was assessed using a magnitude-estimation method. A curvilinear relationship was observed for sensation magnitude that was dependent on sensation level. Compared to audition, the growth of the perceived magnitude was steeper, particularly at low sensation levels. Considering that weak vibrations will be coupled with loud sounds in the context of music perception, this perceptual difference can be compensated for using compression algorithms.

4. *Definition of a new perceptually motivated measurement for the perceived vibration magnitude*
Perceived vibration magnitude M was defined to represent human vibration intensity perception, comparable to auditory loudness N. The measurement unit of the perceived vibration magnitude M is called Vip, with the unit symbol vip. The perceived vibration magnitude of 1 vip was proposed to be equivalent to the perceived vibration intensity of a 20 Hz vibration with a sensation level of 15 dB.

5. *Curves of equal vibration magnitude were measured for WBVs*
Equal-vip contours were constructed from magnitude estimation data, as well as by matching the intensities of test stimuli to reference vibrations using a method of adjustment. Those curves were observed to parallel each other, indicating that in contrast to audition, tactile suprathreshold intensity perception does not depend on frequency, at least at less than 250 Hz.

6. *Cross-modal intensity matching for sinusoids*
   In a cross-modal matching experiment, tones of equal loudness were matched to WBVs with equal vibration magnitude M, even if the acoustic frequency was different from the frequency of vibration. However, large inter-individual variations were observed. This finding might have resulted in large differences in preferred vibration levels for tactile music perception.

7. *Verification of an auditory-tactile loudness illusion*
   An influence of WBVs on loudness perception was proven using a magnitude-production experiment. The participants were instructed to focus solely on loudness and to ignore the vibrations. Nevertheless, low-frequency sounds were perceived to be approximately 1 dB louder if WBVs were present. Interestingly, no systematic influence of vibration frequency or acceleration level was observed. Although the effect was small, it was hypothesized that the vibrations might influence the preferred bass equalization for music reproduction.

8. *Determination of the preferred bass equalization for simultaneous vibration reproduction*
   WBVs were found to have no significant effects on preferred bass equalization for naive listeners. However, the subjects often reported perceiving the bass more intensely with WBVs present. This finding supports the hypothesis of increased loudness because of the vibrations. The two conclusions are not necessarily in conflict with each other because the listeners might have expected a higher level of bass under this condition, which would explain the negligible effect on bass equalization.

9. *Development and evaluation of algorithms to generate music-driven vibrations using knowledge about frequency perception*
   Frequency information is one of the fundamental components of music, resulting in pitch perception. The question was asked whether it was possible and necessary to transform this information into accompanying vibrations. Different approaches were evaluated that modified the frequency content of the audio signal to generate suitable vibrations. Therefore, four different music sequences were selected. The low-pass filter approach was most similar to the vibrations potentially perceived in actual concert halls, and this approach resulted in significantly better quality ratings compared to concert reproduction without vibrations. The approach is not computationally intensive and can be used in reproduction systems with limited processing power. Because the differences with a low-pass filter between 100 and 200 Hz were small, the lower cutoff frequency is recommended to avoid unwanted sound generation. Applying the fundamental knowledge gained previously, the sound generated by the vibration system was further reduced while preserving the good quality scores for concert reproduction. One successful approach involves compression in

the frequency range, e.g., using octave shifting. Surprisingly, even strong frequency compression to an amplitude-modulated sinusoid, e.g., 40 Hz, appears to be applicable. This process allows for much simpler and less expensive vibration-reproduction systems with a narrow frequency range, such as in home cinema scenarios. However, some signal processing power is necessary, e.g., to extract the envelope of the original signal.

10. *Development and evaluation of algorithms to generate music-driven vibrations using knowledge of intensity perception*
The differing growths in sensation intensity in the auditory and tactile modalities were discussed earlier. This difference resulted in a mismatch between the expected and perceived vibration magnitude for music-induced WBVs. Therefore, algorithms were developed and evaluated to better adapt the vibration signals using dynamic compression. The listeners reported that as a result, a more consistent cross-modal intensity match was achieved, which made the adjustment of the vibration level easier. However, the effect depended significantly on the musical material and did not result in an improved overall quality rating. In this study, source signals with a high dynamic range were used. Further evaluation using audio data that are already compressed with unknown parameters is necessary.

11. *Determination of the preferred vibration level for auditory-tactile music reproduction*
The preferred acceleration levels were found to be significantly higher than the corresponding sound pressure levels. Differences of 20-30 dB were common in the experiments, exceeding the values measured in the opera and church. However, as predicted in the cross-modality matching experiment, there was considerable variation between test participants. Thus, it is recommended that listeners be given the ability to adjust the vibration magnitude individually in a reproduction scenario.

In summary, music-induced WBVs significantly influenced the quality of the concert experience. This effect was demonstrated for one pop music sequence and three classical music sequences. In future studies, additional music examples and genres could be added. Because the test participants typically preferred higher acceleration levels in the reproduction scenario, such as those that are typically excited in a concert venue, it is hypothesized that real halls could benefit from amplifying the vibrations in the auditorium. This hypothesis should be verified with in situ experiments. This study focused on vertical seat vibrations; thus, it would be interesting to investigate the influence of vibrations transmitted through other contact surfaces (e.g., the ground or arm rests) or directly via the air. In a pilot experiment, time approaches appeared to be promising and should thus be further explored in a subsequent study. Additionally, the combination of the algorithms described here will be evaluated in the future in greater detail.

The above results illustrate the importance of multimodal effects on quality perception. The example of an auditory-tactile concert experience was used. Further research could extend this work by adding more modalities, such as vision.

A

Measurement Positions in the Semperoper Dresden and the Lutherkirche Radebeul

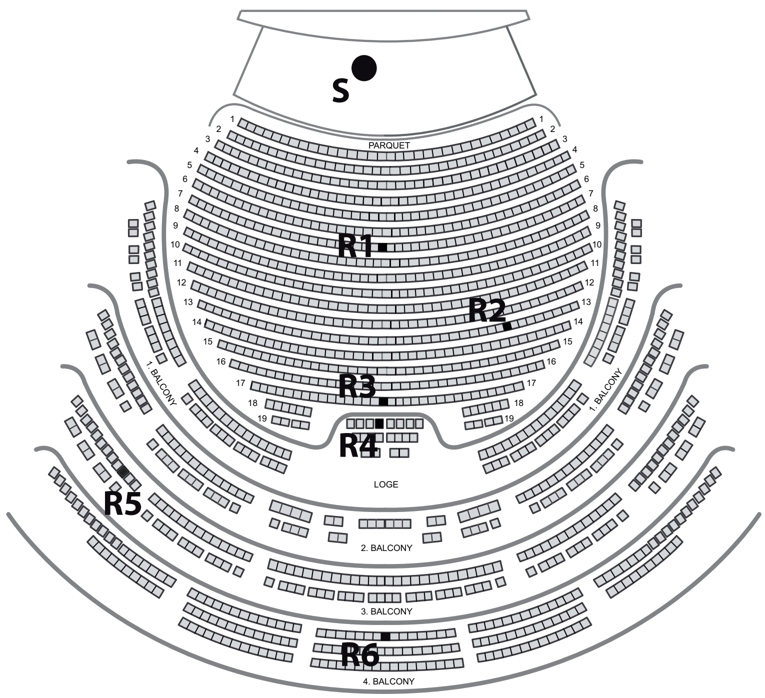


Fig. A.1. Seating plan of the Semperoper Dresden with receiver positions R1 to R6 and position S of the sound source.

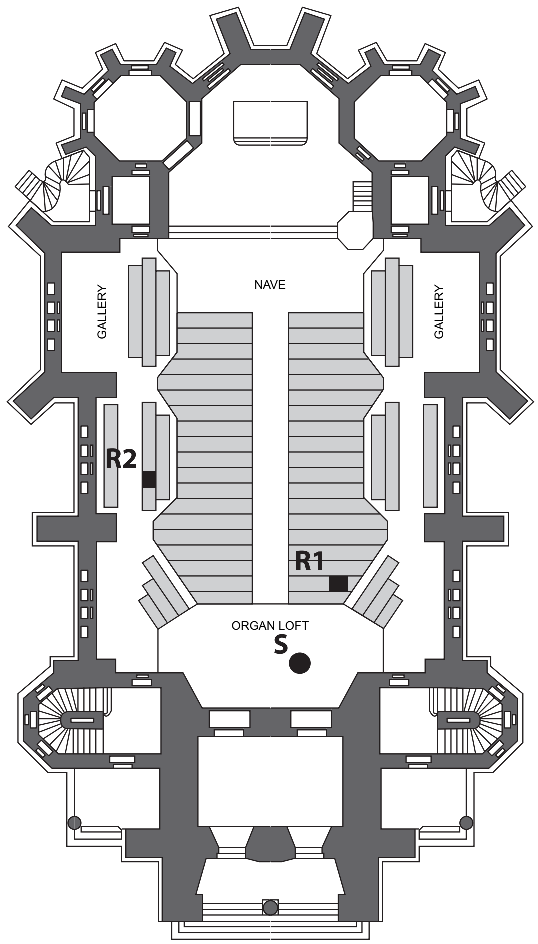


Fig. A.2. Floor plan of the Lutherkirche Radebeul with the positions of the receivers and the sound source.

B

Music-Induced Vibrations in the Lutherkirche Radebeul

An organ sequence with a significant low-frequency content (Max Reger, Introduktion d-Moll) was performed by the organ player Gottfried Trepte in the Lutherkirche Radebeul. Sound pressure levels of 90 dB(A) were obtained. The resulting seat vibrations in the gallery are plotted here. The signal was frequency weighted before plotting by low-pass filtering with 6 dB per octave above 150 Hz to account for the rising perception threshold toward higher frequencies (refer to Section 2.2.1).

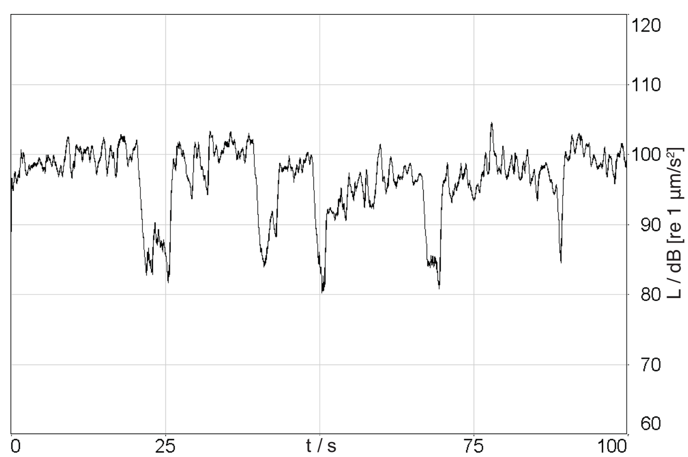


Fig. B.1. Acceleration level (time constant 125 ms) on the seat surface at position R2, measured during a church organ performance (Max Reger, Introduktion d-Moll). The signal was frequency weighted before plotting by low-pass filtering, with 6 dB per octave above 150 Hz to account for the rising threshold of perception toward higher frequencies (refer to Section 2.2.1). During the measurement time, sound-induced vibrations excited clear-to-very-strong perception.

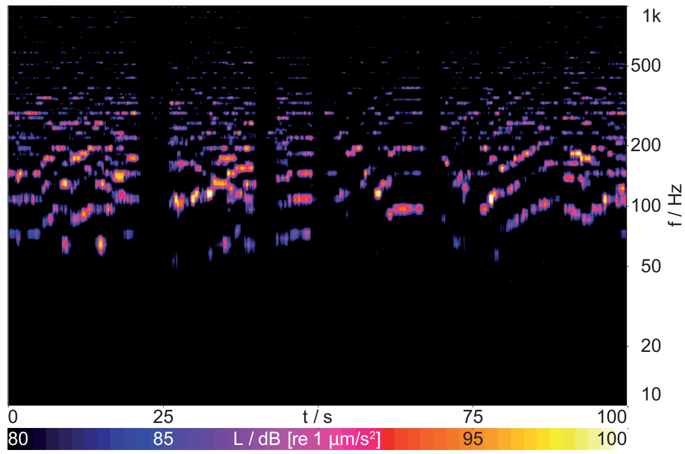


Fig. B.2. Spectrogram (FFT 8192, 10% overlapping Hann windows) illustrating the frequency-weighted (low-pass with 6 dB per octave above 150 Hz) acceleration on the seat surface for the same measurement shown in Figure B.1. The lower end of the dynamic range was adjusted to 80 dB, approximating the threshold of perception. Frequency components between approximately 50 and 500 Hz excited perceivable vibrations.

C

Preferred Level of Audio-Generated Vibrations

This appendix presents the acceleration levels vs. time plots of the first 60 s of each sequence used in Chapter 7. The plotted curves correspond to the preferred intensity averaged over all of the participants. A short time constant of 125 ms was used to calculate the level corresponding to fast time-weighting in acoustics. This constant was chosen because both modalities exhibit strong similarities in their temporal processing, as discussed in Chapter 2. The corresponding sound pressure levels are discussed in Section 7.4.1.

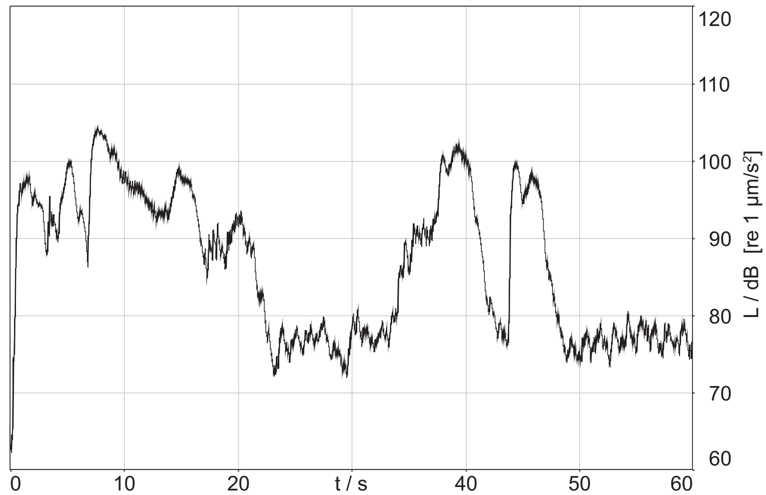


Fig. C.1. Acceleration level vs. time plot for the first 60 s of the BACH sequence, low-passed at 100 Hz. The plotted level corresponds to the preferred intensity averaged over all of the participants.

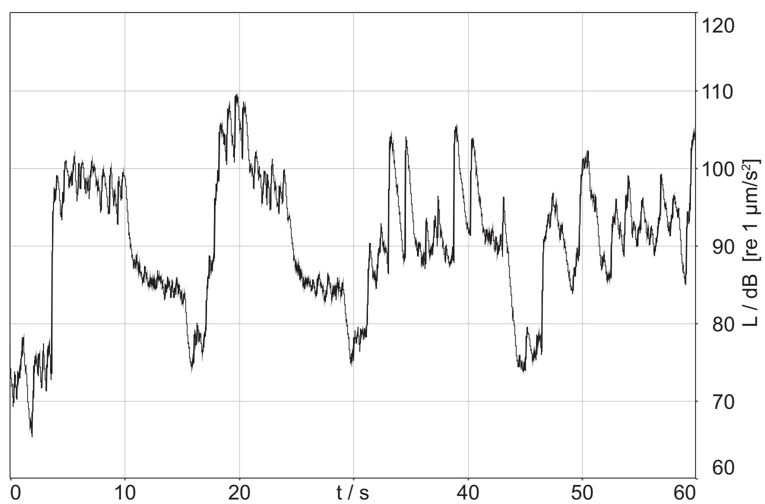


Fig. C.2. Acceleration level vs. time plot for the first 60 s of the VERDI sequence, low-passed at 100 Hz. The plotted level corresponds to the preferred intensity averaged over all of the participants.

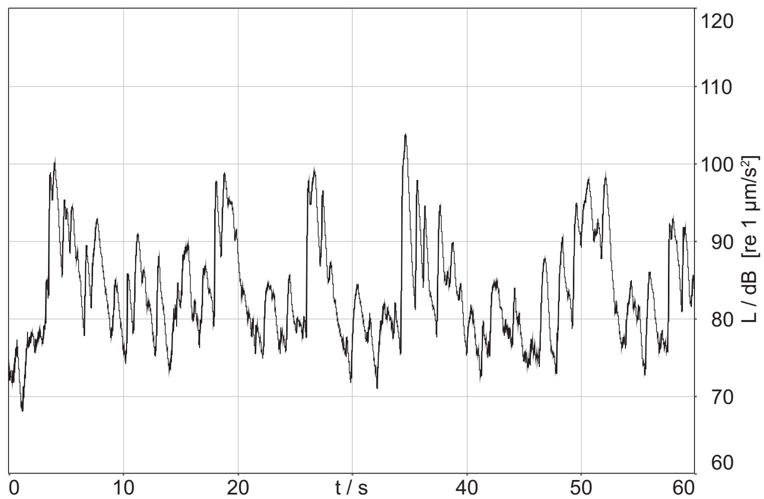


Fig. C.3. Acceleration level vs. time plot for the first 60 s of the DVORAK sequence, low-passed at 100 Hz. The plotted level corresponds to the preferred intensity averaged over all of the participants.

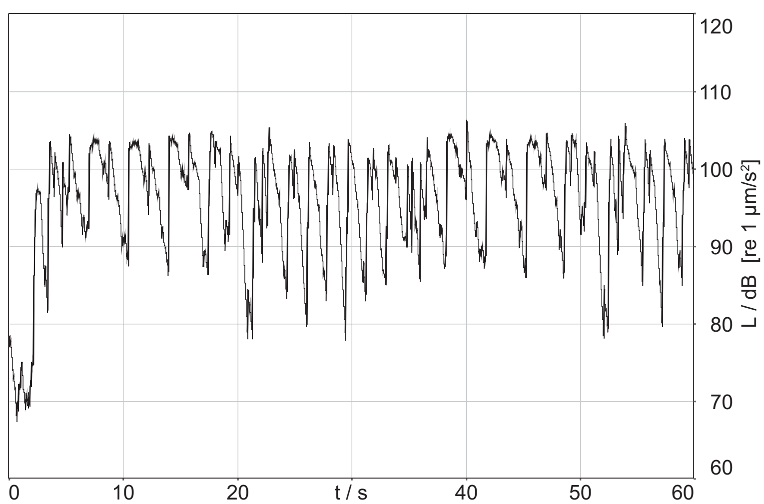


Fig. C.4. Acceleration level vs. time plots for the first 60 s of the BMG sequence, low-passed at 100 Hz. The plotted level corresponds to the preferred intensity averaged over all of the participants.

D

Bandwidth Reduction with Substitute Signal Approach

In this appendix, the reduction of vibration bandwidth, using the substitute signal approach, is illustrated. The BMG sequence is used as an example. The resonant BRTF of a low-cost inertial shaker is shown for comparison.

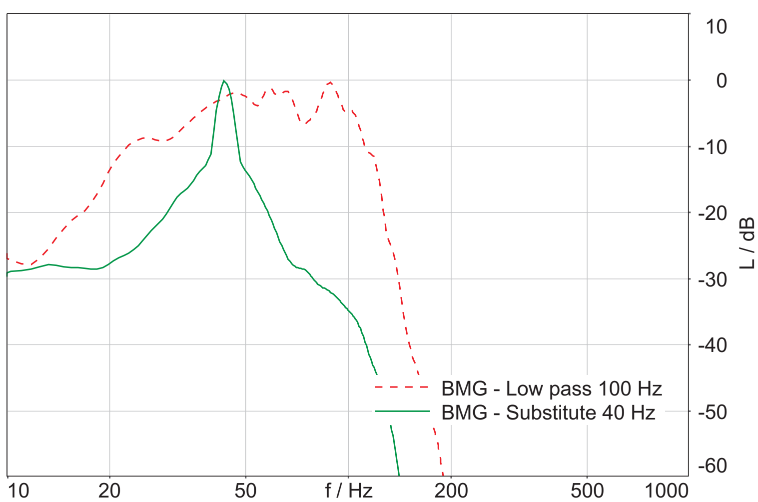


Fig. D.1. Averaged spectrum of the vibration signal for the BMG sequence (FFT 65536, 1/24th octave intensity averaging). The dotted line resulted from filtering with the low-pass 100 Hz approach. The solid line corresponds to the 40 Hz sinusoid, which was amplitude modulated with the envelope of the original signal.

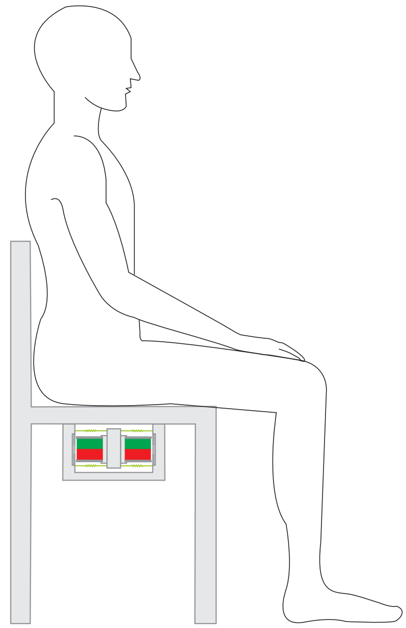


Fig. D.2. Low-cost vibration chair with inertial electrodynamic exciter.

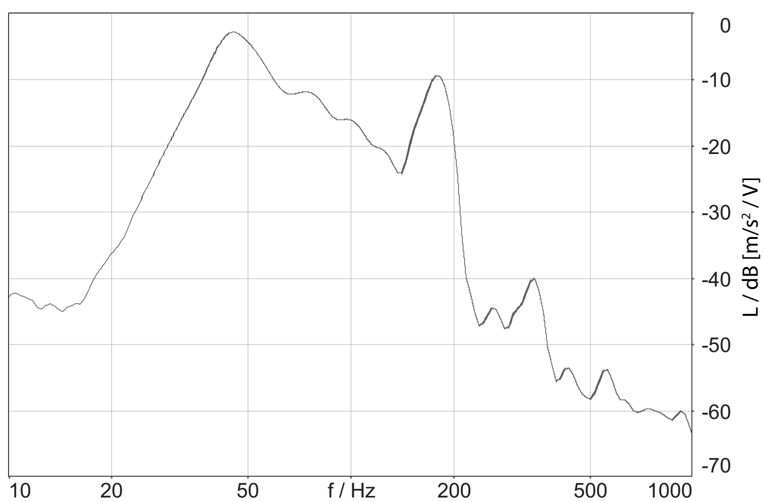


Fig. D.3. Body-related transfer function of a low-cost inertial shaker (FFT 65536, 1/24th octave intensity averaging) at the seat surface in the vertical direction. A Monacor BR-25 was mounted below a wooden chair for this measurement. The transfer function exhibits a distinct resonance at approximately 40 Hz and steep roll-off below the resonance frequency.

List of Abbreviations

ANOVA	analysis of variance
BACH	Sequence Bach: Toccata in D minor
BMG	Sequence Blue Man Group: The Complex, Sing Along
BRTF	body-related transfer function
DVD	digital versatile disc
DVORAK	Sequence Dvořák: Slavonic Dance No. 2 in E minor, op. 72
FA	fast adapting
FFT	fast Fourier transform
JNDF	just-noticeable difference in frequency
JNDL	just-noticeable difference in level
KS-test	Kolmogorov-Smirnov test
LFE	low-frequency effects
MAA	minimum audible angle
Pd	Pure Data
PSE	point of subjective equality
PSS	point of subjective simultaneity
RMS	root mean square
SA	slow adapting
SL	sensation level
SPL	sound pressure level
STD	standard deviation
STFT	short-time Fourier transform
TTS	temporary threshold shift
VERDI	Sequence Verdi: Messa Da Requiem, Dies Irae
WBV	whole-body vibration
WGN	white Gaussian noise

References

[1] The frequencies of music. *Stereo Review*, April, 1980.

[2] Abercrombie C. L. and Braasch J. Auralization of audio-tactile stimuli from acoustic and structural measurements. *J. Audio Eng. Society*, 58 (10):818–827, 2010.

[3] Abercrombie C. L. and Braasch J. Perceptual dimensions of stage-floor vibration experienced during a musical performance. In *Proceedings of Audio Eng. Society 129th Conv.*, San Francisco, USA, 2010.

[4] Altinsoy M. E. Perceptual aspects of auditory-tactile asynchrony. In *Proceedings of the Tenth International Congress on Sound and Vibration*, Stockholm, Sweden, 2003.

[5] Altinsoy M. E. *Auditory-tactile interaction in virtual environments* Phd thesis, Shaker Verlag, 2006.

[6] Altinsoy M. E. and Merchel S. Cross-modal frequency matching: sound and whole-body vibration. In *Haptic and Audio Interaction Design*. Springer, Berlin, Germany, 2010.

[7] Altinsoy M. E. and Merchel S. BRTF (body related transfer function) and whole-body vibration reproduction systems. In *Proceedings of Audio Eng. Society 130th Conv.*, London, UK, 2011.

[8] Altinsoy M. E., Jekosch U., Landgraf J., and Merchel S. Progress in auditory perception research laboratories - Multimodal Measurement Laboratory of Dresden University of Technology. In *Proceedings of Audio Eng. Society 129th Conv.*, San Francisco, USA, 2010.

[9] Bellmann M. A. *Wahrnehmnung von Schall und Vibrationen im tieffrequenten Bereich* Diploma thesis, Carl von Ossietzky - University Oldenburg, 1999.

[10] Bellmann M. A. *Perception of whole-body vibrations: From basic experiments to effects of seat and steering-wheel vibrations on the passengers comfort inside vehicles* PhD thesis, Carl von Ossietzky - University Oldenburg, 2002.

[11] Blauert J. *Spatial hearing - The psychophysics of human sound localization* MIT Press, Cambridge, Massachusetts, 1997.

[12] Blue Man Group Records *The Complex Rock Tour Live (DVD)* Warner Music Group Company, 2003.

[13] Brodsky W. Post-exposure effects of music-generated vibration and whole-body acoustic stimulation among symphony orchestra musicians. *Psychology of Music*, 28(98):98–115, 2000.

[14] Burke M. and Thomas K. Use of physioacoustic therapy to reduce pain during physical therapy for total knee replacement patients over age 55. In Wigram T. and Dileo C., editors, *Music vibration and health*. Jeffery Books, New Jersey, 1997.

[15] Caetano G. *Brain mechanisms of audiotactile and audiomotor interactions* Phd thesis, Helsinky University of Technology, 2007.

[16] Cerdá S., Giménez A., and Cibrián R. M. An objective scheme for ranking halls and obtaining criteria for improvements and design. *J. Audio Eng. Society*, 60(6):419–430, 2012.

[17] Cholewiak R. W. Spatial factors in the perceived intensity of vibrotactile patterns. *Sensory Processes*, 3(2):141–156, 1979.

[18] Cohen L. and Lindley S. Studies in vibratory sensibility. *Am. J. Psychol.*, 51:44–63, 1938.

[19] Corso J. Age and sex differences in pure-tone thresholds. *J. Acoust. Soc. Am.*, 31(4):498–507, 1959.

[20] Craig J. C. Difference threshold for intensity of tactile stimuli. *Perception and Psychophysics*, 11(2):150–152, 1972.

[21] Daub M. *Kreuzmodale Korrelation - Zusammenhang zwischen musikalisch produzierten Ganzkörpervibrationen und der auditiven Wahrnehmung* Diploma thesis, Ruhr-University, Bochum, 2003.

[22] Daub M. Audiotactile simultaneity perception of musical-produced whole-body vibrations. In *Proceedings of CFA/DAGA*, Strasbourg, France, 2004.

[23] Dijk E. O., Weffers-Albu A., and de Zeeuw T. A tactile actuation blanket to intensify movie experiences with personalised tactile effects. In *Proceedings of the 3rd International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN)*, Amsterdam, The Netherlands, 2009.

[24] Dijk E. O., Nijholt A., van Erp J. B. F., Kuyper E., and van Wolferen G. Audio-tactile stimuli to improve health and well-being - A preliminary position paper. In *Proceedings of EuroHaptics*, Amsterdam, The Netherlands, 2010.

[25] DIN 55350 *Begriffe der Qualitätssicherung und Statistik, Grundbegriffe der Qualitätssicherung - Teil 11* Beuth, Berlin, 1987.

[26] Dooley G. J. and Moore B. C. J. Duration discrimination of steady and gliding tones : A new method for estimating sensitivity to rate of change. *J. Acoust. Soc. Am.*, 84(4):1332–1337, 1988.

[27] Drake-Lee A. B. Beyond music: auditory temporary threshold shift in rock musicians after a heavy metal concert. *J. R. Soc. Med.*, 85(10): 617–619, 1992.

[28] Egan J. P. and Hake H. W. On the masking pattern of a simple auditory stimulus. *J. Acoust. Soc. Am.*, 22(5):622–630, 1950.

[29] Ehmer R. Masking patterns of tones. *J. Acoust. Soc. Am.*, 31(8):1115–1120, 1959.

[30] Ehmer R. H. Masking by tones vs noise bands. *J. Acoust. Soc. Am.*, 31 (9):1253–1256, 1959.

[31] Ekman G. A simple method for fitting psychophysical power functions. *Journal of Psychology*, 51(2):343–350, 1961.

[32] Elliott L. Backward and forward masking of probe tones of different frequencies. *J. Acoust. Soc. Am.*, 34(8):1116–1117, 1962.

[33] Elliott L. Backward masking: Monotic and dichotic conditions. *J. Acoust. Soc. Am.*, 97(5):38–44, 1962.

[34] Fastl H. Frequency discrimination for pulsed versus modulated tones. *J. Acoust. Soc. Am.*, 63(1):275–277, 1978.

[35] Fastl H. and Hesse A. Frequency discrimination for pure tones at short durations. *Acustica*, 56(1):41–47, 1984.

[36] Fastl H. and Zwicker E. *Psychoacoustics: Facts and models* Springer, Berlin, Germany, 3rd edition, 2007.

[37] Fearn R., Carter P., and Wolfe J. The perception of pitch by users of cochlear implants: possible significance for rate and place theories of pitch. *Acoustics Australia*, 27(2):41–43, 1999.

[38] Fitzgibbons P. J. and Wightman F. L. Gap detection in normal and hearing-impaired listeners. *J. Acoust. Soc. Am.*, 72(3):761–765, 1982.

[39] Fleming D. and Griffin M. A study of the subjective equivalence of noise and whole-body vibration. *Journal Sound Vib.*, 42(4):453–461, 1975.

[40] Florentine M., Buus S., and Mason C. Level discrimination as a function of level for tones from 0.25 to 16 kHz. *J. Acoust. Soc. Am.*, 81(5):1528–1541, 1987.

[41] Florentine M., Fastl H., and Buus S. Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking. *J. Acoust. Soc. Am*, 84(1988):195–203, 1988.

[42] Florentine M., Buus S., and Geng W. Psychometric functions for gap detection in a yes-no procedure. *J. Acoust. Soc. Am.*, 106(6):3512–3520, 1999.

[43] Forrest T. G. and Green D. M. Detection of partially filled gaps in noise and the temporal modulation transfer function. *J. Acoust. Soc. Am.*, 82(6):1933–1943, 1987.

[44] Forta N. G. *Vibration intensity difference thresholds* PhD thesis, University of Southampton, 2009.

[45] Fothergill L. C. and Griffin M. J. The subjective magnitude of whole-body vibration. *Ergonomics*, 20(5):521–533, 1977.

[46] Foxe J. J., Morocz I. A., Murray M. M., Higgins B. A., Javitt D. C., and Schroeder C. E. Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Cognitive Brain Research*, 10(1-2):77–83, 2000.

[47] Foxe J. J., Wylie G. R., Martinez A., Schroeder C. E., Javitt D. C., Guilfoyle D., Ritter W., and Murray M. M. Auditory-somatosensory multisensory processing in auditory association cortex: An fMRI study. *J. Neurophysiol.*, 88(1):540–543, 2002.

[48] Garner W. The effect of frequency and spectrum on temporal integration of energy in the ear. *J. Acoust. Soc. Am.*, 19(5):808–815, 1947.

[49] Gelfand S. A. *Hearing - An introduction to psychological and physiological acoustics* Marcel Dekker, New York, USA, 3rd edition, 1998.

[50] Gelfand S. A. *Essentials of audiology* Thieme Medical Publishers, New York, USA, 2001.

[51] Gescheider G. and Wright J. Effects of vibrotactile adaptation on the perception of stimuli of varied intensity. *J. Exp. Psychol.*, 81(3):449–453, 1969.

[52] Gescheider G. A. Cutaneous sound localization. *J. Exp. Psychol.*, 70: 617–625, 1965.

[53] Gescheider G. A. Auditory and cutaneous temporal resolution of successive brief stimuli. *J. Exp. Psychol.*, 75(4):570–572, 1967.

[54] Gescheider G. A. Evidence in support of the duplex theory of mechanoreception. *Sensory Processes*, 1(1):68–76, 1976.

[55] Gescheider G. A. and Joelson J. M. Vibrotactile temporal summation for threshold and suprathreshold levels of stimulation. *Perception & Psychophysics*, 33(2):156–162, 1983.

[56] Gescheider G. A. and Verrillo R. T. Contralateral enhancement and suppression of vibrotactile sensation. *Perception & Psychophysics*, 32 (1):69–74, 1982.

[57] Gescheider G. A. and Verrillo R. T. Effects of the menstrual cycle on vibrotactile sensitivity. *J. Acoust. Soc. Am.*, 76, 1984.

[58] Gescheider G. A. and Wright J. H. Effects of sensory adaptation on the form of the psychophysical magnitude function for cutaneous vibration. *J. Exp. Psychol.*, 77(2):308–13, 1968.

[59] Gescheider G. A., Verrillo R. T., and van Doren C. L. Prediction of vibrotactile masking functions. *J. Acoust. Soc. Am.*, 72(5):1421–1426, 1982.

[60] Gescheider G. A., Bolanowski S. J., and Verrillo R. T. Vibrotactile masking: Effects of stimulus onset asynchrony and stimulus frequency. *J. Acoust. Soc. Am.*, 85(5):2059–2064, 1989.

[61] Gescheider G. A., Zwislocki J. J., and Rasmussen A. Effects of stimulus duration on the amplitude difference limen for vibrotaction. *J. Acoust. Soc. Am.*, 100(4.1):2312–2319, 1996.

[62] Gescheider G. A., Berryhill M. E., Verrillo R. T., and Bolanowski S. J. Vibrotactile temporal summation: probability summation or neural integration? *Somatosensory & Motor Research*, 16(3):229–242, 1999.

[63] Gescheider G. A., Bolanowski S. J., and Chatterton S. K. Temporal gap detection in tactile channels. *Somatosensory & Motor Research*, 20 (3-4):239–247, 2003.

[64] Gilson R. Vibrotactile masking: Some spatial and temporal aspects. *Perception & Psychophysics*, 5:176–180, 1969.

[65] Goff G. D. Differential discrimination of frequency of cutaneous mechanical vibration. *J. Exp. Psychol.*, 74(2.1):294–299, 1967.

[66] Green D. M. and Forrest T. G. Temporal gaps in noise and sinusoids. *J. Acoust. Soc. Am.*, 86(3):961–970, 1989.

[67] Griffin M. *Handbook of human vibrations* Elsevier Academic Press, London, UK, 1990.

[68] Gunther E. and O'Modhrain S. Cutaneous grooves: Composing for the sense of touch. *Journal of New Music Research*, 32(4):369–381, 2003.

[69] Hahn J. F. Vibrotactile adaptation and recovery measured by two methods. *J. Exp. Psychol.*, 71(5):655–658, 1966.

[70] Hahn J. F. Low-frequency vibrotactile adaptation. *J. Exp. Psychol.*, 78 (4):655–659, 1968.

[71] Harris J. Loudness discrimination. *Journal of Speech & Hearing Disorders - Monograph Supplement Number II*, 1963.

[72] He N. J., Horwitz A. R., Dubno J. R., and Mills J. H. Psychometric functions for gap detection in noise measured from young and aged subjects. *J. Acoust. Soc. Am.*, 106(2):966–978, 1999.

[73] Hellbrück J. and Ellermeier W. *Hören - Physiologie, Psychologie und Pathologie* Hogrefe, Göttingen, Germany, 2nd edition, 2004.

[74] Hellman R. and Zwislocki J. J. Loudness determination at low sound frequencies. *J. Acoust. Soc. Am.*, 43(1):60–64, 1968.

[75] Hellman R., Miśkiewicz A., and Scharf B. Loudness adaptation and excitation patterns: effects of frequency and level. *J. Acoust. Soc. Am.*, 101(4):2176–85, 1997.

[76] Hempstock T. and Saunders D. Cross-modality determination of the subjective growth function for whole-body vertical, sinusoidal, vibration. *J. Sound and Vib.*, 46(2):279–284, 1976.

[77] Hirsh I. J. and Sherrick C. E. Perceived order in different sense modalities. *J. Exp. Psychol.*, 62(5):423–32, 1961.

[78] Hollins M., Goble A. K., Whitsel B. L., and Tommerdahl M. Time course and action spectrum of vibrotactile adaptation. *Somatosensory & Motor Research*, 7(2):205–221, 1990.

[79] Howarth H. and Griffin M. The relative importance of noise and vibration from railways. *Applied Ergonomics*, 21(2):129–134, 1990.

[80] Howarth H. V. C. and Griffin M. J. The frequency dependence of subjective reaction to vertical and horizontal whole-body vibration at low magnitudes. *J. Acoust. Soc. Am.*, 83(4):1406–1413, 1988.

[81] IEC 60651 Sound level meters. *International Electrotechnical Commission*, 1979.

[82] Imbery C., Biberger T., van de Par S., and Weber R. Influence of vibration on physiological and subjective reactions to vibro-acoustic stimuli. In *Proceedings of AIA-DAGA*, Meran, Italy, 2013.

[83] ISO 226:2003 Acoustics - Normal equal-loudness-level contours. *International Organization for Standardization*, 2003.

[84] ISO 2631-2:1989 Evaluation of human exposure to whole-body vibration - Part 2: Continous and shock-induced vibrations in buildings (1 Hz to 80 Hz). *International Organization for Standardization*, 1989.

[85] ISO 2631-2:2003 Evaluation of human exposure to whole-body vibration - Part 2: Vibrations in buildings (1 Hz to 80 Hz). *International Organization for Standardization*, 2003.

[86] ITU-R BS.775-1 Multichannel stereophonic sound system with and without accompanying picture. *International Telecommunication Union*, 1992.

[87] ITU-T P.800 Methods for objective and subjective assessment of quality. *International Telecommunication Union*, 1996.

[88] Jekosch U. *Voice and speech quality perception - Assessment and evaluation* Springer, Berlin, Germany, 2005.

[89] Jesteadt W., Wier C., and Green D. Intensity discrimination as a function of frequency and sensation level. *J. Acoust. Soc. Am.*, 61(1):169–177, 1977.

[90] Jesteadt W., Bacon S. P., and Lehman J. R. Forward masking as a function of frequency, masker level, and signal delay. *J. Acoust. Soc. Am.*, 71(4):950–62, 1982.

[91] Jones A. J. and Saunders D. J. A scale of human reaction to whole body, vertical, sinusoidal vibration. *J. Sound and Vib.*, 35(4):503–520, 1974.

[92] Kaczmarek K. A., Webster J. G., Bach-Rita P., and Tompkins W. J. Electrotactile and vibrotactile displays for sensory substitution systems. *IEEE Transactions on Biomedical Engineering*, 38(1):1–16, 1991.

[93] Karam M., Russo F., and Fels D. Designing the model human cochlea: An ambient crossmodal audio-tactile display. *IEEE Transaction on Haptics*, 2(3):1–10, 2009.

[94] Kaufmann A., Bellmann M., and Weber R. "Cross-Modality-Matching" zwischen Schall- und Vibrationssignalen. In *Proceedings of DAGA - 33th German Annual Conference on Acoustics*, Stuttgart, Germany, 2007.

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

[96] Klemm O. Untersuchungen über die Lokalisation von Schallreizen IV: Über den Einfluss des binauralen Zeitunterschieds auf die Lokalisation. *Archiv für die gesamte Psychologie*, 40:117–145, 1920.

[97] Koppehele M. & G. (Producer) Mirow B. (Director) *Messa da Requiem - Giuseppe Verdi conducted by Placido Domingo (DVD)* Glor Music Production, 2006.

[98] Kraak W. Bericht zur akustischen Erprobung der Semperoper Dresden. *TU Dresden*, 1984.

[99] LaMotte R. H. and Mountcastle V. B. Capacities of humans and monkeys to discriminate vibratory stimuli of different frequency and ampli-

tude: A correlation between neural events and psychological measurements. *Journal of Neurophysiology*, 38(3):539–559, 1975.

[100] Lange S. *Multimodale Einflüsse auf die auditorische Wahrnehmnung* Diploma thesis, RWTH Aachen, 2004.

[101] Löfvenberg J. and Johansson R. S. Regional differences and interindividual variability in sensitivity to vibration in the glabrous skin of the human hand. *Brain Research*, 301(1):65–72, 1984.

[102] Makous J. C., Friedman R. M., and Vierck C. J. A critical band filter in touch. *Journal of Neuroscience*, 15(4):2808–2818, 1995.

[103] Makous J. C., Gescheider G. A., and Bolanowski S. J. Decay in the effect of vibrotactile masking. *J. Acoust. Soc. Am.*, 99(2):1124–1129, 1996.

[104] Marks L. E. Summation of vibrotactile intensity: An analog to auditory critical bands? *Sensory Processes*, 3(2):188–203, 1979.

[105] Marozeau J. and Florentine M. Loudness growth in individual listeners with hearing losses: A review. *J. Acoust. Soc. Am.*, 122(3):EL81–EL87, 2007.

[106] Martens W. L. and Woszczyk W. Perceived synchrony in a bimodal display: Optimal intermodal delay for coordinated auditory and haptic reprodution. In *Proceedings of ICAD*, Sydney, Australia, 2004.

[107] Martens W. L., Sakanashi H., and Woszczyk W. Whole-body vibration associated with low-frequency audio reproduction influences preferred equalization. In *Proceedings of Audio Eng. Society 36th Int. Conf.*, Dearborn, USA, 2009.

[108] Matsumoto Y., Maeda S., and Oji Y. Influence of frequency on difference thresholds for magnitude of vertical sinusoidal whole-body vibration. *Industrial Health*, 40(4):313–319, 2002.

[109] Matsumoto Y., Maeda S., Iwane Y., and Iwata Y. Factors affecting perception thresholds of vertical whole-body vibration in recumbent subjects: Gender and age of subjects, and vibration duration. *J. Sound and Vib.*, 330:1810–1828, 2010.

[110] Matthews K. Music for bodies. *www.musicforbodies.net*, (last accessed 19.2.2013).

[111] Merchel S. and Altinsoy M. E. 5.1 oder 5.2 Surround - Ist Surround taktil erweiterbar? In *Proceedings of DAGA 2008 - 34th German Annual Conference on Acoustics*, Dresden, Germany, 2008.

[112] Merchel S. and Altinsoy M. E. Vibratory and acoustical factors in multimodal reproduction of concert DVDs. In *Haptic and Audio Interaction Design*. Springer, Berlin, Germany, 2009.

[113] Merchel S. and Altinsoy M. E. Cross-modality matching of loudness and perceived intensity of whole-body vibrations. In *Haptic and Audio Interaction Design*. Springer, Berlin, Germany, 2010.

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

[115] Merchel S., Leppin A., and Altinsoy M. E. Hearing with your body: The influence of whole-body vibrations on loudness perception. In *Proceedings of ICSV - 16th International Congress on Sound and Vibration*, Kraków, Poland, 2009.

[116] Merchel S., Altinsoy M. E., and Stamm M. Equal intensity contours for whole-body vibrations compared with vibrations cross-modally matched to isophones. In *Haptic and Audio Interaction Design*. Springer, Berlin, Germany, 2011.

[117] Merchel S., Altinsoy M. E., and Stamm M. Just-noticeable frequency differences for whole-body vibrations. In *Proceedings of Internoise*, Osaka, Japan, 2011.

[118] Merchel S., Schwendicke A., and Altinsoy M. E. Feeling the sound: audio-tactile intensity perception. In *Proceedings of 2nd Polish-German Structured Conference on Acoustics, The 58th Open Seminar on Acoustics*, Jurata, Poland, 2011.

[119] Meyer J. *Acoustics and the performance of music* Springer, Berlin, Germany, 2009.

[120] Miller J. Effects of noise on people. *J. Acoust. Soc. Am.*, 56(3):729–764, 1974.

[121] Miwa T. Evaluation methods for vibration effect, Part 1: Measurements of threshold and equal sensation contours of whole-body for vertical and horizontal vibrations. *Industrial Health*, 5:183–205, 1967.

[122] Miwa T. Evaluation methods for vibration effect, Part 4: Measurement of vibration greatness for whole-body and hand in vertical and horizontal vibration. *Industrial Health*, 6:1–10, 1968.

[123] Miwa T. and Yonekawa Y. Evaluation methods for vibrations. *Applied Acoustics*, 7(2):83–101, 1974.

[124] Møller H. and Pedersen C. S. Hearing at low and infrasonic frequencies. *Noise & Health*, 6(23):37–57, 2004.

[125] Moore B. C. J. Frequency difference limens for short-duration tones. *J. Acoust. Soc. Am.*, 54(3):610–619, 1973.

[126] Moore B. C. J. *An introduction to the psychology of hearing* Academic Press, San Diego, USA, 5th edition, 2003.

[127] Moore B. C. J. and Glasberg B. Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency modulation. *J. Acoust. Soc. Am.*, 86(5):1722–1732, 1989.

[128] Moore B. C. J. and Glasberg B. R. Gap detection with sinusoids and noise in normal, impaired, and electrically stimulated ears. *J. Acoust. Soc. Am.*, 83(3):1093–1101, 1988.

[129] Moore B. C. J., Glasberg B. R., Donaldson E., McPherson T., and Plack C. J. Detection of temporal gaps in sinusoids by normally hearing and hearing-impaired subjects. *J. Acoust. Soc. Am.*, 85(3):1266–1275, 1989.

[130] Moore B. C. J., Peters R. W., and Glasberg B. R. Detection of temporal gaps in sinusoids by elderly subjects with and without hearing loss. *J. Acoust. Soc. Am.*, 92(4.1):1923–1932, 1992.

[131] Moore B. C. J., Peters R. W., and Glasberg B. R. Detection of temporal gaps in sinusoids: effects of frequency and level. *J. Acoust. Soc. Am.*, 93(3):1563–1570, 1993.

[132] Morioka M. and Griffin M. Magnitude-dependence of equivalent comfort contours for fore-and-aft, lateral and vertical whole-body vibration. *J. Sound and Vib.*, 298(3):755–772, 2006.

[133] Morioka M. and Griffin M. Absolute thresholds for the perception of fore-and-aft, lateral, and vertical vibration at the hand, the seat, and the foot. *J. Sound and Vib.*, 314(1-2):357–370, 2008.

[134] Morioka M. and Griffin M. J. Difference thresholds for intensity perception of whole-body vertical vibration: Effect of frequency and magnitude. *J. Acoust. Soc. Am.*, 107(1):620–624, 2000.

[135] Mountcastle V. B. *The sensory hand: Neural mechanisms of somatic sensation* Harvard University Press, Cambridge, USA, 2005.

[136] Navarra J., Soto-Faraco S., and Spence C. Adaptation to audiotactile asynchrony. *Neuroscience Letters*, 413(1):72–76, 2007.

[137] Olson H. F. *Music, physics and engineering* Dover Publications, Mineola, USA, 2nd edition, 1967.

[138] Oxenham A. J. and Moore B. C. J. Modeling the additivity of nonsimultaneous masking. *Hearing Research*, 80(1):105–118, 1994.

[139] Parsons K. and Griffin M. Whole-body vibration perception thresholds. *J. Sound and Vib.*, 121(2):237–258, 1988.

[140] Penner M. J., Leshowitz B., Cudahy E., and Ricard G. Intensity discrimination for pulsed sinusoids of various frequencies. *Perception & Psychophysics*, 15(3):568–570, 1974.

[141] Plack C. J., Oxenham A. J., Fay R. R., and Popper A. N. *Pitch: Neural coding and perception* Springer, Berlin, Germany, 2005.

[142] Plomp R. Rate of decay of auditory sensation. *J. Acoust. Soc. Am.*, 36 (2):277–282, 1964.

[143] Plomp R. and Bouman M. Relation between hearing threshold and duration for tone pulses. *J. Acoust. Soc. Am.*, 31(6):749–758, 1959.

[144] Poulton E. C. *Bias in quantifying judgments* Psychology Press, London, UK, 1989.

[145] Reed C. M., Durlach N. I., and Delhorne L. A. Historical overview of tactile aid research. In *Proceedings of the Second International Conference on Tactile Aids, Hearing Aids and Cochlear Implants*, Stockholm, Sweden, 1992.

[146] Rosenblith W. A. and Stevens K. N. On the DL for frequency. *J. Acoust. Soc. Am.*, 25(5):980–985, 1953.

[147] Rothenberg M., Verrillo R. T., Zahorian S. A., Brachman M. L., and Bolanowski S. J. J. Vibrotactile frequency for encoding a speech parameter. *J. Acoust. Soc. Am.*, 66(4):1003–1012, 1977.

[148] Rudmose W. The case of the missing 6 dB. *J. Acoust. Soc. Am.*, 71(3): 650–659, 1982.

[149] Sandover J. Some effects of a combined noise and vibration environment on a mental arithmetic task. *J. Sound and Vib.*, 95(2):203–212, 1984.

[150] Schmidt R. F. and Lang F. *Physiologie des Menschen mit Pathophysiologie* Springer, Berlin, Germany, 30th edition, 2007.

[151] Schurmann M., Caetano G., Jousmaki V., and Hari R. Hands help hearing: Facilitatory audiotactile interaction at low sound-intensity levels. *J. Acoust. Soc. Am.*, 115:830–832, 2004.

[152] Shailer M. J. and Moore B. C. J. Gap detection as a function of frequency, bandwidth, and level. *J. Acoust. Soc. Am.*, 74(2):467–473, 1983.

[153] Shailer M. J. and Moore B. C. J. Gap detection and the auditory filter: phase effects using sinusoidal stimuli. *J. Acoust. Soc. Am.*, 81(4):1110–1117, 1987.

[154] Simon G., Olive S., and Welti T. The effect of whole-body vibrations on preferred bass equalization of automotive audio systems. In *Proceedings of Audio Eng. Society 127th Conv.*, New York, USA, 2009.

[155] Spoor A. Presbycusis values in relation to noise induced hearing loss. *International Journal of Audiology*, 6(1):48–57, 1967.

[156] Stamm M., Altinsoy M. E., and Merchel S. Frequenzwahrnehmung von Ganzkörperschwingungen im Vergleich zur auditiven Wahrnehmung I. In *Proceedings of DAGA 2010 - 36th German Annual Conference on Acoustics*, Berlin, Germany, 2010.

[157] Stein B. E. and Meredith M. A. *The merging of the senses* MIT Press, Cambridge, USA, 1993.

[158] Stevens S. A scale for the measurement of a psychological magnitude: loudness. *Psychological Review*, 43(5):405–416, 1936.

[159] Stevens S. S. *Psychophysics: Introduction to its perceptual, neural, and social prospects* Wiley, New York, USA, 1975.

[160] Stuart M., Turman A. B., Shaw J., Walsh N., and Nguyen V. Effects of aging on vibration detection thresholds at various body regions. *BMC Geriatrics*, 3(1):1–10, 2003.

[161] Summers I. R. *Tactile aids for the hearing impaired* Whurr, London, UK, 1992.

[162] Tamura Y., Hiraga M., Usagawa T., and Ebata M. The effect of low frequency vibration on sound quality of the automobile audio system - Proposal of a new automobile audio system. *Society of Automotive Engineers Technical Paper 850165*, 1985.

[163] Tobias J. V. Low frequency masking patterns. *J. Acoust. Soc. Am*, 61 (2):571–575, 1977.

[164] Vallbo A. B. and Johansson R. S. Properties of cutaneous mechanoreceptors in the human hand related to touch sensation. *Human Neurobiology*, 3(1):3–14, 1984.

[165] van de Par S. and Kohlrausch A. Sensitivity to auditory-visual asynchrony and to jitter in auditory-visual timing. In *Proceedings of SPIE : Human Vision and Electronic Imaging V*, San Jose, USA, 2000.

[166] van Doren C. L., Gescheider G. A., and Verrillo R. T. Vibrotactile temporal gap detection as a function of age. *J. Acoust. Soc. Am.*, 87 (5):2201–2206, 1990.

[167] van Eijk R. L. J., Kohlrausch A., Juola J. F., and van de Par S. Audio-visual synchrony and temporal order judgments: Effects of experimental method and stimulus type. *Perception & Psychophysics*, 70(6):955–968, 2008.

[168] VDI 2057 *Human exposure to mechanical vibration: Whole-body vibration* 2002.

[169] Verrillo R. T. Effect of contactor area on the vibrotactile threshold. *J. Acoust. Soc. Am.*, 37:843–846, 1963.

[170] Verrillo R. T. Temporal summation in vibrotactile sensitivity. *J. Acoust. Soc. Am.*, 37(5):843–846, 1965.

[171] Verrillo R. T. Effect of spatial parameters on the vibrotactile threshold. *J. Exp. Psychol.*, 71:570 – 575, 1966.

[172] Verrillo R. T. Vibrotactile thresholds for hairy skin. *J. Exp. Psychol.*, 72:47–50, 1966.

[173] Verrillo R. T. Comparison of vibrotactile threshold and suprathreshold responses in men and women. *Perception & Psychophysics*, 26(1):20–24, 1979.

[174] Verrillo R. T. Change in vibrotactile thresholds as a function of age. *Sensory Processes*, 3:49–59, 1979.

[175] Verrillo R. T. Age related changes in the sensitivity to vibration. *Journal of Gerontology*, 35(2):185–193, 1980.

[176] Verrillo R. T. Bioresponse to vibration. In Havelock D., Kuwano S., and Vorländer M., editors, *Handbook of Signal Processing in Acoustics*, pages 1185–1244. 2009.

[177] Verrillo R. T. and Chamberlain S. C. The effect of neural density and contactor surround on vibrotactile sensation magnitude. *Perception & Psychophysics*, 11:117–120, 1972.

[178] Verrillo R. T. and Gescheider G. A. Enhancement and summation in the perception of two successive vibrotactile stimuli. *Perception & Psychophysics*, 18(2):128–136, 1975.

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

[180] Verrillo R. T., Bolanowski S. J., and Gescheider G. A. Effect of aging on the subjective magnitude of vibration. *Somatosensory & Motor Research*, 19(3):238–244, 2002.

[181] von Békésy G. Sensations on the skin similar to directional hearing, beats, and harmonics of the ear. *J. Acoust. Soc. Am.*, 841:830–841, 1957.

[182] von Békésy G. Neural volleys and the similarity between some sensations produced by tones and by skin vibrations. *J. Acoust. Soc. Am.*, 29(10): 1059–1069, 1957.

[183] von Békésy G. Funneling in the nervous system and its role in loudness and sensation intensity on the skin. *J. Acoust. Soc. Am.*, 30(5):399–412, 1958.

[184] Walker K. and Martens W. L. Perception of audio-generated and custom motion programs in multimedia display of action-oriented DVD films. In *Haptic and Audio Interaction Design*. Springer, Berlin, Germany, 2006.

[185] Walker K., Martens W. L., and Kim S. Perception of simultaneity and detection of asynchrony between audio and structural vibration in multimodal music reproduction. In *Proceedings of Audio Eng. Society 120th Conv.*, Paris, France, 2006.

[186] Ward W. D. Temporary threshold shift in males and females. *J. Acoust. Soc. Am.*, 40(2):478–485, 1966.

[187] Weber E. *On the tactile senses* Psychology Press, London, UK, 2nd edition, 1996.

[188] Wedell C. H. and Cummings, S. B. J. Fatigue of the vibratory sense. *J. Exp. Psychol.*, 22(5):429–438, 1938.

[189] Wegel R. and Lane C. The auditory masking of one sound by another and its probable relation to the dynamics of the inner ear. *Phys. Rev.*, 23:266–285, 1924.

[190] Weinstein S. Intensive and extensive aspects of tactile sensitivity as a function of body part, sex, and laterality. In Kenshalo D. R., editor, *The skin senses*, pages 195–218. Charles C Thomas, 1968.

[191] Weinzierl S. *Handbuch der Audiotechnik* Springer, Berlin, Germany, 2008.

[192] Whittle L., Collins S., and Robinson D. The audibility of low-frequency sounds. *J. Sound and Vib.*, 21(4):431–448, 1972.

[193] Wier C., Jesteadt W., and Green D. Frequency discrimination as a function of frequency and sensation level. *J. Acoust. Soc. Am.*, 61(1): 178–184, 1977.

[194] Winckel F. Nachrichtentechnik unter kybernetischen Aspekten. In *Handbuch für HF- und E-Techniker Bd. 8*. Berlin, Germany, 1969.

[195] Wischmann C. (Director) Smaczny P. and Atteln G. (Producers) *Ton Koopman plays Bach (DVD)* EuroArts Music International, 2000.

[196] Wübbolt G. (Director) Smaczny P. (Producer) *Kurt Masur - Eine Geburtstagsgala* MDR Fernsehen & EuroArts Music International, 2007.

[197] Zwislocki J. J. Theory of temporal auditory summation. *J. Acoust. Soc. Am.*, 3(8):1046–1060, 1960.

[198] Zwislocki J. J. and Sokolich W. G. On loudness enhancement of a tone burst by a preceding tone burst. *Perception & Psychophysics*, 16(1): 87–90, 1974.

Biography

Sebastian Merchel was born in Grimma, Germany, in 1980. He received a Diploma degree in electrical engineering from the Dresden University of Technology, Germany, in 2006. During his studies he visited the Aalborg University in Denmark and participated in the "Master of Science programme in Acoustics". He is now working as a research associate at the Institute of Acoustics and Speech Communication, Chair of Communication Acoustics (TU Dresden). His research interests include multimodal user interfaces, audio-tactile perception, spatial audio and digital signal processing. Moreover he is interested in spatial audio and academic knowledge transfer. Sebastian Merchel is member of the Audio Engineering Society, DEGA German Acoustical Society and VDE Association for Electrical, Electronic & Information Technologies.

Kurzfassung

Schall und Vibrationen werden oft zeitgleich sowohl auditiv als auch taktil wahrgenommen. Typische Beispiele hierfür sind Autofahrten oder Rockkonzerte. Sogar in einem Konzertsaal oder einer Kirche können überschwellige Boden- oder Sitzschwingungen durch Schall erzeugt werden. Werden Konzertaufnahmen über Kopfhörer wiedergegeben, fehlen diese vibratorischen Informationen bisher. Das Gleiche gilt in vielen Fällen für die Musikwiedergabe über Multimedia- oder HiFi-Systeme.

Ziel der vorliegenden Arbeit ist das bessere Verständnis der gekoppelten Wahrnehmung von Schall und Vibrationen am Beispiel der auditiv-taktilen Musikwahrnehmung. Zu Beginn werden dazu beide Modalitäten miteinander verglichen. Da das bisherige Wissen über die Wahrnehmung von Ganzkörperschwingungen bei niedrigen Vibrationspegeln sehr beschränkt ist, wird in verschiedenen Experimenten die menschlichen Frequenzauflösung und Intensitätswahrnehmung untersucht. Der offensichtlichste Unterschied zwischen beiden Modalitäten ist die deutlich eingeschränktere Fähigkeit mit Hilfe des Tastsinns verschiedene Frequenzen zu unterscheiden. Ein weiterer wichtiger Unterschied ist der steilere Anstieg der wahrgenommen Intensität in der taktilen Modalität. Zur proportionalen Abbildung der menschlichen Vibrationswahrnehmung wird eine neue perzeptiv motivierte Größe definiert: die wahrgenommene Vibrationsintensität M. Sie ist vergleichbar mit der Lautheit N. Zusätzlich werden intermodale Effekte berücksichtigt, z.B. der Einfluss von Ganzkörperschwingungen auf die Lautheitswahrnehmung.

Der zweite Teil der Arbeit untersucht ob durch Schall verursachte Ganzkörperschwingungen die Qualität eines Konzerterlebnisses beeinflussen. Es wird festgestellt das Vibrationen eine bedeutende Rolle bei der Wahrnehmung von Musik spielen. Verschiedene wahrnehmungsoptimierte Ansätze werden entwickelt und evaluiert um Körperschwingungen aus Musiksequenzen zu generieren. Dabei wird das im ersten Teil gewonnene Grundlagenwissen angewendet. Mit Hilfe der Ergebnisse können Audiowiedergabesysteme oder sogar Konzertsäle verbessert werden.