Chapter 7
Auditory-Tactile Experience of Music

Sebastian Merchel and M. Ercan Altinsoy

Abstract We listen to music not only with our ears. The whole body is present in a concert hall, during a rock event, or while enjoying music reproduction at home. This chapter discusses the influence of audio-induced vibrations at the skin on musical experience. To this end, sound and body vibrations were controlled separately in several psychophysical experiments. The multimodal perception of the resulting concert quality is evaluated, and the effect of frequency, intensity, and temporal variation of the vibration signal is discussed. It is shown that vibrations play a significant role in the perception of music. Amplifying certain vibrations in a concert venue or music reproduction system can improve the music experience. Knowledge about the psychophysical similarities and differences of the auditory and tactile modality help to develop perceptually optimized algorithms to generate music-related vibrations. These vibrations can be reproduced, e.g., using electrodynamic exciters mounted to the floor or seat. It is discussed that frequency shifting and intensity compression are important approaches for vibration generation.

7.1 Introduction

Several chapters in this book discuss the influence of haptic cues provided by instruments to musicians. Usually, the forces and vibrations at the skin are directly excited by a physical contact with the instrument. However, the radiated sound itself can stimulate the surface of the human body too. This is true for musicians and music listeners alike. The main hypothesis to be evaluated in this chapter is that vibrations at the listeners skin might be important for the perception of music. If the vibratory component is missing, the perceived quality might change, e.g., for a concert experience. From another perspective, the perceived quality of a concert hall or a

S. Merchel (✉) · M. E. Altinsoy
Institut für Akustik und Sprachkommunikation, Technische Universität Dresden,
Helmholtzstr. 18, 01069 Dresden, Germany
e-mail: sebastian.merchel@tu-dresden.de

M. E. Altinsoy
e-mail: ercan.altinsoy@tu-dresden.de

conventional audio reproduction system might be improved or impaired 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 seat vibrations, such as those that can be perceived in a classical chamber concert hall. Measurements in an exemplary concert hall and a church confirmed the existence of seat vibrations during real music performances [27]. If a kettledrum is hit or the organ plays a tone, the ground and chair vibrate. The vibratory intensity and frequency spectra are dependent on various factors, e.g., room modes or construction parameters of the floor. Nevertheless, in many cases, the concert listener may not recognize the vibrations as a separate feature because the tactile percept is integrated with the other senses (e.g., vision and hearing) into one multimodal percept. 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—parameters that are of vital importance in determining the quality of concert halls [8].

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.

Two pilot experiments were conducted and described by Merchel et al. [24, 25], who investigated the influence of seat vibrations on the overall quality of the reproduction of concert DVDs. Low-pass-filtered audio signals were used for vibration generation through a shaker mounted to a seat. In many cases, participants preferred when vibrations were present, instead reporting that something was missing if seat vibration was turned off. However, different complaints were reported: It was stated that the high-frequency vibrations were sometimes prickling and therefore unpleasant; several participants 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 (indeed, a side-effect) 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 or they are not powerful enough for the present application.

Our work aims to broaden the 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 described that were developed and evaluated to improve music-driven vibration generation, taking into account the above questions and complaints. The content is based on several papers [3, 27, 28] and the dissertation of the first author with the title 'Auditory-Tactile Music Perception' [23] with kind permission from Shaker Verlag.

7.2 Experimental Design

In this section, the applied music stimuli, the experimental setup, participants, and procedure are described. Different vibration-generation approaches will be discussed and evaluated in the following section.

7.2.1 Stimuli

To represent typical concert situations for both classical and modern music, four sequences were selected from music DVDs [7, 21, 45, 46] that included significant low-frequency content. A stimulus duration of approximately 1.5 min was chosen to ensure that the participants had sufficient time to become familiar with it before providing quality judgments. 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 Fig. 7.1a, which shows a rising and falling succession of notes covering a broad frequency range, as well as steady-state tones with a rich overtone spectrum that dominate the composition. Strong vibrations would be expected in a church for this piece of music [27]. The second sequence, Dies Irae, abbreviated as VERDI, is a dramatic composition for double choir and orchestra. A spectrogram is plotted in Fig. 7.1b: Impulsive fortissimo sections with a concert bass drum, kettledrum, and tutti orchestra alternate quickly with sections 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, and is a calm orchestral piece, dominated by bowed and plucked strings. Contrabasses and cellos continuously generate low frequencies at a low level (see spectrogram in Fig. 7.2a). The fourth sequence, Sing Along, is a typical pop music example 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 Fig. 7.2b. Additionally, a bass line can be easily identified.

To generate a vibration signal from these sequences, the sum was calculated of the low-frequency effects (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) was used for this purpose. During the process, several signal processing parameters were varied: A detailed description of the different approaches is presented in Sect. 7.3.
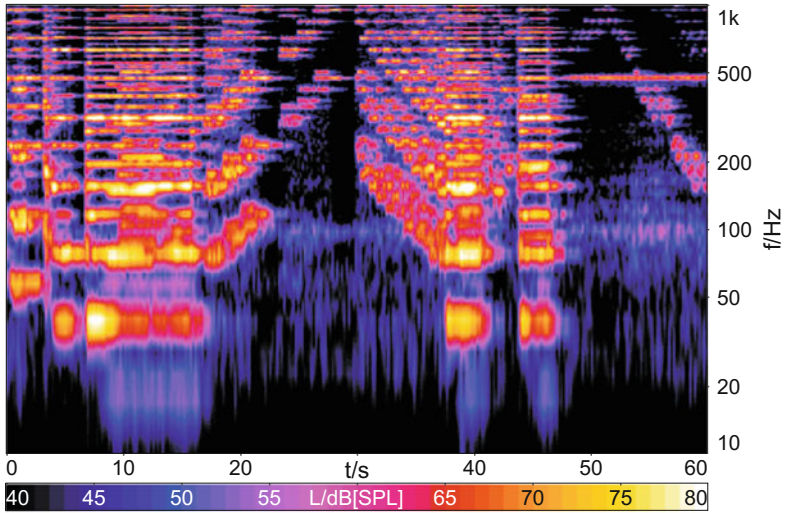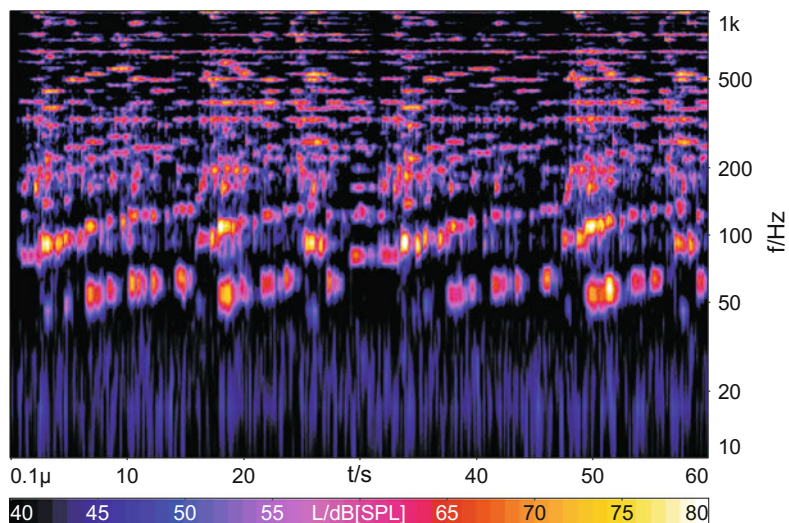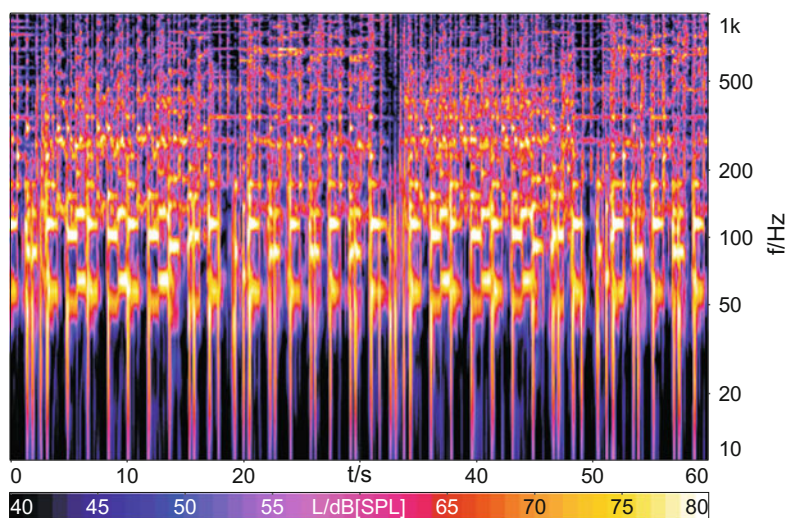
(a) BACH sequence
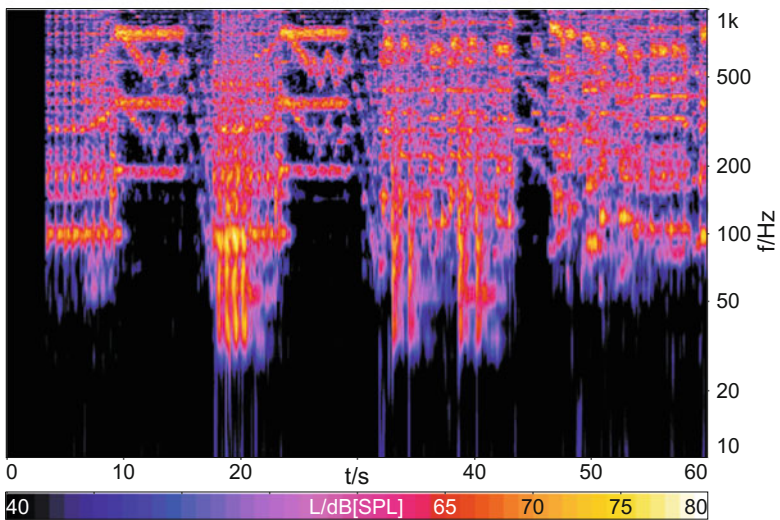
(b) VERDI sequence

Fig. 7.1 Spectrograms of the mono sums for 60 s from the BACH and VERDI sequences. The short-time Fourier transforms (STFTs) were calculated with 8192 samples using 50% overlapping Hann windows

(a) DVORAK sequence

(b) BMG sequence

Fig. 7.2 Spectrograms of the mono sums for 60 s from the DVORAK and BMG sequences. The short-time Fourier transforms (STFTs) were calculated with 8192 samples using 50% overlapping Hann windows

7.2.2 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., [12, 38]). 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 seat vibration), 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 are more difficult to detect than audio advances. Hirsh and Sherrick [17] 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 [1] using broadband noise bursts reproduced via head-phones and broadband vibration bursts at the fingertip: Stimuli with audio delays of approximately 50 to −25 ms were judged to be synchronous, and 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 seat vibrations, audio delays from 63 to −47 ms were found to be synchronous [2]. Using the same setup, audio delays from 79 to −58 ms were judged to be synchronous regarding sound and seat vibrations from a car passing a bump [2].

For musical tones, the PSS 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 [9, 43]. In contrast, PSS values as low as −2 ms for kick drum or −7 ms for piano tones were obtained [43]. Similarly, low PSS values were obtained using impact events reproduced via a vibration platform [22].

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 [37].

7.2.3 Setup

A reproduction system was developed that is capable of separately generating seat vibrations and sound. A surround setup was used, according to ITU-R BS.775-1 [18], 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 participant in a standard multimedia reproduction context, an accompanying movie from the DVD was projected onto a silver screen. The video sequence showed the stage, conductor, or individual instrumentalists while playing.

Vibrations were reproduced using a self-made seat based on an RFT Messelektronik Type 11076 electrodynamic shaker connected to a flat, hard wooden board (46 cm × 46 cm). Seat vibrations were generated vertically, as shown in Fig. 7.3.

The participants were asked to sit on the vibration seat, with both feet flat on the ground. If necessary, wooden plates were placed beneath the participant's feet to adjust for different lengths of legs. The transfer characteristic of the vibrating chair

Fig. 7.3 Vibration chair with electrodynamic exciter
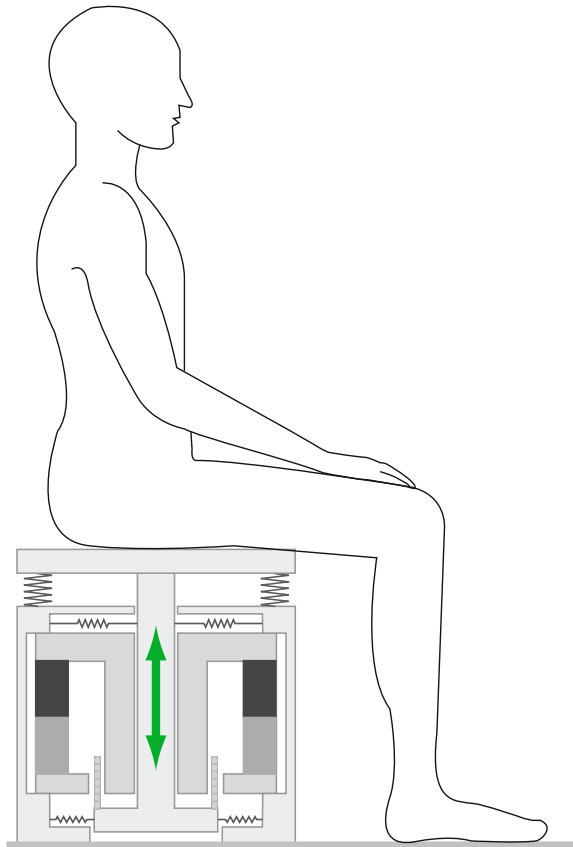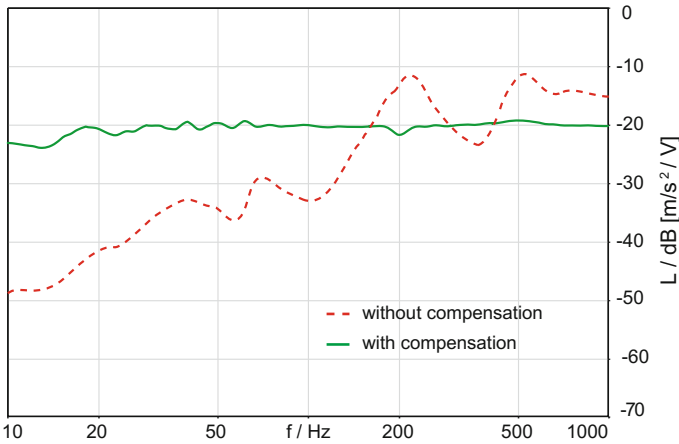
Fig. 7.4 Body-related transfer functions measured at the seat surface of the vibration chair, with and without compensation plotted with 1/24th octave intensity averaging

(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 participants [5]. Considering the just-noticeable difference in thresholds for vertical seat vibrations, which is approximately 1 dB [6, 13, 36], the individual BRTFs should be compensated for during perceptional investigations. The BRTF of each participant was individually monitored and equalized during all experiments. Participants 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 by means of inverse filtering in MATLAB. This procedure resulted in a flat frequency response over a broad frequency range (± 2 dB from 10 to 1000 Hz). An exemplary BRTF, with and without individual compensation, is shown in Fig. 7.4.

7.2.4 Participants

Twenty participants 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 participants were members of bands. The preferred music styles varied, ranging from

rock and pop to classical and jazz. Fifteen participants had not been involved in music-related experiments before, whereas five had already participated in two similar pilot experiments [24, 25].

7.2.5   Procedure

The concert recordings were played back to each participant using the audio setup described above, while vibrations were reproduced using the vibration chair. The vibration intensities were initially adjusted so that the peak acceleration levels reached approximately 100 dB dB (re 10^{-6} m/s^2), which were clearly perceptible. However, perception thresholds can vary heavily between participants [32]; therefore, each participant was asked to adjust the vibration amplitude to the preferred level. This adjustment was typically performed within the first 5–10 s of a sequence. Subsequently, the participant 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 [19]. Figure 7.5 presents the rating scale that was used.

To prevent dissatisfaction, the participants could interrupt the current stimulus as soon as they were confident with their judgment. The required time varied from 30 s to typically no more than 60 s. After rating the overall quality, the participants were encouraged to briefly formulate reasons for their judgments.

Each participant was asked to listen to 84 completely randomized stimuli, 21 for each music sequence. The stimuli were divided into blocks of eight. After each block, the participant 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 participant to complete the experiment.

Before starting the experiment, the participants had to undergo training with three stimuli to become familiar with the task and stimulus variations. The stimuli consisted in the first 90 s 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).

Fig. 7.5 Rating scale for evaluation of the overall quality of the concert experience

Overall Quality

Excellent   Good   Fair   Poor   Bad
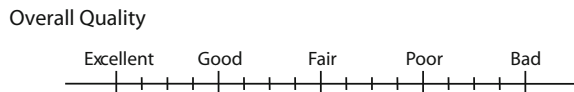
7.3 Vibration Generation: Approaches and Results

Five different approaches to generating vibration stimuli from the audio signal are described in this section. The first four approaches were implemented to modify mainly the frequency content of the signal. The main target was to reduce higher frequencies in order to eliminate tingling sensations and to avoid high-frequency sound radiation. In Sect. 7.3.1 the effect of simple low-pass filtering is evaluated. Reduction of the vibration signal to the fundamental frequency is discussed in Sect. 7.3.2. A frequency shifting algorithm is applied in Sect. 7.3.3, and substitution with artificial vibration signals is discussed in Sect. 7.3.4. In contrast to these frequency-domain algorithms, the last approach (described in Sect. 7.3.5) targets the dynamic range, thus affecting the perceived intensity of the vibration signal.
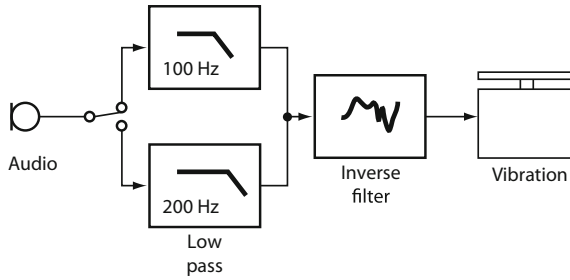
7.3.1 Low-Pass Filtering

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 vibration acceleration measured in real concert venues [27]. However, participants typically chose higher vibration levels in the laboratory, which resulted in significant sound generation from the actuator, especially in the high-frequency range. To address this, the signal was low-pass-filtered using a steep 10th-order Butterworth filter with cutoff frequency set to either 100 or 200 Hz, as illustrated in Fig. 7.6. However, the spurious sound produced by the vibration system could not be completely suppressed. 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 0 to 100, respectively corresponding to 'bad' and '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,[1] which also checks for the homogeneity of variances. The two factors were the played music *sequence* and the applied *treatment*. Averaged results (20 participants) for the overall quality evaluation are plotted in Fig. 7.7 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 < 0.01), 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 Fig. 7.5. The effect seems to be strongest for the BMG pop

---

[1] https://en.wikipedia.org/wiki/SPSS. Last accessed on Nov 10, 2017.

Fig. 7.6 Signal processing chain 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 individually

Fig. 7.7 Mean overall quality evaluation for no-vibration and low-pass filtering vibration-generation approaches, plotted with 95 % confidence intervals.
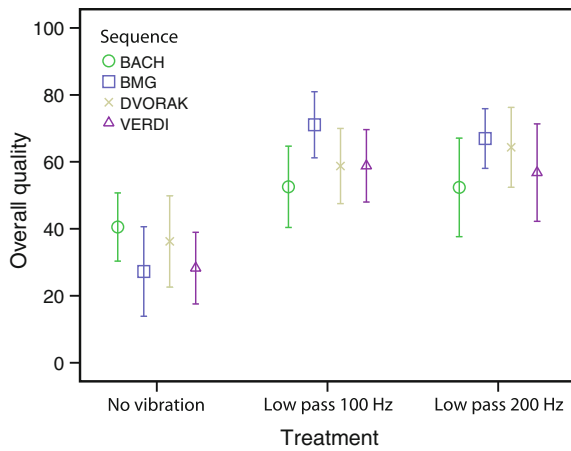
music sequence; however, no significant effects for differences between sequences or interactions between sequences and treatments are observed.

Using the 200 Hz cutoff frequency, the participants occasionally reported tingling sensations on the buttocks or thighs, which only few of them liked. This finding could explain the slightly larger confidence intervals for this treatment.

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 is in agreement with earlier results [25].

7.3.2   Reduction to Fundamental Frequency

In the previous section, low-pass-filtered vibrations were found to be effective for multimodal concert reproduction. However, especially for the low-pass 200 Hz condition, some spurious sound was generated by the vibration system. This fact is particularly critical if the audio signal is reproduced for one person via headphones,

as a second person in the room would be quite disturbed by only hearing the sound generated by the shaker. An attempt was undertaken to further reduce such undesired sound. This goal could be accomplished, e.g., by insulating the vibrating surfaces as much as possible. Because good insulation is difficult to achieve in our case, one effective approach would be to reduce the vibration signal to the fundamental spectral component contained in the signal.

A typical tone generated by an instrument consists of a strong fundamental frequency and several higher-frequency harmonics. If different frequencies are presented simultaneously, strong masking effects toward higher frequencies can be observed in the tactile domain [14, 41]. 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 Fig. 7.8. The fundamentals below 200 Hz of the summed audio signals were tracked using the Fiddle algorithm [39] in Pd, which detects spectral peaks. The cutoff frequency of a first-order low-pass filter was then adaptively adjusted to the lowest frequency peak (i.e., the fundamental). If no fundamental was 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 Fig. 7.9. The statistical analysis was executed in the same manner as in the previous section. Again, the overall quality of the concert experience improved when
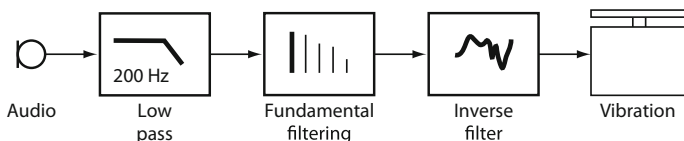
Fig. 7.8 Signal processing chain to generate vibration signals from the audio sum. The fundamental below 200 Hz was tracked, and an adaptive low-pass filter was adjusted to this frequency to suppress all harmonics. If no fundamental was detected, the low-pass filter was set to 100 Hz

Fig. 7.9 Mean overall quality evaluation for no-vibration and the fundamental component vibration approach, plotted with 95 % confidence intervals
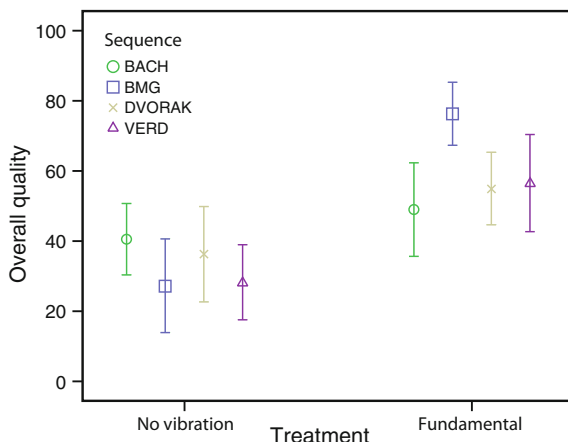
vibrations were added (very significant, $p < 0.01$). At the same time, the generation of high-frequency components could be reduced, except for conditions in which the fundamental frequency approached 200 Hz, e.g., in the VERDI sequence (see Fig. 7.1b). For VERDI and DVORAK, some participants again reported tingling sensations. For BMG and DVORAK, the participants 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, $p < 0.05$). The spectrogram in Fig. 7.2b 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 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 Fig. 7.7).

In addition, Fig. 7.2b 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 Fig. 7.1a. 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 [34]. 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 the intensities appears to be a crucial factor and will be further evaluated in Sect. 7.3.5.

7.3.3 Octave Shift

Another approach would be to shift down the frequency spectrum of the vibration signal. In this manner, the spurious high-frequency sound could be further reduced and the tingling sensation eliminated.
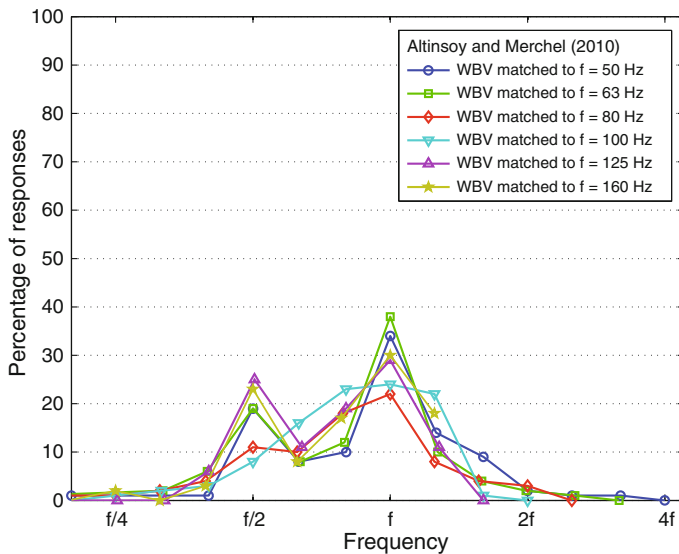
Fig. 7.10 Distribution of crossmodal frequency-matched seat vibrations to acoustical tones with various frequencies f, according to Altinsoy and Merchel [4]

The frequency resolution of the tactile sense is considerably worse than that of audition [31]; therefore, it might be acceptable to strongly compress vibration signals in the frequency domain while still preserving perceptual integration with the respective sound. Earlier experiments have been conducted to test whether participants can match the frequencies of sinusoidal tones and vibrations presented through a seat [4]. The results are summarized in Fig. 7.10. The participants were able to match the frequencies of both modalities with some tolerance. In most cases, the participants 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 down one octave, i.e., dividing their original values by two. This shift corresponds to compression in the frequency range, with stronger compression toward higher frequencies. As shown in Fig. 7.11, before pitch-shifting the original summed audio signal was prefiltered via one of the methods described above (i.e., low-pass filtering or reduction to fundamental frequency). Pitch-shifting was performed in Pd using a granular synthesis approach: The signal was cut into grains of 1000 samples, which were slowed by 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 set at 100 Hz. The resulting low-pass-shifted vibration signals were evaluated as described above. Results are plotted in Fig. 7.12. Again, the statistical analysis was performed using ANOVA after testing the preconditions.

For the BACH sequence, shifting the lowest fundamental even farther down resulted in generally poor-quality ratings. The occasionally weak fundamental components in this sequence caused crossmodal intensity mismatch between vibration
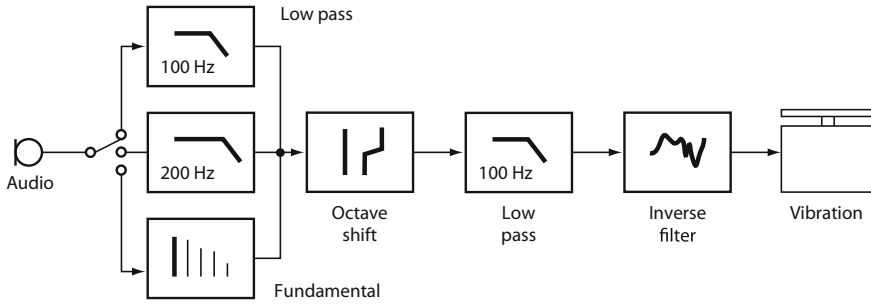
Fig. 7.11 Signal processing chain to generate vibration signals from the audio sum. Compression was applied in the frequency range by shifting all of the frequencies down one octave using granular synthesis. To suppress high-frequency artifacts, a 100 Hz low-pass filter was subsequently inserted
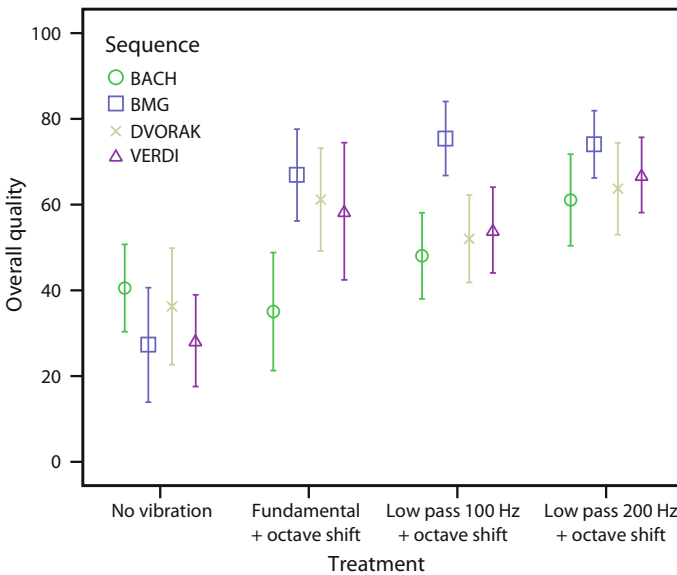
Fig. 7.12 Mean overall quality evaluation for no-vibration and various octave-shift vibration-generation approaches, plotted with 95 % confidence intervals

and sound, which was perceived as louder. However, the perceived quality increases with the bandwidth of the signal, i.e., when using pre-filtering with higher cutoff frequency, 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 participants described the vibrations as 'wavy' or 'bumpy' rather than as 'humming,' as they had previously done. However, many

participants liked the varied vibration character, and the averaged quality ratings did not change significantly compared with Figs. 7.7 and 7.9. No further improvement was found for broader bandwidth of the pre-filtered signal, for the reasons already discussed in the previous section.

Results were significantly different for the DVORAK and VERDI sequences. In Sect. 7.3.1, no preference for one of the two low-pass conditions was observed. However, when these sequences are additionally shifted in frequency, an increase in quality for the 200 Hz low-pass treatment is found, as shown in Fig. 7.12. This could be explained by considering the periods during which the lowest frequency component is greater than 100 Hz (e.g., VERDI second 10–17). By octave-shifting these components while retaining their acceleration levels, they become perceptually more intense due to the decreasing equal-intensity contours for seat vibrations [30]. 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 ($p < 0.01$) 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 ($p < 0.01$). 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 ($p < 0.05$) 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 respective 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.

7.3.4 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 is discussed in this section. Figure 7.13 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 just-noticeable differences (JNDs) for seat vibrations [31]. Additionally, a condition was included using white Gaussian noise (WGN) low-pass-filtered at 100 Hz. These substitute signals were then multiplied with the amplitude envelope of the original low-pass-filtered signal to retain its timing information. An envelope follower was implemented, which calculated the RMS amplitude of the input signal using successive analysis windows. Hann windows

were applied of size equal to 1024 samples, corresponding to approximately 21 ms, to avoid smearing the impulsive content. The period for successive analysis was half of the window size.
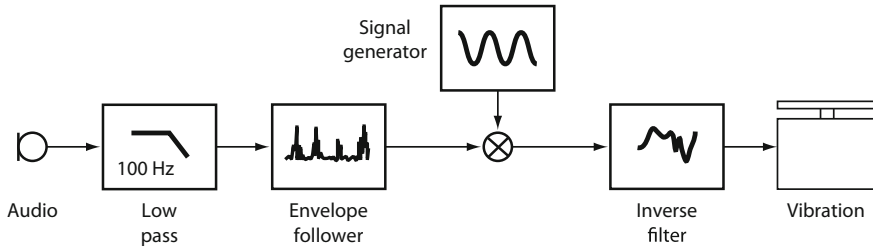
Fig. 7.13 Signal processing chain to generate vibration signals from the audio sum. The envelope of the low-pass-filtered signal was extracted and multiplied with substitute signals, such as sinusoids at 20, 40, 80, and 160 Hz or white noise
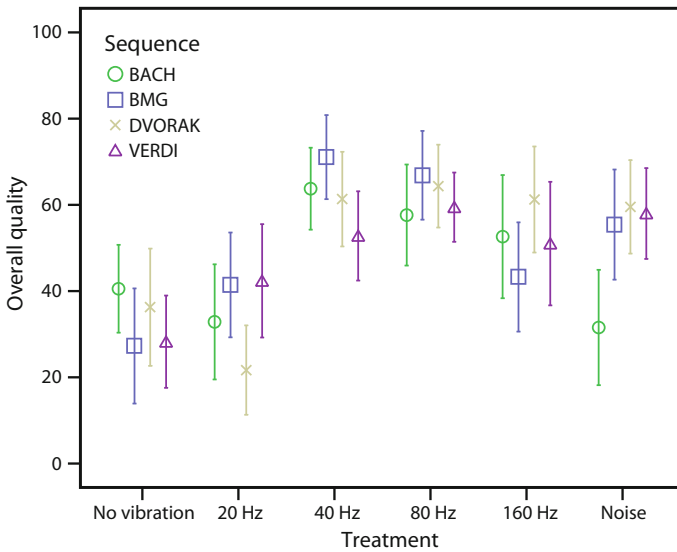
Fig. 7.14 Mean overall quality evaluation for no-vibration and various substitute vibration-generation approaches, plotted with 95 % confidence intervals

The quality scores are presented in Fig. 7.14. 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 < 0.01). 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 participants 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 Fig. 7.7.

Notably, even the 160 Hz vibration resulted in fair-quality ratings. However, compared with the 80 Hz condition, a trend toward worse judgments was observed ($p \approx 0.11$). A much stronger effect was expected because this vibration frequency is relatively high, and tingling effects can occur. There was some disagreement between participants, 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, $p < 0.05$). The effect was strongest for the BACH sequence, which resulted in poor-quality judgments (very significant interaction between sequence and treatment, $p < 0.01$). 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. Specifically, in the BMG sequence 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.

Given these observations, it appears that even simple vibration signals can result in good reproduction quality. For the tested sequences, amplitude-modulated sinusoids at 40 and 80 Hz worked well.

7.3.5 Compression of Dynamic Range

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 kept constant. In the pilot experiments [25], it was reported that expected vibrations were sometimes missing. This might be because of the differing frequency-dependent thresholds and growth of sensations for the auditory and tactile modality [30]. Therefore, an attempt was undertaken to better adapt the signals to the different dynamic ranges.

To better match crossmodally the growth of auditory and tactile sensation with increasing sound and vibration intensity, the music signal is compressed in the vibration-generation process, as illustrated in Fig. 7.15. As one moves toward lower

frequencies, the auditory dynamic range decreases gradually and the growth of sensation with increasing intensity rises more quickly [44]. In the tactile modality, the dynamic range is generally smaller than for audition; however, no strong dependence on frequency between 10 and 200 Hz was found [30]. Accordingly, there was not much variation between frequencies in the growth of sensation of seat vibrations with increasing intensity. Therefore, less compression seems necessary toward lower frequencies. However, a frequency-independent compression algorithm was implemented for simplicity.
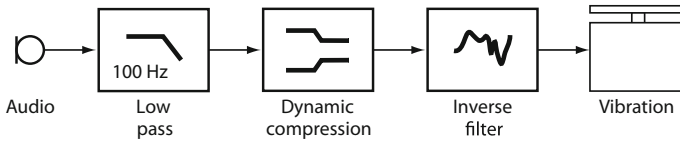
Fig. 7.15 Signal processing chain to generate vibration signals from the audio sum. The low-pass-filtered signal was compressed using different compression factors
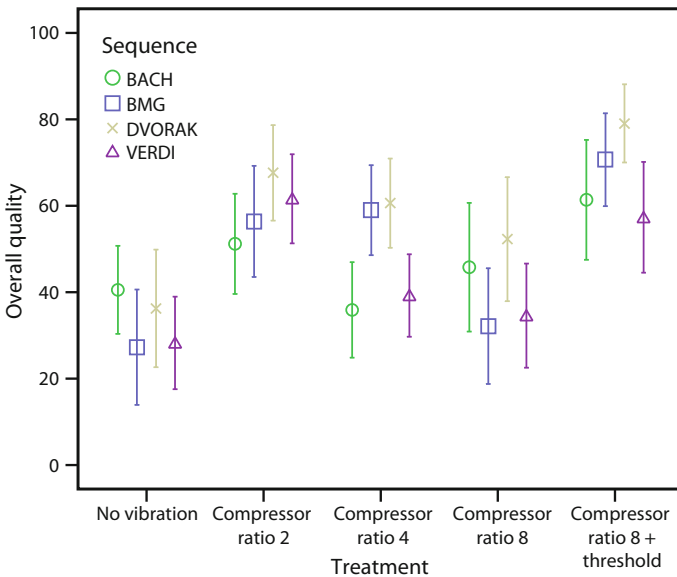
Fig. 7.16 Mean overall quality evaluation for no-vibration and different dynamics compression vibration-generation approaches, plotted with 95 % confidence intervals

The amount of compression needed for ideal intensity matching between both modalities was predicted using crossmodal matching data [26]. 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 [16] and tactile domains [35]. This finding might be important because loud music typically excites weak vibrations. The effect can be accounted for by using higher compression ratios. Therefore, 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 Fig. 7.16. Again, the no-vibration condition was used as a reference. Compressing the audio signal by a ratio of 2 resulted in significantly improved quality perception as compared to the no-vibration condition (average difference = 26, $p < 0.01$). Although the ratings were not statistically better than the 100 Hz low-pass condition in Sect. 7.3.1, 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 4 or 8 reduced the averaged quality scores (average difference between 2 and 4 ratios = 11, $p < 0.05$; average difference between 2 and 8 ratios = 18, $p < 0.01$). 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, such passages are found in BACH and VERDI. This fact would explain the bad ratings for these sequences already with a compression ratio equal to 4. To check this hypothesis, the compression ratio was set to 8, this time using a threshold, and tested again. Loud sounds above the 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 Fig. 7.16. The quality was judged to be significantly better compared with the no-vibration condition (average difference = 34, $p < 0.01$) and with compression ratios of 4 and 8 without a threshold (average difference = 18 and 26, respectively, $p < 0.01$). However, there was no significant difference compared with a compression ratio of 2. These findings indicate that even strong compression might be applied to music-induced vibrations without impairing the perceived quality of a 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. Setting such threshold appears possible if the source signal has a wide dynamic range, 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.

7.3.6 Summary

Various audio-induced vibration-generation approaches have been developed based on fundamental knowledge about auditory and tactile perception. The perceived quality of concert reproduction using combined loudspeaker sound and seat vibrations was evaluated. It can be summarized that seat vibrations can have a considerably positive effect on the experience of music. Since the test participants evaluated all approaches 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 Fig. 7.17 (all judged very significantly better than without vibrations, $p < 0.01$).
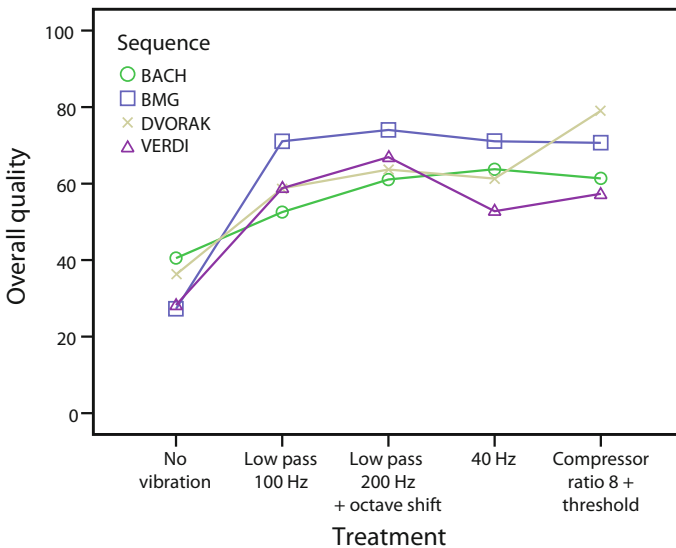
Fig. 7.17 Mean overall quality evaluation for music reproduction using selected vibration-generation approaches. For better illustration, individual data points have been connected with lines

The low-pass filter approach is most similar to vibrations potentially perceived in real concert halls and resulted in good-quality ratings. The approach is not computationally intensive and can be recommended for reproduction systems with limited processing power. Because the differences between a low-pass filter of 100 Hz and 200 Hz were small, the lower cutoff frequency is recommended to minimize sound generation from the vibration system. With additional processing, the unwanted sound can be further reduced while preserving good-quality scores. To this end, one successful approach involves compression in the frequency range, e.g., using octave shifting. Surprisingly, even strong frequency limitation to a simple amplitude-modulated sinusoidal signal seems to be applicable. This allows for much simpler and cheaper vibration reproduction systems, e.g., in home cinema scenarios. However, some signal processing power is necessary, e.g., to extract the envelope of the original signal. Furthermore, it seems useful to apply some dynamic compression, which makes it easier to adjust the vibration level. In this study, source signals with a high dynamic range have been used as a starting point. Further evaluation using audio data whose dynamics are already compressed with unknown parameters is necessary.

Participants usually chose higher acceleration levels in the laboratory compared to measurements in real concert situations. It can be hypothesized that the absolute acceleration level influences the perceived quality of a concert experience. This question should be examined in a further study.

In summary, test participants seemed to be relatively tolerant to 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 expectations are therefore not strictly determined. For example, the intensity of audio-related vibrations might vary heavily between different concert venues. Additionally, various aspects of tactile perception are less refined than for audition. In particular, frequency resolution and pitch perception are strongly restricted [42] for touch, which allows the modification of frequency content within a wide range.

The effect of additional vibration reproduction depended to some extent on the selected music sequence. For example, the BMG rock music sequence was judged significantly better in most of the cases including vibrations than the classical compositions (see Fig. 7.17). This seems plausible because we expect strong audio-induced vibrations at rock concerts. However, adding vibrations seems to clearly increase the perceived concert quality, even for classical pieces of music.

7.4 Conclusions

It has been shown in this chapter that there is a general connection between vibrations and the perceived quality of music reproduction. However, in this study only seat vibrations have been addressed, and a 5.1 surround sound setup was used.

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 test participants preferred generally higher acceleration levels, it is hypothesized that real halls could benefit from amplifying the vibrations in the auditorium. This could be achieved passively, e.g., by manipulating floor construction, or actively using electrodynamic exciters as in the described experiments. Indeed, in future experiments it would be interesting to investigate the effect of additional vibration in a real concert situation. Also, the vibration system could be hidden from participants in order to avoid possible biasing effects.

During the experiments, the test participants sometimes indicated that the vibrations felt like tingling. This effect could be reduced by removing higher frequencies or shifting them down. However, this processing also weakened the perceived tactile intensity of broadband transients. The question arises, what relevance do transients have for the perceived quality of music compared with steady-state vibrations? One approach to reduce the tingling sensations for steady-state tones and simultaneously keep transients unaffected would be to fade continuous vibrations with a long attack and a short release using a compressor. This type of temporal processing appears to be promising based on an unpublished pilot study and should be further evaluated.

Another approach for conveying audio-related vibration 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, tongue, or back using multiple vibration actuators. This frequency-to-place transformation approach is usually applied in the context of tactile hearing aids, in which the tactile channel is used to replace the corrupt auditory perception [20, 40]. However, in such sensory substitution systems, the transformation code needs to be learned. It has been shown in this study that it might not be necessary to code all available auditory information into the tactile channel to improve the perceived quality of music. Still, there is creative potential using this approach, which was applied in several projects [10, 11, 15].

Another interesting effect is the influence of vibrations on loudness perception at low frequencies, the so-called auditory-tactile loudness illusion [33]. It was demonstrated that tones were perceived to be louder when vibrations were reproduced simultaneously via a seat. This illusion can be used to reduce the bass level in a discotheque or an automobile entertainment system [29] and might have an effect on the ideal low-frequency audio equalization in a music reproduction scenario.

References

1. Altinsoy, M.E.: Perceptual aspects of auditory-tactile asynchrony. In: Proceedings of the 10th International Congress on Sound and Vibration, Stockholm, Sweden (2003)
2. Altinsoy, M.E.: Auditory-tactile interaction in virtual environments. Ph.D. thesis, Shaker Verlag (2006)

3. Altinsoy, M.E., Merchel, S.: BRTF–Body related transfer functions for whole-body vibration reproduction systems. In: Proceedings of the NAG/DAGA. Rotterdam, The Netherlands (2009)
4. Altinsoy, M.E., Merchel, S.: Cross-modal frequency matching: sound and whole-body vibration. In: Nordahl R., Serafin S., Fontana F., Brewster S. (eds.), Haptic and Audio Interaction Design (HAID). Lecture Notes in Computer Science, vol. 6306, pp. 37–45. Springer, Berlin, Heidelberg (2010)
5. Altinsoy, M.E., Merchel, S.: BRTF (body related transfer function) and whole-body vibration reproduction systems. In: Proceedings of the Audio Engineering Society 130th Convention, London, UK (2011)
6. Bellmann, M.A.: Perception of whole-body vibrations: from basic experiments to effects of seat and steering-wheel vibrations on the passenger's comfort inside vehicles. Ph.D. thesis, Carl von Ossietzky—University Oldenburg (2002)
7. Blue Man Group Records: The Complex Rock Tour Live (DVD). Warner Music Group Company (2003)
8. Cerdá, S., Giménez, A., Cibrián, R.M.: An objective scheme for ranking halls and obtaining criteria for improvements and design. J. Audio Eng. Soc. **60**(6), 419–430 (2012)
9. Daub, M.: Audiotactile simultaneity perception of musical-produced whole-body vibrations. In: Proceedings of the CFA/DAGA. Strasbourg, France (2004)
10. Dijk, E.O., Nijholt, A., van Erp, J.B.F., Kuyper, E., van Wolferen, G.: Audio-tactile stimuli to improve health and well-being—A preliminary position paper. In: Proceedings of the EuroHaptics, Amsterdam, The Netherlands (2010)
11. Dijk, E.O., Weffers-Albu, A., 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)
12. van Eijk, R.L.J., Kohlrausch, A., Juola, J.F., van de Par, S.: Audiovisual synchrony and temporal order judgments: effects of experimental method and stimulus type. Percept. Psychophys. **70**(6), 955–968 (2008)
13. Forta, N.G.: Vibration intensity difference thresholds. Ph.D. thesis, University of Southampton (2009)
14. Gescheider, G.A., Verrillo, R.T., van Doren, C.L.: Prediction of vibrotactile masking functions. J. Acoust. Soc. Am. **72**(5), 1421–1426 (1982)
15. Gunther, E., O'Modhrain, S.: Cutaneous grooves: composing for the sense of touch. J. New Music Res. **32**(4), 369–381 (2003)
16. Hellman, R., Zwislocki, J.J.: Loudness determination at low sound frequencies. J. Acoust. Soc. Am. **43**(1), 60–64 (1968)
17. Hirsh, I.J., Sherrick, C.E.: Perceived order in different sense modalities. J. Exp. Psychol. **62**(5), 423–32 (1961)
18. Itu-R, B.S.: 775-1: Multichannel stereophonic sound system with and without accompanying picture. International, Telecommunication Union (1992)
19. Itu-T, P.: 800: Methods for objective and subjective assessment of quality. International Telecommunication Union (1996)
20. Karam, M., Russo, F., Fels, D.: Designing the model human cochlea: an ambient crossmodal audio-tactile display. IEEE Trans. Haptics **2**(3), 1–10 (2009)
21. Koppehele, M. & G. (Producer) Mirow B. (Director): Messa da Requiem—Giuseppe Verdi conducted by Placido Domingo (DVD). Glor Music Production (2006)
22. Martens, W.L., Woszczyk, W.: Perceived synchrony in a bimodal display: optimal intermodal delay for coordinated auditory and haptic reproduction. In: Proceedings of the International Conference on Auditory Display (ICAD), Sydney, Australia (2004)
23. Merchel, S.: Auditory-tactile Music Perception. Shaker Verlag, Aachen, Germany (2014)

24. Merchel, S., Altinsoy, M.E.: 5.1 oder 5.2 Surround—Ist Surround taktil erweiterbar? In: Proceedings of the 34th German Annual Conference on Acoustics (DAGA), Dresden, Germany (2008)
25. Merchel, S., Altinsoy, M.E.: Vibratory and acoustical factors in multimodal reproduction of concert DVDs. In: Altinsoy M.E., Jekosch U., Brewster S. (eds.), Haptic and Audio Interaction Design (HAID). Lecture Notes in Computer Science, vol. 5763, pp. 119–127. Springer, Berlin, Heidelberg (2009)
26. Merchel, S., Altinsoy, M.E.: Cross-modality matching of loudness and perceived intensity of whole-body vibrations. In: Nordahl R., Serafin S., Fontana F., Brewster S. (eds) Haptic and Audio Interaction Design (HAID). Lecture Notes in Computer Science, vol. 6306, pp. 1–9. Springer, Berlin, Heidelberg (2010)
27. Merchel, S., Altinsoy, M.E.: Music-induced vibrations in a concert hall and a church. Arch. Acoust. **38**(1), 13–18 (2013)
28. Merchel, S., Altinsoy, M.E.: The influence of vibrations on musical experience. J. Audio Eng. Soc. **62**(4), 1–15 (2014)
29. Merchel, S., Altinsoy, M.E., Kaule, D., Volkmar, C.: Vibro-acoustical sound reproduction in cars. In: Proceedings of the 22nd International Congress on Sound and Vibration, Florence, Italy (2015)
30. Merchel, S., Altinsoy, M.E., Stamm, M.: Equal intensity contours for whole-body vibrations compared with vibrations cross-modally matched to isophones. In: Cooper, E.W., Kryssanov, V.V., Ogawa, H., Brewster, S. (eds.) Haptic and Audio Interaction Design (HAID). Lecture Notes in Computer Science, vol. 6851, 71–80. Springer, Berlin, Heidelberg (2011)
31. Merchel, S., Altinsoy, M.E., Stamm, M.: Just-noticeable frequency differences for whole-body vibrations. In: Proceedings of the Internoise, Osaka, Japan (2011)
32. Merchel, S., Leppin, A., Altinsoy, M.E.: Hearing with your body: the influence of whole-body vibrations on loudness perception. In: Proceedings of the 16th International Congress on Sound and Vibration (ICSV), Kraków, Poland (2009)
33. Merchel, S., Schwendicke, A., Altinsoy, M.E.: Feeling the sound: audio-tactile intensity perception. In: Proceedings of the 2nd Polish-German Structured Conference on Acoustics, The 58th Open Seminar on Acoustics, Jurata, Poland (2011)
34. Meyer, J.: Acoustics and the Performance of Music. Springer, Berlin, Germany (2009)
35. Morioka, M., Griffin, M.: Magnitude-dependence of equivalent comfort contours for fore-and-aft, lateral and vertical whole-body vibration. J. Sound Vib. **298**(3), 755–772 (2006)
36. Morioka, M., 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)
37. Navarra, J., Soto-Faraco, S., Spence, C.: Adaptation to audiotactile asynchrony. Neurosci. Lett. **413**(1), 72–76 (2007)
38. van de Par, S., Kohlrausch, A.: Sensitivity to auditory-visual asynchrony and to jitter in auditory-visual timing. In: Proceedings of the SPIE: Human Vision and Electronic Imaging V. San Jose, USA (2000)
39. Puckette, M.S., Apel, T., Zicarelli, D.D.: Real-time audio analysis tools for Pd and MSP. ICMC **74**, 109–112 (1998)
40. Reed, C.M., Durlach, N.I., 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)
41. Stamm, M., Altinsoy, M.E., Merchel, S.: Frequenzwahrnehmung von Ganzkörperschwingungen im Vergleich zur auditiven Wahrnehmung I. In: Proceedings of the 36th German Annual Conference on Acoustics (DAGA), Berlin, Germany (2010)
42. Summers, I.R.: Tactile Aids for the Hearing Impaired. Whurr, London, UK (1992)
43. Walker, K., Martens, W.L., Kim, S.: Perception of simultaneity and detection of asynchrony between audio and structural vibration in multimodal music reproduction. In: Proceedings of the Audio Engineering Society 120th Convention, Paris, France (2006)
44. Winckel, F.: Nachrichtentechnik unter kybernetischen Aspekten. In: Handbuch für HF- und E-Techniker Bd. 8. Verlag für Radio-, Foto-, Kinotechnik, Berlin, Germany (1969)

45. Wischmann, C. (Director), Smaczny, P., Atteln, G. (Producers): Ton Koopman plays Bach (DVD). EuroArts Music International (2000)
46. Wübbolt, G. (Director), Smaczny, P. (Producer): Kurt Masur - Eine Geburtstagsgala. MDR Fernsehen & EuroArts Music International (2007)