Preprint of: Merchel, S. and Altinsoy, M.E. (2014) 'The Influence of Vibrations on Musical Experience', J. Audio Eng. Soc., Vol. 62, No 4.
Original available: http://www.aes.org/journal/

The Influence of Vibrations on Musical Experience*

SEBASTIAN MERCHEL, *AES Member* **AND M. ERCAN ALTINSOY,** *AES Member*
(sebastian.merchel@tu-dresden.de) (ercan.altinsoy@tu-dresden.de)

Department of Communication Acoustics, Dresden University of Technology, 01062 Dresden, Germany

The coupled perception of sound and vibration is a well-known phenomenon during live pop or organ concerts. However, even during a symphonic concert in a concert hall, sound can excite perceivable vibrations on the surface of the body. This study analyzes the influence of audio-induced vibrations on the perceived quality of the concert experience. Therefore, sound and seat vibrations are controlled separately in an audio reproduction scenario. Because the correlation between sound and vibration is naturally strong, vibrations are generated from audio recordings using various approaches. Different parameters during this process (frequency and intensity modifications) are examined in relation to their perceptual consequences using psychophysical experiments. It can be concluded that vibrations play a significant role in the perception of music.

0 INTRODUCTION

Sound is a mechanical wave that propagates through compressible media such as gases (air-borne sound) or solids (structure-borne sound). The acoustic energy is transported via *vibrating* molecules to the *vibrating* hair cells in the cochlea. However, in this paper the term *vibration* will be used for mechanical stimuli, which excite the surface of the body and stimulate the tactile sense. The term whole-body vibration (WBV) is defined in VDI 2057 [2] as 'mechanical vibrations within the frequency range of 0.1 Hz 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 towards a broader frequency range and a more general excitation area: *Whole-body vibrations are defined as mechanical vibrations, which excite large parts of the body via sound waves or vibrations of a contact surface.*

The main hypothesis to be evaluated in this study is that whole-body vibrations might be important for the perception of music. If the vibratory component is missing, the perceived quality for the concert experience might change. From another perspective, the perceived quality of a concert hall or a 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 whole-body vibrations for a seated person, such as those that can be perceived in a classical chamber concert hall.

Measurements in a concert hall and a church confirmed the existence of whole-body vibrations during music performances [3, 4, 5]. 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 [6].

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

Two pilot experiments were conducted and described in [7, 8], which investigated the influence of WBV on the overall quality of the reproduction of concert DVDs. Low-pass-filtered audio signals were used for vibration generation. It was found that subjects preferred vibrations in many cases. In particular, if the vibration seat was turned off, subjects reported that something was missing. However, during the execution of the pilot experiments, different complaints were noted. Some subjects stated that high-

*This article is a revised and updated version of AES 134th Convention paper 8818 [1].

frequency vibrations caused a prickling sensation that was unpleasant. Several subjects reported that some vibrations were too strong and that others were too weak or were completely missing. It was also noted that the sound generated by the vibration chair at higher frequencies could be disturbing.

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

This work aims at broadening our understanding of the coupled perception of music and vibration by addressing the following questions.

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

In this study, algorithms have been developed to evaluate music-driven vibration generation, taking into account the above questions. To test the different approaches, four music sequences were selected. The experimental setup, subjects, and experimental procedure will be described. Next, different vibration generation approaches will be discussed and evaluated.

1 STIMULI

To represent typical concert situations for both classical and modern music, four sequences were selected from music DVDs [9, 10, 11, 12] that include low-frequency content. A stimulus length of at least 1.5 minutes was chosen to ensure that the participants had sufficient time to become familiar with a stimulus before giving their quality judgement. 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 (*electric bass, percussion, kick drum*)

The first piece, Toccata in D minor (BMV 565), will be referred to as BACH. The full length of the composition was used in the experiments. A spectrogram of the first 60 s is plotted in Figure 1. A rising and falling succession of notes covering a broad frequency range can be seen in the figure. Additionally, steady-state tones dominate the composition. Strong vibrations would be expected in a church for this piece of music [4] containing low-frequency pedal notes.

The second sequence, Dies Irae, will be abbreviated as VERDI. It is a dramatic composition for double choir and orchestra. The excerpt starts with the beginning of the second section of the requiem and was truncated after 1.5 minutes. A partial spectrogram is plotted in Figure 2. Impulsive fortissimo sections with a concert bass drum, a 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 stimuli, Slavonic Dance No. 2 in E minor, will be referred to as DVORAK. The selected sequence includes the first 1.5 minutes of the composition. Again, a spectrogram is plotted in Figure 3. 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

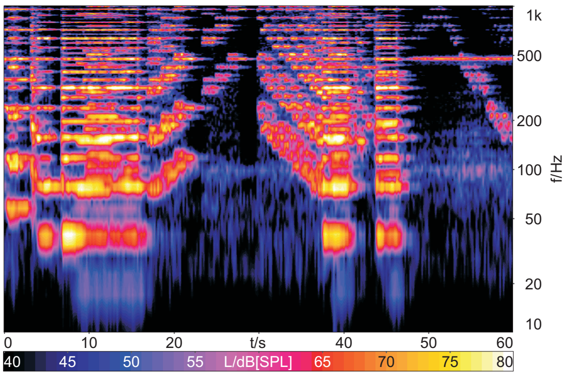

Fig. 1. Spectrogram of the mono sums for 60 s from the BACH sequence. The short-time Fourier transforms (STFTs) were calculated with 8192 samples using 50% overlapping Hanning windows.

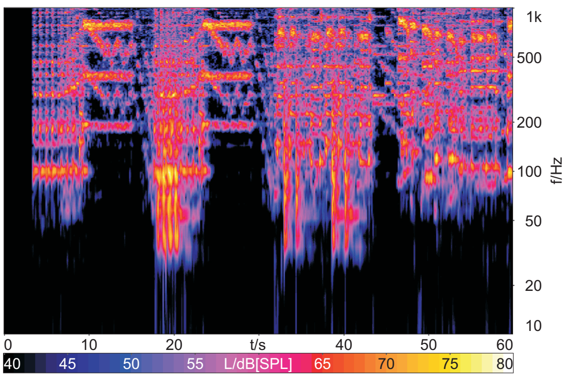

Fig. 2. Spectrogram (STFTs 8192 samples, 50% overlapping Hanning windows) of the mono sums for 60 s from the VERDI sequence.

the Blue Man Group, which will be further shortened to BMG. The sequence starts at 32:24 on the DVD and lasts 1.5 minutes. It is characterized by the heavy use of drums and percussion. These generate transient content at low frequencies, which can be seen in the corresponding spectrogram in Figure 4. Additionally, a bass line can be easily identified.

To generate a vibration signal from these sequences, the sum of the LFE channel and the three respective frontal channels was calculated. No low-frequency content was contained in the surround sound channels in any situation. A real-time audio processing program (Pd - Pure Data) was used to generate stimuli and process the vibration signal. During the process of vibration generation, several signal processing parameters were varied. A detailed description of the different approaches will be presented together with the experimental results in Section 6 (frequency approaches) and Section 7 (intensity approaches). It was decided to present the diverse vibration generation approaches together with the experimental results in order to avoid confusion.

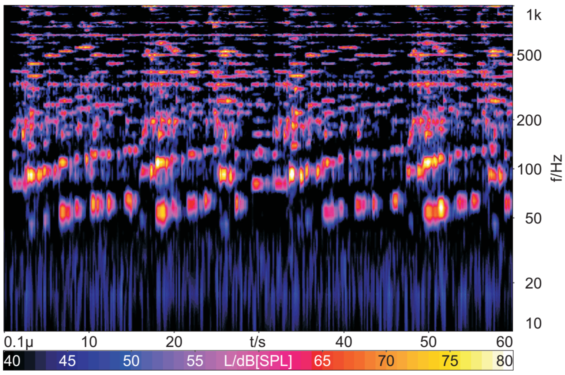

Fig. 3. Spectrogram (STFTs 8192 samples, 50% overlapping Hanning windows) of the mono sums for 60 s from the DVORAK sequence.

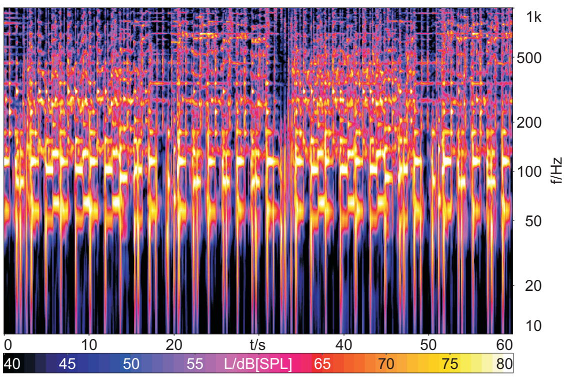

Fig. 4. Spectrogram (STFTs 8192 samples, 50% overlapping Hanning windows) of the mono sums for 60 s from the BMG sequence.

2 SYNCHRONISATION

For a good multisensory concert experience, it is plausible 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., [13]). Few studies have focused on temporal aspects of acoustical and vibratory stimuli. These studies differ in the type of reproduced vibration (vibrations at the hand, forearm, or WBV), type of stimuli (sinusoidal bursts, pulses, noise, instrumental tones, or instrumental sequences), and experimental procedure (time-order judgements or 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 [14] 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 the asynchrony. A similar asymmetry was found by Altinsoy [15] using broadband noise bursts reproduced via headphones and broadband vibration bursts at the fingertip. He 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) is shifted toward an audio delay of approximately 7 ms. It seems that detection thresholds for auditory-tactile asynchrony also depend on the type of stimuli. In an experiment reproducing broadband noise and sinusoidal WBVs, audio delays from 63 ms to -47 ms were found to be synchronous [16]. However, using the same setup, Altinsoy reported that audio delays from 79 ms to -58 ms were judged synchronous for sound and whole-body vibrations from a car passing a bump [16].

For musical tones, the point of subjective simultaneity seems to vary considerably for instruments with different attack or decay times. For example, PSS values as high as -135 ms for a pipe organ or -29 ms for a bowed cello have been reported [17, 18]. In contrast PSS values as low as -2 ms for a kick drum or -7 ms for a piano tone were found [18]. Similarly, low PSS values have been found using impact events reproduced via a vibration platform [19].

It is concluded that auditory-tactile asynchrony detection seems to depend on the reproduced signal. Impulsive content is obviously more prone to delay between modalities. Because music often contains transients, the delay between sound and vibration was set to 0 ms in this study. However, for a real-time implementation of audio-generated vibration reproduction, a slight delay seems to be tolerable or in some cases even advantageous. Additionally, the existence of perceptual adaptation mechanisms, which are able to widen the temporal window for auditory-tactile integration after prolonged exposure to asynchronous stimuli, has been demonstrated [20].

3 SETUP

To investigate the perception of vibrations and sound, it was necessary to develop a reproduction system that is capable of separately generating WBVs and sound. A surround setup was used according to ITU-R BS.775-1 [21] with five Genelec 8040A loudspeakers and a Genelec 7060B subwoofer. All loudspeakers were placed in ear height on a circle with a diameter of 5 m. The system was equalized to a flat frequency response at the listener position in the center of the room. The reverberation time of the reproduction room was below 0.3 s at all frequencies above 100 Hz. To place the subject in a standard multimedia reproduction context, the accompanying picture from the DVD was projected on a silver screen. The video sequence showed the stage, the conductor, or the individual instrumentalists while playing.

To generate whole-body vibrations, usually pneumatic, hydraulic, or electrodynamic systems are used. These systems differ in terms of their usable frequency range, displacement, or load. Pneumatic and hydraulic systems can handle large loads and produce large displacements, but they create some noise and are limited to low frequencies [16]. Currently, new hydraulic systems with wider bandwidths have been developed [22]. For investigating audio-driven vibrations, frequencies from 10 Hz to 200 Hz or more are important. In this frequency range, electrodynamic exciters are advantageous. Such systems have a good frequency response and compact construction. Due to the low manufacturing costs and ease of installation, most consumer products are based on electrodynamic reproduction technology.

This study was conducted using an electrodynamic shaker (self-made based on an RFT Messelektronik Type 11076 with an Alesis RA 150 amplifier). Whole-body vibrations were generated vertically as shown in Figure 5. The system was able to reproduce acceleration levels up to 140 dB. A common measurement unit for vibrations is the acceleration level L_{acc}. It is defined as the logarithmic ratio of the rms acceleration a and a reference value $a_0 = 1\,\mu\mathrm{m\,s}^{-2}$

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

In contrast to sound pressure level, 0 dB acceleration level is not related to the perception threshold, which is above 80 dB for vertical seat vibrations [23].

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 below the subjects feet. The transfer characteristic of the vibrating chair (acceleration at the surface of the seat versus voltage input) was strongly dependent on the individual person. This phenomenon is referred to as the body-related transfer function (BRTF). Differences up to approximately 10 dB have been measured for different subjects. Taking into account the just noticeable difference in thresholds for vertical whole-body vibrations, which are approximately 1 dB [24, 25, 26], the individual BRTFs should be compensated for during perceptual investigations. For a detailed discussion, see [27]. 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 compensated using inverse filters in MATLAB. This resulted in a flat frequency response over a broad frequency range (±2 dB from 10 Hz to 1000 Hz). An exemplary BRTF with and without individual compensation can be seen in Figure 6.

However, although the frequency response in the vertical direction is interesting, the crosstalk to the horizontal vibration axis also has to be taken into account. This crosstalk was minimized using hard, nonslotted disk springs to center the shaker. This 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 can be seen in Figure 7. 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 [28]. In addition, the perception threshold for horizontal seat vibrations is approximately 10 dB higher than for vertical WBVs [23]. There was also crosstalk between the different sys-

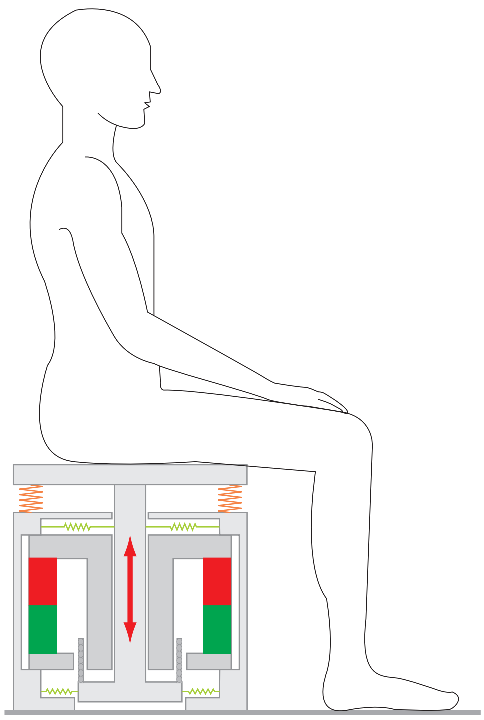

Fig. 5. Vibration chair with electrodynamic exciter.

tems. For example, the subwoofer excited some seat vibrations. They were measured to be below 55 dB acceleration level. Because this is far below the threshold of perception, it was concluded that this type of crosstalk is not critical. Neither the subwoofer nor the shaker generated perceivable vibrations at the feet via the linoleum covered floor.

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

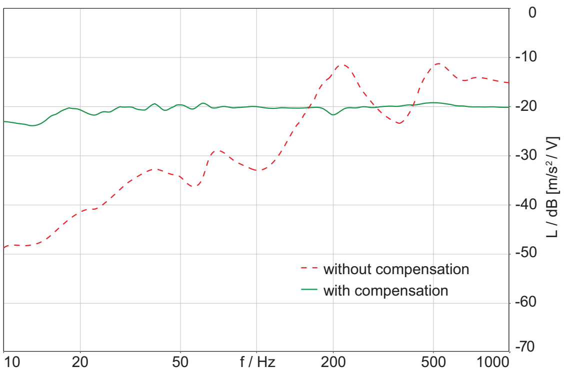

Fig. 6. Body-related transfer functions (FFT 65536, 1/24th octave intensity averaging) measured at the seat surface of the vibration chair with and without compensation.

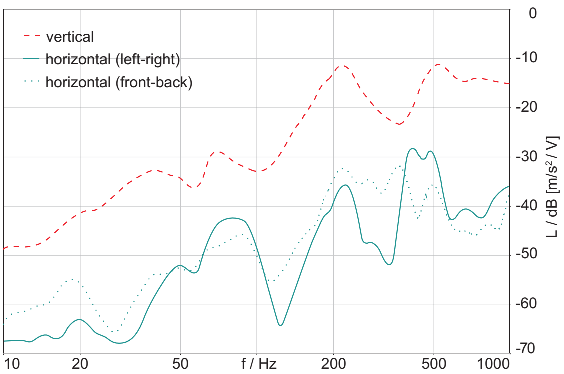

Fig. 7. Crosstalk between vertical and horizontal vibration axes for body-related transfer functions (FFT 65536, 1/24th octave intensity averaging) of the vibration chair.

4 SUBJECTS

Twenty subjects voluntarily participated in this experiment (14 male and 6 female). They were between 20 and 55 years old (mean 24 years) and between 58 and 115 kg (mean 75 kg). All stated that they had no known hearing or spine damage. The preferred music styles varied, ranging from rock and pop to classic and jazz. Fifteen subjects had not been involved in music-related experiments before, while five had already taken part in two similar pilot experiments [7, 8].

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 maximum acceleration levels reached 100 dB and were thus clearly perceptible. However, if such a reproduction system were to be implemented at home, the vibration level could be varied easily. Additionally, the perception threshold can vary heavily between subjects [29]. Therefore, each subject was asked to adjust the vibration amplitude individually to the preferred level. This was usually achieved within the first 5 s to 10 s of a sequence. Subsequent, the subject had to judge the *overall quality of the concert experience* using a quasi-continuous scale. Verbal anchor points from bad to excellent have been added, similar to the method described in ITU-T P.800 [30]. Figure 8 shows the used rating scale. The subject evaluated one stimuli at a time without turning vibrations on and off during the judgment phase. No pairwise comparison method was used.

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

Each subject was asked to listen to 84 completely randomized stimuli, 21 for each music sequence. They were divided into blocks of 8 stimuli. After each block, the subject had the opportunity to relax before continuing with the experiment. Usually, it took approximately 35 minutes to complete 3 to 4 blocks. After 45 minutes at most, the experimental session was interrupted and continued on the next day (and the next, if necessary). Thus, 2-3 sessions were needed 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 the stimuli variations. The used stimuli were

Fig. 8. Rating scale for evaluation of the overall quality of the concert experience.

the first 1.5 minutes 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 BRTF, guided user interface, and collection of data).

In the following sections, different vibration generation approaches will be described together with the perceptual results. Each subsection consists of a motivation for the evaluated approach followed by the corresponding experimental data and a short discussion. During the experiment all approaches were mixed with each other in completely randomized order including a reference condition without vibrations. No information about the different approaches was given to the test participants.

6 FREQUENCY APPROACHES

Four different approaches to generating vibration stimuli from the audio signal will be described in this section. In contrast to the intensity approaches (described in Section 7) the following approaches 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 Section 6.1 the effect of a simple low pass will be evaluated. Reduction of the vibration signal to the fundamental frequency will be discussed in Section 6.2. A frequency shifting algorithm is applied in Section 6.3; and substitution with artificial vibration signals is discussed in Section 6.4.

6.1 LOW PASS APPROACH

The simplest approach would be to directly route the sound (sum of the three frontal channels and the LFE channel) to the vibration actuator. This would, with some deviations, correspond to the approximately linear transfer functions between sound pressure and acceleration measured in real concert venues [4]. However, participants usually 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 Hz and 200 Hz), as illustrated in Figure 9. However, 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.

The corresponding results from the perception experiment will be discussed in the following. For the statistical analysis, the individual quality ratings were interpreted as numbers on a linear scale from 0 to 100, with 0 corresponding to 'bad' and 100 to 'excellent'. Data were checked for a sufficiently normal distribution with the Kolmogorov-Smirnov test (KS-test). A two-factorial analysis of variance (ANOVA) with repeated measures was carried out using IBM SPSS Statistics, which also checks for 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 10 as the mean and 95% confidence intervals. The quality ratings for the concert reproduction *without vibration* are shown as a reference condition on the left.

It can be seen that reproduction with vibration was judged better than reproduction without vibration. Post hoc pairwise comparisons confirmed that both low-pass treatments were judged better than the reference condition at a very significant level ($p<1\%$), both with an average difference of 27 scale units, using Bonferroni correction for multiple testing. This corresponds approximately to one unit on the five point scale shown in Figure 8. The effect seems to be strongest for the pop music sequence BMG; however, no significant effect for differences between sequences or interactions between sequences and treatments could be found.

Using the 200 Hz cutoff frequency, subjects sometimes reported tingling sensations on the buttocks or thighs, which some liked and others did not. This could explain the slightly larger confidence intervals for this treatment.

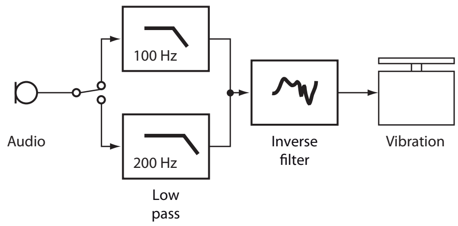

Fig. 9. Signal processing to generate vibration signals from the audio sum. The signal was filtered with a variable low-pass filter, and the BRTF of the vibration chair was individually compensated.

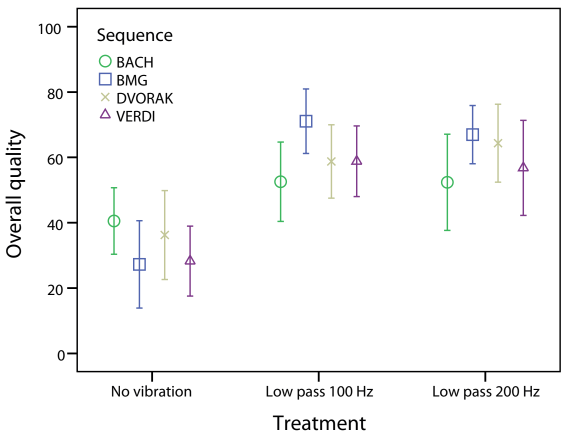

Fig. 10. Mean overall quality evaluation for reproduction using different vibration generation approaches plotted with 95 % confidence intervals. The two low-pass conditions are both judged significantly better than reproduction without vibrations.

The positive effect of reproducing vibrations generated by simple low-pass filtering and the negligible difference between the low-pass frequencies of 100 Hz and 200 Hz is in agreement with earlier results [8].

6.2 FUNDAMENTAL FREQUENCY APPROACH

In the previous section, it was found that low-pass-filtered vibrations are effective for multimodal concert reproduction. However, it would be advantageous to further reduce the sound generated by the vibration system. This is especially 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 made to further reduce the sound generated by the vibration system. This could be done, e.g., by insulating the vibrating surfaces as much as possible. Because this is difficult in this case, it would be advantageous to reduce the sound generated by the shaker by modifying the vibration signal. One effective approach is to reduce the vibration signal to the lowest frequency contained in the signal. This approach will be evaluated in this section.

A typical tone generated by an instrument consists of a strong fundamental frequency and several decreasing harmonics with multiples of the fundamental frequency. If different frequencies are presented simultaneously, strong masking effects toward higher frequencies have been found in the tactile domain [31, 32]. It can be assumed that the fundamental component considerably masks higher frequencies. Therefore, it might be possible to completely remove the harmonics in the vibration generation process, without noticeable effect. This approach is illustrated in Figure 11. The fundamental of the summed audio signal was 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 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 Figure 12. The statistical analysis was executed in the same way 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 2). For VERDI and DVORAK, some subjects again reported tingling sensations.

The average difference for the perceived quality with and without vibrations was 26 scale units. Interestingly, the differences between sequences increased. The strongest effect was found for the BMG sequence compared with the other sequences (significant interaction between treatment and sequence, $p<5\%$). The spectrogram in Figure 4 reveals that for the BMG sequence, the fundamental always lies below 100 Hz and the first harmonic in almost all cases lies above 100 Hz. Therefore, the fundamental filtering, as implemented here, corresponds almost to the low-pass-filtering condition with a cutoff at 100 Hz. As expected, the resulting overall quality is judged similar in both cases (compare with Figure 10).

In addition, Figure 4 reveals that the first harmonic of the electric bass is slightly stronger than the fundamental. However, the intensity balance between fundamental and harmonics is constant over time, resulting in a good match between sound and vibration. This is not the case for the BACH sequence, plotted in Figure 1. The intensity of the lowest-frequency component is high within the first 10 seconds and then suddenly weakens, while the intensities of higher frequencies increase at the same time. If only the lowest frequency is reproduced as vibration, this change in balance between frequencies might result in a mismatch between auditory and tactile perception, which could explain the bad 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 [33]. The fundamental does not necessarily need to be the most intense component or can even be completely missing. Still, the auditory system integrates all harmonics into one tone, in which all partials contribute to the overall intensity. In addition, different simultaneous

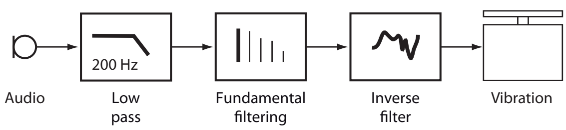

Fig. 11. Signal processing to generate vibration signals from the audio sum. The fundamental below 200 Hz was tracked, and an adaptive low-pass filter was adjusted to this frequency to suppress all harmonics. If no fundamental was detected, the low-pass filter was set to 100 Hz.

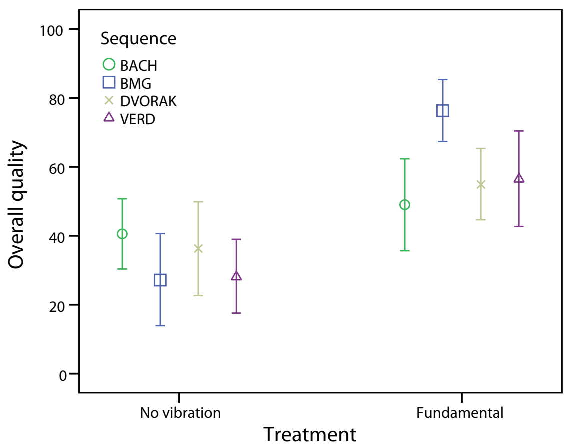

Fig. 12. Mean overall quality evaluation for reproduction using different vibration generation approaches plotted with 95 % confidence intervals.

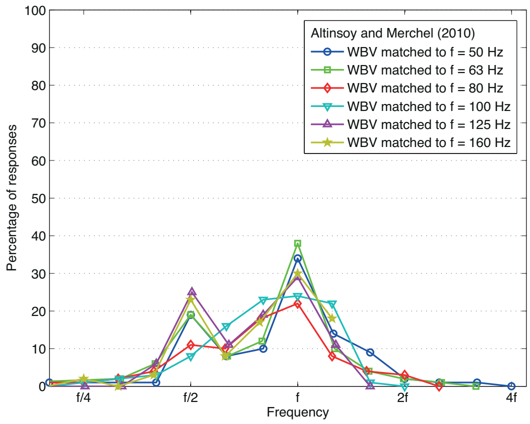

Fig. 13. Distribution of cross-modality frequency-matched whole-body vibrations to acoustical tones with various frequencies f after Altinsoy and Merchel [34].

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 depend on the overall loudness within a specific frequency range. This way, a good match between both modalities might be achieved. However, the processing is complex and would require much computing capacity. Better matching the intensities seems to be a crucial factor and will be further evaluated in Section 7.

6.3 OCTAVE SHIFT APPROACH

Another approach would be to shift down the frequency of all signals. This way, the high-frequency sound generation could be even further reduced. If low-pass-filtered signals were used as a starting point, more partials could be maintained. If the filtered fundamental is used as a source, the tingling sensations could be eliminated.

The frequency resolution of the tactile sense is considerably worse than that of audition [35]. Additional masking effects have been described above. 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 are able to match the frequencies of sinusoidal tones and vibrations presented through a vibration seat [34]. The results are summarized in Figure 13. It can be seen that the subjects are 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 frequencies down one octave relative to their original frequency in this study. This corresponds to compression in the frequency range, as illustrated in Figure 14, with stronger compression toward higher frequencies. A 50 Hz component would become 25 Hz, and a 100 Hz partial, 50 Hz. Again, Pd was used for pitch-shifting using a granular synthesis approach. The signal was cut into grains of 1000 samples, which were slowed down by half and summed again using overlapping Hanning windows. Using this method, some high-frequency artifacts occurred, which were filtered afterward by 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 15. Again, the statistical analysis was performed using an ANOVA after testing the preconditions.

For the BACH sequence, shifting the lowest fundamental even farther down, resulted in poor quality ratings. The sometimes-weak fundamental vibrations in this sequence (as discussed in the last section) still do not cross-modally match the loudness of the sound. Therefore, no improvement in perceived quality can be expected. If more and more partials are added using low-pass filters with increasing bandwidths, the perceived quality increases for BACH. This is 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 dif-

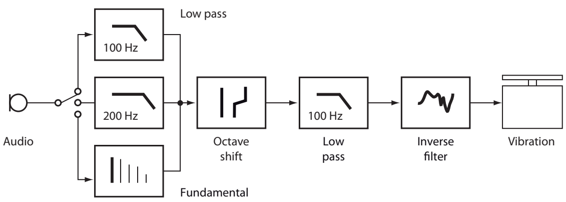

Fig. 14. Signal processing to generate vibration signals from the audio sum. Compression was applied in the frequency range by shifting all frequencies down one octave using granular synthesis. To suppress high-frequency artifacts, a 100 Hz low-pass filter was inserted afterward.

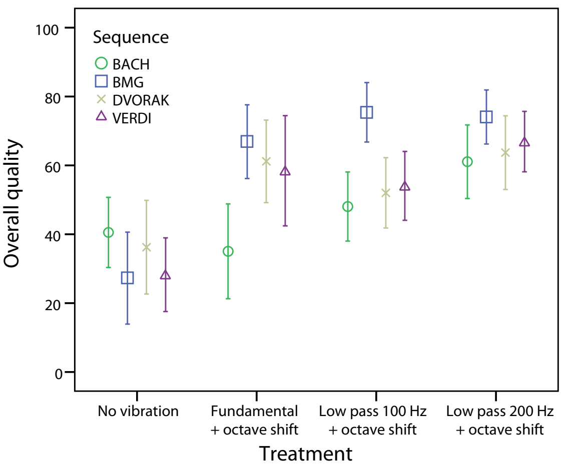

Fig. 15. Mean overall quality evaluation for reproduction using different vibration generation approaches plotted with 95 % confidence intervals.

ference 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 much compared with Figure 10 and 12. Adding more signal content above 100 Hz resulted in no further improvement for BMG for the reasons already discussed in the last section.

The results were different for the DVORAK and VERDI sequences. In Section 6.1, no preference for one of the two low-pass conditions was found. However, if these sequences are shifted in frequency, Figure 15 reveals an increase in quality for the 200 Hz low-pass treatment. This could be explained by considering the periods in which the lowest-frequency component is above 100 Hz (e.g., VERDI second 10 to 17). By octave-shifting these components while retaining their acceleration level, they become perceptually more intense due to the decreasing equal intensity contours for whole-body vibrations [36]. In addition, the vibrations were reported to cause less tingling. The same holds true for octave-shifting the fundamental.

The dependence of the quality scores on the music sequence and 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 treatment conditions were judged better than without vibrations on a very significant level ($p<1\%$). No statistically significant difference between the 'fundamental' and the 'low pass 100 Hz' condition was found. 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) treatment with octave-shifting. As explained above, these main effects have to be interpreted in the context of the differences between sequences.

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

6.4 SUBSTITUTE SIGNALS APPROACH

It was hypothesized in the previous section that the variance of the vibration character that resulted from the frequency shift did not influence the quality scores in a negative way. It might therefore be possible to compress the frequency range even more. This approach was evaluated using several substitute signals and will be discussed in the following section. Figure 16 shows the signal-processing chain. A signal generator was implemented in Pd to produce continuous sinusoidal tones at 20 Hz, 40 Hz, 80 Hz, and 160 Hz. The frequencies were selected to span a broad frequency range and to be clearly distinguishable, taking into account the just noticeable differences in frequency for WBVs [35]. Additionally, a condition using white Gaussian noise (WGN), low-pass-filtered at 100 Hz, was included. These substitute signals have been 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. Hanning windowing was applied, and the window size was set to 1024 samples, which corresponded to approximately 21 ms, to avoid smearing the impulsive signal content. The period for successive analysis was half the window size.

The quality scores using simple vibration signals can be seen in Figure 17. Again, an ANOVA was applied for statistical analysis. It was found that all substitute vibrations, except the 20 Hz condition, were judged better than reproduction without vibration on a very 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. Subjects indicated that the 20 Hz vibration was too low in frequency and did not fit with the audio content. In contrast, 40 Hz and 80 Hz seemed to fit well. No complaints about a mismatch between sound and vibration were noted. The resulting overall quality was judged comparable to the low-pass conditions in Figure 10.

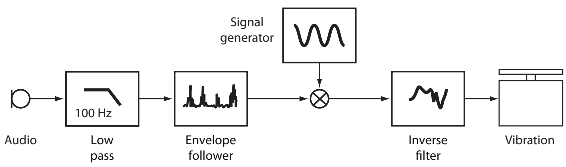

Fig. 16. Signal processing to generate vibration signals from the audio sum. The envelope of the low-pass-filtered signal was extracted and multiplied with substitute signals, such as sinusoids with 20 Hz, 40 Hz, 80 Hz, and 160 Hz or white noise.

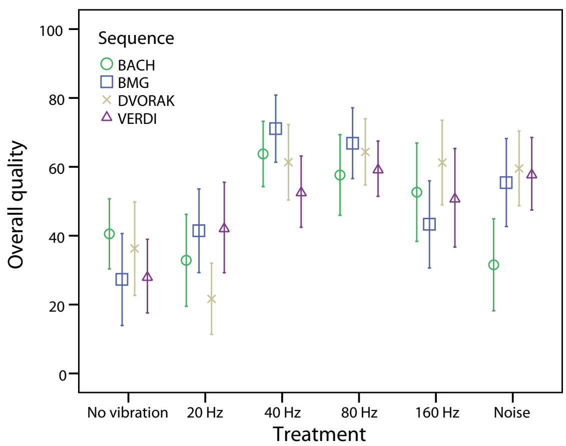

Fig. 17. Mean overall quality evaluation for reproduction using different vibration generation approaches plotted with 95 % confidence intervals.

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

Even more interesting, the reproduction of white Gaussian noise resulted in fair quality ratings. Still, this condition was judged slightly worse than the 40 Hz and 80 Hz vibrations (average difference = 11, $p < 5\%$). The effect is strongest for the BACH sequence, which resulted in poor quality judgements (very significant interaction between sequence and treatment, $p < 1\%$). The BACH sequence contained long tones that lasted for several seconds that 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 did feel less like 'rattling'. Nevertheless, the character of the bursts was different from sinusoidal excitation. Especially for the BMG sequence, the amplitude of the transient vibrations generated by the bass drum varied depending on the random section of the noise. This was most likely one of the reasons why the quality judgement for BMG in the noise condition tended to be worse, e.g., compared 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 Hz and 80 Hz worked well.

7 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 kept constant. In the pilot experiments [8], 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 [36]. Therefore, an attempt was made to better adapt the signals to the different dynamic ranges. In Section 7.1, the influence of dynamic compression will be evaluated, and in Section 7.2, the differences between thresholds will be compensated for.

7.1 COMPRESSION OF DYNAMIC RANGE

To better cross-modally match the auditory and tactile growth of sensation with increasing intensity, the music signal will be compressed in the vibration generation process, as is illustrated in Figure 18. As one moves toward lower frequencies, the auditory dynamic range decreases gradually, and the growth of sensation with increasing intensity rises faster [37]. In the tactile modality, the dynamic range is generally smaller compared with audition; however, no strong dependence on frequency between 10 Hz and 200 Hz was found [36]. Accordingly, there is not much variation between frequencies in the growth of sensation of WBVs with increasing intensity. Therefore, slightly less compression seems necessary toward lower frequencies. However, for the sake of simplicity, a frequency-independent compression algorithm was implemented.

The amount of compression needed for ideal intensity matching between both modalities was predicted using cross-modal matching data [38]. For moderate sinusoidal signals at 50 Hz, 100 Hz, 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 2. Further, the curve of sensation growth versus sensation level flattens toward higher sensation levels in the auditory [39] and in the tactile domain [28]. This might be important because loud music usually excites weak vibrations. The effect can be taken into account by using higher compression ratios. Therefore, altogether, three compression ratios (2, 4, and 8) were selected for testing. Attack and release periods of 5 ms were chosen to quickly follow the source signal.

Statistical analysis was applied as described above using ANOVA with repeated measures and post hoc pairwise comparisons using Bonferroni correction. The quality scores for the concert experience using the three compression ratios are plotted in Figure 19. Again, the no-vibration condition is used as a reference. It can be seen that compressing the audio signal using a ratio of 2 results in sig-

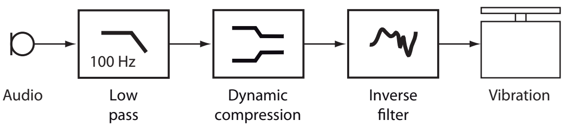

Fig. 18. Signal processing to generate vibration signals from the audio sum. The low-pass-filtered signal was compressed using different compression factors.

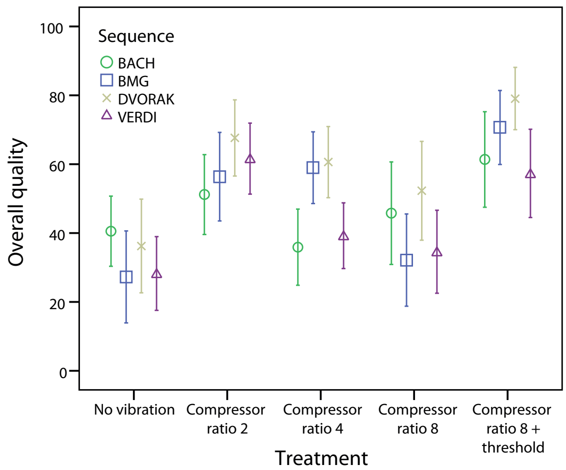

Fig. 19. Mean overall quality evaluation for reproduction using different vibration generation approaches plotted with 95 % confidence intervals.

nificantly improved quality perception compared with the no-vibration condition (average difference = 26, $p<1\%$). However, the results are statistically not better than the 100 Hz low-pass condition in Section 6.1. Still, some test participants reported that the initial level adjustment was easier, especially for the DVORAK sequence. This seems 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 ratio 2 and 4 = 11, $p<5\%$; average difference between ratio 2 and 8 = 18, $p<1\%$). The reason for this decrease in quality seemed to be the noise floor of the audio signal, which was also amplified by the compression algorithm. This vibration noise was mainly noticeable and disturbing during passages of the music with little or no low-frequency content. To check this hypothesis, the compression algorithm with a ratio of 8 was tested again using a threshold. Loud sounds above this threshold were compressed, while quieter sounds remain 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 19. The quality was judged to be significantly better compared with the no-vibration condition (average difference = 34, $p<1\%$) and the compression ratios of 4 and 8 without a threshold (average difference = 18 and 26, respectively, $p<1\%$). However, there was no significant difference compared with the compression ratio of 2. This means that even strong compression might be applied to music-induced vibrations without impairing the perceived quality of the concert experience. In contrast, compression seems to reduce the impression of missing vibrations and therefore makes it easier to adjust the vibration level. However, for strong compression ratios, a suitable threshold needs to be selected. This seems possible if the source signal is uncompressed, which is typically the case for classical recordings. In contrast, modern music or movie sound tracks are sometimes already highly compressed with unknown compression parameters, which could be problematic in terms of vibration generation.

7.2 CORRECTION FOR CROSS-MODAL MATCHING

As discussed earlier, the thresholds and equal-intensity contours for the perception of sound and whole-body vibrations differ. If a sweep is played from 20 Hz 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 decreases toward higher frequencies. Problems with the intensity mismatch between modalities were already discussed in Section 6.2. Therefore, an attempt will be made to match the intensities between both modalities in this section. This is implemented by frequency weighting the low-pass-filtered signal in the processing chain, as illustrated in Figure 20.

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 usually associated with vibrations, it seems reasonable to use a B- or C-weighting for compensation, representing medium- or high-level sounds, respectively [40]. B-weighting was removed from the current measurement standards; however, it will be 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 whole-body vibrations. In the tactile domain, equal-intensity contours seem to be rather parallel [36]. Because the contours are rather flat in terms of acceleration level, no filtering was necessary below 100 Hz. Above this frequency, the increasing contours were approximated by amplifying the signal starting with 0 dB and increasing by

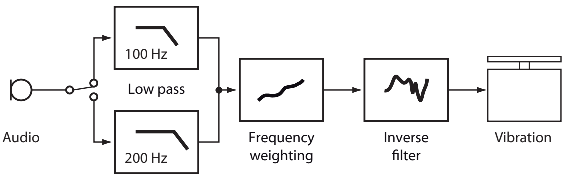

Fig. 20. Signal processing to generate vibration signals from the audio sum. The low-pass-filtered signal was frequency-weighted using the difference between the auditory and the tactile perception threshold.

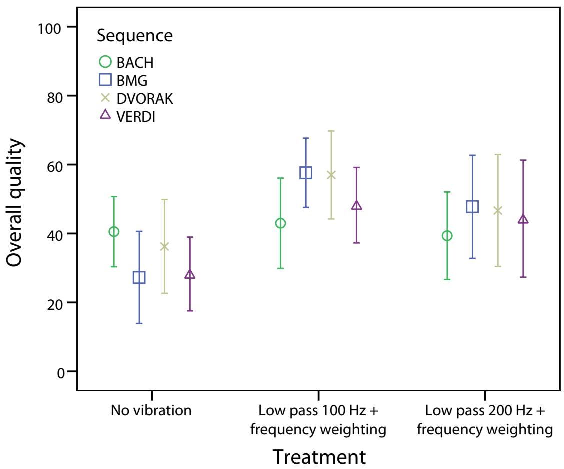

Fig. 21. Mean overall quality evaluation for reproduction using different vibration generation approaches plotted with 95 % confidence intervals.

6 dB per octave toward higher frequencies. In the signal processing chain for vibration generation, both weightings sum up to an overall amplification of higher frequencies.

The quality judgments for these conditions are plotted in Figure 21. It can be seen that reproductions with weighted vibrations are perceived to be better than those without vibrations. This effect is strongest for weighted low-pass-filtered vibrations below 100 Hz (average difference = 18, $p<1\%$). Increasing the low-pass frequency to 200 Hz degrades the resulting quality scores toward the no-vibration condition; however, statistically, only a trend was found (average difference = 11, $p\approx9\%$). This perception of a poor-quality experience can be explained by the tingling sensations that were reported by many test participants. Only two subjects stated that the vibration intensity better matched the audio signal. Because the low-pass conditions without frequency weighting that are plotted in Figure 10 were judged to be better, the frequency weighting implemented here cannot be recommended.

8 SUMMARY AND OUTLOOK

Various audio-induced vibration generation approaches have been developed using fundamental knowledge about auditory and tactile perception. The perceived concert quality using combined sound and vibration reproduction was evaluated. It can be summarized that seat vibrations can have a considerably positive effect on the experience of music. Because the test participants evaluated all approaches 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 22 (all judged very significantly better than without vibrations, $p<1\%$).

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 avoid unwanted sound generation. By adding more processing, the sound generated by the vibration system can be further reduced while preserving the good quality scores for the 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, 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 that 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 the 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, the tactile sense is rather limited compared with audition in some aspects. In particular, the frequency resolution and the perception of pitch are strongly restricted [41], which allows the modification of frequency content within a wide range.

The effect of additional vibration reproduction depended to some extend 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 Figure 22). This seems plausible because we expect strong audio-induced vibrations at rock concerts. However, adding vibrations clearly seems to increase the perceived quality, even for classical pieces of music.

During the experiments, the test participants sometimes indicated that the vibrations felt like tingling sensations. This effect could be reduced by removing higher frequencies or shifting the frequencies down in the frequency range. However, this processing also weakened the perceived tactile intensity of broadband transients. The question arises, what perceptual relevance do transients have for the perceived quality of music perception compared with steady-state vibrations? One approach to reduce the tingling sensations for steady-state tones and simultane-

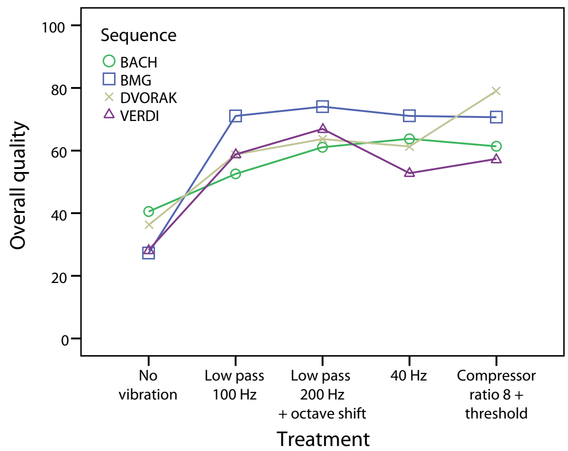

Fig. 22. Mean overall quality evaluation for music reproduction using selected vibration generation approaches. For better illustration, individual data points have been connected with lines.

ously 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 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 [42, 43]. 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 art installations [44, 45, 46] and could be further investigated in a subsequent study.

It has been shown in this paper that there is a general connection between vibrations and the perception of music. However, in this study, only seat vibrations have been 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 test participants preferred generally 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 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 a real concert situation. Additionally, the amplification of vibrations could be hidden from participants in order to avoid possible biasing effects.

9 REFERENCES

[1] S. Merchel and M. E. Altinsoy, “Vibration in Music Perception”, in *Proceedings of Audio Eng. Society 134th Conv.* (Rome, Italy) (2013).

[2] VDI 2057, *Human exposure to mechanical vibration: Whole-body vibration* (2002).

[3] S. Merchel and M. E. Altinsoy, “Der Konzertsaal bebt - Vibroakustische Messungen in der Dresdner Semperoper”, in *Proceedings of DAGA 2012 - 38th German Annual Conference on Acoustics* (Darmstadt, Germany) (2012).

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

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

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

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

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

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

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

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

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

[13] S. van de Par and A. Kohlrausch, “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).

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

[15] M. E. Altinsoy, “Perceptual aspects of auditory-tactile asynchrony”, in *Proceedings of the Tenth International Congress on Sound and Vibration* (Stockholm, Sweden) (2003).

[16] M. E. Altinsoy, “Auditory-tactile interaction in virtual environments”, Phd thesis, Shaker Verlag (2006).

[17] M. Daub, “Audiotactile simultaneity perception of musical-produced whole-body vibrations”, in *Proceedings of CFA/DAGA* (Strasbourg, France) (2004).

[18] K. Walker, W. L. Martens, and S. Kim, “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).

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

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

[21] ITU-R BS.775-1, “Multichannel stereophonic sound system with and without accompanying picture”, International Telecommunication Union (1992).

[22] M. E. Altinsoy, U. Jekosch, J. Landgraf, and S. Merchel, “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).

[23] M. Morioka and M. Griffin, “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**, 357–370 (2008).

[24] M. A. Bellmann, “Perception of whole-body vibrations: From basic experiments to effects of seat and steering-wheel vibrations on the passengers comfort inside vehicles”, Ph.D. thesis, Carl von Ossietzky - University Oldenburg (2002).

[25] N. G. Forta, “Vibration intensity difference thresholds”, Ph.D. thesis, University of Southampton (2009).

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

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

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

[29] S. Merchel, A. Leppin, and M. E. Altinsoy, “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).

[30] ITU-T P.800, “Methods for objective and subjective assessment of quality”, International Telecommunication Union (1996).

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

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

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

[34] M. E. Altinsoy and S. Merchel, “Cross-modal frequency matching: sound and whole-body vibration”, in *Haptic and Audio Interaction Design* (Springer, Berlin, Germany) (2010).

[35] S. Merchel, M. E. Altinsoy, and M. Stamm, “Just-noticeable frequency differences for whole-body vibrations”, in *Proceedings of Internoise* (Osaka, Japan) (2011).

[36] S. Merchel, M. E. Altinsoy, and M. Stamm, “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).

[37] F. Winckel, “Nachrichtentechnik unter kybernetischen Aspekten”, in *Handbuch für HF- und E-Techniker Bd. 8* (Berlin, Germany) (1969).

[38] S. Merchel and M. E. Altinsoy, “Cross-modality matching of loudness and perceived intensity of whole-body vibrations”, in *Haptic and Audio Interaction Design* (Springer, Berlin, Germany) (2010).

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

[40] IEC 60651, “Sound level meters”, International Electrotechnical Commission (1979).

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

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

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

[44] E. O. Dijk, A. Weffers-Albu, and T. de Zeeuw, “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).

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

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

THE AUTHORS

Sebastian Merchel

M. Ercan Altinsoy

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 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.

●

M. Ercan Altinsoy studied mechanical engineering at the Technical University of Istanbul and became a research and teaching assistant at the chair of mechanical vibrations and acoustics. In 2000, he was accepted to the Insitute of Communication Acoustics, Ruhr-University Bochum, Germany as a Ph. D. student. He also participated in the International Graduate School for Neuroscience of the Ruhr-University Bochum. In 2005, he completed his Ph.D. thesis: "Auditory-Tactile Interaction in Virtual Environments". After his Ph.D., Ercan Altinsoy worked at HEAD acoustics as NVH (Noise Vibration Harshness) project engineer. In 2006, he started Lecturing at the Dresden University of Technology. He currently holds the chair of communication acoustics. His research interests include psychoacoustics, vibroacoustics, product sound and vibration design, vehicle acoustics, auditory and haptic interfaces for virtual environments. He is Lothar-Cremer medalist of the Acoustical Society of Germany, DEGA.