# Touch the Sound: Audio-Driven Tactile Feedback for Audio Mixing Applications

*Sebastian Merchel, M. Ercan Altinsoy and Maik Stamm*

Chair of Communication Acoustics, Dresden University of Technology, Germany
sebastian.merchel@tu-dresden.de

## Abstract

In this study, experiments were conducted to determine if a person could distinguish percussive audio loops by their fingertips using audio-driven tactile feedback. The audio signal was adapted to generate a vibration signal (tactile feedback) taking into account the limited capabilities of the tactile modality. A systematic approach to find the different adaptation parameters is discussed. The vibrations were created by an electro-dynamic shaker mounted behind a touch-sensitive screen. Results indicate percussive loops are best distinguished if the source features (e.g., frequency spectrum) and sequence features (e.g., rhythm) are maintained. Also, the pragmatic and hedonic qualities of an interactive, multimodal system using audio-driven tactile feedback were evaluated using a semantic differential.

**Index Terms**: audio-driven vibration feedback, tactile percussion instrument recognition, perceptual quality, tangible user interfaces

## 1. Introduction

A groovebox produces live loop-based music characterized by a high degree of control that allows a user to improvise. The performer can manipulate different audio tracks playing simultaneously using a control surface. Today, touch screens more and more replace the physical buttons, knobs and sliders of traditional grooveboxes. Touch screens allow the interface elements to be easily arranged to fit the user's needs; however, if space and time constraints prevent the interface elements from being labeled (see Figure 1), then the performer must remember all the connections between audio tracks and groovebox channels. Alternatively, he might identify a specific loop by watching its VU meter or by pre-listening to a track using headphones. The latter is not feasible as the auditory modality is already in use during a live musical setup. This paper evaluates a system that uses the tactile modality for audio loop identification using the example of percussion instruments. In our daily life, perceived sounds and vibrations are often coupled; thus, associating a specific vibration with a specific sound might be possible.

There have been attempts incorporating audio-driven tactile feedback with motorized faders. Beamish et al. [2] developed the tangible Q-Slider to control the playback position of a sound track. Anderson et al. [1] controlled the position of a lever using the amplitude envelope of an audio signal. However, these systems were limited by the slow mechanics of the fader. No experimental data exists that examines the ability to distinguish between different audio signals using audio-driven tactile feedback.

## 2. Auditory and Tactile Perception

The auditory and tactile modalities have different capabilities and restrictions (e.g., different frequency and intensity range), which must be considered when generating audio-driven tactile feedback. The perception of sound and vibration is a complex area that has been studied for several decades. For a detailed comparison of both modalities see the recent publication by Merchel et al. [3]. The frequency range of the auditory sense is much larger than for the tactile sense. The ability to discriminate between frequencies for perceived sounds is better compared with perceived vibrations. The auditory system works within a dynamic range up to $100$ dB and can discriminate intensity differences less than $1$ dB. In comparison, the tactile system has a much smaller dynamic range, operating between $35$ dB and $50$ dB. The reported values for vibrotactile intensity discrimination vary from $0.4$ dB to $2.3$ dB [4]. In addition, tactile sensitivity depends strongly on time of exposure, the size of the contact area [5] and the individual subject [6].

However, the auditory and tactile frequency range overlaps up to several hundred Hertz. Similar psychophysical effects have been observed in both modalities (e.g., masking or iso-perception contours). If the audio signal is adapted accordingly, enough information could be transferred through the tactile modality.

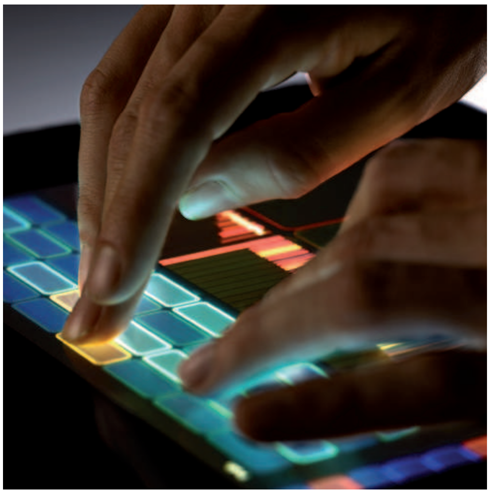


Figure 1: *Touch sensitive audio controllers with tight placement of interface elements (e.g., JazzMutant Lemur) can benefit from audio driven tactile feedback.*

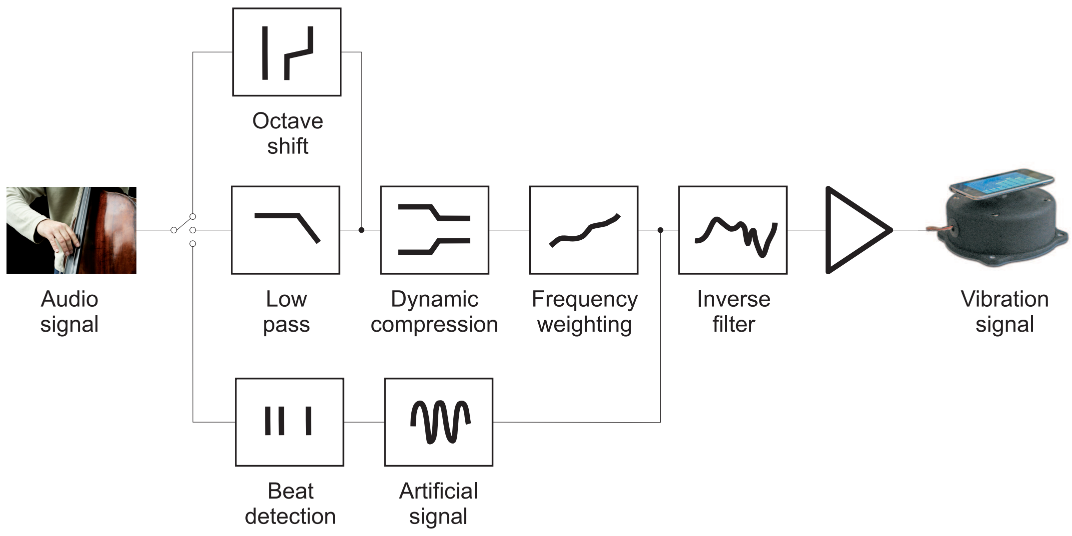



Figure 2: *Three different approaches to generate audio-driven tactile feedback. Top: Frequencies were shifted down one octave using granular synthesis. Middle: The audio signal was low-pass filtered at* 1 *kHz. Bottom: Attacks were detected, which triggered an artificial signal (*100 *Hz,* 80 *ms).*

## 3. Audio-Driven Tactile Feedback

The vibration signal, acting as a sign carrier, should relate to the audio sequence from which it was generated. In this study, the goal is to identify a specific audio loop (e.g., a rattling shaker) by touch feedback. To achieve this aim, identification features characteristic of the audio track must be maintained in the adaptation process. These features can be related to the *sound source* itself or to the *sound sequence* that was generated by a sound source. The following are important identification features of a *sound source* (e.g., kick drum, kettledrum, or artificial source):

- frequency and level structure: e.g., harmonic / inharmonic tone structure, noise background
- time structure: e.g., starting transients or decay times of individual tones

Most percussive instruments are unpitched (e.g., snare), while others excite pitch perception (e.g., kettledrum). In the latter case, the fundamental frequency might not be the strongest component in the spectrum. Still, the auditory perception will recognize the fundamental as pitch due to the harmonics. No virtual pitch effect is known for the tactile domain.

The following are some exemplary features of a *sound sequence*:

- melody
- rhythm
- dynamics

These features must be abstracted to convert the acoustical signal into a suitable vibration signal. The level of abstraction should be chosen so the system incorporates the following factors:

- ease (learnable, effective, efficient and intuitive)
- joy (aesthetics, system personality) [7]

The design process should strive for a balance between the above factors. For example, while the learnability and intuitivity might improve with less abstract tactile signs, the efficiency might suffer. The following three user-related categories of abstraction are known from semiotics [8]:

- index
- icon
- symbol

An *index* connects the user directly to its physical source. The simplest example for tactile feedback is a low-pass filtered sound signal, applied in experiment one and four with a low pass frequency of 1 kHz (10th order Butterworth). Experiment two investigated whether loop identification is improved when information is shifted from high to low frequencies. Altinsoy [9] has shown that good integration between auditory and tactile information occurs when the acoustical frequency is a harmonic of the vibration frequency. Thus, all audible frequencies were shifted down one octave using granular synthesis with a grain size of 22 ms. This maintained accurate timing but added some artifacts, especially at higher frequencies; which was acceptable because only low frequency vibrations were felt.

An *icon* is based on similarity. For example, the artificial signal for tactile feedback would contain reduced characteristics of the audio track. In experiment three the beat (attacks in the amplitude envelope) was extracted from the audio loop. The detected attacks triggered sinusoidal pulses chosen to be easily perceived (100 Hz, 80 ms).

A *symbol* is based on conventions. An arbitrary signal might be selected for tactile feedback and the user is trained to associate the abstract signal with a specific audio loop or instrument. This was not implemented in this study because it is the least intuitive approach and requires substantial training.

The flow chart to generate audio-driven tactile feedback using the three described approaches can be seen in Figure 2. The dynamic range (see Section 2) was compressed by a factor of two (with 20 ms attack and release). The difference between the frequency dependent auditory and tactile perception threshold was compensated. Also, the transfer function of the electrodynamic shaker was compensated with an inverse filter.

# 4. Identification Experiments

## 4.1. Setup

The setup can be seen in Figure 3. Tactile feedback was reproduced using an electro-dynamic vibration actuator (Monacor, BR-25) coupled with a touch sensitive device (Apple, iPod). The device was connected to the computer using TouchOSC, an application that can send Open Sound Control messages over a Wi-Fi network. The user interface was divided into six buttons. Each button corresponded to a specific audio signal. When the finger of the participant came in contact with a button, tactile feedback for the respective channel was rendered in real time using Pure Data, while simultaneously, the sum of all six audio signals was played on closed headphones (Sennheiser, HDA 200). The task of the participant was to associate the vibrating buttons to the specific audio signals.

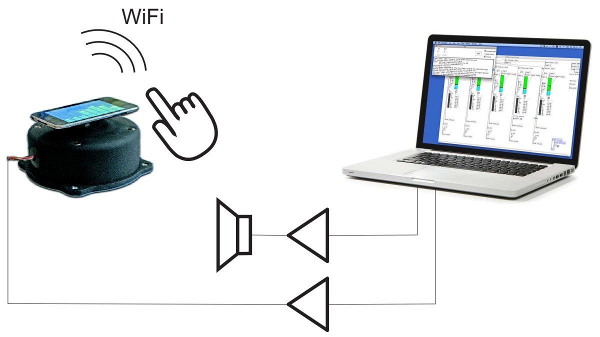



Figure 3: *The touch screen was mounted on an electro-dynamic shaker to reproduce vibrations.*

## 4.2. Participants

Twenty subjects voluntarily participated in the experiments (16 male and 4 female). Their ages ranged from 20 to 40 years. None had participated in previous audio-tactile experiments, and all indicated they had no hearing damage or hand disorders.

## 4.3. Results and Discussion

### *4.3.1. Loop - Low Pass*

In the first experiment the six vibration signals were generated by low pass filtering the audio loops at 1 kHz. The association data (stimulus and response plot) between the six stimuli and the responses are shown in Figure 4 a). The area of each circle is proportional to the number of answers given for a particular combination of stimulus and response. A full circle [e.g., at (BASS LINE, BASS LINE)] corresponds to the total number of subjects (20 in this case). With the low pass filtering, most answers lie on the diagonal, indicating correct answers. Some errors are seen, particularly for percussion instruments, which generate mainly higher frequencies. The participants reported that time structure and frequency content was important. Snare and kick drum were also differentiated using intensity cues.

### *4.3.2. Loop - Octave Shift*

In the second experiment the vibration was generated by shifting the frequencies down one octave before low pass filtering the audio loops at 1 kHz. Figure 4 b) shows that stimulus identification improved, perhaps due to a better perception of rhythm because more of the shifted signal content can be perceived through the tactile sense. The results improved for the kick drum and shaker, but there were slightly more errors between the hihat and snare, perhaps because the hihat was perceived more intense than before as its dominant high frequency energy was shifted toward lower frequencies.

However, it is unclear whether features of the sequence (e.g., rhythm) or features of the source (e.g., frequency content) or both influenced the results; therefore, the two subsequent experiments focused on separating the sequence and source features.

### *4.3.3. Loop - Beat Detection*

In the third experiment the vibration was generated by detecting the beat of the individual loops, which triggered an artificial vibration signal; thus, source features were removed from the vibration signal. The results are shown in Figure 5 a). Good classification is still possible (strong main diagonal); thus, indicating that the sequence feature, “rhythm”, is an important factor for loop identification; however, the overall detection rate decreased. This decrease is primarily due to two participants that seemed to lack a good sense of rhythm.

### *4.3.4. Hit - Low Pass*

In the fourth experiment rhythm (sequence) information was removed to test whether a percussion instrument could be identified with only source features; thus, only a single hit was replayed. The bass line and tambourine were removed from the stimuli set and other characteristic percussion sounds with distinct source features (guiro and handclap) were added. The vibration was generated by low pass filtering the hit at 1 kHz.

As seen in Figure 5 b), the kick drum and snare were identified with 100 percent accuracy due to their characteristic frequency content, which results in different tactile perceptual qualities. Of the remaining instruments, the guiro had the highest number of correct identifications, perhaps because of its typical time structure (rattle like) that distinguishes it from the instruments with different time structures (bang like). The high frequency percussive sounds were not differentiated well. Subsequent experiments revealed that the detection rate did not improve with octave shifting the vibration signal or participant training.

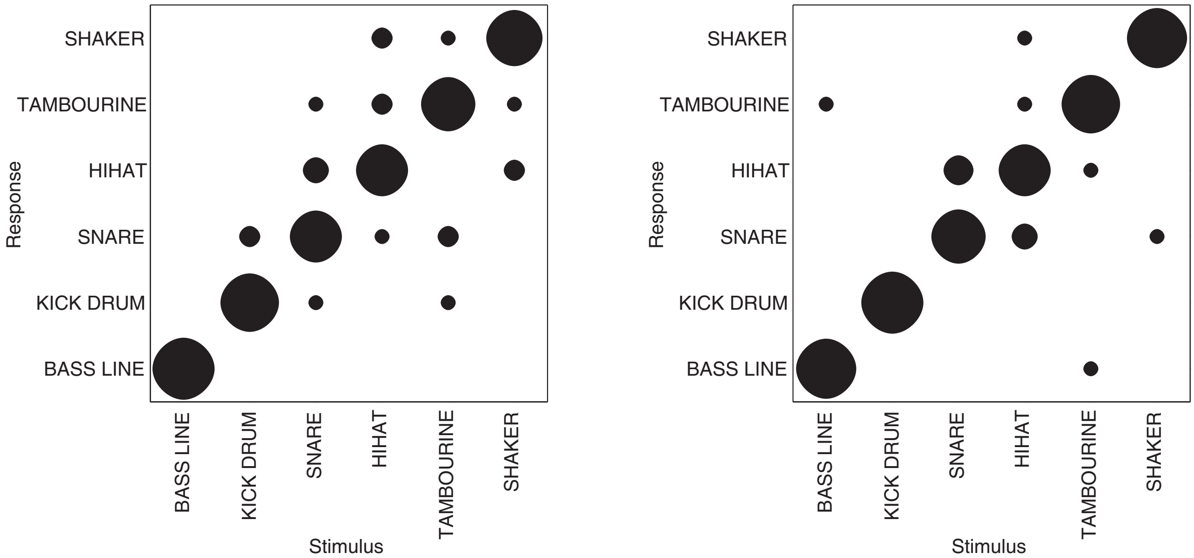
The results show that some differentiation is possible using source features only; however, if the frequency content or time structure of different signals is similar, it is difficult to link the tactile sense and auditory sense.

### *4.3.5. Summary*

The best identification rates are obtained if the source and sequence features are combined (index: octave shifted loop). The loop identification using the sequence feature rhythm (icon: beat detection), was observed to be time consuming and varied between subjects. Participants needed an average identification time of approximately 10 s per loop in the third experiment. In comparison only 6 s per loop were needed in the first and second experiment.

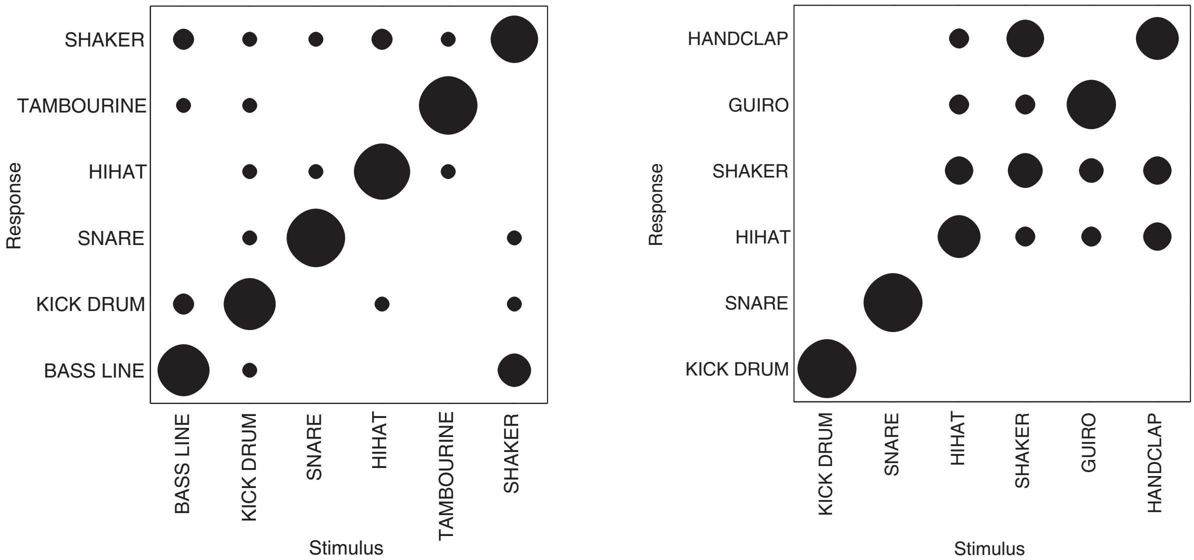
# 5. Attractiveness

Before and after the above experiments, participants were asked to mix six audio loops into a 1.5 minute composition. They used the setup described above. Instead of buttons, six faders were




**(a)** **(b)**

Figure 4: *Association results for six loop stimuli. The vibration was generated using* **(a)** *a* 1 *kHz low pass and* **(b)** *the octave shifted signal using granular synthesis. The area of each circle is proportional to the number of answers given for a particular combination of stimulus and response.*




**(a)** **(b)**

Figure 5: *Association results for the third and fourth experiment. The vibration was generated using* **(a)** *sequence features (loop stimuli and beat detection) and* **(b)** *source features (low-pass filtered percussive hits.)*

used to blend the different audio signals. In the first set (before all the experiments) a conventional groovebox was simulated. In the second set (after all the experiments) audio-driven tactile feedback was rendered using the octave shift approach. When the finger of the user came in contact with a fader, vibration for the respective channel was rendered. After completion, participants were asked to judge the usability and attractiveness of the groovebox using the AttrakDiff [10] semantic differential. This method uses pairs of bipolar adjectives (see Figure 6) to evaluate the pragmatic and hedonic qualities of interactive products.

Figure 6 shows the mean values for all subjects. The pragmatic quality is on average better without tactile feedback. This was due to the participants who experienced some difficulty with audio-tactile association in the prior experiments. The individual ratings for the tactile feedback set varied, indicating disagreement between subjects. Thus, the difference in pragmatic quality is marginally not statistically significant at the $5\%$ level (dependent t-test for paired samples).

On average, the hedonic quality is better with tactile feedback, especially for the aspect "stimulation" (significant at the $5\%$ level). The hedonic category "stimulation" refers to the ability of a product to support the user to further personal devel-

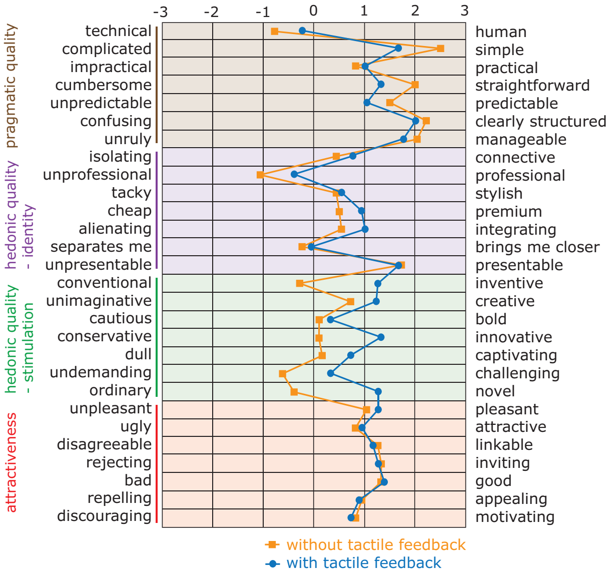



Figure 6: *Mean values of the AttrakDiff semantic differential for seven items for each of the four dimensions: pragmatic quality, hedonic quality - identity, hedonic quality - stimulation and attractiveness.*

opment. The groovebox with audio-driven tactile feedback was rated as more innovating, captivating and challenging. These results are in agreement with other studies that evaluated multimodal feedback [11].

The overall attractiveness of the groovebox remains the same with or without audio-driven tactile feedback. This result is reasonable if the attractiveness is understood based on the hedonic and pragmatic quality, where each contributes in equal parts to the attractiveness of a product [10].

Note that the presented data are only valid for the specific exercise and the laboratory conditions described above. The results might change depending on task and context. For example, in a real live set it might be more important to know the finger is on the correct fader. Tactile feedback might also help a DJ match beats between different tracks, influencing his pragmatic quality perception; thus, conclusions should be drawn carefully.

## 6. Conclusions

The results show percussive instruments can be identified to some degree with audio-driven tactile feedback. The detection rate is highest when the source and sequence features are maintained (octave shifted loop). Though source features (e.g., frequency content) are observed faster, the sequence feature "rhythm" is more reliable. Using the described methods, identification errors for different instruments cannot be avoided. The inherent limitations might have caused the poor ratings for the pragmatic quality of the groovebox. The limitations are not surprising because the tactile sense is more limited than the auditory sense; however, the method seems promising, and there is opportunity for improvement.

## 7. Outlook

For further studies, an alternative approach could relate the auditory features to particular tactile capabilities, like the perception of touch location. To implement this, different vibration reproduction methods (e.g., a Braille display) would be necessary.

This study used closed headphones for audio reproduction. If this is not the case, audio radiation of the vibration reproduction system might be problematic, especially in quiet environments like a studio. Thus, alternative, quieter tactile reproduction methods should be considered, such as using horizontal instead of vertical vibrations or electro-tactile stimulation [9, 12].

## 8. Acknowledgements

The authors wish to thank Prof. U. Jekosch for her support and informative discussions.

## 9. References

[1] Andersen, T.H., Huber, R., Kretz, A., Fjeld, M., "Feel the Beat: Direct Manipulation of Sound during Playback", TableTop: First International Workshop on Horizontal Interactive Human-Computer Systems, Adelaide, Australia, 2006.

[2] Beamish, T., Maclean, K., Fels, S., "Manipulating Music: Multimodal Interactions for DJs", CHI Letters 6, no. 1, pp. 327-334, Vienna, Austria, 2004.

[3] Merchel, S., Altinsoy, E. and Stamm, M., "Tactile Music Instrument Recognition for Audio Mixers", Proceedings of 128th AES Convention, London, UK, 2010.

[4] Verrillo, R.T. and Gescheider, G.A., "Perception via the Sense of Touch", Tactile Aids for the Hearing Impaired, Whurr Publishers, London, England, 1992.

[5] Verrillo, R.T., "Effect of Contactor Area on the Vibrotactile Threshold", J. Acoust. Soc. Am. 37, 843-846, 1963.

[6] Merchel, S., Leppin, A. and Altinsoy, E., "Hearing with your Body: The Influence of Whole-Body Vibrations on Loudness Perception", Proceedings of ICSV16, Kraków, Poland, 2009.

[7] Möller, S., Engelbrecht, K.-P., Kühnel, C., Naumann, A., Wechsung, I. and Weiss, B., "Evaluation of Multimodal Interfaces for Ambient Intelligence" in H. Aghajan, R. López-Cózar Delgado and J. C. Augusto, eds., Human-Centric Interfaces for Ambient Intelligence, Chapter 14, Academic Press, London, pp. 347-370, 2010.

[8] Jekosch, U., "Assigning Meaning to Sounds - Semiotics in the Context of Product-Sound Design" in Blauert, J., ed., Communication Acoustics, Springer, Berlin, pp. 193-221, 2005.

[9] Altinsoy, M.E., "Auditory-Tactile Interaction in Virtual Environments", Shaker Verlag, Aachen, 2006.

[10] Hassenzahl, M., Burmester, M., Koller, F., "AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität" in Ziegler, J., Szwillus, G., eds., Mensch & Computer 2003, Interaktion in Bewegung, pp. 187-196, B.G. Teubner, Stuttgart, 2003.

[11] Wechsung, I., Naumann, A.B., "Evaluation Methods for Multimodal Systems: A Comparison of Standardized Usability Questionnaires", PIT 2008 LNAI 5078, pp. 276-284, Heidelberg: Springer, 2008.

[12] Loviscach, J., "Music at Your Fingertips: An Electrotactile Fader", Proceedings of the 123rd AES Convention, New York, USA, 2007.