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

SEBASTIAN MERCHEL, *AES Student Member,* **M. ERCAN ALTINSOY,** *AES Member,* **AND**
(Sebastian.Merchel@tu-dresden.de) (Ercan.Altinsoy@tu-dresden.de)

MAIK STAMM, *AES Student Member*
(Maik.Stamm@tu-dresden.de)

Chair of Communication Acoustics, Dresden University of Technology, Dresden, Germany

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.

0 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 Fig. 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. [1] developed the tangible Q-Slider to control the playback position of a soundtrack. Anderson et al. [2] 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. Few experimental data exists that examines the ability to distinguish between different audio signals using audio-driven tactile feedback. Hoggan and Brewster [3] designed auditory-tactile alerts using headphones and three vibrotactile actuators around the waist. Different messages were encoded in both modalities using rhythm, roughness, and spatial location. They showed that if users are trained to understand messages in one modality, they are able to identify them in the other. Related research exists in the field of sensory substitution system for people with impairments [4]. Often pin arrays are used to encode speech information into vibratory patterns. However, essential training is required to use such systems [5]. In this study a single vibration actuator will be used.

To generate intuitive tactile feedback from audio signals, it is essential to understand the capabilities and limitations of the auditory and the tactile senses.

1 AUDITORY PERCEPTION

The perception of sound is a complex area that has been studied for several decades. The basic physical attributes of sound (e.g., intensity, frequency or location of a sound source) have been correlated to perceptual attributes like loudness, pitch or distance. Different effects like adaptation to loud signals or masking characterize the auditory system. However, there is much more complex processing before a listener assigns meaning to a perceived sound. Integration with other senses and the experience of the user plays an important role. All those factors should be taken into account when audio-driven tactile feedback is rendered. However, this paper focuses on the basic frequency perception

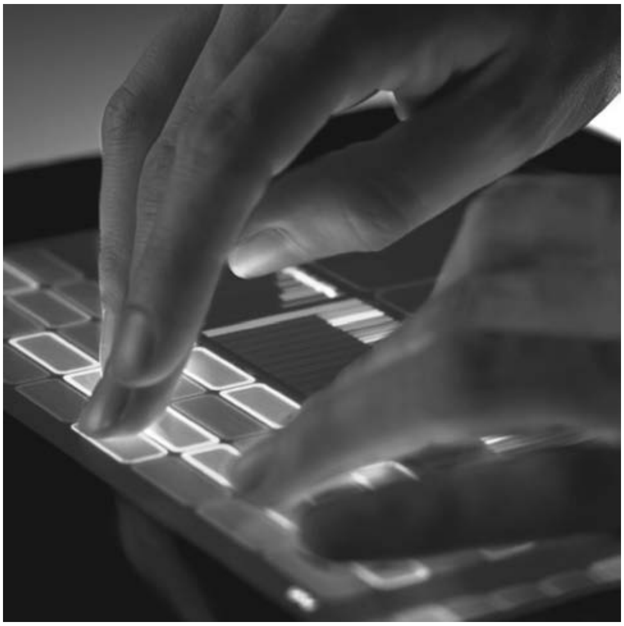


Fig. 1. Touch sensitive audio controllers with tight placement of interface elements can benefit from audio-driven tactile feedback.

and intensity perception of the auditory and tactile sensory systems.

1.1 Frequency Perception

The lowest frequency at which sound is perceived as a tone is around 16 Hz. For even lower frequencies it is possible to follow the time structure of a signal [6]. The perceived character of the sound changes and pitch perception fades. The upper frequency limit is around 20 kHz. It depends strongly on the age of the subject. Music perception takes place in the frequency range up to approximately 10 kHz.

One of the fundamental characteristics of the auditory system is the ability to discriminate between different frequencies. Differences as small as 1 Hz can be perceived at 1 kHz and below. The Weber fraction ($\frac{\Delta f}{f}$) in the frequency range from 400 to 2000 Hz is approximately 0.2% for tones with 40 dB above threshold according to [7]. For lower or higher frequencies the difference limen for frequency discrimination increases up to several percent.

1.2 Intensity Perception

Fig. 2(a) shows that the hearing is most sensitive to sound pressure between 300 and 7000 Hz. It becomes less sensitive for decreasing and increasing frequency. In addition, the figure shows estimates for the pain threshold and the annoyance threshold after Winckel [8].

The curves of equal subjective intensity (equal loudness contours) are plotted according to ISO 226:2003 [9]. They follow the threshold curve to some degree. It can be seen that they get closer toward lower frequencies. The relevant dynamic range is, thus, frequency dependent from 40 dB to more than 100 dB.

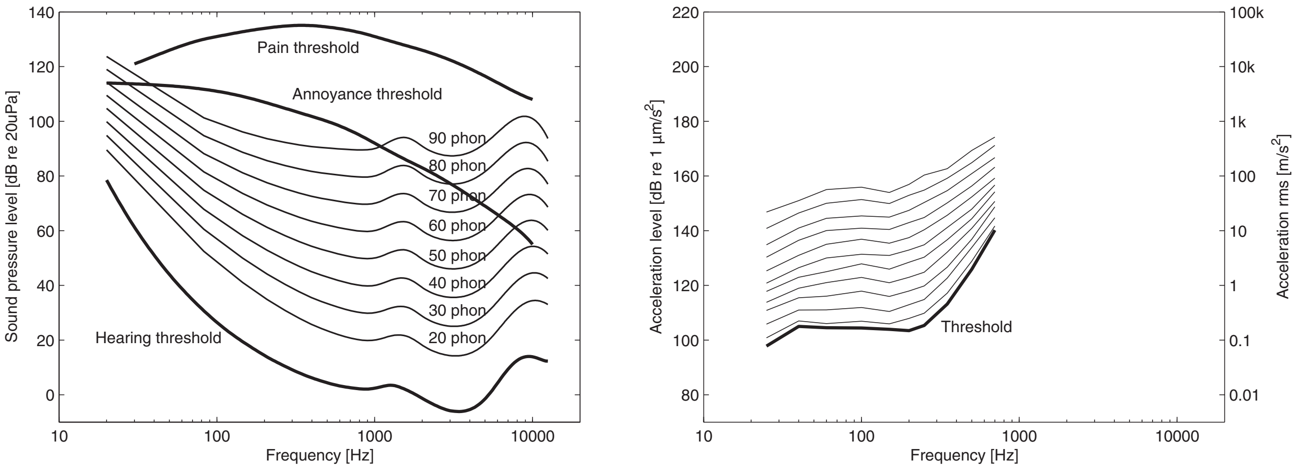
On the other hand, the auditory system is able to discriminate intensity differences between 2 and 0.2 dB, depending on sound pressure level [10].

2 TACTILE PERCEPTION

2.1 Frequency Perception

In comparison to the auditory modality, the tactile sense is rather limited. Only frequencies between a few Hertz and approximately 1 kHz can be perceived via the mechanoreceptive system. It has been reported that the quality of sensation changes with frequency [11].

The ability to discriminate between frequencies is also quite limited if compared to the auditory system. A Weber fraction ($\frac{\Delta f}{f}$) of about 20% to 50% was found by [12] for sinusoidal vibrations at the finger, depending on frequency. This is much lower than the frequency discrimination ability of the ear. However, Rothenberg [13] showed that auditory pitch and vibrotactile frequency can be associated by hearing subjects without training. This was tested for variations



(a) (b)

Fig. 2. Curves of equal subjective intensity plotted as a function of frequency for (a) sounds (according to ISO 226:2003 [9] and Winckel [8]) and (b) vibrations on the thenar eminence (adapted from Verrillo [14]).

in voice fundamental frequency (intonation patterns with moderate to strong stress patterns) using a short sentence.

2.2 Intensity Perception

Similar to the ear, the vibration sensitivity of the skin depends on frequency. In addition, the sensitivity depends strongly on the size of the contact area [11]. With increased contactor size, perception threshold decreases for higher frequencies. For very low frequencies, no influence was found.

Fig. 2(b) shows the frequency dependent perception threshold on the thenar eminence adapted from [14]. It can be seen that the glabrous skin becomes more sensitive to the acceleration of a surface with decreasing frequency. The curves of equal subjective intensity follow the threshold to some degree. Again a frequency dependence can be seen. The curves get closer toward higher frequencies. The dynamic range can thus be quantified between approximately 35 dB to 50 dB. Vibrations more than 55 dB above threshold become very unpleasant or painful. Because thresholds vary between subjects, the usable dynamic range is even smaller.

There is not much knowledge of vibrotactile intensity discrimination. The reported values in the literature range from 0.4 dB to 2.3 dB. See [11] for a review.

Many more factors, like time of exposure or multiple simultaneous stimuli, have an influence on frequency and intensity perception but will not be discussed here.

2.3 Summary

It can be seen that the frequency range of the auditory and tactile senses overlap up to approximately 1 kHz. Still there are some differences in the perception. The limited ability of the skin to discriminate frequency differences is important if audio-driven tactile feedback is generated. In addition, the different dynamic ranges of both modalities have to be adapted.

3 AUDIO-DRIVEN TACTILE FEEDBACK

In this study the goal is to identify a specific percussive audio loop (e.g., a tambourine) by touch feedback. The vibration signal, acting as a sign carrier, should relate to the audio sequence from which it was generated. 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*:

- Frequency and level structure: e.g., harmonic/inharmonic tone structure, noise background; and
- Time structure of individual notes: e.g., starting transients or decay times.

In this study different sound sources were selected, which can be distinguished in the auditory modality (bass line, kick drum, snare, hihat, tambourine, and shaker). For example, the sound from a hihat hit and snare hit have a similar time structure but can be easily discriminated by their frequency content. Hihat and tambourine have similar frequency content but can be well separated because of their characteristical time structure.

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

- Rhythm
- Dynamics.

In this study an audio loop was generated using the six above-mentioned sound sources. For each source a distinct rhythm was selected. When played together, a groove emerged.

The described source and sequence features must be abstracted to convert the acoustical signal into a suitable vibration signal. The simplest example for tactile feedback generation would be to play the sound signal directly through the vibration actuator. However, this would generate unnecessary sound radiation at higher frequencies that are imperceptible through the tactile sense (see Fig. 2). Thus, a low-pass filter with 1 kHz (10th order Butterworth) was implemented as a first approach.

In a second approach more information is shifted from high to low frequencies. Altinsoy [18] has shown that auditory and tactile information integrates well into a single multimodal percept 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 at higher frequencies, which was acceptable because only low frequency vibrations were felt. Using this method, source and sequence features are maintained.

In a third approach the beat (attacks in the amplitude envelope) was extracted from each audio loop. The detected attacks triggered sinusoidal pulses chosen to be easily perceived (100 Hz, 80 ms). In this case, the artificial signal for tactile feedback contains only characteristical features of the sequence. Any source features are removed.

The flow chart to generate audio-driven tactile feedback using the three described approaches can be seen in Fig. 3. The dynamic range (see Section 1) 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 electro-dynamic shaker was compensated with an inverse filter. The described processing did not introduce a perceivable latency between the modalities. For guidelines on acceptable delays refer to Altinsoy [18].

4 IDENTIFICATION EXPERIMENTS

4.1 Setup and Procedure

Tactile feedback was reproduced using an electrodynamic vibration actuator (Monacor, BR-25) coupled with

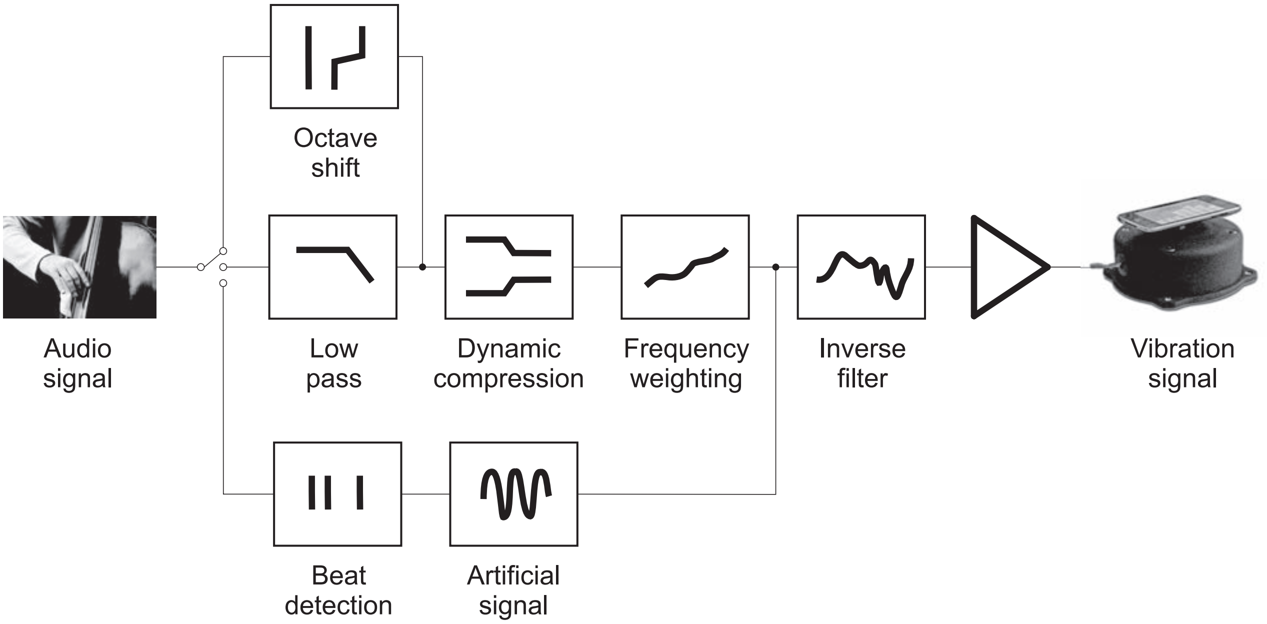


Fig. 3. Three different approaches to generate audio-driven tactile feedback. Top: Frequencies were shifted down one octave using granular synthesis. Middle: The audio signal was low-pass filtered at 1 kHz. Bottom: Attacks were detected, which triggered an artificial signal (100 Hz, 80 ms).

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. No time limit was applied and participants were allowed to listen to each stimulus as many times as they liked. There was no training session before the experiments.

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 Fig. 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. Fig. 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 Fig. 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

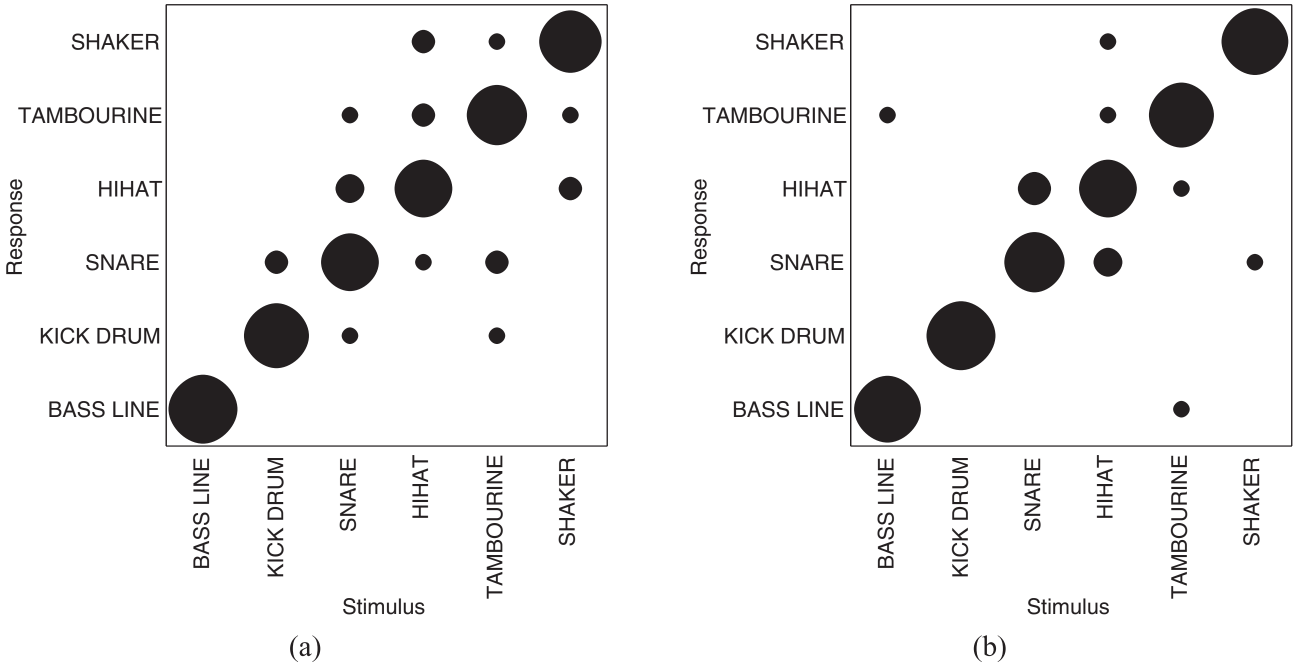


Fig. 4. Association results for six loop stimuli. The vibration was generated using (a) a 1 kHz low pass and (b) the octave shifted signal using granular synthesis. The area of each circle is proportional to the number of answers given for a particular combination of stimulus and response.

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 Fig. 5(b), the kick drum and snare were identified with 100 percent accuracy due to their characteristic frequency content. 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

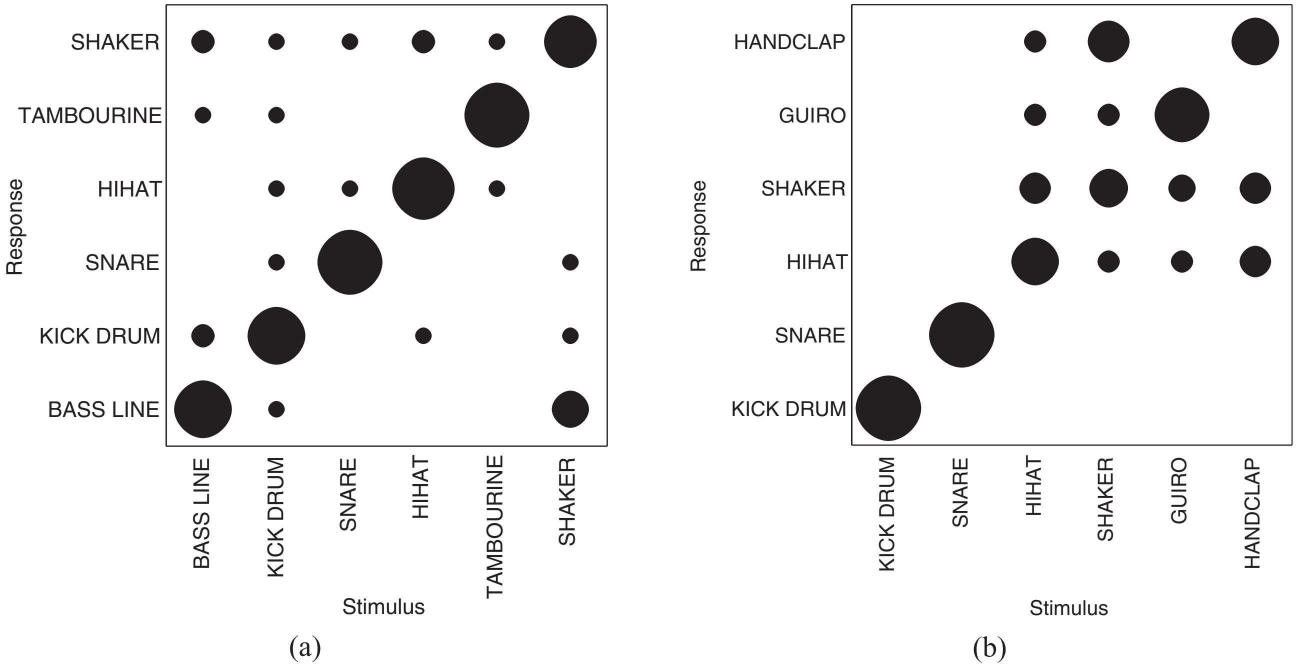


Fig. 5. Association results for the third and fourth experiment. The vibration was generated using (a) sequence features (loop stimuli and beat detection) and (b) source features (low-pass filtered percussive hits).

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, differentiation becomes more difficult.

5 CONCLUSIONS

The results show percussive loops 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, because one single hit might suffice to identify the instrument, the sequence feature "rhythm" is more reliable. Using the described methods, identification errors for different percussive instruments cannot be avoided. Similar results were found for tactile identification of non-percussive musical instruments [16]. 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.

6 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. In addition, many modern touch screens are able to simultaneously register multiple distinct touch positions. In this case separate tactile feedback (e.g., using bending wave technology [17]) for each finger should be investigated.

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 electrotactile stimulation [18, 19, 20].

7 ACKNOWLEDGMENTS

The authors wish to thank Prof. U. Jekosch for her support and informative discussions. A preprint version of this paper has been presented at the 3rd International Workshop on Perceptual Quality of Systems, *PQS*, Bautzen, Germany, September 2010.

8 REFERENCES

[1] T. Beamish, K. Maclean, and S. Fels Maclean, "Manipulating Music: Multimodal Interactions for DJs," *CHI Letters* 6, no. 1, pp. 327–334 (2004.

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

[3] E. Hoggan and S. A. Brewster, "*Designing Audio and Tactile Crossmodal Icons for Mobile Devices,*" *Proceedings of ACM ICMI*, Nagoya, Japan (2007).

[4] C. Lenay, S. Canu, and P. Villon, "*Technology and Perception: The Contribution of Sensory Substitution Systems,*" *Proceedings of ICCT*, IEEE (1997).

[5] T. Ifikube, "Discrimination of Synthetic Vowels by Using Tactile Vocoder and a Comparison to that of an Eight-Channel Cochlear Implant," *IEEE Trans., BME* 36(11), pp. 1085–1091 (1989).

[6] H. Møller and C. S. Pedersen, "Hearing at Low and Infrasonic Frequencies," *Noise and Health*, 6;23, pp. 37–57 (2004).

[7] C. C. Weir, W. Jesteadt, and D. M. Green, "Frequency Discrimination as a Function of Frequency and Sensation Level," *J. Acoust. Soc. Am.* 61, pp. 178–183 (1977).

[8] F. Winckel, "Nachrichtenverarbeitung unter kybernetischen Aspekten," *Handbuch für HF- und E-Techniker*, Bd. 8, Berlin, Germany (1969).

[9] ISO 226:2003, "Acoustics - Normal Equal-Loudness-Level Contours."

[10] H. Fastl and E. Zwicker, "Psychoacoustics—Facts and Models," *Springer Series on Information Sciences*, Berlin, Germany (1999).

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

[12] M. Rothenberg, R. T. Verrillo, S. A. Zahorian, M. L. Brachman, and S. J. Bolanowski Brachman, "Vibrotactile Frequency for Encoding a Speech Parameter," *J. Acoust. Soc. Am.* 62, 1003–1012 (1977).

[13] M. Rothenberg and R. D. Molitor, "Encoding Voice Fundamental Frequency into Vibrotactile Frequency," *J. Acoust. Soc. Am.* 66, 1029–1038 (1979).

[14] R. T. Verrillo, "Effect of Contactor Area on the Vibrotactile Threshold," *J. Acoust. Soc. Am.* 37, 843–846 (1963).

[15] R. T. Verrillo A. Fraioli, and R. Smith, "Sensation Magnitude of Vibrotactile Stimuli," *Perception and Psychophysics* 6(6A), 366–372(1969).

[16] S. Merchel M. E. Altinsoy, and M. Stamm, "*Tactile Identification of Non-Percussive Music Instruments,*" *Proceedings of Forum Acusticum*, Aalborg, Denmark (2011).

[17] N. Harris, "*Applications of Bending Wave Technology in Human Interface Devices,*" *Proceedings of the 126th Convention* of the Audio Engineering Society, paper 7658 (2009).

[18] M. E. Altinsoy, *Auditory-Tactile Interaction in Virtual Environments,* Shaker Verlag, Aachen (2006).

[19] J. Loviscach, "*Music at Your Fingertips: An Electrotactile Fader,*" Proceedings of the *123rd Convention of the Audio Engineering Society*, paper 7306 (2007).

[20] S. Merchel and M. E. Altinsoy, "Vorrichtung zur Verarbeitung und Wiedergabe von Signalen in elektronischen Systemen zur elektrotaktilen Stimulation," *German Patent* DE 10 2009 020 796 (2010).

THE AUTHORS

Sebastian Merchel

M. Ercan Altinsoy

Maik Stamm

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 and participated in the International Graduate School for Neuroscience at the Ruhr-University Bochum. In 2005, he completed his Ph.D. thesis, "Auditory-Tactile Interaction in Virtual Environments." From 2004 to 2006, Ercan Altinsoy worked at HEAD acoustics as NVH Consulting Engineer. In 2006 he started lecturing at the Dresden University of Technology, Chair of Communication Acoustics. He is leading the "Audiotactile Interaction" group at the same university. His research interests include auditory and haptic interfaces for virtual environments, tactile psychophysics and psychoacoustics, vibroacoustics, product sound, and vibration design.

●

Maik Stamm began the study of electrical engineering at the Dresden University of Technology in 2004 and received his Diploma degree in 2009. During his studies, he worked at the Fraunhofer IPMS and the Institute of Acoustics and Speech Communication in Dresden as a research student. Furthermore, he spent an internship at the Entiretec AG participating in the development of self-managing networks for communication and multimedia. In 2009 Maik Stamm joined the Chair of Communication Acoustics at the Dresden University of Technology as a scientific assistant. His research interests include hapto-audio interfaces, multimodal interaction, sound-design, and psychophysics. He is a member of the Audio Engineering Society (AES) and the German Acoustical Society (DEGA).