Mapping Monophonic MIDI Tracks to Vibrotactile Stimuli Using Tactile Illusions *

Byron Remache-Vinueza[1,2][0000−0003−1656−7821], Andrés Trujillo-León[1,3][0000−0003−4798−1777], Maria-Alena Clim[4][0000−0001−7469−9713], Fabián Sarmiento-Ortiz[2][0000−0001−8961−9229], Liliana Topon-Visarrea[2][0000−0001−5307−5450], Alexander Refsum Jensenius[4][0000−0001−6171−8743], and Fernando Vidal-Verdú[1,3][0000−0002−5459−8306]

[1] Departamento de Electrónica, Universidad de Málaga, 29071 Málaga, Spain
paulremacheing@gmail.com
[2] SISAu Research Group, Facultad de Ingeniería y Tecnologías de la Información y la Comunicación, Universidad Tecnológica Indoamérica, Quito 170103, Ecuador
[3] Instituto de Investigación Biomédica de Málaga (IBIMA), 29010 Málaga, Spain
[4] RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, Department of Musicology, University of Oslo, Norway

Abstract. In this project, we propose an algorithm to convert musical features and structures extracted from monophonic MIDI files to tactile illusions. Mapping music to vibrotactile stimuli is a challenging process since the perceptible frequency range of the skin is lower than that of the auditory system, which may cause the loss of some musical features. Moreover, current proposed models do not warrant the correspondence between the emotional response to music and the vibrotactile version of it. We propose to use tactile illusions as an additional resource to convey more meaningful vibrotactile stimuli. Tactile illusions enable us to add dynamics to vibrotactile stimuli in the form of movement, changes of direction, and localization. The suggested algorithm converts monophonic MIDI files into arrangements of two tactile illusions: "phantom motion" and "funneling". The validation of the rendered material consisted of presenting the audio rendered from MIDI files to participants and then adding the vibrotactile component to it. The arrangement of tactile illusions was also evaluated alone. Results suggest that the arrangement of tactile illusions evokes more positive emotions than negative ones. This arrangement was also perceived as more agreeable and stimulating than the original audio. Although musical features such as rhythm, tempo, and melody were mostly recognized in the arrangement of tactile illusions, it provoked a different emotional response from that of the original audio.

Keywords: Audio tactile rendering · MIDI · Monophonic music · Tactile illusion · Vibrotactile stimuli.

* Supported by the Spanish Government under contract PID2021-125091OB-I00.

1 Introduction

Experiencing music means more than just perceiving sound. It is a multi-sensory experience as well as a complex cognitive process where tactile and visual modalities play important roles. In fact, musical sound is produced through vibrations may reach our somatosensory system to complement the experience. Recently, researchers have found that it is possible to convert some musical features such as rhythm, frequency, and intensity into tactile stimuli presented to the user through vibrating devices attached to the skin (vibrotactile stimuli)[8]. Different methods have been proposed to convert musical features to perceptible vibrotactile stimuli (see [9] for a review). The strategies used range from complex signal processing to conceptual implementations.

One approach to such mappings consists of placing different frequencies on different parts of the body, the so-called Frequency Model [5]. With this technique, audio signals are processed and separated into different frequency bands to generate sine tones with frequencies found in the skin perception range. Dynamics can be added by extracting the envelope and transients from control signals, which are then arranged in different locations in a music haptic installation that presents vibrotactile stimuli through actuators. The concept of low pitches being "down" and high pitches being "up" is replicated from the bottom to the top of the installation. Even sitting still, participants reported an enhanced feeling of rhythm and sensations of movement [5]. However, researchers agree that some musical information is lost with the Frequency Model when presenting discrete frequencies instead of the whole frequency range of the audio file.

Another method is the Track Model, in which a musical composition is divided into tracks containing different instruments, each routed to different actuators touching the skin. In [4], the effectiveness of the Track Model was evaluated. They report a better performance of the Track Model than the Frequency Model when conveying emotional content, but underline the requirement of multi-track audio files, which are not available for every piece of music.

Moreover, researchers agree that the perceptible frequency range of the skin (3Hz–1kHz) is far more limited than that of the auditory system (20Hz–20kHz). Therefore, musical information embedded in the higher frequencies is usually lost in the mapping process. In addition, the emotional response to vibrotactile stimuli rendered from musical features remains unclear, thus making it difficult to establish a unified rendering method.

To help overcome these limitations, we explore a resource known as *tactile illusions* (*TIs*) [3][6]. A *TI* may be defined as the discrepancy between a tactile stimulus and what is perceived. What we propose is to use *TIs* to convey more meaningful vibrotactile stimuli. *TIs* enable us to add dynamics; something more than just vibrations, pulses, or beats. *TIs* create sensations of movement, direction, and localization; which are also features used in music [7].

In section 2, the proposed algorithm and the experimental methodology are detailed. In section 3, the results of the main experiments are described. Finally in section 4, the discussion and conclusions are elaborated.

2 Method

2.1 Proposed algorithm

Audio Audio was played back from monophonic MIDI files to avoid overlapping notes. One note played at a time is desired since each tactile illusion can be presented just one at a time (i.e., each *TI* corresponds to one note only), see Fig. 3. Musical features derived from the MIDI files are shown in Table 1. Each musical note produces a vector of musical information (i.e., vibrations that can be perceived by the ear using standard audio technology but not easily perceived through touch) as follows: $V_{Ai}(N_i,F_i,t_i,D_i)$.

Table 1: Musical information derived from the MIDI files.

Feature	**Representation**	**Description**
Number of notes	n	Total number of notes
Notes	N_i	Note played in position i, with i from 1 to n
Frequency (Hz)	F_i	Frequency of note N_i
Onset time (s)	t_i	Time when note N_i is played
Duration (s)	D_i	How long note N_i is played
Maximum frequency (Hz)	F_{max}	Highest frequency from notes N_1 to N_n
Minimum frequency (Hz)	F_{min}	Lowest frequency from notes N_1 to N_n

Tactile Illusions The proposed *TIs* to implement are *Phantom Motion* (PM) or *Apparent Movement* [3] and *Funneling* (FUN) or *Illusory Actuator* [6]. These illusions can be created using a pair of actuators similar to stereo headphones. With *PM*, the participant perceives a linear illusory movement of a vibrating point from one actuator to the other, where the frequency of vibrations, duration, and direction of movement can be controlled. *FUN* evokes the sensation of an illusory actuator vibrating in a specified location between the physical actuators, where the frequency of vibrations, duration, and location can also be controlled. Besides the versatility of these illusions, movement of *PM* may be associated with the movement of hands when playing instruments such as violin, piano, or guitar, while *FUN* allows adding dynamics and rhythm when locating short-duration illusory actuators on opposite sides when playing different notes.

Technical and psychophysical implications Technical limitations are especially associated with the design of the actuators. During pilot tests, it was determined that the vibrotactors designed for this project had a satisfactory response from 90Hz and above while overheating was negligible for arrangements with durations of around 60s.

Regarding psychophysical limitations, the skin can perceive vibrations from about 3Hz up to 1kHz. Therefore, the effective range of frequencies is from 90Hz to 1000Hz; it means notes from F2# (92.5Hz) to B5 (987.8Hz).

Audio–Tactile mapping As a first step, MIDI-files with notes out of the effective range are transposed in steps of 1 semitone until F_{max} and F_{min} fall between the allowed range. The direction of transposition depends on the limit that is surpassed. If F_{max} exceeds the upper limit then the transposition is towards a lower tone. On the other hand, if F_{min} is below the lower limit then the transposition is towards a higher tone. If after transposing these frequencies remain out of the effective range, the MIDI file is rejected. The number of notes n from the musical excerpt corresponds to the number of *TIs*. The fundamental frequency F_i of every note N_i in the MIDI file corresponds to the frequency of vibration of each TI_i. Onset time t_i and duration D_i correspond to the onset time and duration of TI_i, respectively. Depending on the value of D_i, a different *TI* is selected. In pilot studies, and in [1] and [3], it was found that perceptible *PM* can be obtained for the whole effective range of frequencies (90Hz to 1kHz) and durations from 0.5s up to 2.5s. Longer durations produce an effect of adaptation while shorter ones break the apparent movement into two independent pulses. *FUN* can be obtained with confidence for the whole effective range of frequencies and durations as low as 0.1s [2]. Although over 2.5s the effect of illusory actuator is weak, it remains perceptible. The composition used to validate this algorithm did not have notes longer than 6.8s. The final conditions are as follow: *PM* when 0.5s$\leq D_i \leq$2.5s; *FUN* when 0s$<D_i<$0.5s $\wedge$ $D_i>$2.5s.

The spatial feature SF_i of tactile illusion TI_i (i.e., direction of illusory movement in the case of *PM* and the location of illusory actuator in the case of *FUN*) is assigned depending on the direction of the previous *TI*. If *TI* PM_{i-1} was from left to right (L-R) then PM_i will be from right to left (R-L), and vice versa. If FUN_{i-1} was on the left side (L) (i.e., between the left actuator and the center) then FUN_i is assigned to the right side (R) (i.e., between the right actuator and the center) and vice versa (See Fig. 2b). If the previous *TI* was PM_{i-1} from L-R (or R-L) and the following *TI* is FUN_i, then FUN_i is assigned to the L (or R). Finally, if the previous *TI* was FUN_{i-1} on the L (or R) and the following *TI* is PM_i, then PM_i is assigned movement from L-R (or R-L). The initial value SF_1 is assigned randomly.

The final vector in the tactile domain (i.e., vibrations that can be perceived through touch; although some of them may also be audible) is: V_{Ti} (F_i,t_i,D_i,TI_i,SF_i), where TI_i is the *TI* assigned to note N_i and SF_i is its spatial feature. The flowchart of the algorithm is shown in Fig. 1.

2.2 Experimental evaluation

To evaluate the effectiveness of the proposed algorithm a randomized, blind, controlled experiment was designed - each participant went through a practice session and then was assigned to either experiment A or B.

Participants Twenty volunteers from Universidad Tecnológica Indoamérica (students, lecturers, and staff) were recruited; 50% male and 50% female, 19–52 years old (*M*=28.9, *SD*=9.3). A total of 70% of the subjects described themselves as music-loving non-musicians, 10% as amateur musicians, other 10% as

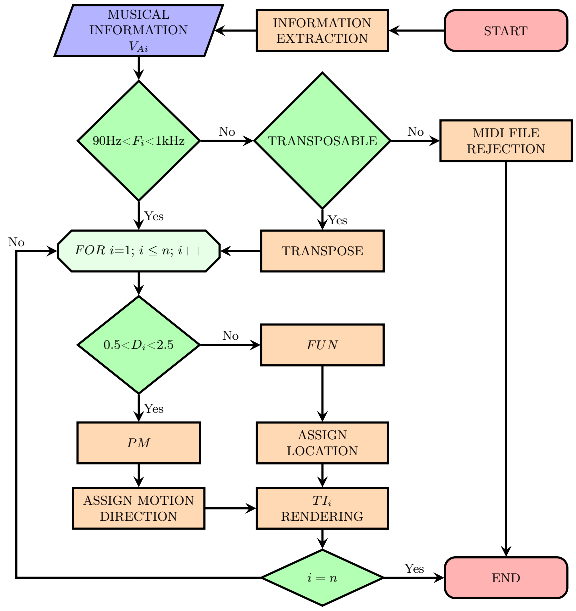


Fig. 1: Flowchart of the proposed algorithm for a MIDI file with n notes (refer to Table 1).

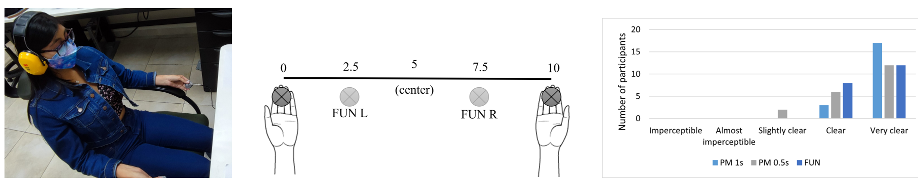
semiprofessional musicians, and only 1 as non music-loving non-musician. Ten of the participants were randomly assigned to experiment A and the rest to experiment B. All of them took part in a practice session.

Practice session The purpose of the practice session was to familiarize the participants with *TIs*. Their recognition is crucial for subjects to be able to evaluate the arrangement of *TIs*.

Apparatus Two analog outputs of a National Instruments DAQ model USB 6002 were routed to a Fosi Audio stereo amplifier model TDA7498E and then to a pair of voice coil actuators built by the authors based on the design proposed in [11] and named "hap-phones". Each actuator was made with a neodymium cylindrical 10mm diameter by 10mm height magnet, 4 layers with 12 loops of 0.5mm diameter copper wire (each coil), and a 32mm diameter 3D printed enclosure.

Stimuli The algorithms to create *TIs* were implemented in Matlab. The model used in [3] was selected to generate *PM* with two durations: 0.5s and 1s. In a pilot study it was observed that *PM* with these durations produced easily perceptible continuous smooth illusory movement. The model proposed in [6] was used to create *FUN* with a duration of 0.5s. Illusory actuators were moving from one side to the other in small separated jumps. *PM* and *FUN* were presented at 250 Hz, the skin's highest sensitivity for vibrations suggested in [9].

Procedure Participants were first instructed to read and sign a consent letter where the general objective and experimental procedure were detailed. After accepting to voluntarily participate in the study, they filled in general information in a questionnaire. Participants were seated in a chair with armrests wearing earmuffs to avoid auditory feedback, and held an actuator in each hand as shown in Fig. 2a. They were asked to close their eyes to avoid visual stimuli and favor concentration. *PM* was presented alternating from one actuator to the other in a loop until the subject reported the perception of movement for each duration (i.e., 0.5s and 1s). The participant then evaluated the clarity of the perceived *TI* on a 5-point scale where 0 is imperceptible, 1 almost imperceptible, 2 slightly clear, 3 clear, and 4 very clear. The procedure was replicated for *FUN*.



(a) (b) (c)

Fig. 2: a) Participant during experimentation, b) Relative locations of illusory actuators for left (2.5) and right(7.5) sides, c) Levels of clarity for *PM* and *FUN*.

The results show that every participant felt both *TIs* (See Fig. 2c). *PM* with both durations was mostly perceived as very clear (*Mo*=4, *Md*=4); 17 participants perceived 1s *PM* as very clear and 3 as clear, while 12 participants perceived 0.5s *PM* as very clear, 6 participants as clear, and just 2 as slightly clear. *FUN* was also perceived mostly as very clear (*Mo*=4, *Md*=4); 12 participants reported very clear and 8 clear perception of the *TI*. Although changes in the location of the illusory actuator were recognized with ease, their location was not precisely identified. Therefore, only two positions were presented on alternating sides: left and right (explained in section 2.1).

Experimental session A The purpose of experiment A was to determine the effect on the emotional response of adding the tactile modality, in the form of a synchronized arrangement of *TIs*, to a musical excerpt.

Apparatus In addition to the setup used in the practice session, an audio interface Focusrite Scarlett 4i4 was integrated to route the different outputs. Outputs 1 and 2, corresponding to vibrotactile stimuli, were sent to the amplifier and then to the actuators, while outputs 3 and 4, corresponding to audio, were routed to a pair of Sennheiser headphones model HD 240 pro.

Stimuli The 4^{th} Movement of "Summer" by Vivaldi played by a violin was selected as the musical composition for this experiment. Using the algorithm proposed, an arrangement of *TIs* was obtained. Figure 3 shows eleven seconds of the synchronized session in Audacity. The top stereo channel contains the signals to prompt the *TIs* which are sent to each actuator, while the bottom channel represents the monophonic MIDI file.

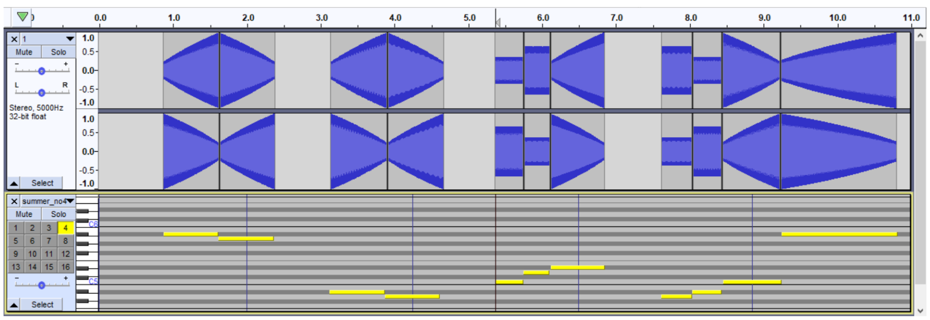


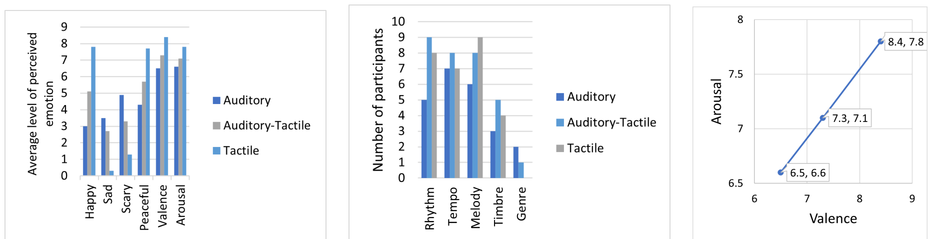
Fig. 3: An 11s excerpt of the MIDI track synchronized with the *TIs*.

Procedure Participants were instructed to assess the extent to which they felt "Happy", "Sad", "Scary" and "Peaceful" on a scale from 0 absent to 9 present (as done in [10]), as well as the valence and arousal of the excerpt from 0 to 9 (unpleasant-pleasant and relaxing-stimulating, respectively). All participants also reported recognition of musical features: rhythm, tempo, melody, timbre, and genre; a quick and clear description of each feature was presented to avoid confusion. After that, the musical composition synced with the arrangement of *TIs* was presented to the participants, and the evaluation was repeated following the same previous instructions.

Experimental session B This experiment aimed to determine the emotional response to the arrangement of *TIs* alone, and to then ascertain any significant difference from the emotional response to the audio excerpt. The apparatus and setup were similar to the practice session. In this experiment, the arrangement of *TIs* was presented without audio.

3 Results

Since a 90% confidence interval overlapped for every emotion, it was not possible to assign a specific emotion to the musical composition [10] - suggesting that the composition elicited blended emotions. Therefore, the statistical analysis focused on each emotion per condition: auditory, auditory-tactile, and tactile (Fig 4a).



(a) (b) (c)

Fig. 4: a) Average emotional evaluation, valence and arousal , b) Recognition of musical features, c) Relation of increase in average valence and arousal for auditory (left), auditory-tactile (center), and tactile (right) conditions.

Using Q-Q graphics it was found that data does not present a normal distribution. Hence, to compare the auditory and auditory-tactile conditions (performed in the same group of participants) a non-parametric Wilcoxon's T-test was used. Only for the emotion labeled as "Happy" a significant increment was found when adding the tactile modality to the auditory, with $T = 0$, $p \leq .05$. For the emotions "Sad", "Scary", and "Peaceful" no significant difference was found (T=19, p>.05; T=33, p>.05; T=6.5, p>.05). Regarding valence and arousal no significant effect was found (T=10, p>.05; T=11.5, p>.05).

To evaluate the difference between the emotional response for auditory and tactile conditions a non-parametric Mann-Whitney U-test was used. Results show that compared to the auditory condition, the tactile condition was perceived significantly happier and more peaceful (U=61, p<.05; U=70.5; p<.05), significantly less sad and scary (U=134, p<.05; U=137, p<.05), and significantly more agreeable and stimulating (U=66.5, p<.05; U=77, p<.05).

To compare the auditory-tactile and tactile conditions a non-parametric Mann-Whitney U-test was used. The tactile condition was significantly happier than the auditory-tactile (U=75, p<.05). For the emotions "Sad", "Scary", and "Peaceful" no significant difference was found (U=129, $p \geq$.05; U=130.5, $p \geq$.05; U=80.5, $p \geq$.05). Regarding valence and arousal, no significant effect was found (U=80.5, $p \geq$.05; U=104, $p \geq$.05). The relation of increase in valence and arousal for auditory, auditory-tactile, and tactile conditions is shown in Fig. 4c.

Rhythm, tempo, and melody were musical features mostly recognized in the auditory-tactile (90%, 80%, 80%, respectively) and tactile (80%, 70%, 90%, respectively) conditions. The less recognized element was genre with 20% for auditory, 10% for auditory-tactile, and 0% for tactile. The best recognized feature in the auditory condition was tempo (70%), in the auditory-tactile condition was rhythm (90%) and in the tactile condition was melody (90%). Overall, musical features were recognized to a grater extent in the auditory-tactile condition, followed by the tactile condition, and finally the auditory condition (see Fig. 4b).

4 Discussion and conclusions

We have proposed an algorithm to map monophonic MIDI files to arrangements of *PM* and *FUN* tactile illusions (*TI*). Each note of a MIDI file was assigned a *TI* based on various aspects: onset time, duration, and fundamental frequency. Transposition of notes was required when the maximum and/or minimum frequencies of the musical composition were out of the effective range (90Hz - 1kHz); a limitation that causes the rejection of some musical compositions. Changes in the direction of *PM* and the location of *FUN* were used to control the dynamics of the *TIs* arrangement. This was supported by a significant increase in perceived arousal when comparing the auditory with the tactile condition. The auditory-tactile condition was found to be "happier" than the auditory condition with no effect identified for valence and arousal. However, the tactile condition presented significant changes in the emotional response compared to the auditory condition as well as in valence and arousal. This was probably because of the novelty of the stimuli and because participants were more focused on the tactile modality (auditory and visual feedback were blocked).

Although the tactile condition was perceived as more positive, it was not possible to assign a specific emotion to any condition, becoming a challenge to define whether the arrangement of *TIs* evoked similar or different emotional responses. Surprisingly, the arrangement of *TIs* conveyed rhythm, tempo, and melody to most of the participants and enhanced their perception when presented along with the audio. This suggests that even considering the technical and psychophysical limitations, key elements of music can be embedded into arrangements of *TIs*. Regardless, experimentation with more participants is required to properly support our suggestions. Currently, we are exploring the emotional response to arrangements of *TIs* rendered from musical excerpts with tested and confirmed emotional labels (happy, sad, scary, and peaceful) in [10]. In future work we hope to find an algorithm that combines various resources to properly translate music to vibrotactile stimuli.

References

[1] Angelina Bellicha, Andres Trujillo-Leon, and Wael Bachta. “Phantom Sensation: When the phantom escapes the bounds of the actuators and the end-point is sensed in the air”. In: *2019 IEEE World Haptics Conference (WHC)*. IEEE, July 2019, pp. 91–96. ISBN: 978-1-5386-9461-9. DOI: 10.1109/WHC.2019.8816176. URL: https://ieeexplore.ieee.org/document/8816176/.

[2] Ali Israr and Ivan Poupyrev. “Tactile brush”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, May 2011, pp. 2019–2028. ISBN: 9781450302289. DOI: 10.1145/1978942.1979235. URL: https://dl.acm.org/doi/10.1145/1978942.1979235.

[3] Jongman Seo and Seungmoon Choi. "Initial study for creating linearly moving vibrotactile sensation on mobile device". In: *2010 IEEE Haptics Symposium*. IEEE, Mar. 2010, pp. 67–70. ISBN: 978-1-4244-6821-8. DOI: 10.1109/HAPTIC.2010.5444677. URL: http://ieeexplore.ieee.org/document/5444677/.

[4] Maria Karam, Frank A. Russo, and Deborah I. Fels. "Designing the Model Human Cochlea: An Ambient Crossmodal Audio-Tactile Display". In: *IEEE Transactions on Haptics* 2.3 (2009), pp. 160–169. ISSN: 19391412. DOI: 10.1109/TOH.2009.32.

[5] Maria Karam et al. "Towards A Model Human Cochlea: Sensory Substitution for Crossmodal Audio-Tactile Displays". In: *Graphics Interface Conference*. 2008, pp. 267–274. DOI: 10.5555/1375714.1375759.

[6] Jaedong Lee, Youngsun Kim, and Gerard Jounghyun Kim. "Rich Pinch: Perception of Object Movement with Tactile Illusion". In: *IEEE Transactions on Haptics* 9.1 (Jan. 2016), pp. 80–89. ISSN: 1939-1412. DOI: 10.1109/TOH.2015.2475271. URL: https://ieeexplore.ieee.org/document/7234937/.

[7] Neil P. McAngus Todd. "The kinematics of musical expression". In: *The Journal of the Acoustical Society of America* 97.3 (June 1998), p. 1940. ISSN: 0001-4966. DOI: 10.1121/1.412067. URL: https://asa.scitation.org/doi/abs/10.1121/1.412067.

[8] Sebastian Merchel and M. Ercan Altinsoy. "Auditory-Tactile Experience of Music". In: *Musical Haptics*. Springer, Cham, 2018, pp. 123–148. DOI: 10.1007/978-3-319-58316-7{_}7. URL: https://link.springer.com/chapter/10.1007/978-3-319-58316-7_7.

[9] Byron Remache-Vinueza et al. "Audio-Tactile Rendering: A Review on Technology and Methods to Convey Musical Information through the Sense of Touch". In: *Sensors 2021, Vol. 21, Page 6575* 21.19 (Sept. 2021), p. 6575. ISSN: 14248220. DOI: 10.3390/S21196575. URL: https://www.mdpi.com/1424-8220/21/19/6575/htm%20https://www.mdpi.com/1424-8220/21/19/6575.

[10] Sandrine Vieillard et al. "Happy, sad, scary and peaceful musical excerpts for research on emotions". In: *https://doi.org/10.1080/02699930701503567* 22.4 (2008), pp. 720–752. DOI: 10.1080/02699930701503567. URL: https://www.tandfonline.com/doi/abs/10.1080/02699930701503567.

[11] Hsin-Yun Yao and Vincent Hayward. "Design and analysis of a recoil-type vibrotactile transducer". In: *The Journal of the Acoustical Society of America* 128.2 (Aug. 2010), pp. 619–627. ISSN: 0001-4966. DOI: 10.1121/1.3458852. URL: http://asa.scitation.org/doi/10.1121/1.3458852.