Effect of audio-tactile congruence on vibrotactile music enhancement

Aker, Scott C.; Innes-Brown, Hamish; Faulkner, Kathleen F.; Vatti, Marianna; Marozeau, Jeremy

Published in:
Journal of the Acoustical Society of America

Link to article, DOI:
10.1121/10.0016444

Publication date:
2022

Document Version
Peer reviewed version

Citation (APA):
Effect of audio-tactile congruence on vibrotactile music enhancement

Scott C. Aker\textsuperscript{1,a}, Hamish Innes-Brown\textsuperscript{3}, Kathleen F. Faulkner\textsuperscript{2}, Marianna Vatti\textsuperscript{2}, and Jeremy Marozeau\textsuperscript{1}

\textsuperscript{1}Music and Cochlear Implant Lab, Department of Health Technology, Technical University of Denmark, Kongens Lyngby, 2800, Denmark
\textsuperscript{2}Oticon Medical, Smørum, 2765, Denmark
\textsuperscript{3}Eriksholm Research Centre, Snekkersten, 3070, Denmark

Music listening experience can be enhanced with tactile vibrations. However, it is not known which parameters of the tactile vibration must be congruent with the music to enhance it. Devices which aim to enhance music with tactile vibrations often require coding an acoustic signal into a congruent vibrotactile signal. Therefore, understanding which of these audio-tactile congruences are important is crucial. Participants were presented with a simple sine wave melody through supra-aural headphones and a haptic actuator held between the thumb and forefinger. Incongruent versions of the stimuli were made by randomizing physical parameters of the tactile stimulus independently of the auditory stimulus. Participants were instructed to rate the stimuli against the incongruent stimuli based on preference. It was found making the intensity of the tactile stimulus incongruent with the intensity of the auditory stimulus, as well as misaligning the two modalities in time, had the biggest negative effect on ratings for the melody used. Future vibrotactile music enhancement devices can use time alignment and intensity congruence as a baseline coding strategy, which improved strategies can be tested against.

\textsuperscript{a} Also at: Oticon Medical, Smørum, 2765, Denmark.
\textsuperscript{b} Also at: Hearing Systems Section, Department of Health Technology, Technical University of Denmark, Kongens Lyngby, 2800, Denmark.
I. INTRODUCTION

Tactile vibrations can enhance music listening experiences (Merchel & Altinsoy, 2014). This phenomenon might not surprise anyone who enjoys loud concerts, which can elicit vibrations perceivable with the tactile system (Merchel & Altinsoy, 2013). Naturally, one might assume this benefit depends on the tactile vibrations being congruent with its sound source. However, it is not known which parameters of the tactile vibration, such as its intensity or frequency, need to match those of the music to enhance it. In this study, we investigated if tactile vibrations must be congruent with acoustic parameters of a melody to enhance it, and if so, along which parameters.

Both the auditory and tactile system can perceive vibrations (Von Bekesy, 1959). While the auditory system is specialized to perceived vibrations in the air between 20 Hz and 20 kHz, the tactile system is specialized to perceive low frequency vibrations through solid surfaces up to 1 kHz. Regardless, both modalities are sensitive to many of the same physical parameters. Temporal discrimination of sinusoids is similar for both modalities (Van Doren et al., 1990; Moore et al., 1993), as is intensity discrimination (Craig, 1972; Florentine et al., 1987). One difference is the frequency selectivity of both modalities. Apart from a wider frequency range, the auditory system also has smaller just-noticeable differences (JND) in frequency, at less than 0.59 Hz for a 125 Hz sinusoid (Moore, 1973) compared to the tactile system’s frequency JNDs around 40 Hz for a 100 Hz sinusoid (Goff, 1967). There are also differences in how each physical parameter is perceived. A sound’s perceived intensity is described as its loudness, and its fundamental frequency as pitch. A tactile vibration’s perceived intensity is commonly described as its vibrotactile intensity. As described by (Verrillo et al., 1969), vibrotactile intensity is proportional to the amplitude of a sinusoidal vibration, and therefore its velocity and acceleration, and is frequency dependent. While “vibrotactile pitch” does not exist as it does in the auditory system, perceived vibrotactile frequency is defined as the subjective frequency of a sinusoidal vibration when vibrations are intensity matched (detailed by Bensmaïa et al., 2005).
For a full review comparing the psychophysics of auditory and tactile perception, see Merchel & Altinsoy, 2020).

In short, the perception of vibrations in the auditory and tactile system are similar, but far from identical. However, by restricting stimuli to pure tone sinusoids and equalizing the vibrotactile intensity across frequency, one can equivalate auditory parameters (loudness, frequency, timing) to tactile parameters (vibrotactile intensity, perceived vibrotactile frequency, timing) to create series of audio-tactile congruences which can be explicitly coded into an audio-tactile set-up. In this study, congruent parameters are defined as agreeing in movement or directionality. The congruences, out of these equivalencies, which confer a benefit for vibrotactile music enhancement can then be identified. Interestingly, musicians have been shown to have better vibrotactile frequency discrimination than non-musicians (Sharp, Houde, Maheu, Ibrahim, & Champoux, 2019). It consequently is important to consider the effect of musicianship as a possible interaction as well. Given what is known about multisensory integration, it is reasonable to assume a tactile vibration must be congruent with music for it to enhance it. A key principle of multisensory integration is temporal synchrony (Holmes & Spence, 2005; Stein & Meredith, 1993). Auditory stimulation can influence the number of perceived tactile taps, an effect which is affected by temporal synchrony (Bresciani et al., 2005). Tactile stimulation has been known to increase auditory loudness perception (Schürmann et al., 2004), an effect which is also dependent on time synchrony (Gillmeister & Eimer, 2007). Wilson et al. found the increase in performance of an audio-tactile detection task due to audio-tactile integration to be dependent on both temporal synchrony (2009) and a similar frequency (2010) between the auditory and tactile stimuli. Yau et al. (2010) have also shown the effect of tactile distractors on an auditory intensity and frequency discrimination task. However, the effect of the distractors was time dependent only in the auditory intensity discrimination task. Therefore, while
timing alignment, intensity congruence, and frequency congruence all have a role in audio-tactile integration, the perceptual mechanisms that guide them are likely different.

However, it is also possible full multisensory integration is not necessary for vibrotactile music enhancement. The tactile system is shown to have beat perception (Brochard et al., 2008). Additionally, it has been shown that tactile rhythms can be integrated into auditory rhythms to detect meter (Huang et al., 2012). There is evidence that rhythm perception is shared between the auditory and haptic systems (Bernard et al., 2022). Vibrotactile music effect could emerge from perceiving rhythm separately in the auditory and tactile systems, then integrating them at a higher level as two complementary rhythms.

Understanding which audio-tactile congruences affect vibrotactile music enhancement is beneficial for the design of wearable, vibrotactile devices for music listening. These devices are often created, used, or tested for enhancing music for music enthusiasts or people with hearing impairments (Fletcher, 2021) and can be easily described in an audio-tactile congruence framework. Common to most vibrotactile music enhancement devices is audio-tactile intensity congruence and timing alignment. The devices differ on the location of stimulation, and how to code the frequency information of the acoustic signal.

Devices which utilize frequency congruence often use voice coils (VCMs), linear resonant actuators (LRAs), or contact speakers, which can manipulate the intensity and frequency of the vibration independently. These include three commercial devices, the Woojer Vest, Woojer Strap (Woojer Incorporated, 2020), and the SUBPAC X1 (SUBPAC, 2022), and the Haptic-Chair (Nanayakkara et al., 2009), all of which are intended to vibrate the chest or back. The SUBPAC has also been shown to convey certain musical emotions to deaf users ((Schmitz et al., 2020)). The mosaicOne_B (Fletcher et al., 2020), shown to improve pitch perception for cochlear implant users, and the Pump and Vibe (Haynes et al., 2021), use an array of ERM actuators along the forearm. ERMs cannot
manipulate frequency and intensity independently, each actuator is coded based on fundamental frequency for congruence between the acoustic fundamental frequency and tactile location on the arm. The Pump and Vibe also utilizes an upper arm squeeze, which is synchronized to the lowest fundamental frequencies of the audio. The Emoti-Chair (Karam et al., 2010), run by the Model Human Cochlea (Karam et al., 2009), is a chair configuration with an array of VCMs on the back. The Emoti-Chair was tested with two models, one in which the rows of VCMs corresponded to frequency bands of the music, and one in which the rows of VCMs corresponded to different instrument tracks. The Emoti-Chair can therefore have frequency congruence and acoustic frequency to tactile location congruence in the frequency model. (For a full review on devices which convey music through haptics, see Remache-Vinueza et al., 2021).

It is difficult to compare the benefit of each acoustic to tactile congruence directly across devices. It is possible a device could have a relatively poor coding strategy but is still preferred due to a more comfortable set-up or desirable stimulation location. To test a coding strategy without this confound, a device’s audio-tactile coding strategy can be tested against audio with unrelated tactile vibrations but with similar physical parameters. The benefit of a specific coding strategy can then be disentangled from the benefits of a stimulation location or physical aspect of the device. Specific audio-tactile congruences can also be tested by randomizing only specific parameters of the tactile stimulation. Identifying which audio-tactile congruences directly contribute to vibrotactile music enhancement would allow these relationships to be explicitly coded into a device, while audio-tactile congruences that do not contribute to music enhancement could be reconfigured to convey other musical information. Alternatively, cheaper ERMs that cannot convey all tactile parameters independently could be used over VCMs or LRAs.

In the current investigation it was hypothesized that completely arbitrary tactile vibrations would not enhance music enjoyment ratings to the same extent as tactile vibrations that are congruent to the
music. Subsequently, it was hypothesized that the importance of congruence would depend on which tactile parameters were congruent with the auditory parameters of the music. The audio-tactile congruences tested include auditory and tactile intensity congruence, auditory and tactile frequency congruence, and timing alignment. The relatively poor ability to perceive frequency with the tactile system could also influence the vibrotactile music enhancement. If so, by stretching the frequency range of the tactile vibrations, the importance of frequency congruence between the auditory and tactile stimuli could be increased. Additionally, we will test for an interaction with musicianship.

The hypotheses were tested in a rating task, in which participants could use an interface to present four audio-tactile stimuli at will and rate them between zero and 100 based on preference. The stimuli contained a 20 second melody with a simultaneous auditory and tactile component. In each stimulus, a different tactile parameter (intensity, frequency, timing) was made congruent or incongruent with the associated auditory parameter to test its effect on participant ratings. Six conditions were tested in total with one repetition, for 24 unique audio-tactile stimuli.

As certain tactile stimuli might be innately more pleasant than another, the distributions of the physical parameters of the tactile components were kept as similar as possible. This helped ensure the only difference between each stimulus was the congruence between its auditory and tactile components.

The experiment was also preceded by a short training regimen to familiarize participants with the audio-tactile stimuli, and the Goldsmiths Musical Sophistication Index (Gold-MSI) questionnaire to measure the effect of musical training and general music sophistication on the results (Müllensiefen et al., 2014).
II. METHODS

A. Participants

Participants consisted of twenty adult listeners (8 women, 12 men) ranging in age from 21 to 30, (median = 25) with hearing thresholds equal to or below 20 dB HL at 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz, and 6000 Hz. Participants had a wide range of musical training based on the Gold-MSI. Participants provided informed consent and all experiments were approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391) and were compensated financially for their time unless they voluntarily declined. All relevant regulations were followed.

B. Stimuli

1. Auditory Stimuli

Calibrated auditory stimuli were presented over a pair of Sennheiser HD200 headphones through a RME Fireface UCX soundcard and Phonitor mini headphone amplifier. Stimuli was controlled through a MATLAB interface.

The auditory stimulus was a 20 second clip of a musical signal based on a pop song from the Hungarian band “Chalga” titled “szárad a száj.” The musical signal was created by recreating the bass track of the song into midi format using Ableton Live 10 at 100 beats per minute (BPM). The midi file was then transposed so that the lowest note of the file corresponded to a frequency of 73.42 Hz (D2) and the highest corresponded to a frequency of 164.81 Hz (E3). The velocities of the midi notes were either set to 127 or 64. The velocities, or intensity, of the notes were determined manually based on the original bass track. The midi file was then transformed into an acoustic signal using MATLAB. Each midi note was transformed into a sine wave with a 31 ms linear onset and offset. The frequency of each note was determined by the pitch of the corresponding midi note, or
in cases when the pitch was fixed, by the lowest frequency of 73.42 Hz. The amplitudes of the notes were calibrated to 70 dB SPL for 127-velocity notes, or 58 dB SPL for 64-velocity notes.

In tasks without musical stimulus, pink masking noise was presented at 70 dB SPL to ensure the sound of the haptic actuator was inaudible. In tasks with musical stimulus, pink masking noise was presented at a constant level to ensure an overall SNR of 30 dB. An SNR of 30 dB was decided informally to sufficiently mask the noise while keeping the auditory stimulus completely audible.

2. Tactile Stimuli

Tactile stimuli were presented through a Lofelt L5 haptic actuator through an RME Fireface UCX sound card, Phonitor mini headphone amplifier, and a K4001 7W amplifier. Tactile stimuli were controlled through a MATLAB interface. The actuator was held between the participants left-hand thumb and forefinger inside a fabric-lined cardboard box, which was used to both hide the actuator visually as well as dampen the sound.

The tactile stimulus was based on a midi file identical to the one used for the auditory stimulus. The amplitude of the sine wave was adjusted based on the velocity of the midi note. A velocity of 127 corresponded to a vibrotactile intensity perceptually equivalent to a 100 Hz sine wave with an RMS voltage of 600 mV at the actuator, and a velocity of 64 corresponded to the vibrotactile perceptual equivalent of a 100 Hz sine wave with an RMS voltage of 300 mV at the actuator.

The frequency of each note was determined in one of three ways: a fixed frequency of 73.42 Hz, the fundamental frequency of the pitch of the corresponding midi note, or the pitch of the corresponding midi note with a stretching formula (1) applied. The frequency stretching formula took the frequency of the corresponding midi note and multiplied it by 1.5 over the lowest possible note. In other words, the distance between each note was 1.5 times bigger than the distance without the stretching formula applied. The formula can be seen in (1). $f_{\text{orig}}$ is the corresponding midi note frequency, $\alpha$ is the stretching factor equal to 1.5, and $f_{\text{min}}$ is the lowest frequency at 73.42 Hz (D2).
A factor of 1.5 was chosen as the highest number that still ensured the tactile stimulus was still within the Lofelt’s L5’s effective frequency range.

\[ f_{\text{stretch}} = a \times (f_{\text{orig}} - f_{\text{min}}) + f_{\text{min}} \]  

The vibrotactile stimulus could also be randomized in one of three ways: timing, intensity, and frequency. Randomization was always performed on the midi file prior to waveform generation. Intensity randomized by shuffling the order of the midi velocities of the midi file prior to waveform generation. Frequencies were similarly randomized by shuffling the order of the midi pitch numbers of the midi file prior to waveform generation. Both methods ensured randomization did not overly affect the physical properties of the tactile vibration sequence.

Timing was randomized by manually dividing the midi file into musical phrases based on the original melody. Permutations of the note onsets and durations were generated for each phrase which maintained the same number of total notes, and the same total note duration as the original midi file. One permutation was then randomly selected for each phrase of the tactile stimulus with randomized timing. In other words, each phrase of the randomized rhythm had the same number of notes and total note duration as the original rhythm. This randomized the rhythm of the tactile stimulus without overtly affecting its physical properties, while still ensuring the timing of the tactile stimulus was incongruent with the timing of the auditory stimulus.

The method of randomization used a sixteenth note as the smallest possible denomination, equally to 150 ms in length. Therefore, notes in the tactile stimulus were only ever fully aligned or fully unaligned with the auditory stimulus.

C. Procedure

The testing procedure consisted of four components: the Gold-MSI questionnaire, a vibrotactile intensity matching task, a training session, and the rating task. Each component took approximately 10-20 minutes, for a total of around one hour.
1. **Goldsmiths Musical Sophistication Index**

Participants completed the Gold-MSI questionnaire to gather background data on their musical experience. The Gold-MSI is a questionnaire consisting of five categories: active engagement, perceptual abilities, musical training, emotions, and singing abilities, and when added together, general sophistication. Questions range from lifestyle questions (“I don’t spend much of my disposable income on music.”) to questions on training and expertise (“I engaged in regular, daily practice of a musical instrument [including voice] for ___ years.”). Each category can be scored separately or together. The Gold-MSI was administered to capture two scores: general sophistication and musical training.

2. **Intensity Matching**

A vibrotactile intensity matching task was conducted with test frequencies 50, 75, 100, 125, 150 and 180 Hz and a reference vibration of 100 Hz at 600 mV for a high-intensity level and 300 mV for a low-intensity level. Participants were instructed to adjust a virtual knob, which controlled the amplitude of the test tone, so that the perceived intensity of the test vibration and reference vibration were the same. The results for each intensity level were then interpolated linearly in dB to obtain an equal-intensity curve for the high-intensity and low-intensity reference level for each participant. All presented tactile vibrations afterwards were equalized in perceived intensity based on each participant’s individualized results.

3. **Training**

Participants also completed a three-part training session to ensure they could discriminate between tactile stimuli. First, participants were presented with four buttons which presented sinusoidal tactile vibrations which differed in intensity and frequency. Two tactile vibrations had a frequency of 85 Hz, two with a frequency of 135 Hz, two with an equalized “strong” intensity and two with an equalized “weak” intensity. Participants were asked to identify the differences between them. If the participant
could not identify the differences, the experimenter explained the difference. The participant was then encouraged to use the buttons feel the stimuli as many times as they wanted until they felt they could identify the difference.

The second part consisted of a 3-AFC structured task with a three-tone tactile sequence as its three stimuli. In one of the three intervals, either the frequencies of the tactile sequence or the intensities of the tactile sequence would be randomized. Participants were then asked to identify the stimuli that had been altered. The task was fixed in difficulty, with no tracking. Once the participant's running score of the last 10 trials exceeded 80%, the training task ended.

Finally, the participant was then presented with three buttons, each which presented a short audio-tactile musical stimulus. In all three stimuli, the auditory stimulus was the same and was generated by a timing randomization of the first two phrases of the original midi stimuli. Each tactile stimulus was different and provided plausible alternative strategies in which an acoustic signal could be coded into a vibrotactile signal. These alternative strategies were intended to disguise the true intent of the experiment, so participants would assume they were rating different methods to convert an acoustic signal into a vibrotactile signal.

In the first coding strategy, the tactile stimulus was generated from the same midi file as the auditory stimulus. In the second coding strategy, the frequency of the tactile stimulus was rounded either to the highest frequency or the lowest frequency. The frequency and intensity of the stimulus were then swapped: notes with low frequency were assigned a high intensity and vice versa, and notes with a high frequency were assigned a low intensity. The third coding strategy then had the timing of its tactile stimulus altered by selecting another random permutation of the timing of the original in which the number of notes and total duration of notes were kept constant.

Participants were asked to identify the difference between the three stimuli. If the participant could not identify the differences, the experimenter explained the difference, but did not assign a value
judgement to any of them. For this reason, no stimuli were described as random, only that each had different processes applied to convert the acoustic signal into a vibrotactile signal. The participant was then encouraged to use the buttons present the stimuli as many times as they wanted until they felt they could identify the difference.

4. Rating Task

The rating task used a GUI similar to a MUSHRA experiment. Participants controlled four buttons (Fig. 1), in which each button presented an audio-tactile stimulus to the participant. The button assigned to each stimulus was randomized for each screen. Each stimulus consisted of the same auditory stimulus but a different tactile stimulus. Unlike a MUSHRA, there was no reference stimulus.

Participants were instructed to rate each audio-tactile stimulus with an on-screen slider between zero and one hundred based on how much they liked it. Participants were instructed to take both the auditory and tactile stimulus into account for their judgement, however, they were also told the auditory stimulus would be the same in each audio-tactile combination. Participants were not told that any tactile stimuli would be randomized, only that alternative strategies to convert an acoustic signal into a vibrotactile signal would be applied. However, they were encouraged to consider carefully what aspects of each alternative processing they preferred. Was not included to ensure every stimulus had a similar distribution of tactile parameters. Like a MUSHRA, participants were instructed that they could play the stimuli in any order and as many times as desired, to only compare stimuli within a screen, and to be as specific as possible with their rating. When participants were satisfied with their ratings, they could advance to the next condition.
FIG. 1. The screen presented to participants in the rating task. Each button played the same auditory stimulus but a different tactile stimulus. Each tactile stimulus had different parameters made congruent or incongruent to the auditory stimulus. The effect of audio-tactile congruence on participant ratings could then be measured.

Each screen in the experiment was designed to test the effect of a specific audio-tactile congruence on vibrotactile music enhancement. There were twelve stimuli included in the final analysis divided into three conditions. One screen tested the effect of intensity congruence between the auditory and tactile stimuli, called the intensity test condition. One screen tested the effect of frequency congruence between the auditory and tactile stimuli, called the frequency test condition. Finally, one screen tested the effect of frequency congruence between the auditory and tactile stimuli, but with the frequency range of the tactile stimuli stretched, called the stretched frequency test condition. Stimuli pairs which were misaligned in time were present in every screen as an anchor. Screen order was randomized, and the final screen was a duplicated version of the first presented screen to test
reliability for four screens in total. Originally, six screens were used, including intensity-frequency test conditions, however they were ultimately disregarded due to a programming error.

a. Intensity Test Condition

The first screen was designed to test the effect of audio-tactile intensity congruence on participant preference. The screen presented four different audio-tactile stimuli, in which the auditory stimulus was always the same but the tactile stimulus was different. The auditory and tactile stimulus were versions in which the frequency was kept constant. The tactile stimuli varied in congruence, with three different levels of congruence distributed amongst the four audio-tactile stimuli. Examples of the stimuli in the intensity test condition are included in Figure 2.

In the first level of congruence, (Stimulus 1), the audio-tactile stimulus had the tactile stimulus time aligned with the auditory stimulus and the high and low intensities of the tactile stimulus aligned with the high and low intensities of the auditory stimulus. The frequency of the tactile stimulus was kept fixed to ensure it had no effect on the participant’s vibrotactile intensity perception and to stay congruent with the frequency of the auditory stimulus.

In the second congruence level, (Stimuli 2 and 3) the timing and frequency of the tactile stimuli were kept congruent with the auditory stimuli, however the intensity of the tactile stimulus was randomized. Two different stimuli, 2 and 3, were generated with different randomization.

In the third level, (Stimulus 4), both the timing and intensity of the tactile stimulus was randomized.
FIG. 2. An example of the stimuli on the screen that tests intensity congruence. The x-axis represents time, and the y-axis represents frequency, organized like notes on a piano. Each column represents one of the four audio-tactile stimuli. In the first stimulus, the high and low intensities of the tactile component and congruent with the high and low intensities of the auditory component, and the two components are aligned in time. In the second and third stimuli, the high and low intensities of the tactile stimulus are incongruent with the high and low intensities of the auditory stimulus. In the fourth stimuli pair, the tactile component is incongruent in intensity and time to the auditory component.

b. Frequency Test Condition

The frequency test condition was designed to test the effect of frequency congruence between the auditory and tactile stimuli on participant preference. Examples of the stimuli pairs in the second screen can be seen in Figure 3. Like the first screen, the second screen presented four different audio-tactile stimuli with four identical auditory stimuli and four different tactile stimuli. The frequency of the auditory stimulus matched that of fundamental frequency of the transposed original song. Each tactile stimulus varied in congruence to the auditory stimuli.

In the first level of congruence (Stimulus 1), the tactile stimulus was time aligned with the auditory stimulus and the high and low intensities of the tactile stimulus were aligned with the high and low intensities of the auditory stimulus. In addition, the frequency of the tactile stimulus was matched to the frequency of the auditory stimulus.

In the second level of congruence (Stimulus 2), the tactile stimulus was time aligned with the auditory stimulus, however the intensity of the tactile stimulus was randomized independently of the auditory stimulus. This was done so that vibrotactile intensity could not be used as a vibrotactile frequency cue.
In the third level of congruence (Stimulus 3), the intensity and the frequency of the tactile stimulus were both randomized independently of the auditory stimulus.

In the fourth level of congruence (Stimulus 4), intensity, frequency, and timing of the tactile stimulus was randomized independently of the auditory stimulus.

FIG. 3. An example of the stimuli on the screen that tests frequency congruence. The x-axis represents time, and the y-axis represents frequency, organized like notes on a piano. Each column represents one of the four audio-tactile stimuli. In the first stimulus, the intensity and frequency of the tactile component is congruent with the intensity and frequency of the auditory component, respectively. In the second stimulus, the intensities of the tactile component are incongruent with the intensities of the auditory component. In the third stimulus pair, the frequencies of the tactile component are incongruent with the frequencies of the auditory component. In the fourth stimulus pair, the intensities, frequencies, and timing of the tactile component are all incongruent with the auditory component.

c. **Stretched Frequency Test Condition**

The third screen was designed to test auditory frequency and tactile frequency congruence in more favorable conditions for vibrotactile frequency discrimination. Examples of the stimuli on the third screen can be seen in Fig. 4. Again, the third screen presented four audio-tactile stimuli with identical auditory stimuli and tactile stimuli which varied in its congruence to the auditory stimuli.

The frequency of the auditory stimulus matched the fundamental frequency of the original song.
In the first level of congruence (Stimulus 1), the tactile stimulus was time aligned with the auditory stimulus and the high and low intensities of the tactile stimulus were aligned with the high and low intensities of the auditory stimulus. The frequencies of the tactile stimulus were the frequencies of the auditory stimulus with the stretching formula applied (1). The three remaining congruence levels were the same as the frequency test condition. The second congruence level (Stimulus 2) had the intensity of the tactile stimulus randomized. The third congruence level (Stimulus 3) had the intensity and frequency of the tactile stimulus randomized. The frequency of the tactile stimulus was randomized after the stretched formula was applied (1). The fourth congruence level (Stimulus 4) had the intensity, frequency, and timing of the tactile stimulus randomized.

FIG. 4. An example of the stimuli on the screen that tests frequency congruence with a stretched vibrotactile frequency range. The x-axis represents time, and the y-axis represents frequency, organized like notes on a piano. Each column represents one of the four audio-tactile stimuli. The congruence of the stimuli follows a similar pattern to those seen in Figure 3. Note, the frequencies of the tactile component have a greater distance between them than in the auditory component, differentiating the stretched frequency congruence test condition from the frequency congruence test condition.
III. RESULTS

A. Goldsmiths Musical Sophistication Index

The Gold-MSI was administered to capture two scores: general sophistication and the musical training sub-score. The general sophistication scored by the Gold-MSI did not show any significant ability to categorize participants. A histogram of the musical training sub-score of the Gold-MSI is shown in Figure 5. The musical training sub-score reflects more musical training with a higher score. Participants could be classified into two groups based on the Musical Training sub-score of the Gold-MSI with a k means clustering. The higher scoring participants can be labelled “musicians” and lower scoring participants “non-musicians.”
FIG 5: A histogram of the participants’ musical training sub score from the Goldsmith Musical Sophistication Index (Gold-MSI). Participants could be easily divided into two groups of 12 and 9 using k means clustering based on music training. Centroids of the clusters are shown in dotted lines. The dark gray coloring shows participants with lower scores, labelled “non-musicians,” and the light gray coloring shows participants with higher scores, labelled “musicians.”

B. Training Task

FIG 6: A histogram of the participants’ number of trials in the training task. After ten trials, the task took the percent correct score of a participants last ten trials. If the score ever exceeded 80%, the training task ended.
The training task was analyzed by determining how many trials were required for each participant to reach a running score of 80%. Figure 6 shows a histogram of the participants’ number of trials completed, where a lower number means the participant reached 80% faster with a minimum of ten. The participant with the greatest number of trials required to complete was 46, and the median 23.5 (22 and 25). There was no significant difference in performance for the trials which had randomized intensity and the trials which had randomized frequency.

C. Rating Task

Ratings were analyzed through a Friedman test, which analyzes non-parametric data based on rank. Post-hoc analysis was done with a Wilcoxon signed rank test with Holm correction and a linear model with arcsine transformed data. The linear model was used to investigate the effect size of each audio-tactile congruence on participant ratings. A larger estimated coefficient signifies a larger effect size on participant ratings.

1. Intensity Test Condition

Results for the intensity test condition are shown in Figure 7. In the intensity test condition, which measured the effect of intensity congruence, participant preferences were reduced when only acoustic intensity and vibrotactile intensity were made incongruent (middle two columns) and lowest when, in addition, the stimuli were misaligned in time (right column). A Friedman test showed a significant effect of audio-tactile congruence ($p = 2.56e-6^{***}$). A post-hoc analysis with a Wilcoxon Signed Rank Test showed a significant difference in means between all vibrotactile stimuli except the two, intensity-only incongruent stimuli ($p=0.867$).
FIG. 7. Ratings of the melody on the screen which tested intensity congruence, where a higher rating means greater preference. A statistically significant change in ratings was found when the intensity of the vibrotactile stimulus was made incongruent with the intensity of the acoustic stimulus, and when the vibrotactile stimulus was misaligned in time with the auditory stimulus based on a Wilcoxon signed rank test.
Ratings from the intensity test condition were arcsine transformed and a linear model was constructed with intensity congruence, timing alignment, musicianship, and an intensity congruence musicianship interaction as fixed effects. A summary of the model can be seen in Table 1. Both timing alignment and an interaction between intensity congruence and musicianship were found to have statistically significant effects and a large, positive estimated coefficient for the ratings. No effect of participant was found. A plot of the model coefficients can be seen in Figure 8.

FIG. 8. Estimated, arcsine transformed rating coefficients for the screen which tested intensity congruence using the model shown in Table 2.

| Coefficient          | Estimate  | Std. Error | T Value | Pr(>|t|)  |
|----------------------|-----------|------------|---------|-----------|
| (Intercept)          | 0.45256   | 0.06595    | 6.862   | 1.68e-09  *** |
| Intensity Congruence | 0.10367   | 0.09220    | 1.124   | 0.26445   |
| Timing Alignment     | 0.47916   | 0.07306    | 6.558   | 6.17e-09  *** |
| Musicianship         | -0.10502  | 0.07030    | -1.494  | 0.13941   |
| Musicianship:Intensity Congruence | 0.40993 | 0.14061 | 2.915 | 0.00468 ** |

Table 1: A summary of a linear model on arcsine transformed ratings from the intensity test condition, with non-musician, intensity congruence and timing alignment as a reference condition.

2. Frequency Test Condition

Results for the frequency test condition are shown in Figure 9. In the frequency test condition, which tested the effect of frequency congruence, participant preferences were similarly reduced when acoustic intensity and vibrotactile intensity were made incongruent with or without acoustic frequency and vibrotactile frequency congruence (middle two columns). Preferences were lowest when the stimuli were misaligned in time (right column), as well as intensity and frequency incongruent. A Friedman test showed a significant effect of audio-tactile congruence on the ratings (p=6.87e-9***). A post-hoc analysis with a Wilcoxon test showed a significant difference in the means between all pairs except the stimuli pair with intensity incongruence, and intensity and frequency incongruence (p=0.341, middle two columns).
FIG. 9. Ratings of the melody on the screen which tested frequency congruence, where a higher rating means greater preference. A statistically significant drop in ratings was found when the intensity of the tactile stimulus was made incongruent with the intensity of the auditory stimulus, and when the tactile stimulus was misaligned in time with the auditory stimulus based on a Wilcoxon signed rank test. However, no statistically significant change was found when the
frequency of the tactile stimulus was made incongruent with the frequency of the auditory stimulus.

A linear model was constructed to predict the arcsine transformed ratings of the frequency test condition with intensity congruence, frequency congruence, timing alignment, musicianship, and an intensity congruence musicianship interaction as fixed effects. A summary of the model can be seen in Table 2 and is plotted in Figure 10. Like the intensity test condition, timing alignment, musicianship, and an interaction between intensity congruence and musicianship was found to have a significant effect on the frequency test condition. In addition, an effect of intensity congruence alone was found. No effect of frequency congruence or participant was found.

![Frequency Test Condition Model](image)

**FIG. 10.** Estimated arcsine transformed rating coefficients for the frequency test condition using the model shown in Table 2.

| Coefficient     | Estimate | Std. Error | T Value | Pr(>|t|)   |
|-----------------|----------|------------|---------|------------|
| (Intercept)     | 0.40133  | 0.06198    | 6.476   | 9.19e-09 *** |
Table 2: A summary of a linear model on arcsine transformed ratings from the frequency test condition.

### 3. Stretched Frequency Test Condition

Results for the stretched frequency test condition can be seen in Figure 11. In the stretched frequency test condition, participant preferences were similarly reduced when acoustic intensity and vibrotactile intensity were made incongruent with or without acoustic frequency and vibrotactile frequency congruence (middle two columns). Preferences were again lowest when the stimuli were misaligned in time (right column), as well as intensity and frequency incongruent. A Friedman test showed a significant effect of audio-tactile congruence ($p=3.04e^{-7}$). A post-hoc analysis with a Wilcoxon test showed a significant difference between all means except the intensity incongruent stimuli and the intensity incongruent and frequency incongruent stimuli ($p=0.36$).
FIG. 11. Ratings of the melody on the screen which tested frequency congruence with a stretched dynamic range, where higher ratings mean greater preference. No statistically significant change was found when the frequency of the tactile stimulus was made incongruent with the frequency of the auditory stimulus with the Wilcoxon Signed Rank test.
A linear model constructed from the arcsine transformed ratings showed a significant effect of intensity congruence, frequency congruence, time alignment, musicianship, and an interaction between musicianship and frequency congruence. In the case of musicianship, frequency congruence, and the interaction coefficients, a negative coefficient was estimated. No effect of participant was found. A representation of the model can be seen in Figure 12. A summary of the model can be seen in Table 3.

FIG. 12. Estimated arcsine transformed ratings for the stretched frequency test condition using the model shown in Table 3.

| Coefficient              | Estimate | Std. Error | T Value | Pr(>|t|) |
|--------------------------|----------|------------|---------|---------|
| (Intercept)              | 0.49071  | 0.06621    | 7.411   | 1.67e-10 *** |
| Intensity Congruence     | 0.27768  | 0.08109    | 3.424   | 0.001009 ** |
| Frequency Congruence     | -0.19374 | 0.09364    | -2.069  | 0.042042 *  |
| Timing Alignment         | 0.50192  | 0.08109    | 6.189   | 3.06e-08 *** |
| Musicianship             | -0.28496 | 0.08277    | -3.443  | 0.000950 *** |
Table 3: A summary of a linear model on arcsine transformed ratings from the stretched frequency test condition. Note the negative coefficient for frequency congruence.

| Musicianship:Frequency-Frequency Congruence | 0.42763 | 0.11705 | 3.653 | 0.000481 *** |

IV. DISCUSSION

The results of the study show that audio-tactile congruence had a robust effect on vibrotactile music enhancement. Additionally, the level of vibrotactile music enhancement was dependent on which vibrotactile parameters were congruent with the music. Time alignment and intensity congruence between the auditory and tactile stimuli were shown to influence participant ratings in a variety of conditions, whereas frequency congruence was not. An interaction with musicianship was shown for certain congruent parameters, particularly when the dynamic range of the vibrotactile frequency was artificially stretched.

The time alignment of the tactile stimulus and the auditory stimulus had the most robust effect on participant ratings. In every condition and in each statistical test, a drop in ratings was found when the tactile stimulus was misaligned in time with the auditory stimulus compared to when it was aligned. Timing alignment had a large, estimated coefficient in all three linear models as well, implying that timing alignment is not only consistently important among participants but also has a large effect on their enjoyment of the stimuli pairs.

The intensity congruence of the stimuli also had a robust effect of participant ratings, however it varied by participant group. While there was a drop in ratings in the intensity test condition, the linear model showed an interaction with musicianship, which increased the size of the effect. In other words, while a significant number of the participants found intensity congruence important for enjoyment, participants who were classified as musicians had their enjoyment more strongly affected by it.
In the frequency test condition, a change in the frequency congruence showed no significant change in ratings based on the Wilcoxon sign test, nor was frequency congruence used as a coefficient in the reduced linear model. The only case in which frequency congruence was found to have an effect was the linear model on the stretched frequency test condition, and only for participants classified as musicians. These results imply that in most cases, vibrotactile music enhancement can occur when the audio-tactile stimuli are only time-aligned and congruent in intensity. As musicians have shown to have better vibrotactile frequency discrimination than non-musicians (Sharp et al., 2019), it is likely that the limiting factor for frequency congruence is the relatively poor vibrotactile frequency discrimination of humans compared to acoustic frequency discrimination. The difference becomes especially apparent when compared to vibrotactile intensity discrimination and vibrotactile time discrimination, which have similar discrimination thresholds to their auditory counterparts. However, this does not explain why frequency congruence had a negative coefficient in non-musicians.

A. Limitations

There are several limitations for the study. First is regarding what participants considered for their ratings. Participants were instructed to rate the audio-tactile stimuli based on their preference and to take both auditory and tactile stimuli into account. Within a condition, certain tactile stimuli were randomized along specific parameters to make it incongruent with the auditory stimuli. However, the overall distribution of parameters in each tactile stimulus was kept constant. This was done to ensure the only difference between each audio-tactile stimulus was how the tactile stimulus compared to the auditory stimulus. Additionally, each run of the experiment re-randomized all random components of the stimuli. If participants considered factors other than the congruence of the tactile stimulus to the auditory stimulus in their ratings, it would be averaged out over all the
participants. However, it is still impossible to truly verify what participants were considering with their ratings. Therefore, this can be considered a limitation of the study.

Another weakness is that the auditory and tactile stimuli were based off only one melody, which itself was based off the bass track of a “world pop” song. The song was chosen as a piece of music which would have a generally wide appeal as a pop song, had a driving rhythm in which the bass track generally followed the main melody, and would be unknown by most participants. Pilot studies done with multiple melodies showed similar results, which were removed in the final experiment due to concerns about participant fatigue. Regardless, with only one melody used in the final experiment, it is difficult to be confident in how generalizable the results are. Even if the results could be generalized to other pop songs, considering the vast variety of musical styles, it is impossible to truly generalize the results to every piece of music, even within the Western world. While the selection of music stimuli is a limitation innate to any study that uses musical stimuli, it is still critical to consider a more robust selection of stimuli for future studies. Additionally, within the song, the bass track was chosen to be used as the stimulus as the most likely track to be perceived through the tactile system in a realistic situation. However, by using the bass track, there might have been an emphasis on rhythm and intensity variations, as opposed to pitch variations, in the stimulus itself. Future studies should expand the stimuli to include other melodies which are based on other genres, music cultures or instruments.

In this experiment, timing alignment and intensity congruence were considered as separate parameters. However, as both timing and intensity are components of rhythm, timing alignment and intensity congruence together could be conflated as “rhythm” congruence. In other words, changing the intensity of the note could have also changed how the timing of the vibrotactile stimuli was organized as rhythm. Therefore, time alignment and intensity congruence may not have been completely independent.
On a similar note, while the randomization of the vibrotactile stimulation controlled for the physical parameters of the vibration to ensure that certain physical patterns did not feel innately more pleasing than others, it did not control for the musicality of the vibration itself. In other words, vibrotactile stimulus made incongruent with the auditory stimulus might also be made less musically pleasing. It is possible therefore, that the musicality of the vibration itself is the source of the music enhancement, not the physical congruence of the auditory and tactile stimuli. For more clarity, consider the vibration as an additional instrument to a multi-track piece of acoustic music. The additional instrument would be more appealing if it is completely synchronized with another instrument than if it is completely random and distracting, as was shown in the present study. However, the additional instrument might also be more appealing if it is not completely synchronized but still musically on-beat and artistically cohesive with the other instruments than if it is completely random and distracting. The vibration would then accompany the music like an additional instrument, rhythm, or “backing track.” In some cases, the random vibration might offer an improved experience. There is evidence musical meter can be perceived through vibrotactile stimulation ([Huang et al., 2012]). Therefore, the musical congruence between the auditory and tactile stimuli is a confounding variable with the physical congruence between the auditory and tactile stimuli.

Musicians have shown to have better vibrotactile frequency discrimination than non-musicians [Sharp et al., 2019]. Therefore it is possible through rigorous training, one could increase the importance of frequency congruence for vibrotactile enhancement of music. Training vibrotactile frequency would mirror the implicit vibrotactile training which happens when playing a musical instrument. Such training could then potentially change the results of the study. However, the requirement of such rigorous training then also brings into question the applicability of such a result. If a device which utilizes audio-tactile frequency congruence requires a lifetime of vibrotactile
frequency discrimination training to use properly, it would likely be infeasible for most use-cases. Further studies could explore the amount of training that could be needed to significantly reduce a person’s vibrotactile frequency JND.

The method of presenting vibrations in the study is also a limitation of the study. The Lofelt L5 actuator was chosen to represent a possible device which might use vibrotactile stimulation to enhance music. The benefits of a Lofelt L5 include being small, quiet, and the ability to vibrate with a relatively large frequency range and be frequency and amplitude independent through some minor signal processing. However, there are theoretically infinite methods of presenting vibrotactile information one could use: different sizes of actuators, different frequency ranges, and different locations on the body all present largely untested variations on the same experiment. It is difficult to predict how such changes might affect the results, especially considering a lot of psychophysical tactile research utilizes either the fingers and thenar, or in more application-based research, the wrist and back. A larger tactile interface, such as a vibrating chair, is both significantly different from the finger-based method in the present study, as well as feasible for some realistic application of a musical haptic device. It is therefore worth considering how it might change the results.

Similarly, in the present experiment, the auditory stimulus was presented over headphones while the tactile stimulus was presented through the fingers. This set-up is consistent with many psychophysical audio-tactile experiments. However, as spatialization has shown to be an important factor in audio-visual integration, it is possible that localizing the audio closer to the tactile vibrations could affect the results. Spatial discrimination of tactile air puffs have been shown to be faster when accompanied by a spatial congruence sound versus a spatially incongruent sound (Teramoto et al., 2013). Another variation of the experiment could include using a loudspeaker set-up or visual cues which better associate the spatialization of the auditory and tactile stimuli. In some ways, different spatialization set-ups of the experiment simply add to the infinite possible variations along with
location, size, and frequency range of the vibrotactile stimulation. However, an experiment which tests the effect of localization congruence on vibrotactile music enhancement would be interesting for future applications.

B. Implications

1. Vibrotactile Music Devices

The lack of evidence for frequency congruence being relevant for vibrotactile music enhancement has implications for future music-enhancing vibrotactile devices. For one, haptic actuators which can control frequency and intensity independently, such as LRAs, may not be necessary for vibrotactile music enhancement. An ERM, in which the intensity of the tactile vibration is dependent on its frequency, could have a similar benefit at a cheaper cost. Of course, an ERM may have other disadvantages, such as being more power intensive. However, if vibrotactile frequency is coded to auditory pitch, there might not be a perceptual difference between the two. The vibrotactile frequency could also have potential to be made congruent to other aspects of the sound which have not been tested, such as its timbre. Without the need to code the auditory pitch to the vibrotactile frequency, many other possibilities open, such as coding auditory brightness or other timbral features to vibrotactile frequency. Verma et al. (in press) have shown vibrotactile stimulation can affect the perception of auditory roughness, and Russo et al. (2012) have shown strong discrimination abilities for auditory timbres presented through vibrotactile stimulation. The frequency of the vibration could also be used to change the timbre of the vibration itself, allowing for more flexibility in how the vibrotactile stimulation is perceived or associated with different instruments. Consider, for example, that an 80 Hz vibration is discriminable from a 135 Hz vibration. While there might not be enough discriminable steps to convey a continuum of pitches from 80 Hz to 135 Hz, one could still utilize both frequencies as a binary, associating one with one instrument and the other with other instruments.
The fact that vibrotactile music enhancement can be beneficial with only timing alignment and intensity congruence also has benefits for certain hearing-impaired listeners, such as cochlear implant users. In general, multisensory integration is stronger in people with weaker senses (Stein & Meredith, 1993). However, for certain audio-tactile multisensory effects, a similarity in frequencies between the tactile and auditory system is required (Wilson et al., 2010). It is not known how this might interact with a cochlear implant user, who generally have frequency-to-place mismatches based on the insertion of their device (Canfarotta et al., 2020; Landsberger et al., 2015). In other words, the same frequency stimulates a different part of the basilar membrane for two different cochlear implant users. If the same auditory and tactile stimulation frequency was required for vibrotactile music enhancement to occur, the effect might not be as strong for a cochlear implant listener as a normal hearing listener. Cochlear implant users are shown to be less susceptible to multisensory effects (Landry et al., 2013); however, the effect is dependent on the length of time the user has been implanted (Landry et al., 2014). Regardless, if frequency congruence is not required, a cochlear implant user should have the same music enhancing benefits as people without cochlear implants. However, given the variation in outcomes for cochlear implant users, this should be tested in a future study.

Regardless of the use of the vibrotactile frequency, evidence that intensity congruence and timing alignment improves vibrotactile music enhancement also allows for a new baseline audio-tactile coding strategy. It is almost certain that further improvements can be made to future audio-tactile coding strategies. With some exceptions, such as in (Haynes, Lawry, Kent, & Rossiter, 2021), new audio-tactile coding strategies and music-enhancing devices are often tested against a “no vibration” condition. While it is still beneficial to show how a specific device or coding strategy can improve music perception, it makes it difficult to compare different strategies and their respective benefits and drawbacks. Now, new devices and coding strategies can be tested against a baseline, intensity-
congruent time-aligned coding strategy, which is shown to have some benefit, to quantify what additional benefits a novel device or coding strategy might bring. Deviations from this baseline can also be tested individually for their effect on vibrotactile music enhancement.

2. Mechanisms Behind Vibrotactile Music Enhancement

The results show further insight into the possible mechanisms behind vibrotactile music enhancement. Two possible explanations initially presented themself for vibrotactile music enhancement: first, that vibrotactile music enhancement is due to a multisensory integrative effect in which the tactile stimulation enhances pleasurable aspects of the audio perceptually, or second, that vibrotactile music enhancement is a top-down, musical effect that is possible due to a relatively high rhythm perception in the tactile system (Bernard et al., 2022; Brochard et al., 2008; Huang et al., 2012). Participants preferring timing aligned stimuli suggests a level of multisensory integration is necessary for vibrotactile music enhancement, although as stated in the limitations, the study did not control for the musicality of the tactile stimulation.

With that said, the study did not show a consistent benefit when the frequencies of the tactile and auditory stimuli were matched in value. The stretched frequency test condition showed some benefit for vibrotactile frequency congruence, but the frequency was only congruent directionally and not exactly matched in value. This is contrasted with audio-tactile studies done in Wilson et al., 2010, which showed matched frequencies were necessary for certain audio-tactile integration mechanisms. However, Yau et al., 2010 showed audio-tactile frequency interactions, unlike loudness interactions, were not dependent on time synchrony. Yau et al., 2010 use this as evidence for two different perceptual mechanisms behind audio-tactile frequency and audio-tactile loudness interactions. As timing alignment was shown to be important in the present study for participant preferences, it is possible that the mechanism controlling audio-tactile loudness interactions, not frequency interactions, was a more relevant factor.
One explanation for vibrotactile music enhancement is that tactile vibrations increase the perceived loudness of bass frequencies of music. A common audio-tactile multisensory effect is an increase in perceived auditory loudness with tactile stimulation (Gillmeister & Eimer, 2007; Schürmann et al., 2004). The effect has been expanded to music as well, where participants listening to music in a car with added tactile stimulation preferred music with lower bass levels (Simon et al., 2009). One explanation is that the tactile stimulation was increasing the perceived bass levels in the music. McCown et al. (1997) showed certain demographics prefer music with boosted bass frequencies. It is possible that people prefer music with tactile stimulation because it increases the perceived levels of bass in the music and therefore makes it more preferential.

V. CONCLUSION

An experiment was designed which could separate vibrotactile music enhancement into a series of audio-tactile congruences. It was found that having the auditory and tactile stimuli time-aligned consistently contributed to vibrotactile music enhancement. Having the intensity of the auditory and tactile stimuli congruent was also important and had a larger effect for musicians. Congruence of the frequency of the auditory and tactile stimuli had the least effect, but the effect increased for musicians when the vibrotactile frequency changes were more extreme. Future work should include using more genres and melodies as stimuli and testing the effect of the musicality of the vibrotactile stimulation on the musical experience.

ACKNOWLEDGEMENTS

This work was supported by the William Demant Foundation and the Innovation Fund Denmark.

REFERENCES (BIBLIOGRAPHIC)


https://doi.org/10.1097/AUD.0000000000000163


https://doi.org/10.3390/s21196575


https://doi.org/10.1037/a0029046


