


Effect of Vibrotactile Stimulation on Auditory Timbre Perception for Normal-Hearing Listeners and Cochlear-Implant Users

Trends in Hearing
Volume 26: 1–11
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/23312165221138390
journals.sagepub.com/home/tia


Tushar Verma¹ , Scott C. Aker^{1,2} and Jeremy Marozeau¹ 

Abstract

The study tests the hypothesis that vibrotactile stimulation can affect timbre perception. A multidimensional scaling experiment was conducted. Twenty listeners with normal hearing and nine cochlear implant users were asked to judge the dissimilarity of a set of synthetic sounds that varied in attack time and amplitude modulation depth. The listeners were simultaneously presented with vibrotactile stimuli, which varied also in attack time and amplitude modulation depth. The results showed that alterations to the temporal waveform of the tactile stimuli affected the listeners' dissimilarity judgments of the audio. A three-dimensional analysis revealed evidence of crossmodal processing where the audio and tactile equivalents combined accounted for their dissimilarity judgments. For the normal-hearing listeners, 86% of the first dimension was explained by audio impulsiveness and 14% by tactile impulsiveness; 75% of the second dimension was explained by the audio roughness or fast amplitude modulation, while its tactile counterpart explained 25%. Interestingly, the third dimension revealed a combination of 43% of audio impulsiveness and 57% of tactile amplitude modulation. For the CI listeners, the first dimension was mostly accounted for by the tactile roughness and the second by the audio impulsiveness. This experiment shows that the perception of timbre can be affected by tactile input and could lead to the developing of new audio-tactile devices for people with hearing impairment.

Keywords

cochlear implants, music, timbre, audiotactile, crossmodal processing

Received 10 February 2022; Revised 18 October 2022; accepted 21 October 2022

Introduction

We perceive the world multimodally. Our brain collects information from all our senses to provide a cohesive perception of our environment. In many cases, the information provided by these modalities reinforces each other. For example, the shape and the sound of potato chips contribute to their taste (Piqueras-Fiszman & Spence, 2011). However, a possible consequence of this phenomenon is that if information from one modality is artificially changed, the perception of another modality may also be affected. A classic example is the McGurk effect (McGurk & MacDonald, 1976), in which the sound of a person saying “BA” combined with a video of a person saying “GA” is perceived as “DA.” The McGurk effect demonstrates that altering visual information can alter our auditory perception.

The McGurk effect has also been extended to study tactile and auditory system links. In a study by Gick and Derrick (2009), listeners were presented with a sound of a person saying two sounds, one aspirated such as “PA” and one

unaspirated such as “TA.” Test listeners were asked to identify which of the two sounds they heard. Simultaneously, puffs of air were applied unknowingly to the listeners' hands and neck. Listeners were more likely to identify the aspirated sound correctly when the puffs of air were applied. The study demonstrated tactile stimulation could affect the perception of sound.

Few other studies have shown that tactile stimuli can affect different aspects of sound. For example, Schürmann et al. (2004) performed a study in which listeners were

¹Music and Cochlear Implant Lab, Department of Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark

²Oticon Medical, Smørum, Denmark

Corresponding Author:

Jeremy Marozeau, Music and Cochlear Implant Lab, Department of Health Technology, Technical University of Denmark, Kongens Lyngby, 2800, Denmark.
Email: jemaroze@dtu.dk



asked to adjust the intensity of a faint tone to match the loudness of a reference tone while holding a vibrating tube in half of the conditions. They matched the tone intensity to an average of 12% less intense when holding the vibrating tube versus not. The result implies the vibration of the tube led the listeners to perceive the tones as being louder.

Beyond loudness, timbre has clear similarities to aspects of tactile perception. Timbre is a multidimensional sound characteristic, often defined as what makes a sound distinct outside of pitch, loudness, and duration. For example, timbre characterizes what makes a trumpet sound different than a violin. Dimensions of timbre perception can be broadly divided into temporal, spectral, and spectro-temporal characteristics (McAdams et al., 1995). Yau et al. (2010) argued that tactile explorations are similar to timbre and determined by the temporal and spectral content of the vibrations elicited by a surface. Russo et al. (2012) investigated if normal hearing, NH, and hearing-impaired listeners could discriminate between the waveforms of different musical instruments presented through vibrotactile stimulation. The study highlighted that participants could easily discriminate musical instruments presented via vibrotactile stimuli with a 90% accuracy rate, demonstrating the adeptness of the tactile system at differentiating complex vibration waveforms with differing spectral information. They argued that a vibrotactile system could help hearing-impaired people restore their ability to identify a talker voice over the phone.

The ability to integrate different modalities depends on factors such as the strength of each input modality as predicted by the law of inverse effectiveness (Meredith & Stein, 1983). This law states that multisensory stimuli are

more likely to be integrated effectively when the unisensory responses are relatively weak. We often can perceive sound through our tactile and auditory systems in our everyday lives, but we ignore the former because the latter is much more efficient. On the other hand, people with hearing impairment need to rely more on the tactile sense to supplement their weaker auditory sense.

Different studies have shown that vibrotactile information can help hearing-impaired people with a cochlear implant, CI, to perceive speech in noise better (Fletcher et al., 2019) and can improve their ability to discriminate musical pitch (Huang et al., 2020; Luo et al., 2012). The CI is a surgically implanted device that acts as an “electric ear,” stimulating the auditory nerve directly with electric pulses based on the input from a microphone. After implantation, CI users see a dramatic improvement in sound perception. However, certain aspects of sound, such as timbre, are not conveyed well by the CI (Marozeau & Lamping, 2019). Poor timbre perception could be a possible cause of reported lower music enjoyment by CI users. If a vibrotactile stimulation could alter timbre perception, CI users could potentially have a new path to improve their music experience.

Once the tactile input effect is established, we can consider the nature of the relationship between tactile and auditory senses. As Spence et al. (2009) proposed, it is important to distinguish between multimodal integration and crossmodal processing. Multimodal integration describes a situation in which the brain combines information from two senses to form a single percept, for example, watching a foreign film with subtitles. The emotion of the character’s voices and the dialogue information in the text combine to inform the viewer of the sentence’s meaning and emotion. Crossmodal processing refers to a situation in which stimuli presented in one sense influence our perception of stimuli presented in another. For example, in the ventriloquist effect, our perception of spatial audio is modified by the sight of the puppet.

Russo et al. (2012) showed that instrument identification scores can be achieved with only tactile stimulation. This result indicates that listeners can integrate sensory information from two modalities to identify an auditory object. However, it remains unclear if the listeners use cues from both modalities (as in multimodal integration) or whether their auditory perception of timbre can be enhanced and clearer (as in crossmodal processing).

In this experiment, we studied how a tactile stimulus can affect the perception of two dimensions of timbre, namely impulsiveness and roughness. NH and CI listeners were asked to rate the sound dissimilarity of audio-tactile stimuli that varied in attack time and amplitude modulation. The data were analyzed using a multidimensional scaling (MDS) technique to investigate the effect of introducing a tactile modality to a timbre MDS space. The analysis of the space informs us whether the tactile stimulation does not manifest in the MDS space, showing no integration, if it

Table 1. Table of CI-Related Statistics for CI User Listeners.^a

Age	Years implanted	CI brand	CI patient type	Type of deafness
67	1	Oticon Medical	Bimodal	Post-Lingual
76	6	Oticon Medical	Bimodal	Post-Lingual
66	3	Oticon Medical	Bimodal	Post-Lingual
65	4	Oticon Medical	Bimodal	Post-Lingual
82	4	Oticon Medical	Bimodal	Post-Lingual
65	5	Oticon Medical	Bilateral	Pre-Lingual
57	8	Med-El	Bilateral	Post-Lingual
55	4	Advanced Bionics	Unilateral	Post-Lingual
64	24	Advanced Bionics	Bilateral	Post-Lingual

^aBimodal refers patient with a CI on one ear and a hearing aid on the other, bilateral to patients with two CIs, and unilateral to patient with a single CI.

manifests as an independent new perceptual dimension, showing multimodal processing, or if it alters an existing auditory dimension, showing crossmodal enhancement.

Methods

Listeners

Twenty NH volunteers (9 female, 11 male) and 9 CI users (6 female, 3 male) participated in this study. The NH listeners were between 25 and 30 years old and were confirmed to have no hearing threshold higher than 20 dB. The CI listeners were between 55 and 82 years old (see Table 1). Listeners provided informed consent before the study, and the Science-Ethics Committee approved all experiments for the Capital Region of Denmark (reference H-16036391). All research was performed following the relevant guidelines and regulations corresponding to the use of human participants. The listeners were compensated at DKK 122 per hour unless they voluntarily declined compensation.

Stimuli

The stimuli were composed of a synchronized 600-ms audio and tactile burst. Each stimulus was composed of two parts (see Figure 1). First, an attack section was a 200-ms pure tone modulated with raised cosine to produce a smooth ramp. Three different attack times were played: 50, 150, and 250 ms to create stimuli with fast, medium, or slow impulsiveness, respectively. The second part was an amplitude-modulated tone with a modulation frequency of 20 Hz. For the NH listeners, the carrier frequency was set at 200 Hz. For the CI users, the carrier was set at 250 Hz to maximize the activation of the apical electrode based on the frequency-electrode map allocation of most of the brands. The difference in frequency results from a decision to better adapt it for the CI after collecting the NHL data. It would have been better to use 250 Hz for both groups for consistency. However, we do not think that this difference can affect the study's outcome. Three modulation depths were used at 15%, 55%, and 75% to create stimuli with low, medium, and high roughness. Each stimulus had a 100-ms raised-cosine offset ramp and a modulated sustain part equal to 500 ms minus the attack time.

Six different audio stimuli were created in MATLAB with the three possible attack times and two AM depths (high and medium). Four tactile stimuli were created with two attack times (slow and fast) and two AM depths (high and low). The 24 stimuli were composed of all possible combinations of the audio and tactile stimuli. All the stimuli were set to an equal root mean square (RMS) level. No differences in loudness were perceivable according to the authors' ears.

Task

The listeners were seated in an audiometric booth in front of a monitor. With their left hand, they held a tactile actuator (see Figure 2) encased in a fabric-lined cardboard box. With the right hand, they were able to interact with a GUI composed of a horizontal slider labeled from "similar" (coded 0) to "dissimilar" (coded as 1) and two buttons: one to listen to the pair again, the other to validate the response.

For each pair, they were instructed to judge whether the sounds were similar or different, using the full scale of the cursor. They were allowed to listen to the pair as many times as they wanted using the "repeat" button. Before each trial, the cursor's position was placed randomly to reduce the influence of the judgment of the previous pair. All possible pairs of the 24 stimuli were presented with no order distinction, totaling 276 pairs. The order within pairs and the order of pairs were randomized for each listener. They were specifically instructed to base their judgment only on the sound while ignoring the tactile stimuli. Before the experiment, the listener was presented with the 24 stimuli in a random order to acquaint them with the range of possible stimuli. They were allowed to take a break at any moment during the experiment. The written instruction can be found in the supplementary material.

Apparatus

The sound was presented to the NH listeners over calibrated Sennheiser HDA200 headphones in an audiometric booth. The calibration was checked using a Brüel & Kjær artificial ear simulator, and the deviation was minimal and within an acceptable range. The level of presentation was kept at 69 dBA, complying with the hearing regulations of the Danish Working Environment Authority for the duration of the test, where the listening period was less than one hour. This comfortable but high level was set to ensure that any potentially audible sound produced by the tactile actuator would be masked. For the CI users, the audio stimuli were provided directly via a CI streaming device, depending on their manufacturers. They could set a comfortable listening level before the test. Bilateral CI users were tested only on their preferred ear.

The tactile vibrations were presented over a Lofelt L5 tactile actuator in a wooden case. The listeners were instructed to hold the actuator, as shown in Figure 2, to maximize the contact area at the fingertips. Both audio and tactile stimuli were presented at suprathreshold levels. The tactile actuator had an input voltage of approximately 860 mV, corresponding to an RMS vibration velocity of approximately 0.063 m/s. The audio and tactile channels were also excited simultaneously by a click and a test sample to measure any delay in response due to mechanical inertia. A delay of 0.024 s was found and had a minor deviation depending on the sample. The audio signal was delayed by this amount

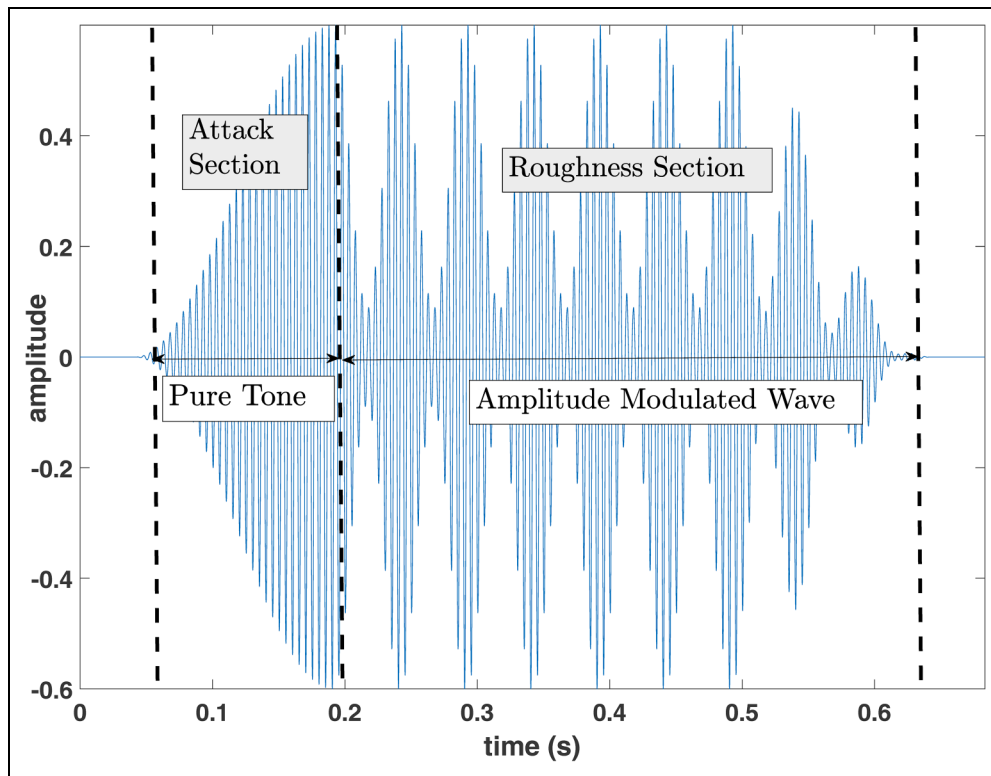


Figure 1. Example of a Stimulus. The section between the first two dashed lines is the attack section with a pure tone. The part after the right side is the section with the amplitude-modulated wave.

to synchronize the two devices. The simultaneity of the stimuli and the passive attenuation of the headphones prevented the listener from perceiving the soft sound emitted by the actuator. The test was run on a MATLAB-based GUI.

Analysis

This study assessed whether a tactile input could affect sound perception with MDS technique. The MDS uses dissimilarity judgments, or “perceptual distances,” between pairs of stimuli to construct an N-dimensional perceptual space. An analysis of the position of each stimulus within this space can then reveal the nature of the modality on which the dissimilarity judgment was based. The MDS representation is particularly well-adapted for this type of study, as it decomposes a listener’s percept into different perceptual orthogonal dimensions without assuming its nature or modality.

First, we extracted the number of dimensions that compose the MDS solution. The number of dimensions is an important first result that outlines the number of modalities the listeners use. To avoid any bias, we have followed the method proposed by Riche and Verheyen (2020) to derive the number of dimensions objectively. In a nutshell, each MDS solution with 1–6 dimensions was fitted to a subset of 80% of the available dissimilarity data. Then, each model was used to predict the dissimilarity of the unused

data. The solution that maximizes performance on the held-out data was selected. Then, an analysis of the projection on those dimensions reveals the contribution of each physical parameter. The individual dissimilarity ratings were processed through an MDS analysis using the INDSCAL (Carroll & Chang, 1970) technique (individual scaling), with the SMACOF package (de Leeuw & Mair, 2009) (using R.4.04). This model allows different weights to be applied to each dimension. As a result, unlike the more traditional MDS model, the INDSCAL solution has a deterministic orientation. A linear model was applied to each dimension with the four physical parameters as fixed effects. This analysis revealed the contribution of each modality to each of the dimensions.

This analysis allowed us to test two different hypotheses plus a null hypothesis:

Null-Hypothesis: No effect of the tactile modality on the auditory judgment.

In this experiment, the listener rated the auditory dissimilarity of stimuli that varied on four physical dimensions (two auditory and two tactile). If the listener can ignore the tactile input, the MDS should result in a two-dimensional space correlated with the two audio parameters (the attack time and the AM). This space should be similar to the one created based on audio stimuli without any

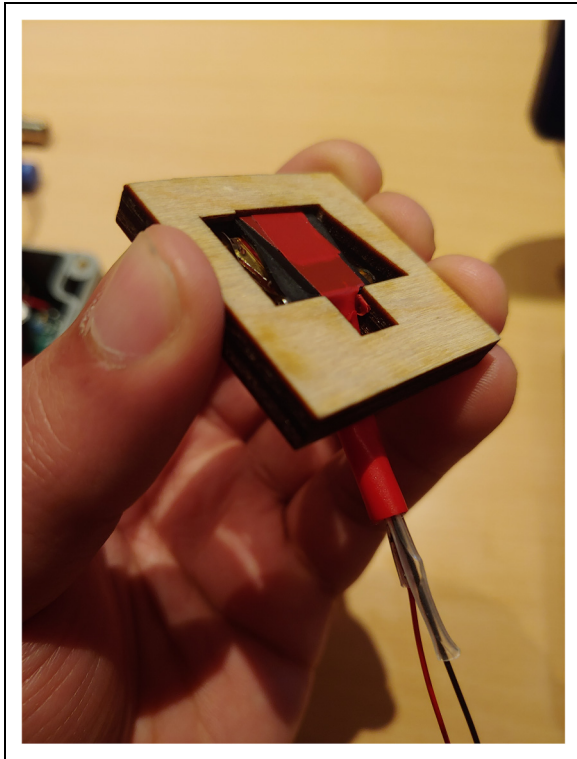


Figure 2. The Lofelt L5 Vibration Motor in a Wooden Casing. The listeners were instructed to hold the tactile actuator with all 5 fingertips as shown.

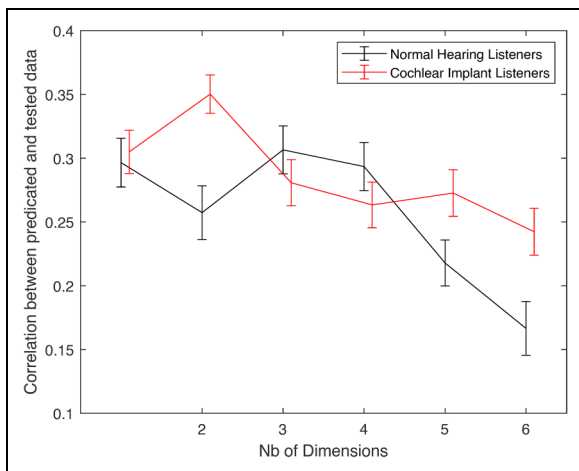


Figure 3. Method to Objectively Assess the Optimal Number of Dimensions as Outlined by Richie and Verheyen (2020).

tactile stimulation, and it will be considered the reference auditory space.

Hypothesis 1: The tactile modality will be perceived as an additional orthogonal dimension. If each parameter has an independent effect (without any interaction), the MDS will

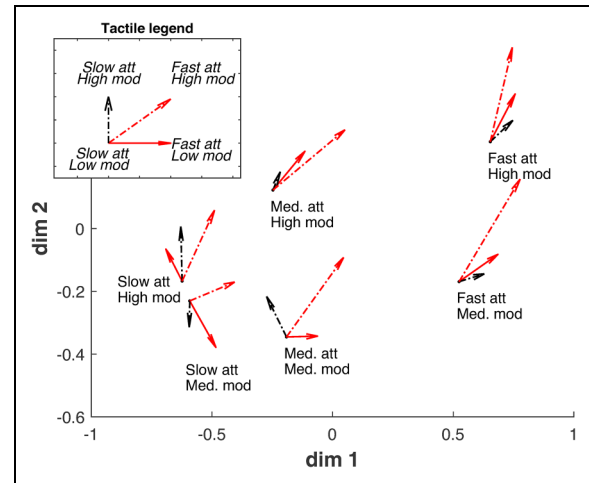


Figure 4. MDS Space of Dimension 1 (x-axis) and Dimension 2 (y-axis) for the NH Listeners. The label represents the acoustic properties of the audio attack time (Slow: 0.05 s, Medium: 0.15 s, and Fast: 0.25 s); and the AM depth (Medium: 55% and High: 75%). The legend in the top-left corner depicts the effect of each haptic sample on the audio sample depending on the stimulus characteristics. For example, the origin of the arrows represents the position of the stimuli with a slow tactile attack time and a low tactile AM; the endpoint of the continuous red arrows the stimuli with a fast tactile attack and low AM.

result in a four-dimensional space, with each dimension perfectly correlated to each parameter. Such a mechanism can be qualified as “Multimodal interaction” (Spence et al., 2009). The two dimensions linked to the auditory modality should be similar to the reference auditory space.

Hypothesis 2: The tactile modality affects the auditory modality. If the tactile modality directly modifies an auditory perceptual dimension, the position along that dimension will be predicated by audio and tactile parameters. The auditory dimensions will then differ significantly from the reference space. Such a mechanism can be qualified as “crossmodal processing” (Spence et al., 2009).

Results

Number of Dimensions

Following the method outlined by Riche and Verheyen (2020), the goodness-of-fit was plotted as a function of the number of dimensions of the MDS solution (Figure 3). The model for the CI listeners shows a clear maxima at two dimensions. For the NH listeners, the maximum was found for three dimensions, followed closely by the solution at four dimensions and one dimension. As an additional analysis of stress (i.e., mean squared error of the model) revealed a large drop from 1 to 2 dimensions, the solution at one dimension was not considered further. The two MDS solutions with 3 and 4 dimensions (3D and 4D, respectively) were extracted,

and each dimension was analyzed separately. The first three dimensions of the 4D solution were very similar to the 3D solution. However, the fourth dimension did not significantly correlate with any of the physical parameters or combinations. This fourth dimension was therefore discarded, and the 3D solution was analyzed. The analysis of the discarded solutions can be found in Supplemental Material (Figures S1–S4).

Multidimensional Scaling Space

Figure 4 shows the projection on the first two dimensions of the 3D solution for the NH listeners. Each stimulus with the same audio signal is grouped with arrows and their acoustical property (see figure caption). Each arrow group's origin represents the labeled audio stimulus' position combined with a slow tactile attack and low AM depth tactile input. The dashed arrows' endpoints represent the stimuli' position with a high AM depth tactile input. The red arrow's endpoint represents the stimuli' position with a fast tactile attack time. The red dashed arrows' endpoint represents the stimuli position with a fast tactile attack time and a high AM depth tactile input.

First, stimuli are ordered mostly according to audio attack time along with the x-axis, indicating that the first perceptual dimension is related to auditory impulsiveness (Marozeau et al., 2003; McAdams et al., 1995). All stimuli with a slow attack are located toward the left side and are perceived as less impulsive. Stimuli with the fast attack are located to the right and are perceived as more impulsive.

The second dimension separates the stimuli with medium AM, at the bottom with the high AM stimuli at the top, indicating that the second perceptual dimension is related to the auditory roughness. Interestingly, the space resembles a slanted rectangle, indicating some interaction between the two audio parameters. As those two audio parameters are

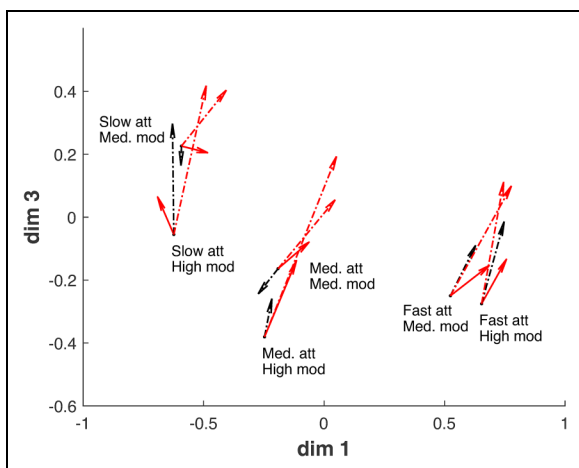


Figure 5. MDS Space of Dimension 1 (x-axis) and Dimension 3 (y-axis) for the NH Listeners (See Caption of Figure 4).

dependent on the temporal envelope, it is not surprising to observe some interactions. Stimuli with slow attack time have shortened sustained part, which contains the AM. Therefore, they are perceived as less rough and are located further down in the dimension. Additionally, the smaller the AM part, the smaller the roughness difference, as can be observed in the smaller distance between the two stimuli with a slow attack than the distance between the stimuli with a fast attack.

The configuration of those dimensions is similar to a control space obtained by asking 12 NH listeners to rate the dissimilarity between the six audio-only stimuli. A two-dimensional solution was obtained through classical MDS and Procrustean rotations. Each dimension of this control space correlated with the average position of the stimuli with the same sound (center of gravity of each group of arrows in Figure 4) with a R^2 of 99.53% ($p < .0001$) for the first dimension, and 93.23% ($p = .0018$) for the second dimension (see Supplemental Material, Figure S5).

To disentangle the null hypothesis (no effect of tactile input: MDS space is similar to control) from hypothesis 2 (the tactile input shifts the position of the stimuli on the auditory dimensions), the arrows' length and directions in Figure 4 can be observed. First, the average length of the arrow was 20% of the average distance between stimuli with the same tactile input. This ratio indicates that while the distance was mostly driven by the audio signal along these first two dimensions, the tactile input significantly contributed to the position of the stimuli. Second, the orientation of the arrows points in a similar direction. Except for stimuli with "slow attack and high AM depth," the solid red arrows point toward the right, indicating that a sound was perceived as more impulsive if the tactile input had a short attack. The dashed black arrows generally point upward, indicating that a sound was rougher if the tactile input had a high AM. Interestingly, the red dashed arrows (fast attack with high

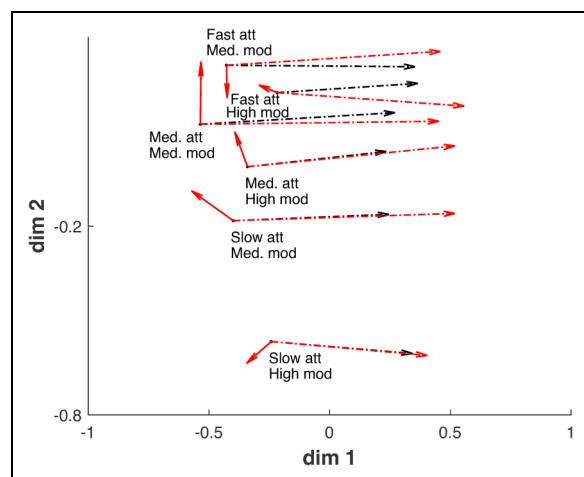


Figure 6. MDS Space of Dimension 1 (x-axis) and Dimension 2 (y-axis) for CI Listeners (See Caption of Figure 4).

AM depth) were longer in length and point in a right-up direction, indicating that the tactile input's attack and AM parameters can have a complex and additive effect.

Figure 5 shows dimensions 1 and 3 of the 3D solution for NH listeners. Unlike the first two dimensions, the projection along with the third dimension does not show a clear order based on the audio characteristics of the stimuli. On the other hand, all the arrows point upward except for two stimuli, indicating that this dimension might be linked to tactile perception. The average length of arrows is about the same (96.3%) as the average distance between stimuli with the same audio input on that dimension.

Figure 6 shows the dimensions 1 and 2 of the 2D solution for CI listeners. The first dimension is related to the tactile roughness, as seen by all the dashed arrows pointing toward the right side. The second dimension is linked to impulsiveness, as stimuli are ordered from low to high according to the attack time.

Correlation of Each Dimension

The MDS reveals a space composed of three independent perceptual dimensions for the NH listeners. We can now evaluate the contribution of the two tactile and two audio parameters on each of those dimensions through a statistical model. The two modalities' contribution ratio will quantify the effect of the tactile input on a timbre perception. For each audio and tactile stimuli, two descriptors were extracted.

First, McAdams et al. (1995) showed that the first dimension of the timbre space for their stimuli was highly correlated with the logarithm of the stimulus attack time. Therefore, a similar descriptor should be correlated with the first dimension of the present timbre space. From now on, we designate these descriptors as "audio attack time" and "tactile attack time." Second, as the present stimuli were designed to vary in AM depth, one dimension should be linked to roughness perception. Auditory roughness is a function of modulation depth, modulation frequency, and length of the stimuli (Fastl & Zwicker, 2007; Meredith & Stein, 1983). As the AM frequency of the present stimuli was fixed at 20Hz, the AM depth multiplied by the duration of the AM signal part after the attack was used as a predictor of roughness. We designate these descriptors as "audio roughness" and "tactile roughness." Note that in this context, tactile roughness refers to amplitude-modulated vibrotactile stimulation, not a rough tactile surface in the colloquial sense.

All descriptors were normalized to have a mean of 0 and a standard deviation of 1, so their contribution could be directly comparable. Because audio attack time and audio roughness were significantly correlated ($r(5) = -0.65$, $p = .0006$), only one of these two descriptors was included in any given model, and only the model with the better fit

was reported. An effect was considered significant for a $\alpha < 0.01$ to account for multiple comparisons.

Normal-Hearing Listeners

The first dimension was modeled as a linear regression of the audio attack time, the tactile roughness, the tactile attack time, and their first-level interactions. The statistical analysis revealed a significant effect of the audio attack time ($p < .0001$) and the tactile attack time ($p = .002$). A reduced model with just those two parameters shows a very strong correlation with the first dimension's stimuli position ($r(23) = 0.99$, $p < .00001$). The physical descriptor was composed of the linear combination of 88% of the audio attack time and 12% of the tactile attack time. Figure 7A shows the scatter plot of the model versus the projection on the first dimension. All the points are well aligned on the regression line (with a slope of 1). Interestingly, all the arrows showing the tactile contribution are also aligned with the regression line. As shown in Figure 7A, the model accurately predicted the first perceptual dimension.

The second dimension position was modeled as a linear regression of the audio roughness, tactile roughness, tactile attack time, and first-level interactions. The statistical analysis revealed a significant effect of audio roughness ($p < .0001$) and tactile roughness ($p = .0005$). A reduced model with just those two parameters shows a very strong correlation with the second dimension's stimuli position ($r(23) = 0.96$, $p < .00001$, see Figure 7B). The physical descriptor was composed of the linear combination of 75% of the audio roughness and 25% of the tactile roughness.

The third dimension was modeled with the audio attack time, the tactile roughness, the tactile attack time, and their first-level interactions. The model reveals a significant effect of the audio attack time ($p = .0054$) and the tactile roughness ($p = .0067$). A linear regression with those two parameters is highly correlated with the third dimension's projection ($r(23) = 0.73$, $p < .00001$, see Figure 7C). The regression coefficients showed a contribution of 43% for the audio attack time and 57% for the tactile roughness, suggesting that the tactile parameter had a stronger contribution.

Cochlear-Implant listeners

The first dimension position was modeled as a linear regression of the audio roughness, tactile roughness, tactile attack time, and first-level interactions. The statistical analysis revealed a significant effect of audio roughness ($p = .005$), tactile roughness ($p < .00001$), and tactile attack time ($p < .00001$). A reduced model with just those three main parameters shows a very strong correlation with the first dimension's stimuli position ($r(23) = 0.99$, $p < .00001$). The physical descriptor was composed of a linear combination of 9% of the audio roughness, 16% of the tactile attack time, and 75% of tactile roughness. Figure 8A shows the

scatter plot of the model versus the projection on the first dimension.

The position on the second dimension was modeled as a linear regression of the audio attack time, the tactile roughness, the tactile attack time, and their first-level interactions. Only the main effect of the audio attack time was significant ($p < .004$). This parameter correlates strongly with the second dimension ($r(23) = 0.77$, $p < .0004$) (see Figure 8B).

Discussion

Testing the Null-Hypothesis

This study aimed to investigate two hypotheses related to how a tactile input can affect our auditory perception of timbre and compared against the null hypothesis that the listeners can ignore the tactile stimuli. Here, this additional sensory input should not affect the obtained MDS space. By asking listeners to rate the auditory dissimilarity between the stimuli while ignoring the

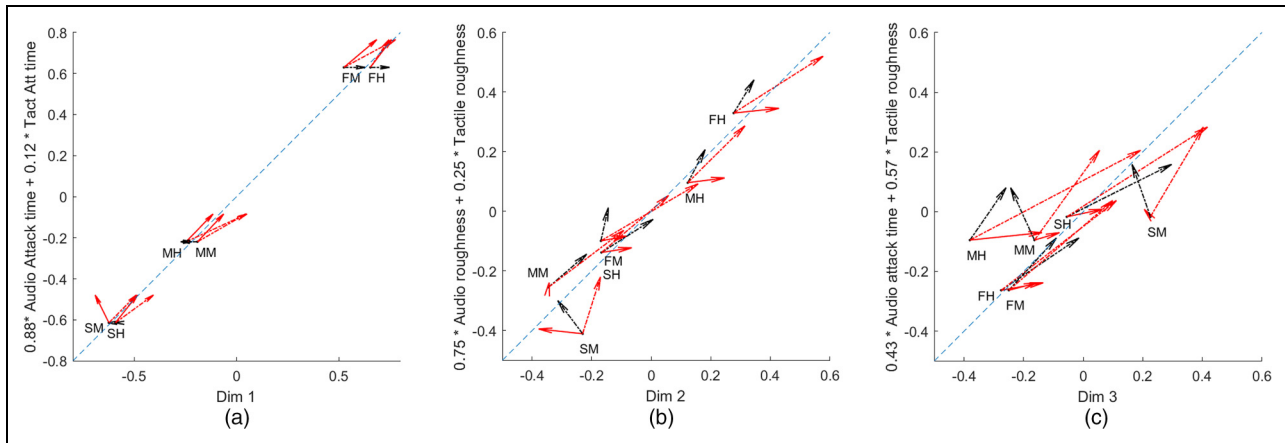


Figure 7. Results for NH Listeners. Panel. (a) Scatter plot between the projection along the first perceptual dimension and a descriptor composed of 88% of the audio attack time and 12% of the tactile attack time. Panel. (b) Scatter plot between the projection along the second perceptual dimension and a descriptor composed of 75% of the audio roughness and 25% of the tactile roughness. Panel. (c) Scatter plot between the projection along the third perceptual dimension and a descriptor composed of 43% of the audio attack time and 57% of the tactile roughness. Each arrow represents the position of the four stimuli with the same acoustic parameter but different tactile input (see caption of Figure 4 for more information). The two letters at the origin of each group of arrows represent the acoustic properties of the stimulus. The first letter represents the value of the attack time (F: Fast, M: Medium, and S: Slow), and the second represents the value of depth of the AM (High and Medium).

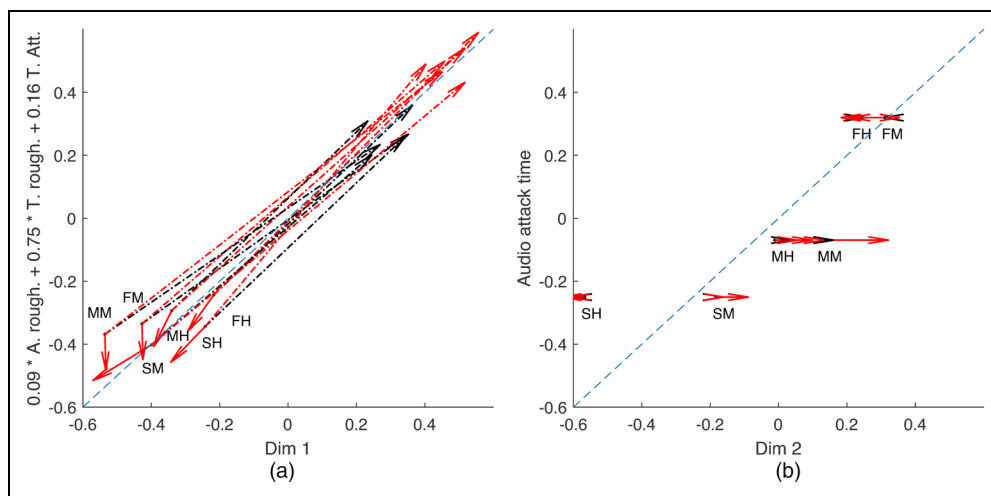


Figure 8. Results for CI Listeners. Panel. (a) Scatter plot between the projection along with the first perceptual dimension and a descriptor composed of 9% of the audio roughness, 16% of the tactile attack time, and 75% of tactile roughness. Panel. (b) Scatter plot between the projection along with the second perceptual dimension and the audio attack time (see caption of Figure 4 for more information).

tactile input, we should obtain a two-dimensional space correlated with the logarithm of the attack time and a model of roughness based on the amplitude modulation depth of the stimuli, as obtained in the control experiment (see supplement). However, as seen in Figures 4–6, the MDS analysis clearly shows the effect of the tactile input. First, for the NH listeners, the goodness-of-fit of the 3D solution was much better than the one for the 2D solution, indicating that the listeners have based their judgment on at least three perceptual dimensions.

Given that the stimuli varied on only two audio and two tactile parameters, it is reasonable to conclude that the listeners integrated some tactile parameters in their judgments. Second, for the CI listeners, the first dimension of their two-dimensional MDS is significantly linked to the tactile input, indicating that they have integrated this additional modality in their auditory dissimilarity judgment. As a consequence, the null hypothesis can be rejected for both groups. The other two hypotheses will be discussed below.

Multimodal Integration or Crossmodal

The MDS analysis is a particularly efficient tool for studying the nature of the audio-tactile interaction. If each modality contributes independently to the perception of timbre difference, then the analysis should reveal a dimension linked to each modality independently, as proposed in hypothesis 1. On the other hand, if the tactile modality enhances auditory perception, then the stimuli positions on the dimensions linked to the acoustics parameters should also be dependent on the values of the tactile modality, as stated in hypothesis 2. Our result shows that, for the NH listeners, the first and second dimensions were linked to specific timbre dimensions (impulsivity and roughness). However, the position of each stimulus was also influenced by the tactile equivalent, as proposed in hypothesis 2. This result indicates that a tactile input can change the perception of impulsivity and the roughness of a sound. To our knowledge, this is the first evidence of crossmodal processing of tactile input on timbre.

On the other hand, the results from CI listeners showed two orthogonal dimensions linked to two different modalities, as stated in hypothesis 1. It is, therefore, possible that CI listeners may have relied on the tactile input to form their rating without combining the two modalities. We can conclude that NH listeners show possible audio-tactile crossmodal processing (Hypothesis 2), while CI users seem to have a multimodal integration (Hypothesis 1). As the third dimension of the space obtained for the NH listeners was linked to the tactile modality, they should also experience some form of multimodal integration.

The Law of Inverse Effectiveness

Although listeners with a CI have a weak sensation of timbre (Marozeau & Lamping, 2019), previous MDS experiments found very similar spaces between NH and CI listeners

(Kong et al., 2012; 2011). However, in this study, the two spaces are widely different, and this discrepancy might be caused by the type of stimuli presented in this experiment: Kong et al. used complex stimuli, and our audio stimuli were limited to two simple dimensions of attack time and AM in this experiment. Although we ensured that the AM would be audible for the CI users, this cue might not have been salient compared to a potentially distracting tactile input.

As the auditory input is degraded for CI users, it is not surprising that they will put more weight on the tactile input as predicted by the law of inverse effectiveness (Meredith & Stein, 1983). This result indicates that an audio-tactile device might be particularly effective for people with degraded auditory perception.

The Third Dimension

The MDS analysis revealed a third dimension that could be modeled as a linear combination of the audio attack time and the tactile roughness. This dimension is linked to a contribution of the tactile input, and it is less clear why it can be modeled by a mixture of audio attack and tactile roughness. However, it is possible that the physical descriptor used to model the tactile input is not optimal and that dimension might be better predicted with a physical descriptor designed based on purely tactile psychophysics experiments.

Caveat and Limitations

This study represents a first step toward understanding the crossmodal enhancement of timbre with tactile stimulation. We, therefore, decided to use simple stimuli that we could control. Additional experiments are needed to test how the effect found can be applied to complex and ecological valid stimuli like recordings of musical instruments.

Using modulated pure tone presented in direct audio stimulation is not ideal with CI listeners. It is difficult to ensure that the processor's input level range was not distorted or compressed too much. For example, if the presentation level is too high, the whole signal would have been presented at the maximum level of stimulation, and as a consequence, the roughness cue would be reduced. This compressive behavior depends on many factors, such as sound coding strategy, input level, and spectral density of the signal. For example, Cochlear© devices have a hard clipping after 40 dB above the input level (set at 25 dB SPL by default). Oticon Medical, used primarily in this study, has a more soft compression behavior up to 100 dB SPL (Vaerenberg et al., 2014). Broadband noises with low internal modulation could have been used to reduce the overall level. Alternatively, the stimuli could have been presented through an experimental interface with direct control of the output current level.

The analysis reveals a contribution of 12% of the tactile attack time on the first dimension and 22% of the audio roughness on the second dimension in NH listeners. These ratios depend directly on the magnitude of our stimuli parameters, and a similar experiment but with other choices of parameters will inevitably result in different ratios. Nevertheless, this study demonstrates that tactile input can contribute to at least 22% of the perception of one auditory dimension.

Due to the constraints of the recruitment process, the two groups of listeners were not matched in age. Therefore, this factor may contribute to the difference observed. Finally, although 6 out of the 9 CI listeners used the same device (Oticon Medical), the other listeners were equipped with a Med-El or an Advanced Bionics device. This discrepancy can increase the variability of the result as different sound processing strategies could result in different perceptions of the same input sound.

Conclusions

A MDS experiment was conducted to test whether vibrotactile feedback can affect timbre perception in a population of normal-hearing listeners and cochlear-implant users. The results reveal that, for the NH listeners, the first dimension of space can be modeled with a linear combination of audio and tactile attack time. The second dimension is correlated with a combination of audio and tactile roughness. Finally, the third dimension is linked to a combination of audio attack time and tactile roughness. The result of the CI listeners reveals a much stronger contribution of the tactile roughness in their auditory judgment. This study clearly shows that at least two dimensions of timbre, impulsivity, and roughness can be modified by tactile stimulation. These results could help design new audio-tactile hearing devices that can help restore the perception of timbre for people with hearing impairment.

Authors' Note

NHL at four dimensions and solution CI at three dimensions.

Acknowledgements

The authors would like to thank Rikke Sørensen for her help testing Danish-speaking listeners, as well as all the listeners for their time and effort.

Authors' Contribution

T.V., S.A., and J.M. conceived the experiment; T.V. and S.A. conducted the experiment; T.V. and J.M. analyzed the results. All authors reviewed the manuscript.


Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Tushar Verma  <https://orcid.org/0000-0002-6678-5627>

Jeremy Marozeau  <https://orcid.org/0000-0002-4505-135X>

Supplemental Material

Supplemental material for this article is available online.

References

- Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. *Psychometrika*, *35*(3), 283–319. <https://doi.org/10.1007/BF02310791>
- de Leeuw, J., & Mair, P. (2009). Multidimensional scaling using majorization: SMACOF in R. *Journal of Statistical Software*, *31*(3), 1–30. <https://doi.org/10.18637/jss.v031.i03>
- Fastl, H., & Zwicker, E. (2007). *Psychoacoustics: Facts and models*. Springer-Verlag.
- Fletcher, M. D., Hadeedi, A., Goehring, T., & Mills, S. R. (2019). Electro-haptic enhancement of speech-in-noise performance in cochlear implant users. *Scientific Reports*, *9*(11428). <https://doi.org/10.1038/s41598-019-47718-z>
- Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature*, *462*(7272), 502–504. <https://doi.org/10.1038/nature08572>
- Huang, J., Lu, T., Sheffield, B., & Zeng, F. G. (2020). Electro-Tactile stimulation enhances cochlear-implant melody recognition: Effects of rhythm and musical training. *Ear and Hearing*, *41*(1), 106–113. <https://doi.org/10.1097/AUD.0000000000000749>
- Kong, Y. Y., Mullangi, A., & Marozeau, J. (2012). Timbre and speech perception in bimodal and bilateral cochlear-implant listeners. *Ear and Hearing*, *33*(5), 645–659. <https://doi.org/10.1097/AUD.0b013e318252caae>
- Kong, Y. Y., Mullangi, A., Marozeau, J., & Epsteina, M. (2011). Temporal and spectral cues for musical timbre perception in electric hearing. *Journal of Speech, Language, and Hearing Research*, *54*(3), 981–994. [https://doi.org/10.1044/1092-4388\(2010\)10-0196](https://doi.org/10.1044/1092-4388(2010)10-0196)
- Luo, X., Padilla, M., & Landsberger, D. M. (2012). Pitch contour identification with combined place and temporal cues using cochlear implants. *The Journal of the Acoustical Society of America*, *131*(2), 1325–1336. <https://doi.org/10.1121/1.3672708>
- Marozeau, J., de Cheveigné, A., McAdams, S., & Winsberg, S. (2003). The dependency of timbre on fundamental frequency. *The Journal of the Acoustical Society of America*, *114*(5), 2946. <https://doi.org/10.1121/1.1618239>
- Marozeau, J., & Lamping, W. (2019). Timbre perception with cochlear implants. In K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper, & R. R. Fay (Eds.), *Timbre: Acoustics, perception, and cognition* (pp. 273–293). Springer.

- McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, *58*(3), 177–192. <https://doi.org/10.1007/BF00419633>
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. <https://doi.org/10.1038/264746a0>
- Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science (New York, N.Y.)*, *221*(4608), 389–391. <https://doi.org/10.1126/SCIENCE.6867718>
- Piqueras-Fiszman, B., & Spence, C. (2011). Crossmodal correspondences in product packaging. Assessing color–flavor correspondences for potato chips (crisps). *Appetite*, *57*(3), 753–757. <https://doi.org/10.1016/J.APPET.2011.07.012>
- Riche, R., & Verheyen, S. (2020). Using cross-validation to determine dimensionality in multidimensional scaling. *Proceedings of the 18th International Conference on Cognitive Modeling*, Retrieved from <http://hdl.handle.net/1765/128988>.
- Russo, F. A., Ammirante, P., & Fels, D. I. (2012). Vibrotactile discrimination of musical timbre. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 822–826. <https://doi.org/10.1037/A0029046>
- Schürmann, M., Caetano, G., Jousmäki, V., & Hari, R. (2004). Hands help hearing: Facilitatory audiotactile interaction at low sound-intensity levels. *The Journal of the Acoustical Society of America*, *115*(2), 830–832. <https://doi.org/10.1121/1.1639909>
- Spence, C., Senkowski, D., & Röder, B. (2009). Crossmodal processing. *Experimental Brain Research*, *198*(2), 107–111. <https://doi.org/10.1007/S00221-009-1973-4>
- Vaerenberg, B., Govaerts, P. J., Stainsby, T., Nopp, P., Gault, A., & Gnansia, D. (2014). A uniform graphical representation of intensity coding in current-generation cochlear implant systems. *Ear and Hearing*, *35*(5), 533–543. <https://doi.org/10.1097/AUD.000000000000039>
- Yau, J. M., Weber, A. I., & Bensmaia, S. J. (2010). Separate mechanisms for audio-tactile pitch and loudness interactions. *Frontiers in Psychology*, *1*(OCT), 160. <https://doi.org/10.3389/fpsyg.2010.00160>