2009 Second International Conferences on Advances in Computer-Human Interactions

Modelling Perceptual Elements of Music in a Vibrotactile Display for Deaf Users: A Field Study

Maria Karam
Centre for Learning Technologies
Ryerson University
Toronto, Ontario, Canada
maria.karam@ryerson.ca

Gabe Nespoli
Department of Psychology
Ryerson University
Toronto, Ontario, Canada
gnespoli@ryerson.ca

Frank Russo
Department of Psychology
Ryerson University
Toronto, Ontario, Canada
russo@ryerson.ca

Deborah I. Fels
Centre for Learning Technologies
Ryerson University
Toronto, Ontario, Canada
dfels@ryerson.ca

Abstract

The Model Human Cochlea (MHC) is a sensory substitution technique for presenting music as multiple discrete channels of vibrotactile stimuli. The MHC prototype was introduced to a group of Deaf senior citizens at a community centre, where they could try the chair, and provide us with feedback about their experience. Preliminary results from this work suggest that the MHC can potentially offer a more effective means of expressing basic emotional information from music in a vibrotactile display when the signals more closely match the perceptual elements of the music such as melody and instrument parts. We describe the MHC prototype and present the results of our field study, which extends our lab experiments with the MHC to include Deaf and hard of hearing participants in a real world setting.

1 Introduction

Film represents a multimodal form of entertainment that draws on the fusion of audio and visual stimuli to tell a story and to create an emotional experience for its audience. Both the audio and visual elements of film are essential in creating the user experience, and both should be universally accessible to all those who seek out this form of entertainment. For deaf movie enthusiasts, closed captioning or sign language translations can make speech more accessible However, music and other forms of non-speech audio such as background noises are not included in captioning. Music, in general, has yet been effectively interpreted using alternative display channels, towards making the experience associated with music listening accessible to those with limited or no hearing ability. The Alternative Sensory Information Display (ASID) project aims to create a crossmodal display that provides alternative approaches to experiencing the emotional elements expressed by music and other audio. The Model Human Cochlea (MHC) is one technique we are exploring for communicating emotional information from music using a tactile display. In this paper, we report on a field study aimed at introducing the MHC to a group of senior citizens in a local Centre for the Deaf.

2 Background

There has recently been a surge in research investigating the use of crossmodal displays to improve accessibility to information across modalities. Information from one modality can be interpreted or translated onto an alternative modality, as when music is substituted by tactile vibrations [5] or visual stimuli [7]. But research into music-based applications that can assist the deaf community tend to focus on the development of task-based applications, such as learning to play instruments [6, 11] or understanding music theory [1, 14]. We have found little work in the area of music enjoyment, towards making the emotional experiences of music listening more accessible to deaf and hard of hearing people. While Eric Gunther has developed a suit that presents audio signals to the body as vibrotactile stimuli, this work aims to create music specifically for the tactile display, rather than on interpreting existing music as vibrotactile stimuli [5].

978-0-7695-3529-6/09 $25.00 © 2009 Crown Copyright
DOI 10.1109/ACHI.2009.64

IEEE computer society

Though it may be possible to express musical emotion by presenting specific musical characteristics such as rhythm, and tempo as vibrations [4], the emotional complexity of a composition often relies on additional elements of music such as timbre and harmony. Because these elements may extend beyond the tactile spectrum, this difference in frequency spectrum of the audio and tactile systems —audible vibrations span frequencies from 20 Hz up to 20 kHz, while our tactile system only detected vibrations ranging in frequency from 10 Hz and 1000 Hz [13] —could leave much of the audio spectrum inaccessible to the tactile system. In addition, instruments with frequency signals in the higher ranges, such as the flute, or violin, may be masked by vibrations from instruments in the lower frequency range, such as bass and drums. Despite these differences in the two perceptual systems, Deaf and hard of hearing people often seek out music by coming in close contact with audio speakers, which enables users to *feel the music*.

Since most audio signals that we find in western harmonic music centre around the middle C note (around 262Hz), and frequencies over 1000Hz that are out of the tactile range occur generally less than 15% of the time [12] it may be possible to feel most of the sounds presented in classical music as vibrotactile stimuli [10]. Thus, to improve tactile detection to the more subtle sounds, it is important to reduce the level of vibrotactile masking that occurs when lower frequency signals of drums or bass are present. We address this problem by isolating discrete sounds from a musical recording or composition for presentation as multiple discrete bands of audio-tactile signals that are spatially distributed along the body.

The MHC draws on some basic characteristics of the human cochlea as a design metaphor, and the tonotopic ordering of hair cells along the basilar membrane of the cochlea [3]. Each of the thousands of hair cells within the cochlea, in very simple terms, vibrate when a specific frequency signal is detected, stimulating that location along the basilar membrane. These vibrations lead to electrical potentials that are further processed by the audio cortex of the brain, leading to audio perception. Using this simplified model of the cochlea, we embarked on designing the MHC as an interface for communicating audio information through a tactile display.

3 The Model Human Cochlea

Drawing on rudimentary functions of the human cochlea, we developed the MHC using the following analogy: the body represents the basilar membrane in terms of the spatial layout of voice coils (VCs) that are used to deliver sound, which causes physical vibrations when they come in contact with the body, simulating the function of the hair cells [10]. Since both audio and vibrotactile signals can be measured in terms of frequency (Hz) and amplitude (dB) units, sensory substitution can easily be implemented since the displacements that the audio signals emit as air waves can easily be detected as physical vibrations when placed on the skins surface. The quality of the audio signal thus directly effects the quality of the tactile signal, and the intensity of the audio signal directly effects the intensity of the vibrations.

Figure 1: The MHC prototype presents 4 discrete audio-tactile channels along each of the four rows of voice coils used in the display.

3.1 Current MHC Prototype

To maximise the number of vibrotactile channels that can be displayed on the body in the MHC, we developed a prototype using voice coils (VCs), which are embedded into the back of a canvas chair in a four by two array configuration. This design leverages the large surface area available on the back to maximise the number of unique vibrotactile channels to increase the resolution of the audio-tactile display (see Figure 1). The design is additionally constrained by the signal range and physical size of the VCs, which must maximise the number of available channels, and produce the signal that is intended to be displayed on each channel. Each of the four rows of VCs presents a discrete band of audio signals that vibrate and create the tactile sensations corresponding to the music.

This is a different arrangement from our previous research [10] which presented unique signals to each of the eight VCs. In this work, we explore a different distribution of signals to create a more symmetrical effect by presenting pairs of signals to each of the four rows of VCs along both sides of the spinal column as an additional condition in our ongoing efforts to improve the design of the MHC.

The MHC aims to display existing musical recordings, and uses multiple, discrete vibrotactile channels to produce vibrations that reflect the music. Vibrotactile channels present vibrations corresponding to the instruments, sections, voices, or melodies of the music to the body, referred to as the track model (TM) of the MHC. Early studies have shown that the TM communicates some of the music's emotional expression better than what is possible when a single output channel is used to display the complete audio signal as vibrations [9]. It is not, however, always possible to gain access to the multi-track recordings or masters of existing music to support the TM. Since our work aims to improve the amount of vibrotactile information that can be detected from music, we developed a second model to support the presentation of existing music using the MHC, referred to as the frequency model (FM). The FM creates multiple bands of audio signals based on the separation of the music according to audio frequency bands rather than on individual track recordings.

In early experiments, we developed a working model of the FM, which distributes notes in music to the multiple vibrotactile channels according to a distribution model based on the frequency-of-occurrence of notes (DFN) commonly found in western harmonic [12]. Frequency signals that represent notes are distributed to vibrotactile channels based on their distance from the mean in the normal distribution [10]. Slight adjustments to the boundaries of the frequency bands produce a distribution that corresponds to octaves of notes that can be mapped onto the rows of VCs for the FM. The FM presents four bands of frequencies based on mapping the top two octave of notes of the piano keyboard to the top row of VCs. The second and third rows of the FM display the 3rd and 4th octaves respectively, and the fourth row at the bottom of the display is maps the lowest two octaves of notes onto the VCs. This supports an approximately proportional distribution of audio signals to each of the output channels of the MHC according to the DFN model. We refer to the vibrotactile versions of music we present on the MHC as *Vibetracks*. In this paper, we continue our research into the exploration of the presentation of music emotion through the MHC with the updated prototype, and a new population of users based in the Deaf and hard of hearing (HOH) community. We moved our lab experiment into a social setting where Deaf and HOH senior citizens meet once a week, and report on this field study and the modifications made to facilitate this new experimental environment.

4 Field Study

Slight modifications were made to our original protocol [10] in order to better meet the needs of the population of Deaf and HOH senior citizens who participated in this study. Modifications primarily involved adapting the written sections of the experiment to a sign-language format, so that questionnaires and rating scales could be replaced by images and sign language discussion. Interviews were also conducted American sign language (ASL), which was the preferred language of this community. Over 20 members of the community tried the chair, however only seven seniors agreed to take part in the formal study, 6 were deaf (5 female, 1 male) and one female was hard of hearing (HOH). Because many of the older members of the group were shy and did not feel comfortable answering questions about their experience in the chair, and did not agree to take part in the formal study, but have provided us with useful and positive comments about their experience. Participants were either residents of the centre, or members of the social club that meets once a week.

In order to emulate the intended interaction scenario for the MHC with an entertainment setting, we conducted the study in a common recreation room where the seniors met weekly to play pool, bingo, and watch television. In this study, we used the set of four midi-based classical music compositions from our previous study [10], consisting of two-30 second segments for each of the joy and sad midi recordings, selected from a collection of 27 classical music pieces that were rated for their emotional expressiveness [8]. We used only two emotions, joy and sadness, to eliminate confusion that could arise from having to rate a larger set of emotions in this novel interface. Like our previous experiment, we eliminated the sound that could be heard by participants through the VCs in the chair by providing those with some level of hearing with headphones and ear plugs. Tracks are listed below:

- Joy1 (H1): Vivaldi, A. Concerto in E major, Op8. Spring
- Joy2 (H2): Schumann, R. (SCA5) Symphony No. 4, Romance min. 1.53–2.23 Pilz CD 239
- Sad1 (S1): Mozart Piano Concerto No.9, Part 2
- Sad2 (S2): Schumann, R. (SCA4) Symphony No. 4, Romanze measures 1–10

For the study, the sign language interpreter explained the procedure of the experiment to each participant and recorded any comments expressed during the trials. Participants were instructed to sit in the chair, to relax, and to comment on any emotional expression they could detect from the vibrations. Participants were instructed to *sign out loud* so that the ASL interpreter could record their comments throughout the sessions. In addition, participants were to indicate the strength of any emotion they could detect in the vibrations as weak, medium, or strong.

Vibetracks were randomly assigned to participants during the trials for each of the FM, TM, and a control models

(CM), and the four midi recordings (H1,H2,S1,S2). Each session lasted approximately 15 minutes to minimize the disruption caused to the group, and to enable more members of the group to experience the chair.

5 Results

Of the 7 seniors who took part in our study, remarkably, all but one participant (female, HOH) used their hands to 'dance' to the vibetracks. All participants used facial expressions to indicate which of the emotions they were feeling, in addition to signing their thoughts to the interpreter. Three of the four deaf women were highly expressive in this way, with one deaf female participant expressing accurate facial expressions that reflected the content of the music in all vibetracks. The participants also exhibited highly expressive body gestures, which included using gestures of playing a violin, and of using a conductors baton. We received many comments throughout the sessions about the vibetracks, the comfort of the chair, and the overall experience of feeling the music. Ratings given to each vibetrack for emotional expressions are presented in Figure 2, and intensity ratings are shown in Figure 3.

As with our previous study, we found that the TM was most accurately interpreted as being joyful in both the H1 and H1 recordings. S1 was correctly rated as expressing sadness by 2 out of 3 participants in the TM, while S2 was incorrectly detected as expressing joy in all three conditions. The intensity of TM-H1 was rated as expressing strong or medium levels of joy across all participants. H2, S1, and S2 in the TM were rated weak in emotional expressiveness. In the FM trials, H1 received 1 sad rating, with other participants rating this track as joyful. H2 was also rated as joyful, while S1 and S2 received sad ratings, in addition to joy. All FM tracks were rated as expressing emotion at a medium intensity level. The CM trials showed that S2 did not express any emotion, while only H1 and S1 received ratings of joy. Ratings for H2 in the CM were mixed. Of all three models, the CM was rated weakest in intensity of emotional expression, except for S1, which was rated a being strong in emotional expression.

5.1 Discussion

Results support that the TM was most effective at expressing emotion as vibetracks, both for the joyful and sad tracks. The FM was less expressive than the TM, but was more expressive than the CM. The CM was rated weakest overall in emotional expressiveness, except in the S1 vibetracks since the CM presented the entire audio signal through all of the VCs, the joy rating for S1 in the CM condition could likely be due to the stronger signal of the display. Because the audio signal is presented as multiple signals in TM and the FM, the sad tracks could lead to a weak signal during quite moments or for sustained notes in some of the channels. Songs with little variation can also lead to interpretations of sadness in the track. Songs that have a wider variation in notes, stronger beat, or faster tempo produce vibetracks could lead to the detection of joy in the vibetracks.

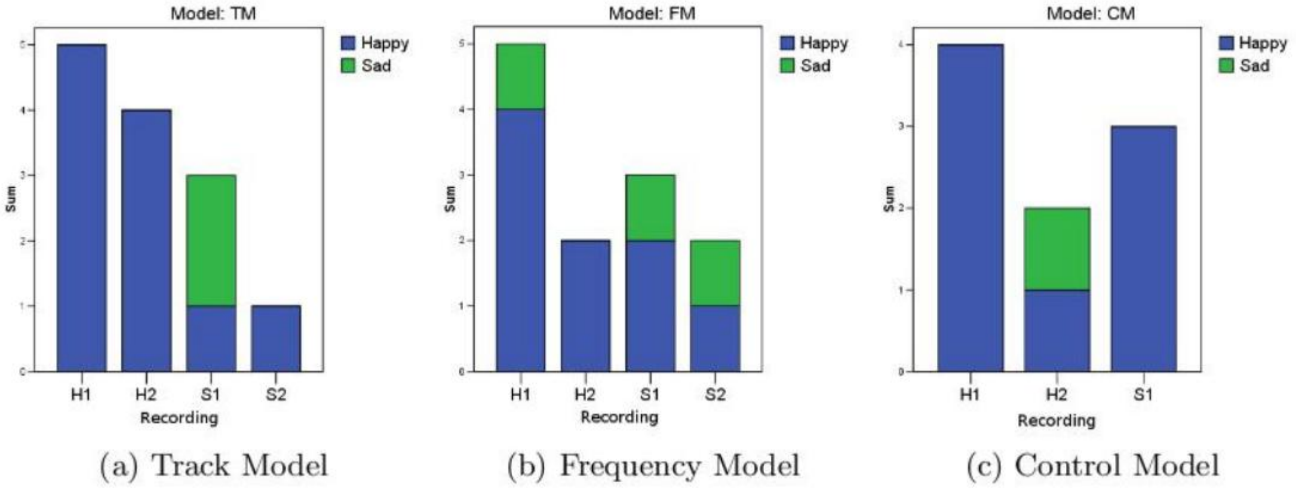
The results show that there is some correlation between the emotional expressiveness of a track based on the level of variation and intensity of the vibrations: higher, more varied signals convey joy when interpreting the emotional content of the vibetracks: the slower, less variable, weaker signals produce vibrations that are difficult to detect, which were either interpreted as sad, or as expressing no emotional information. More intense, varied signals could promote the impression of joy in vibetracks for the TM and FM. We also found that in the CM, where there all the signals are the same, the sad tracks could present a stronger signal in general, which may explain the interpretation of the sad track as being more joyful, with the intensity of the varied joyful tracks lost in the single channel output.

These results raise interesting questions about the how well the MHC could be used to distinguish vibetracks based in joy or angry songs, which both exhibit high energy, varied signals. Pilot studies are promising in comparing anger to joy vibetracks, which we have observed to be stronger and more varied than the joy tracks, suggesting that it is possible to detect more than just two emotions in vibetracks. We are working towards developing a taxonomy for understand the relationships between the different characteristics of the vibrations, and the emotions they are associated with.

From our observation, we note that all of the members of the group who tried the chair expressed some form of movement, facial expression, or hand gestures that suggested they were enjoying the vibrations, could move in time with the vibrations, and often could alter their movements to reflect the type of music they were feeling.

One of the most interesting observation from this study was the dance-like gestures that the participants expressed through hand and finger movement while the vibetracks played. These expressive participants also tended to mimic what they perceived in the vibrations through this specific style of movement, using fingers to match the rhythms of the vibetracks and facial expressions to reflect the intensity of the vibrations. It was this dancing that suggested that the participants were experiencing more emotions from the chair for all of the joy recordings in each model, however, the TM consistently led to more intense expressions than both the FM and the CM.

While the FM performed better than the CM in all cases, it is still not clear if there is a significant difference between the FM and the CM. Further research is underway to improve the FM so that it can better reflect the percep-



(a) Track Model (b) Frequency Model (c) Control Model

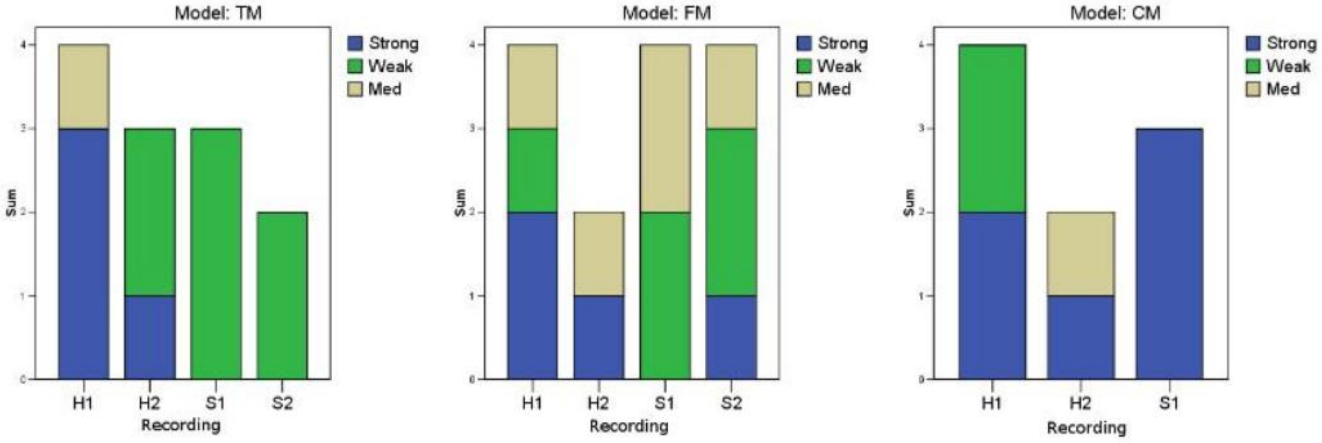
Figure 2: The three graphs present the raw data obtained from our study, divided according to the model (TM,FM,CM), with bars showing the results of the participant's ratings of each vibetrack.

tual elements of the music, such as rhythms, instruments, or melody. We found five of the six women who took part in the study responded more favourably to the vibrotactile sensations than the male participant, but one of the women complained of body aches and ended the study early. Of the over 20 members of the community who tried the chair, the female participants tended to be more expressive in their gestures, while the males were more subtle. This could be attributed to the expressive nature of the individual in general, although everyone who did try the chair exhibited some level of physical or facial expression, mirroring the emotional content of the music they were experiencing.

The observations from this study support and extend our previous empirical findings [10], suggesting that the TM is most effective in expressing joy or sad emotions for participants who are deaf, hard of hearing, as well as for hearing people. We are beginning to form new hypotheses about the nature of the relationship between vibrations and music. For example, there could be a link between the way we perceive sounds and the ability to detect these sounds as vibrations when they more closely map onto the individual sections of the music. As in the TM, where the vibrations closely match the original sounds of the instrument, melody, or rhythm, and we are considering other approaches to creating an FM that can better mirror these in how we perceive them in the vibrotactile display.

6 Conclusions and Future Work

We presented a study which explored the potential benefits of using the MHC as a sensory substitution technique towards creating a higher-resolution vibrotactile display that distributes the audio signal to multiple, discrete vibrotactile channels. One of the main contributions of this work is the modification of the lab-based experiment that could facilitate a field study in a local Centre for the Deaf. Results from this study further support previous findings in that the TM was shown to be most expressive emotionally, closely followed by the FM, with the CM showing least intensity or expression. The field study approach allowed us to access participants that would otherwise not be interested in taking part in a lab experiment. Most of the seniors who participated would never have been able to participate in the empirical version at our lab due to limitations in their mobility, or because many would simply have no interest in a lab experiment. By taking the lab to the community, we were able to obtain end user responses to our research, and to introduce a new technology to users who could benefit from this novel approach to making music accessible. We will continue this study, and explore the effects of the TM on a larger set of stimuli that presents vibetracks for music that has been previously validated as expressing emotions such as joy, anger, fear, sadness [2]. Finally, we hope to contribute a better understanding of the relationship between music and vibrations by using the MHC as a test bed for conducting further studies on crossmodal audio-tactile displays, towards making musical expression a more universal, inclusive form of entertainment for people of all levels of hearing ability.



(a) Track Model (b) Frequency Model (c) Control Model

Figure 3: The three graphs present the raw data obtained from our study for user ratings for the strength of the emotional expression detected in each of the vibetracks.

7 Acknowledgements

Funding for this project has been generously provided by the Natural Sciences and Engineering Council of Canada. We would also like to thank all those at the Bob Rumball Centre for the Deaf who volunteered to participate in our research. In addition, we acknowledge Norma Thompson, Lisa Chan, Carmen Branje, Gabe Nespoli, and Robert Indrigo who contributed to this research.

References

[1] T. Bergstrom, K. Karahalios, and J. C. Hart. Isochords: visualizing structure in music. In *GI '07: Proceedings of Graphics Interface 2007*, pages 297–304, New York, NY, USA, 2007. ACM.

[2] E. Bigand, S. Vieillard, F. Madurell, J. Marozeau, and A. Dacquet. Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition and Emotion*, 19(8):1113–1139, 2005.

[3] G. v. Bksy. *Experiments in Hearing*. McGraw Hill, New York, 1960.

[4] D. P. Ellis. Extracting information from music audio. *Commun. ACM*, 49(8):32–37, 2006.

[5] E. Gunther, G. Davenport, and S. O'Modhrain. Cutaneous grooves: composing for the sense of touch. In *NIME '02: Proceedings of the 2002 conference on New interfaces for musical expression*, pages 1–6, Singapore, Singapore, 2002. National University of Singapore.

[6] R. Hiraga and N. Kato. Understanding emotion through multimedia: comparison between hearing-impaired people and people with hearing abilities. In *Assets '06: Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility*, pages 141–148, New York, NY, USA, 2006. ACM.

[7] R. Hiraga and N. Matsuda. Visualization of music performance as an aid to listener's comprehension. In *AVI '04: Proceedings of the working conference on Advanced visual interfaces*, pages 103–106, New York, NY, USA, 2004. ACM.

[8] K. Kallinen. Emotional ratings of music excerpts in the western art music repertoire and their self-organisation in the kohonen neural network. *Psychology of Music*, 33(4):373–393, 2005.

[9] M. Karam and D. I. Fels. Proceedings of the first international workshop on haptic ambient systems. Quebec City, Canada, 2008.

[10] M. Karam, F. Russo, C. Branje, E. Price, and D. I. Fels. Towards a model human cochlea: Sensory substitution for crossmodal audio-tactile displays (in press). In *GI '08: Proceedings of Graphics Interface 2008*, New York, NY, USA, 2008. ACM.

[11] M. T. Marshall and M. M. Wanderley. Vibrotactile feedback in digital musical instruments. In *NIME '06: Proceedings of the 2006 conference on New interfaces for musical expression*, pages 226–229, Paris, France, France, 2006. IRCAM — Centre Pompidou.

[12] F. A. Russo, L. L. Cuddy, A. Galembo, and W. F. Thompson. Sensitivity to tonality across the pitch range. *Perception*, 36:781–790, 2007.

[13] C. Strumf. *Tonpsychologie*, volume 1. S. Hirzel, Leipzig, 1883.

[14] P. Toiviainen. Visualization of tonal content with self-organizing maps and self-similarity matrices. *Comput. Entertain.*, 3(4):1–10, 2005.