Designing a Model Human Cochlea: Issues and challenges in crossmodal audio-haptic displays

Maria Karam
Ryerson University
Toronto, Ontario
Canada
maria.karam@ryerson.ca

Deborah I. Fels
Ryerson University
Toronto, Ontario
Canada
dfels@ryerson.ca

ABSTRACT

In this paper, we describe a Model Human Cochlea (MHC), a sensory substitution technique aimed at translating some of the emotional content expressed in music onto a haptic ambient display. We present some of the issues and challenges encountered in designing the model. This research is situated within the domain of crossmodal displays, with specific focus on enhancing the entertainment experience associated with film audio for users who are deaf or hard of hearing. The interface, in its final form factor, will be integrated into an EmotiChair, a multi-sensory entertainment interface that supports crossmodal audio-haptic display interactions. To assist with the design of the MHC, we have developed a flexible prototype to support research in crossmodal audio-haptic displays. Details of the multidisciplinary design process that has informed the development of the the MHC prototype, and the evaluation methodology adopted to explore the different configurations of the MHC are presented.

Categories and Subject Descriptors

H.1.2 [**Information Systems**]: User/Machine Systems—*Human factors, human information processing, software psychology*; H.5.2 [**Information interfaces and presentations**]: User Interfaces—*Haptic I/O*;

General Terms

Sensory substitution, model human cochlea, crossmodal displays, haptic interfaces, music, emotion

1. INTRODUCTION

Crossmodal displays represent a rapidly growing field in human-computer interaction research, where information that is designed to be presented to one sensory modality is displayed using an alternative modality. Much of the research focus within this domain lies in the exchange of input and output stimuli between the audio, visual, and tactile modalities, although it is theoretically possible to create an alternative display for any sensory modality. These types of alternative display fall into the class of *crossmodal displays*, with the most prominent areas of research based in the presentation of audio signals using visualizations [17], visual information using haptic displays [21], and audio signals using haptic displays [3]. Crossmodal displays are used for a variety of applications, including mobile computing [3], virtual reality [4], assistive and adaptive technologies [9], and entertainment applications [14]. Often these displays are used to augment a primary display with a source of redundant information that is presented using a different display modality. For example, a driving game that is presented on a primary visual display can also incorporate a secondary audio display to provide the corresponding driving sounds to the user. A a third display can also be used to further extend the multisensory experience by including tactile sensations such as motions that reflect the physical events ocurring in the game environment.

A different approach considers the substitution of one modality using another, where the stimuli from the input modality is replaced by the stimuli from the alternative display modality. This type of display is primarily aimed at providing users who are blind or deaf with an alternative means of experiencing the stimuli from one sensory modality using the other, however it also offers an enhanced sensory experience to all users. Some examples include pin arrays, which provide a tactile replacement for visual information, and music visualizations, which offers a visual interpretation of music audio.

One of the main research challenges in creating such displays is in determining which of the characteristics of the input modality to present using the alternative display. This exchange of stimuli from one modality to another is commonly referred to as *sensory substitution* [24] and involves the identification, selection, and mapping of characteristics between the input and the display modalities. The characteristics or properties of the two modalities must be identified, drawing on perspectives such as the physical, perceptual, or computational nature of the stimuli; visual stimuli can be described using angles, lines, or coordinates as physical characteristics, perceptual characteristics such as colour or type, or computational characteristics such as the pixel activation of the rendered object. Once identified, the properties of the input modality must then be interpreted, translated, or otherwise mapped onto the properties of the alternative display to achieve the desired effect.

HAS 2008, February 11-14, Quebec, Canada

DOI 10.4108/ICST.AMBISYS2008.2837

Mapping the characteristics of the input stimuli to the alternative display can be done using a number of different approaches. A translation would be used when there exists some natural way to map the two modalities, as with pin arrays and visual displays: here, pixels in a visual display can be translated onto the low resolution pins of a tactile array, which serve as tactile pixels. Alternatively, an interpretation would be appropriate when the properties of the input modality do not correspond to those of the display modality, or when a more ambient form of information is intended for the alternative display. For example, music visualizations provide an ambient interpretation of the audio signal that does not aim to accurately recreate the music, but rather, creates an alternative way of experiencing music [17]. Using a different approach, one can adopt a direct mapping, when possible, of properties between the two modalities: frequency and amplitude are examples of properties that can be used to describe both sound and physical sensations. In one example, the vibrations that result during human speech can be detected when touching the throat of the speaking person. This technique, referred to as the Tadoma method [23] and has been shown to improve lip reading comprehension for users who are deaf-blind. Similarly, music audio can also be detected by touching an audio speaker that is producing sound, a technique referred to as speaker listening.

Designing alternative displays to effectively reproduce emotional characteristics conveyed in audio poses many challenges for researchers. Consider closed captioning: although text is presented as a replacement for speech audio we rarely see representations for the other sound components such as music or sound effects incorporated into the captioning [12]. These components may be omitted from captioning due to the insufficient time, screen space, or knowledge available on how to represent these audio features using text: leaving the caption user with an incomplete or even inaccurate version of the production. Because much of the sound information is commonly excluded from captioning, it is of little consequence to those who are not accustomed to interpreting audio, even though important information through music, speech prosody, and sound effects.

Sound is an important element of film and television media, designed to create moods or enhance emotions for its viewers [5]. Films rely on music as a technique to engage the audience and to create additional sensations that affect the deepest subconscious levels of human experience. This relationship between music and film persists in all forms, cultures, and genres, but can be lost to deaf and hard of hearing audiences. While it is possible to detect sound as tactile vibrations, musical often consists of multiple layers of audio signals that are presented as a single source of audio. To illustrate, consider the sound waves presented in Figure 1. Here, we see that there are several individual signals that, when combined, produce the sound wave shown in Figure 2. If each of the signals from Figure 1 are presented as tactile sensations using separate vibrotactile devices, it will be possible to express more of the vibrational content of the music than what is possible with the signal shown in Figure 2. This represents the basic premise behind the MHC.

Another issue that the MHC addresses involves the different physiological and psychophysical properties of the the

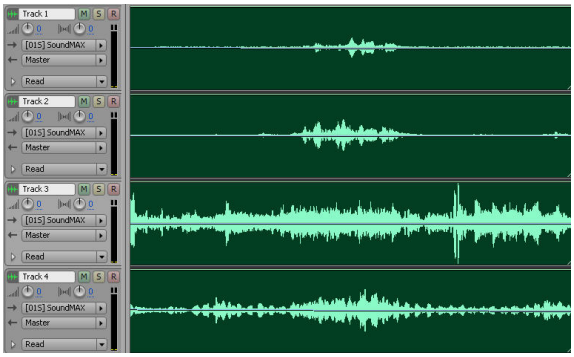

Figure 1: The image shows the sine waveforms for the individual tracks used in a midi recording of a classical composition.

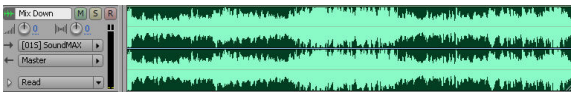

Figure 2: The image shows sine waveforms of the mixed version of the classical midi track from Figure 1.

auditory and cutaneous systems: this introduces important constraints and considerations for designing sensory substitution techniques for audio-haptic displays. First, the auditory system is only capable of detecting a limited range of frequencies then that of the cutaneous system: the human auditory system can detect 20 Hz - 20 kHz, while cutaneous receptors detects 20 Hz - 1000 Hz [22]. Second, the audio system has a difference threshold or *just noticeable difference* (JND) that is much more sensitive than our cutaneous receptors: according to Weber's Law, the detection of stimulus on the skin is proportional to the log of the stimulus intensity. Stimuli can be expressed according to the receptor type, the concentration and quantity of those receptors in a sensory organ, the rate at which receptors respond (sensitivity), and the functional complexity of the sense organ. These factors effect the signal to noise detection ratio of a sensory system, which in turn effects the ability of that system to accurately detect, disambiguate, and process stimuli [2].

To attempt to address these constraints, we have developed a sensory substitution technique for translating audio signals onto a tactile display. We refer to this as the *Model Human Cochlea (MHC)*, which draws on some of the basic features of the cochlea as a design metaphor. The basic premise of the MHC is to present the different layers or segments of music using a series of tactile displays, in this case, we are using audio speakers, which produce sufficient vibrations for our model. In the rest if this paper, we present the MHC and the initial stages in its development: we discuss some of the major challenges encountered in designing the MHC prototype, and describe the experimental methodology we developed to conduct further evaluations and design alternatives of the MHC.

2. RELATED RESEARCH

Crossmodal displays represent a multidisciplinary field of research within the computing sciences, incorporating techniques from psychology that provide insight into the perceptual elements associated with the different modalities, psychophysics, which enables the measurement of sensory stimuli perception, and computing and engineering sciences that support the design and implementation of crossmodal displays. Tactile displays have been adopted as an effective modality for presenting both audio and visual stimuli. For example, pin arrays create a tactile representation of objects that would commonly presented on a visual display. The pins are arranged in a grid formation to produce vibrotactile sensations that stimulate the skin in a pattern corresponding to the low resolution version of the visual display object [21]. In addition to vibrations, there are other forms of tactile stimulation that create pressure, thermal, or electro-cutaneous stimulation, but it is the vibrotactile stimuli that have been most commonly used in tactile tcrossmodal displays [3, 8, 25].

Some crossmodal displays are primarily intended to deliver accurate representations of information from the input modality to the alternative display; for example, pin arrays aim to create an accurate tactile representation of visual information such as braille lettering [13]. Other crossmodal displays have less stringent requirements for precision in information delivery, presenting relative information pertaining to the direction or spatial layout of objects located in the environment or a visual scene [16, 15]. When information must be accurately communicated, this places additional constraints on the types of mappings that are used to associate the properties of the two modalities.

However, ambient forms of crossmodal displays aim to enhance the experience associated with games [14] and other entertainment applications [8], and impose less stringent requirements on accurately representing the input data than is required for information-critical displays: video game chairs are fitted with actuators that create motion to simulate the physical actions of the game, increasing the immersion for the user by increasing the sensory experiences associated with playing the game [18].

Ambient haptic displays can also enhance the experiences associated with music. In one example, vibrotactile devices are embedded into a suit to create spatio-temporal patterns of vibrations on the body [8]. Audio transducers create the vibrotactile sensations in an interface that augments music with tactile vibrations. Music is a highly expressive form of art that communicates different information to different people. Some of the information that is expressed through music, such as beat and rhythm, can easily be detected when we make physical contact with an audio speaker. But although the complete set of vibrations that are expressed in the music are detectable as audio signals, most of this information is lost to human tactile perception. Psychophysical research has shown that human cutaneous receptors, specifically the Pacinian corpuscles, are most effective at perceiving single-point vibration stimuli below 1000 Hz, with optimal detection occurring around 250 Hz [22]. However, our auditory system is capable of detecting frequency vibrations that range from 20Hz to 20kHz. When designing a music-tactile display, this difference in perceptual ability between the auditory and tactile receptors must be addressed.

2.1 Ambient Audio-Haptic Displays

In one approach to designing crossmodal audio-haptic displays, only signals below 1000 Hz are included in the tactile display [8]. One problem with this approach is that it excludes existing musical recordings from being presented in the display. To accommodate the limitations in tactile perception, existing recordings would have to be altered in some way to ensure they could be perceived through touch. Songs could be pitch shifted down so that the maximum frequency output is below 1000Hz, or simply excluded from the display. But altering the music in any way could potentially lead to an incomplete representation of the audio stimuli, potentially excluding important information from expression through the tactile display.

To create a more accurate set of vibrations that represent music, one would need to use audio speakers in order to obtain the most complete set of vibrations from the recording, however, this leads to a different problem. Although we can hear most of the different elements of a musical recording when perceived as audio through a single speaker, only the strongest signals can be detected through touch. Since most of the weaker signals are essentially drowned out by the strong, rhythmic vibrations of instruments in the lower frequency ranges, information contained in the higher frequency ranges are essentially masked by the lower sounds. Though it is possible to strengthen the signal of the higher frequencies in a musical recording, this would again alter the original content of the music

One solution to presenting more information from an existing recording as tactile stimulation is using audio speakers is to adopt the basic functionality of the cochlea to a tactile model. Since the cochlea is designed to detect separate signals, which are then transmitted to the audio cortex to create audio perception, it may also be possible to present separate vibrational signals to the skin in order to simulate the signal dissemination feature of the cochlea. By separating out the different signals contained in a musical piece and present them using individual vibrotactile devices, it may be possible to present more information from music than is possible using a single speaker, while maintaining the original content of the music. This forms the basic premise for the design of the MHC, which uses multiple sensors to present different segments of music through touch.

3. DESIGNING THE MHC

The long term goal for this project is to develop an entertainment chair that can present the emotional content associated with film audio using an ambient haptic display. The tactile display will be embedded in the EmotiChair and used to present some of the emotional content of film audio. While the MHC is intended to assist deaf movie goers in accessing music content, a more general goal is to enhance the overall entertainment experience for all film enthusiasts. The first step in this project is the presentation of music using a tactile display. Creating an alternative tactile display that can effectively communicate some of the emotional content expressed in music is a challenging project. While music is known to be emotionally expressive when perceived through

our auditory channels, there has been little research focus on determining if that emotional content can also be communicated using an alternative display, or what characteristics of the music best represent that emotion in a tactile display. For example, for a song that conveys elements of the happy emotion, it could be possible to identify the individual characteristics that are responsible for creating the emotional content, such as pitch, tempo, and chord progression, which are known to express emotion such as tempo and articulation [6, 10]. However, it is not clear if any of those elements can convey the same information when presented as vibrotactile stimuli. Although some of the characteristics that contribute to the emotional expression of music are identifiable, these components alone are not enough to convey that which is presented as music. Also, altering the music so that we extract only the information that is considered to communicate emotion does not preclude the potential loss of valuable emotional information necessary to create an effective crossmodal representation of the music. Thus, in the initial stages of our research, we wanted to explore the complete set of audio signals and alter only the way that they are presented in a vibrotactile device, rather than the original content of the music.

With this constraint in place, we selected a vibrotactile device that could best be used to present audio information in its most accurate form: audio speakers. We had initially considered using motors as vibrotactile devices, however since motorized vibrotactile devices typically vibrate at frequencies well under 1000 Hz, we would potentially lose much of the audio signal that could be transmitted through vibrations. In addition, vibrations occurring at higher frequencies will naturally generate sound, so audio speakers remain the obvious vibrotactile device for the direct translation of audio to tactile stimuli considered in this research. Although the cutaneous receptors can detect some of the vibrations when an audio speaker is placed on the skin, the different frequency spectrums for audio and tactile perception impose limitations on what we can perceive through touch. To address this problem, we adopt the cochlea as a metaphor for designing our tactile display.

3.1 The Cochlea Metaphor

The cochlea is the organ in the ear that is responsible for the detection of audio signals. There are two fundamental features of the cochlea which we adopt for our metaphor: the individual sensors that detect the different frequency waves, and the linear organization of hairs that detect signals within the cochlea. The cochlea is a spiral structure that is situated within the inner ear, which contains thousands of microscopic hairs that are tuned to detect sound vibrations of different frequencies. The hairs convert the frequency vibrations into electrical potentials that are sent to the audio cortex, and processed as audio perception. We draw on this feature of the cochlea to inform the design of the MHC, which uses multiple speakers to send separate signals to the tactile receptors in the skin.

A second feature of the cochlea is the configuration of the hairs: although the cochlea is a spiral structure, when uncoiled, it reveals a linear structure of the hairs, where low frequencies are detected towards the inner part of the cochlea (the apex), and high frequencies detected towards the opening of the ear (the oval window). This informed the layout of speakers along the body, which presents highest frequencies at the upper most speakers, progressing down to the lowest frequencies at the bottom of the speaker configuration. Because the cochlea uses thousands of hairs to detect individual sound waves, we had to determine an alternative approach to grouping and mapping individual signals to the limited number of speakers we could use in our tactilel display.

3.2 Speaker: Location, Number, and Size

We addressed several issues relating to the distribution and layout of speakers on the body. First, we had to determine the number of speakers that we could use to create the MHC; this was dependent on two interrelated factors: the physical locations on the body where the speakers would be displayed, and the size of speakers that could be used in the display.

Since our approach aimed at distributing musical signals to several individual speakers, we considered two different configurations for placement: one involved distributing the speakers to different parts of the body (distributed model), while the other option was to place the speakers on the same body part (localized model). For our distributed model, we considered the placement of speakers along the legs, and back. Our initial setup placed four-inch speakers on the body: two for the upper and lower back, and one for each of the legs, just below the seat. Early pilot studies with this configuration suggested that the speakers were too big and were mostly uncomfortable and distracting for users.

In a second configuration of the distributed model, we switched to eight two-inch speakers: two speakers were placed on the upper arm, four were attached to the upper and lower back, with two on the back of the upper thigh. Pilot tests conducted on this model suggested the different sensitivity levels of the body proved to be mostly distracting to participants, since the arms were more sensitive than other parts, demanding more of the users attention.

In the localized model, we altered the set up, placing all eight speakers on the back of the user. This model provide to be more conducive to the effect we were aiming for, which would produce a cohesive set of stimuli that could more naturally be associated with the structure of a complete musical composition. Also, the choice to use two-inch speakers enabled us to maximize the number of speakers we could use in the tactile display, while minimizing the area the would take up on the body. While we considered using up to 16 speakers for the MHC display, we chose to limit the number to eight for the early stages of the research so that we could first determine if the MHC approach would produce a more effective method of communicating audio information through vibrations than is possible using a single source of audio signal.

Speakers were placed in rows of two along the back, with each speaker would presenting a different signal, shown in Figure 3. An alternative configuration, which sends the same signal to each of the two speakers in a single row is being explored in current experiments to create a more symmetrical sensation for the user.

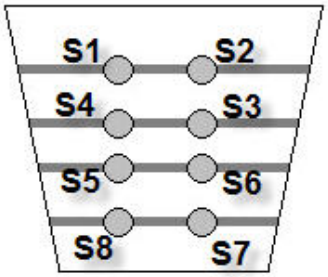

Figure 3: The diagram represents the layout of the speakers on the back, based on the speaker number, which displays signals within predetermined frequency ranges.

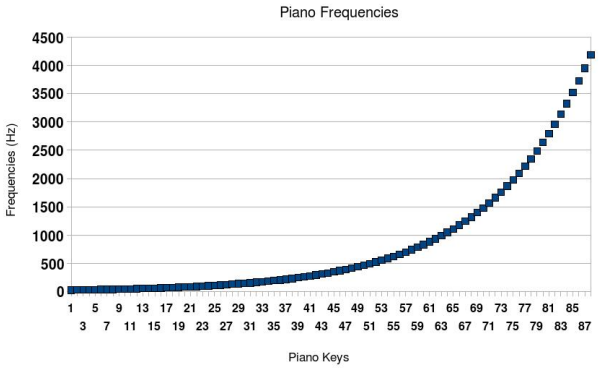

Figure 4: The graph shows the logarithmic distribution of notes on the piano to their corresponding frequency values. This relationship informed the distribution of frequency bands to the different speakers used in the MHC.

3.3 Sensory Substitution Models

One of the more challenging problems we address in this research involves the distribution of signals to the different speakers. Two approaches were considered for the MHC: the frequency model (FM), which separates the audio recordings into separate ranges or bands of frequencies, and the track model (TM), which divides the signal into the separate sources of audio, such as instruments, vocals, percussion, and other identifiable layers of a musical composition. We limited the range of frequencies we would use in the MHC to that which is presented on a piano keyboard, ranging from 27.5Hz to 4186Hz, which also includes the approximate range of frequencies associated with orchestral music.

3.3.1 Frequency Model

Each of the eight speakers used for the MHC were assigned to a unique range of frequencies. While we originally considered distributing the frequencies as equal segments to each speaker, separating the frequency signals of the piano into eight equal segments, this would have resulted in a disproportionate distribution of signals to speakers. Since is has been shown that the largest proportion of notes that occur in western harmonic music centers around them middle of the keyboard, around the D# key, notes that are at either end of the piano occur less frequently in this type of music [20]. Thus, most of the audio signals would be directed to the speakers that displayed the mid ranges of frequencies, while those responsible for the upper and lower notes would receive little or no signal. To create a display with a more equally distributed set of signals, we modeled the frequency distribution of signal to speaker on the natural distribution of notes that occur in western harmonic music. Russo et al. (2005) determined that there is a logarithmic relationship that exists between notes on the piano, and their corresponding frequencies. This relationship between notes and frequencies can be expressed using the normal distribution, which subsequently has an approximate standard deviation of one octave on a keyboard (see Figure 4).

Since we were focusing our research on western classical music, this configuration would be the most natural approach to distributing the signals to the different speakers, while ensuring the most equal distribution of musical signals to each. The final distribution of signal to speakers is presented in Figure 5. The rationale behind the distribution

Speaker	Lowest Hz	Highest Hz	Band Size	Mid-Hz
S1	1480	4186	2307	1655
S2	660	1480	1070	820
S3	467	660	563.5	193
S4	311	466	389	156
S5	220	311	265.5	91
S6	145	220	238	186
S7	69	145	107	76
S8	27.5	69	51	36

Figure 5: The table shows the distribution of frequency bands to each speaker, used in the FM model of the MHC.

of frequencies to the audio speakers used in the MHC is as follows: There is a normal distribution that represents the frequency in which notes of the piano are played in western harmonic music. This influences how we distribute the bands of notes, which are represented by their approximate frequency in Hz. We make an approximate division of notes using the normal curve as a guide. Since 1 standard deviation (sd) of a normal distribution contains roughly 68% of all data points under the normal curve, we round that number down to approximately of 50% of the notes. Thus, 50% of notes should be represented by 50% of the speakers. In our model, four of the eight speakers are used to represent 50% of the notes (1/2 sd each), with the remaining four speakers presenting one full standard deviation each at the higher and lower ends of the frequency spectrum, where fewer of the notes are commonly played. A similar distribution is used to inform the TM, which is discussed next. Figure 6 shows the normal curve, the distribution of notes which are measured as vibrational frequency (Hz), and the speakers on which they are assigned, which match the frequency bands presented in Figure 4.

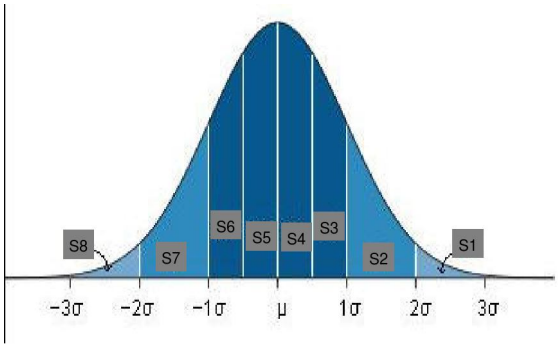

Figure 6: The diagram shows the distribution of frequency bands to each of the speakers used in the MHC. Each standard deviation under the normal curve represents the unique set of notes that are to be presented on individual speakers in the FM model of the MHC.

3.3.2 *Track Model*

The *Track Model (TM)* is similar to the FM in that there is a high to low distribution of frequency bands to the speakers, which is informed by the distribution model used in the FM. However, the TM applies an additional step and separates the audio signals into individual tracks that represent instruments, vocals, percussion, or other identifiable section of music (see Figure 7).

In this distribution, low frequency sounds related to instruments such as a drum or bass, are displayed on the lower speakers, while sounds in the higher frequency ranges, including instruments such as the flute or violin, are presented higher up on the back in the first and second speakers. One

INSTRUMENT	TYPE	Approx. Frequency Range	Avg. Hz.	Speaker
VOCALS	Soprano	250Hz - 1K	625	S3
	Baritone	110Hz - 425Hz	265	S5
	Bass	80Hz - 350Hz	210	S6
WOODWIND	Piccolo	630Hz - 5K	2815	S1
	Flute	250Hz - 2.5K	1375	S2
	Oboe	250Hz - 1.5K	875	S2
	Bass Clarinet	75Hz - 800Hz	438	S4
	Bassoon	55Hz - 575Hz	126	S7
	Saxophone	225Hz - 1K	612	S3
BRASS	Trumpet	170Hz - 1K	585	S3
	Trombone	80Hz - 600Hz	340	S4
	Tuba	45Hz - 375Hz	210	S6
STRINGS	Violin	200Hz - 3.5K	1850	S1
	Viola	125Hz -1K	560	S3
	Cello	63Hz - 630Hz	346	S4
	Double Bass	40Hz - 200Hz	120	S7
	Guitar	80Hz - 630Hz	355	S4
KEYBOARD	Piano	28Hz - 4.1K	2064	S1-S7
	Organ	20Hz - 7K	3510	S1-S8
PERCUSSION	Celeste	260Hz - 3.5K	1880	S1
	Timpani	90Hz - 180Hz	135	S7
	Xylophone	700Hz - 3.5K	2100	S1

Figure 7: The table shows the different instruments, their approximate frequency ranges, and the proposed distribution of each to the different speakers of the MHC.

advantage of the track model is that there is an intuitive level of understanding for this type of distribution. Since we can perceive and identify individual instruments in music as audio stimuli, it seems an intuitive approach to convey the content of music through vibrations by ensuring that the individual sections and layers of a musical track are presented as individual signals in the tactile display. However, there are several problems with this model. First, it is not always possible to obtain original music recordings in an unmixed format, which is what is required to distribute a recording using the TM. Since source separation techniques are not yet able to reliably isolate separate components of musical recordings, this approach is one to consider in the future. But for our research, we gained access to a set of midi recordings that presented each of the individual instruments as separate layers of the recording. A second problem arises from the actual number of layers that are included in a musical recording. Most recordings do not have exactly eight tracks or instruments to present on each speaker, which would require a different distribution model to accommodate for the different number of tracks. In these cases, we would group instruments according to their frequency ranges and send those signals to the appropriate speaker, as in the FM approach. Since the FM already distributes music based on frequency ranges, it may be the more practical approach that be used to incorporate most existing recordings into the MHC.

3.4 MHC Prototype

The early phases of this research required a flexible prototype that could support the testing of different speaker sizes, locations, and the different methods of distributing signals among the speakers. In the prototype, each pair of speakers is attached to nylon straps, which can be secured to the body using Velcro fasteners. The straps are adjustable to support users with different body sizes and shapes. The current model also allows us to explore the placement of speakers on different locations on the body. Audio speakers can easily be changed and reconfigured to support many different experimental conditions. A photograph showing one of the researchers wearing the the MHC in its current form is presented in Figure 8. Speakers are arranged in pairs, which

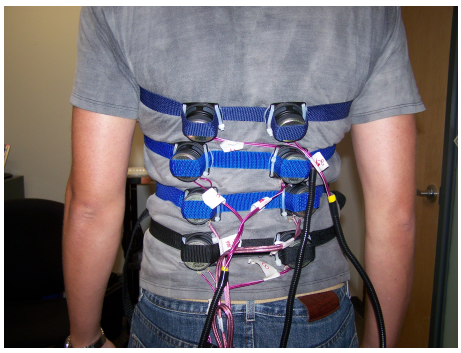

Figure 8: One of the project researchers wearing the MHC in its current form. During experiments, participants are seated in a chair to simulate the intended form factor of the system.

run in parallel rows approximately two apart along the spinal

cord. Each pair of speakers is powered by a two channel 75 watt car audio amplifier. We use four amplifiers to run the eight speakers. Amplifiers are powered by one 12 V computer power supply. Audio signals are generated through a Firepod 10/10 firewire digital audio interface which is capable of operating up to eight channels of full duplex audio recording and playback. We are using the Windows XP operating system, and Adobe Audition 2.0 to control the audio output levels and timing. Audition provided us with a ready-made solution to distributing an audio recording to the different outputs used in the MHC. Individual tracks can be assigned to each of the audio outputs, which are then amplified and directed to their corresponding speakers.

3.5 Vibetracks

Each of the different sensory substitution models of the MHC produce tactile stimuli that we refer to as *vibetracks*. We wanted to test the different models for their effectiveness in presenting some of the emotional content that is expressed through music using vibrotactile stimuli. To do this, we designed an experimental methodology to conduct a series of studies on the MHC.

4. EXPERIMENTAL METHOD

The methodology incorporates qualitative and quantitative methods: quantitative measures are based on the ratings users assign to the vibetracks during the experimental trials, and qualitative data is obtained from the interviews and questionnaires presented to participants. Qualitative methods are gathered before and after the experiment, providing information about the user's perspectives and opinions on the different models they experience, as well as demographics of the participants. In our methodology, qualitative data offers valuable information that can lead to the formation of new hypotheses. Once these hypotheses are formed, we use quantitative methods to test the hypotheses, which can be used to further improve the MHC. Additional qualitative results offer further insight into our models, that often leads to the formation of additional hypotheses, which can be explored in future experiments. This cyclical approach to designing and evaluating the MHC will allow us to experiment with a variety of configurations, and assist in making design decisions for the final version of the MHC. The experimental method is outlined in the following protocol:

1. Pre-study questionnaire: aimed at gathering demographic information about participants.

2. Calibration session: each speaker is set at a decibel level that is most comfortable and detectable by the user. This information is recorded for each participant, and will be analyzed for each study.

3. Experimental trials: participants experience the vibetracks, and rate each on emotion scales designed to reflect the intensity of the emotion, as well as the type of emotion conveyed through the vibetrack.

4. Validation trials: participants also rate the original audio using the same scale used to rate the vibetracks, providing additional internal validation to the results.

5. Post-study questionnaire: One source of qualitative data, intended to gather opinions about the vibetracks, and the general experience of the user.

6. Free-form interview and debriefing: Qualitative approach aimed at obtaining further information about the user's experience and comments on the MHC.

In our first study [11], we tested the MHC using eight speakers. Speakers were placed along the back of participants in the localized configuration. We used two music samples: one was previously rated for its ability to express emotional features associated with happy, while another expressed sad features. Samples of each of the tracks are included in [1]. Results from this study suggested that emotional information could potentially be communicated through both the FM and the TM, however, the TM was shown to be most highly associated with the corresponding emotion. We also included a control condition (CM), which presented the complete audio segments on each of two speakers (S5 and S6). The CM was shown to be least effective in suggesting what the emotional information conveyed by the vibetracks was. A graph depicting the emotional ratings that users assigned to each model is presented in Figure 9. Participants all said that they felt that the MHC could effectively communicate some emotional elements through vibrations. With this pro-

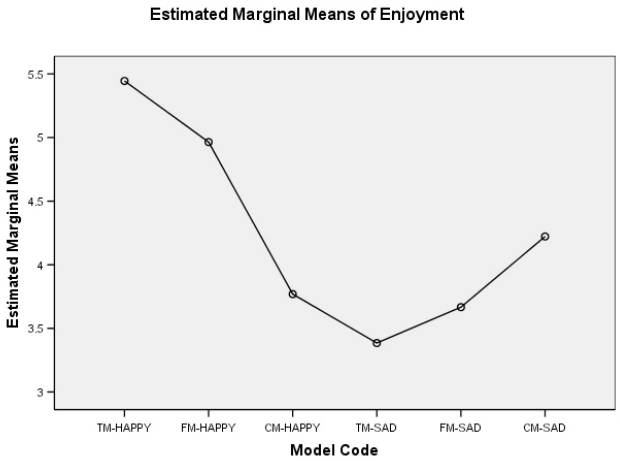

Figure 9: The graph presents the results of the emotional ratings that participants gave each model of the MHC. The most effective model at communicating happy emotions was the TM, followed by the FM. For sad, the TM was most effective, with the CM receiving the least effective ratings.

tocol in place, we are conducting further studies that are exploring different configurations of the MHC, as well as a larger set of vibetracks associated with a larger set of emotions. Independent variables that we are exploring in our experiments can be organized into two main categories:

- Speaker configurations: Speaker size, number of speakers, speaker output, layout, and signal

- Audio stimuli: Stimuli type (music, speech prosody or sound effects), stimuli characteristics, including associated emotional content, mood, and intended affects.

5. NEXT STEPS

There are several directions that this research will explore in the next few months. Currently, we are conducting a set of experiments to assess additional vibetracks and the MHC using five basic emotions, as described by Ekman [7]. For each of the vibetracks, we present a series of scales that participants will use to rate them. Vibetracks will be rated according to the emotional content that participants feel are being communicated through the vibrations, using a scale based on Russel's circumplex model of affect, which measures emotion in terms of valence and arousal [19].

5.1 Psychophysics and the MHC

A second experiment is being conducted to explore psychophysical characteristics of the MHC. Contrary to existing psychophysical research, which states that cutaneous receptors can detect a maximum of 1000Hz from vibrations, in the experiment we previously conducted, some of the frequencies in the highest bands were detectable by participants, which needs to be further explored. Since most of the existing tactile research has been conducted on single-point stimuli, primarily on the fingertips, we wanted to conduct a series of experiments that explores audio speakers as vibrotactile devices. Since speakers make contact with a much larger area of stimuli than is typically considered in the literature, we are exploring these devices to better understand the absolute and relative frequency detection that users have for the vibrotactile stimuli created by the speakers.

5.2 Speaker Size, Number and Placement

The flexibility of the MHC prototype supports the evaluation of different speakers, speaker configurations, and stimuli that is being considered for this research. In our initial prototype, we explored speakers that were placed in two rows running approximately two inches apart along the spine. Eight speakers were used in this experiment, which was shown to have some positive results in expressing emotional content through vibrations. However, there are many other experiments that can be conducted to further explore the relationship between the speakers and the signals. For example, we want to determine if there is an optimal speaker size to distance ratio. We also aim to explore the relationship between speaker size and frequency range in order to ensure that we are using an optimal mapping for signal to speaker in the MHC. These results can be generalized towards gaining a greater understanding about vibrotactile displays and the information they present.

5.3 Speech Prosody

Speech is another area of film audio that we are exploring with the MHC. Prosody refers to the rhythm, intonation, and related attributes in speech, and is one of the elements of sound that we are trying to communicate through a tactile display. Speech prosody is an important element of film, expressing many complex concepts that may be lost through captioning alone. For example, sarcasm, irony, as well as mood and emotion are often expressed through speech prosody and are not always represented in the dialog. Since speech prosody typically occurs as part of the foreground of a film, we are directing this output to the more sensitive areas of the body, namely the forearm and hands. We will employ the experimental methodology to perform these tests, in addition to evaluating the MHC for communicating additional emotional content associated with music.

5.4 Vibrotactile Feedback

Finally, we are looking towards the incorporation of feedback mechanisms such as galvanic skin response and heart rate monitors, into the EmotiChair to support physiological research into the effects of the MHC and the different models used for sensory substitution. Because of the subjective nature of emotional responses to music, we anticipate that these feedback mechanisms will further serve as validation for the MHC and the potential emotional information that can be communicated through the different models used to present music as vibrations.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we presented and discussed some of the design issues encountered in developing a model human cochlea (MHC). The model is intended to be applied as a sensory substitution technique for translating audio signals into tactile sensations, with the aim of communicating emotional information from music. The MHC prototype is in its early development stages, and currently consists of eight audio speakers that are attached to adjustable straps. These straps can then be secured to the body to create our haptic ambient music display. The MHC adopts the cochlea as a metaphor for designing the sensory substitution technique that will be used to translate audio into tactile sensations. Two different approaches were considered for sensory substitution and evaluated in a formative study: the track model, which divided up the audio recording based on the individual instruments, and the frequency model, which distributes the recording to the speakers based on the frequency distribution of the occurrence of notes played on western harmonic music.

We tested these two models and results from our formative study suggest that both the track and the frequency models are effective in communicating basic emotional information through vibetracks. These results have motivated our next study, which will explore additional emotional content associated with music using the MHC. In these early stages of the research, we are focusing on the communication of emotional content from music, with future studies aimed at expanding our current understanding of emotion and vibrotactile stimuli. The prototype also serves as a tool for testing other characteristics of crossmodal displays, such as the communication of speech prosody through vibrotactile stimuli, and for determining psychophysical characteristics of the crossmodal ambient audio-haptic display.

7. ACKNOWLEDGMENTS

Funding for this project and study was generously provide by the Canadian Natural Sciences and Engineering Council and the Canada Council for the Arts. We would like to thank all those who participated in our study and Carmen Branje and Emily Price who assisted with this research.

8. REFERENCES

[1] Emotichair: Emotional music excerpts. http://ryerson.ca/ m2karam/MusicTracks.html, 2007.

[2] R. Blake and R. Sekuler. *Percption.* McGraw-Hill Higher Education, New York, NY, 2005.

[3] S. Brewster and L. M. Brown. Tactons: structured tactile messages for non-visual information display. In *AUIC '04: Proceedings of the fifth conference on Australasian user interface*, pages 15–23, Darlinghurst, Australia, Australia, 2004. Australian Computer Society, Inc.

[4] G. C. Burdea. *Force and touch feedback for virtual reality.* John Wiley & Sons, Inc., New York, NY, USA, 1996.

[5] M. Chion. *Audio-Vision: Sound on Screen.* Columbia University Press, 1994.

[6] S. DiPaola and A. Arya. Emotional remapping of music to facial animation. In *sandbox '06: Proceedings of the 2006 ACM SIGGRAPH symposium on Videogames*, pages 143–149, New York, NY, USA, 2006. ACM Press.

[7] P. Ekman. *The Handbook of Cognition and Emotion*, chapter Basic Emotions, pages 45–60. John Wiley & Sons, Ltd, Sussex, U.K., 1999.

[8] 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.

[9] K. Kahol and S. Panchanathan. Distal object perception through haptic user interfaces for individuals who are blind. *SIGACCESS Access. Comput.*, (84):30–33, 2006.

[10] 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.

[11] M. Karam, C. Branje, E. Price, F. Russo, and D. I. Fels. Towards a model human cochlea: Exploring vibetracks for crosssmodal audio-tactile displays. In *Submitted: GI '08: 34th Graphics Interface conference*, 2007.

[12] D. G. LEE, D. I. FELS, and J. P. UDO. Emotive captioning. *Comput. Entertain.*, 5(2):11, 2007.

[13] V. Lévesque, J. Pasquero, V. Hayward, and M. Legault. Display of virtual braille dots by lateral skin deformation: feasibility study. *ACM Trans. Appl. Percept.*, 2(2):132–149, 2005.

[14] S. R. Livingstone and A. R. Brown. Dynamic response: real-time adaptation for emotion. In *HAPTICS '03: Proceedings of the 11th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (HAPTICS'03)*, page 57, Washington, DC, USA, 2003. IEEE Computer Society.

[15] J. R. Marston, J. M. Loomis, R. L. Klatzky, R. G. Golledge, and E. L. Smith. Evaluation of spatial displays for navigation without sight. *ACM Trans. Appl. Percept.*, 3(2):110–124, l 06.

[16] T. L. McDaniel and S. Panchanathan. A visio-haptic wearable system for assisting individuals who are blind. *SIGACCESS Access. Comput.*, (86):12–15, 2006.

[17] J. B. Mitroo, N. Herman, and N. I. Badler. Movies from music: Visualizing musical compositions. In *SIGGRAPH '79: Proceedings of the 6th annual conference on Computer graphics and interactive techniques*, pages 218–225, New York, NY, USA, 1979. ACM Press.

[18] Pyramat. Pyramat: Sound rocker. Web Resource, 2007.

[19] J. A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, (39):1161–1178, 1980.

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

[21] H. Shin, M. Sohn, and J. Park. A design of cell-based pin-array tactile display. In *ICAT '05: Proceedings of the 2005 international conference on Augmented tele-existence*, pages 251–252, New York, NY, USA, 2005. ACM Press.

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

[23] Tadoma. The tadoma method. Web Resource, 2007.

[24] H. Z. Tan and A. Pentland. Tactual displays for sensory substitution and wearable computers. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Courses*, page 105, New York, NY, USA, 2005. ACM Press.

[25] J. B. F. Van Erp, H. A. H. C. Van Veen, C. Jansen, and T. Dobbins. Waypoint navigation with a vibrotactile waist belt. *ACM Trans. Appl. Percept.*, 2(2):106–117, 2005.