Designing the Model Human Cochlea: An Ambient Crossmodal Audio-Tactile Display

Maria Karam
Ryerson University

Frank Russo
Ryerson University

Deborah I. Fels
Ryerson University

Please Cite:
Karam, M., Russo, F., & Fels, D. I. (2009). Designing the model human cochlea: An ambient crossmodal audio-tactile display. *IEEE Transactions on Haptics, 2*(3), 160-169.
doi:10.1109/TOH.2009.32

Designing the Model Human Cochlea: An Ambient Crossmodal Audio-Tactile Display

Maria Karam, Frank A. Russo, and Deborah I. Fels

Abstract—We present a Model Human Cochlea (MHC), a sensory substitution technique and system that translates auditory information into vibrotactile stimuli using an ambient, tactile display. The model is used in the current study to translate music into discrete vibration signals displayed along the back of the body using a chair form factor. Voice coils facilitate the direct translation of auditory information onto the multiple discrete vibrotactile channels, which increases the potential to identify sections of the music that would otherwise be masked by the combined signal. One of the central goals of this work has been to improve accessibility to the emotional information expressed in music for users who are Deaf or hard of hearing. To this end, we present our prototype of the MHC, two models of sensory substitution to support the translation of existing and new music, and some of the design challenges encountered throughout the development process. Results of a series of experiments conducted to assess the effectiveness of the MHC are discussed, followed by an overview of future directions for this research.

Index Terms—Human factors, user interfaces, models and principles, music, sensory aids.

1 Introduction

Crossmodal displays support the presentation of sensory information using an alternate sensory modality. Such alternative displays offer a novel approach to experiencing audio, visual, and tactile stimuli through translations, interpretations, or other mappings that support feeling music, hearing images, or seeing sound. They can be used to *augment* one modality with *extraneous* information based in a second modality as an approach to increasing the realism of a primary sensory experience. They can also be used to *replace* information intended for one sensory modality using a different modality, e.g., speech audio can be replaced by closed captioning, which serves as an alternative modality for making dialogue more accessible to Deaf or hard of hearing (HoH) people.

Crossmodal displays have effectively been used in information critical applications such as the Optacon, Tactile System Awareness System (TSAS) [6], the Tactile Vision Substitution System (TVSS) [29], or tactile braille displays [15], offering an accurate representation of text or visual information in a tactile form. Alternatively, crossmodal displays can also provide noncritical, secondary task information in ambient entertainment displays such as music visualizations [19] to augment the music experience with the accompanying graphical representation.

- *M. Karam is with the Centre for Learning Technologies, Ryerson University, 7 Redwood Ave, Toronto, ON M4L 2S5, Canada. E-mail: maria.karam@ryerson.ca.*
- *F.A. Russo is with the SMART Lab, Department of Psychology, Ryerson University, 113 Waverley Road, Toronto, ON M4L 3T4, Canada. E-mail: russo@ryerson.ca.*
- *D.I. Fels is with the Centre for Learning Technologies, Ryerson University, 164 Arlington Ave, Toronto, ON MCC 2Z2, Canada. E-mail: dfels@ryerson.ca.*

Manuscript received 15 Nov. 2008; revised 27 May 2009; accepted 29 May 2009; published online 13 July 2009.
Recommended for acceptance by K. Kahol, V. Hayward, and S. Brewster.
For information on obtaining reprints of this article, please send e-mail to: toh@computer.org, and reference IEEECS Log Number THSI-2008-11-0090.
Digital Object Identifier no. 10.1109/ToH.2009.32.

Music is a source of input for sensory substitution that spans a range of applications including information critical representations of theoretical elements of music as visualizations [5], to more ambient interpretations of music as vibrations [11]. Entertainment applications are primarily used to enhance the experience of music using an alternative modality to sound in noncritical, ambient crossmodal displays. These displays are often implemented as entertainment chairs, which provide an effective form factor for experiencing tactile stimulation that accompanies movies or video games.

For film media, closed captioning effectively provides access to sound information originating from dialogue; however, the emotional information that is expressed through the music and soundtrack of a film cannot be communicated through text alone, leaving Deaf and HoH viewers with only an indication that music is playing. Since music is commonly used to enhance moods, invoke emotions, enforce the plot, and strengthen the multimodal nature of film [26], those who cannot hear may miss the full extent of the entertainment experience: suspense, drama, excitement, relief, and other emotional effects that music creates. One of the challenges in making music accessible lies in determining an effective and meaningful mapping that can represent the emotional elements of music through an alternative display modality. Since the audio and visual channel is occupied or not available while watching films, the sense of touch represents a practical channel for displaying sensory information intended for audition.

Although audio signals naturally produce vibrations that can be detected through the sense of touch, only a small portion of the vibration can actually be felt through physical contact with the amplified audio signal. Since there exists a natural mapping between the frequency and amplitude measures of audio signals and tactile vibrations, we commonly find sound vibrations being leveraged by Deaf and HoH people seeking to obtain auditory information.

For example, the Tadoma method supports deaf-blind individuals in detecting vibrations and movements from speech by placing the forefingers and thumb on the lips and throat of the speaking person to improve lip reading comprehension [25]. Similarly, music enthusiasts who are Deaf can feel some of the musical vibrations produced by an amplified audio signal by making physical contact with the loudspeaker cabinet while music is playing, often referred to as speaker listening. But while it is possible to feel some of the vibrations that are produced by sound, most of the complex audio signals that we can hear cannot be accessed using devices intended for auditory perception.

To explore this problem, we considered the characteristics of the human cochlea as a design metaphor for creating a sensory substitution approach to translating music into vibrations. Similar to the vocoder [3], the Model Human Cochlea (MHC) is a sensory substitution technique aimed at increasing the audio-tactile resolution associated with the physically detectable waveforms produced by an amplified audio signal. The MHC offers a direct translation of an audio signal using a series of discrete vibrotactile devices or *channels* that can potentially increase the audio-tactile resolution of physical vibrations on the body. The current version of the MHC uses voice coils as vibrotactile devices, which serve as inexpensive yet effective stimulators for displaying music as vibrotactile energy [18]. Voice coils are embedded into the back of a canvas chair and the audio-tactile signal can be altered to support the investigation of different configurations of the MHC. The chair serves as an effective form factor for presenting the ambient, passive information to the body, and is a natural interface for supporting entertainment activities such as listening to music or watching a film without constraining the hands or arms with unnatural devices. In the following sections of this paper, we present the MHC, the underlying theoretical model used in its design, and some of the major hurdles we addressed throughout the development process. Results from formative evaluations are presented along with an explanation of the experimental methodology used to evaluate different configurations of the model. Finally, we present future directions of this research.

2 Background

The alternative sensory information display (ASID) project aims to develop a multimodal entertainment chair that can express emotional content associated with film audio using a series of ambient tactile displays for people who are Deaf or HoH. For hearing users, these sensations will serve to enhance the existing experiences of the film audio, while Deaf and HoH people may be able to receive some of the emotional content that the soundtrack provides through a different modality. We are exploring ways of improving access to music and the emotion it expresses using a direct translation approach. The MHC prototype allows us to explore sensory substitution of music as vibrotactile stimuli along the back, toward making sound information more universally accessible.

Film music can be highly expressive. Rather than interpreting this expression through extant theories of music or emotion, the MHC aims to increase the amount of music that can be detected as vibration, thus allowing the user to explore and interpret the emotion in a bottom-up manner. Although it would have been possible to interpret specific musical characteristics such as pitch, tempo, etc., these do not encapsulate the full experience of music.

The choice to use the sense of touch as the alternative modality for music is based on growing research, which has demonstrated that the human cutaneous system or skin, is capable of detecting a great deal of information and is a relatively untapped resource that can serve as an effective communication channel for computer interactions [7]. Moreover, recent research in perceptual neuroscience suggests the potential for the audio cortex to process vibrotactile information [13]. The skin is the largest organ in the human body, and contains receptors that respond to a variety of sensations. Our work focuses on using vibrotactile stimuli as an effective approach to communicating information about music to the body [10], [28].

The difference in sensitivity to touch found along the body is largely determined by the type of skin on a particular locus: the skin on the most sensitive parts of the body, the palms, fingers, genitalia, soles of the feet, and lips, are *glabrous* or nonhairy skin, while the remaining parts of the body are nonglabrous or *hairy skin*. The glabrous skin is more sensitive to vibrotactile sensations than the hairy skin [23] largely due to the relative concentration of Pacinian corpuscles. Given the ratio of receptors to skin location and size, the sensitivity of the glabrous skin is much more acute than the hairy skin, but the differences in the two skin types also influence the type of tactile display that each is best suited to receive. The glabrous skin obtains information through touch by *actively* interacting with the outside world, moving along surfaces, exploring textures, and gathering details about the physical features of objects it encounters. In contrast, the hairy skin appears to be more specialized for passive interaction with the world [23].

Other displays that present tactile information to the torso or back include the TVSS [29] and the TSAS [6], which show that the hairy skin is an effective locus for receiving tactile information pertaining to vibrations from images. Similarly, the tactile vocoder uses the hairy skin as a locus for presenting vibrations from speech information [3] using 16 vibrotactile channels that correspond to audio frequencies ranging from 200 Hz to 8,000 Hz. Each of the 16 channels represent 1/3 of an octave of sound from this frequency range, and are activated when a sound in the corresponding range is presented.

The MHC draws on a similar approach to the tactile vocoder, and separates music into multiple vibrotactile channels that are presented along the back. By displaying the vibrotactile signal along the back, we can leverage the passive receptive qualities of the hairy skin, while enabling the user to maintain free use of their hands and arms while experiencing the vibrations. This form of ambient entertainment display uses the direct translation of music signals to create the vibrotactile display. However, in using the complete range of audio signals in the display, we must address the differences in the perceptual abilities of the auditory and tactile systems—we can hear frequencies

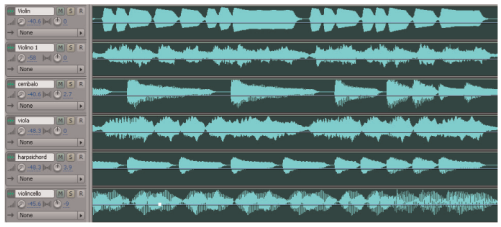

Fig. 1. Waveforms for different instrument tracks that comprise a classical music composition.

within a range of 20 Hz to 20 kHz, but only feel frequencies that range from approximately 5 Hz to 1,000 Hz [7], [17].

Psychophysical research has shown that human cutaneous receptors, specifically the Pacinian corpuscles, are most effective at perceiving single-point vibration stimuli below 1,000 Hz, with optimal detection occurring around 250 Hz [24]. While there is a large area of research concerned with the precise features of tactile perception [16], and although optimal detection and discrimination through touch is essential for information critical displays, the ambient entertainment system we are describing does not rely on the identification of specific objects or information. Rather, we are interested in determining if there are emotional characteristics present in the direct translation of music as spatiotemporal vibrations applied directly to the body.

Given that we are using audio signals to create vibrotactile stimuli, further limitations arise due to the vibrotactile masking that occurs in tactile perception. For example, the speed with which an active braille display may be "read" is limited by the problems of forward and backward masking [8]; that is, spatial patterns occurring in spatial and temporal proximity to one another tend to compete and compromise accuracy of perception. Masking effects have also been demonstrated for recognition of vibrotactile frequency. The range of inter-stimulus intervals (ISIs) over which masking has been observed in vibrotactile frequency identification is comparable to that observed in auditory frequency identification [14]. Moreover, like audition, masking properties are nonuniform, varying with respect to the relative frequency of the competing signals [27]; consider the subtle vibrations of the violin or flute, which are easily overpowered by the low frequency signals of the cello or bass.

As a metaphor, we illustrate the vibrotactile masking problem using Fig. 1, which shows discrete audio waveforms that make up the different tracks of a classical music composition. Once we combine the multitrack signal into a single audio track, as illustrated in Fig. 2, we see that the individual waveforms are no longer discernible. The shape of the composite waveform reflects the rhythmic structure and expressive intensity peaks of the composition rather than the melodic line that each instrumental track has contributed. The MHC approach aims to reduce the amount of auditory-tactile masking that occurs when a large range of vibrotactile stimuli are present in the same space at the same time.

Previous work by Gunther et al. demonstrates an approach to making music more accessible to the sense of touch in a wearable tactile display using audio transducers to create spatiotemporal patterns of vibrations along the body [10]. Individual instruments are presented to different

Fig. 2. Stereo waveforms of the mixed version of the classical recording from Fig. 1.

vibrotactile channels in the suit intended to augment the music listening experience. In Gunther et al.'s work, only sounds that lie within the tactile range of frequencies are presented to the body in novel compositions created specifically for this application. However, in our research, we aim to facilitate the interpretation of any musical composition, recorded or live, by using the MHC to reduce vibrotactile masking, and by investigating the effects of increasing the number of discrete vibrotactile channels, toward making the emotional content of music more accessible to all users, including the Deaf and HoH.

3 Designing the MHC

One of the key research challenges in creating effective crossmodal displays lies in determining which of the characteristics or features of the original modality may be appropriately and accurately transferred to the alternative display. This exchange of characteristics between modalities is called sensory substitution, and requires the identification, selection, and mapping of relevant characteristics from the original modality onto the alternative display modality.

There are a variety of perspectives from which we can select characteristics for sensory substitution, including physical, perceptual, or computational approaches. Once selected, characteristics of the original modality can be mapped onto appropriately selected characteristics of the alternative modality using an interpretation, translation, or other transformation technique. In this work, we refer to translations as mappings that draw on the natural relationships that exist between the selected characteristics of the two modalities, while interpretations use indirect mappings such as representing musical chords and tones as shapes and colors [19].

3.1 Human Cochlea Design Metaphor

The auditory cochlea is the main organ responsible for detecting and processing audio signals, which we use as the design metaphor for the MHC. The cochlea essentially detects individual frequency signals that activate the appropriate microscopic hair cells that line the spiral structure of the cochlea within the inner ear [1]. Rows of hair cells are positioned along the basilar membrane in a precise ordering, which places those hairs that detect highest frequencies toward the base, and the lowest frequencies toward the apex, as illustrated in Fig. 3.

Frequency signals activate specific hairs that generate electrical potentials transmitted along the auditory nerve, eventually reaching the auditory cortex. For our design metaphor, we draw on the tonotopic organization of hair cells located along the basilar membrane of the inner ear. The hair cells are organized with respect to their frequency selectivity, sometimes referred to as a tonotopic organization.

In addition to the frequency selectivity of different hair cells, it is important to note that the hair cells are phase

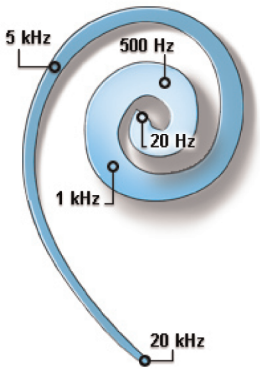

Fig. 3. Arrangement of hair cells in the tonotopic ordering along the basilar membrane of the human cochlea. (From http://www.neuroreille.com.)

locked to the stimulating frequency. Thus, our biological metaphor involves a place coding as well as a time coding of the input frequency. We attempt to simulate this functionality by placing voice coils in a linear spatial order positioned from lowest to highest frequency bands.

In our current prototype, this is implemented as a canvas chair with voice coils arranged in a 2×8 array along the back. This linear spatial order is maintained by placing lower frequency stimulators on the lower back, and higher signals to the upper back of the chair, leveraging the natural spatial orientation of the body and common conceptions of pitch height [20].

Another issue we address in adopting this metaphor is the different frequency discrimination ability of the skin and the cochlea. Thus, it is necessary to perform a mapping of frequency signals onto a smaller set of receptors that present vibrations to the skin while maintaining the place code adopted from the cochlea metaphor. To do this, we group discrete sections or *bands* of frequencies and display these on the individual rows of the vibrotactile channels used in the tactile display.

4 MHC Prototype

The early stages of this research required a flexible prototype to support our initial investigation into the cochlea metaphor used in the tactile display. We implemented a system that enabled us to attach different numbers and sizes of voice coils to the body using nylon straps (see Fig. 4). We ran several exploratory studies to determine an effective configuration of voice coils, and explored different placements around the body, as well as different distributions of signal to voice coil. We began testing a *distributed* model of the MHC, which presented audio signals to different parts of the body using eight voice coils housed in two inch paper cone speakers: two on the upper arms, four on the upper and lower back, and two on the back of the upper thigh. Participants said that the different sensitivity levels of the body was distracting, demanding more attention to make sense of the signals. A second *localized* model was then developed, placing eight 2-inch paper cone speakers along participants' backs, which proved to be the more favorable configuration. The localized model presented a more consistent set of vibrations in a more constrained area along the back, enabling participants to perceive the multiple signals independently, and as an organized whole. These early studies informed the design of our second prototype, which used a 2×4 configuration of voice coils embedded into the back of a canvas chair, shown in Fig. 5.

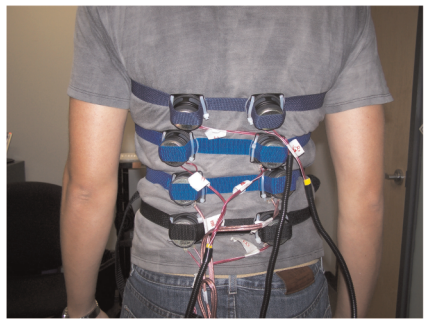

Fig. 4. A project researcher wears the first MHC prototype before being seated in an early experiment.

The hardware we use to run the prototype consists of four 2-channel 75 watt amplifiers and a 12 V power supply. The audio signals are generated through a Firepod 10/10 firewire digital audio interface. We developed a software application using Cycling 74's MAX/MSP environment that runs on both the Mac OS X and Windows XP operating systems. The software is highly configurable and allows control over the frequency range and amplitude displayed in each vibrotactile channel, as well as the number of channels that can be used in the interface. We use the term *vibetracks* to describe the sets of signals that are produced as vibrations through the MHC.

4.1 Sensory Substitution Models

We are investigating two models for distributing audio signals to voice coils using the MHC: the frequency model (FM), which separates audio signals into multiple bands of frequencies, and the track model (TM), which uses a multitrack master recording. Both models are mapped onto

Fig. 5. An early MHC prototype.

TABLE 1
Voice Coils and their Assigned Frequency Bands of the MHC Used in this Experiment

Display	Min Hz	Max Hz	Mid Hz	Piano Keys
S1 - S2	661	4186	2423	F5 - C8
S3 - S4	312	660	486	E4 - E5
S5 - S6	146	311	228.5	D3 - D4#
S7 - S8	27.5	154	86.25	A0 - C3#

the low-high configuration of voice coils. Bands of frequencies are assigned to pairs of voice coils presented in the rows. Each of the coils can represent one of the octaves of a piano keyboard, spanning 27.5 Hz to 4,186 Hz. We assume that most of the signals above 1,000 Hz will not be detected; however, these are preserved in the highest frequency channel to maintain the original signal in our display. This design decision is motivated by our bias toward preservation of the original signal, exploiting the capacity for direct perception of the underlying stimulus organization [9].

Each of the eight octaves within the piano range of frequencies are distributed to the respective rows of voice coils used in the MHC, corresponding to the number of channels in the display. In previous studies, we used four rows of two channels that run parallel along either side of the back of the chair. Musical notes are divided into octaves, and mapped onto each voice coil channel in a distribution based on the frequency of occurrence of notes (DFN) common in Western harmonic music. Following a normal distribution model, notes commonly found in classical music typically center around the middle of the keyboard, at approximately the D# above middle C [22]. This supports a more equal distribution of notes to vibrotactile channels since frequency bands are assigned according to their expected frequency of occurrence [11], as shown in Table 1.

For the TM, each vibrotactile channel presents a discrete track from a multitrack master recording. When there are more instruments than channels, we use an approximate distribution of frequency bands based on the FM to assign signals to channels [11]. The advantage to the TM lies in the natural mapping of sounds from each instrument to the different vibrotactile channels, so that each instrument we hear can be felt as a unique set of vibrotactile signals. Intuitively, the TM can potentially express vibrations that more closely resemble the original music. It is not always possible to implement this model since most music recordings are not accessible in a multitrack format.

5 Experimental Method

To evaluate the many potential configurations of the MHC, we developed an experimental methodology that combines qualitative and quantitative approaches to explore the effectiveness of the MHC in delivering emotional information from music in the vibrotactile display. Qualitative measures include interviews, pre- and postexperiment questionnaires, and observations of participants during the trials. Quantitative measures include ratings given to vibetracks for emotional content and enjoyment. In future studies, we will include biometric measures such as heart rate, respiration, facial electromyography, and galvanic skin response to validate the results on a physiological level. The experimental method we have adopted is described below:

1. Prestudy questionnaire: To gather demographic information about participants.
2. Calibration session: Voice coils are set at decibel levels that can be felt by participants. We send a signal at the midpoint of the frequency band, and adjust the levels to suit participants.
3. Experimental trials: Participants experience a vibetrack, then rate it using one of several different emotional scales.
4. Baseline trials: The audio version of the vibetrack is presented to participants, who rate this using the same scales used in vibetrack trials.
5. Poststudy questionnaire: More qualitative data about user perspectives and opinions on the system, the vibetracks, and the experiment.
6. Interviews and debriefing: To discuss additional observations or comments about the experience with participants for additional qualitative data.

Results from the vibetrack trials are then compared to results from the audio trials to determine if there is a similarity between what participants detect as emotional content for each of the three models. Qualitative data and observations further contribute to the potential forming of new hypotheses, which can be evaluated in subsequent studies.

5.1 Pilot Study

We conducted an experiment using the first prototype (see Fig. 4) to compare the effectiveness of the two sensory substitution techniques (TM and FM) for communicating emotional information from the music as vibetracks [11]. We also created a control model (CM) that presents the entire signal through one pair of voice coils. The CM provides a temporal code (periodicity of the fundamental frequency is available) without the potential benefit of the place code (unique spatial position for each band of frequencies), which is conferred by the MHC and the cochlea itself.

Results from this study suggested that the TM was significantly more effective at communicating emotions from classical music, rated as expressing happy or sad content. While the FM was found to be less expressive than the TM, both the TM and the FM vibetracks were rated more accurately for their respective happy and sad emotions than the CM. The TM and FM also received higher ratings on preference and enjoyment in this study [11]. We note several limitations in this experiment, mainly our use of only two emotions and the prototype itself. To address these limitations, we developed a second prototype (see Fig. 5) using a 2×4 array of voice coils embedded into the back of a canvas chair, with new music samples to be used in our next set of experiments.

6 Evaluating the MHC

Our second experiment used the new prototype (see Fig. 5) and compared the FM to the CM to evaluate the effectiveness of the MHC in expressing emotion through vibrations. The experiment was a randomized $2 \times 4 \times 2$ factorial

TABLE 2
Classical Music Recordings and Corresponding Ratings for the Two Scales Used in this Experiment

ID	Track Title	Arousal	Valence	Emotional Condition
S1	J. Brahms. Violin Concerto, Adagio	Weak	Negative	Sad
S2	D. Scarlatti. Sonata A for Harpsichord, K208	Weak	Negative	Sad
F1	D. Shostakovitch. Symphony 15, Adagio	Strong	Negative	Fear
F2	R. Wagner. Tristan, Act 3	Strong	Negative	Fear
J1	F. Liszt. Les Prludes	Strong	Positive	Joy
J2	L. Beethoven. Symphony 7, Vivace	Strong	Positive	Joy
A1	F. Liszt. Tasso Lamento & Triomfo	Strong	Negative	Anger
A2	R. Strauss. Tod and Verklrung, 7-730	Strong	Negative	Anger

design, investigating two different display models (FM and CM), four emotions (joy (J), sadness (S), fear (F), and anger (A)) and two different music samples for each emotion (T1 and T2). The CM enabled us to determine if the separation of audio signals to different voice coils would be more expressive than conventional audio-tactile displays that do not present the signal using multiple output channels. We also ran an audio condition as a baseline for comparing responses for vibetracks with those of audio version of each of the tracks.

Participants were undergraduate students from Ryerson University. A total of 21 (six female) participants completed the experiment, ranging in age from 18 to 34 years from a variety of disciplines. We selected eight musical excerpts, used previously in research by Bigand et al. [2]. In that study, judgments of emotional similarity of music excerpts were subjected to multidimensional scaling (MDS). The MDS solution yielded two dimensions, which were interpreted as arousal (weak to strong) and valence (positive to negative), consistent with Russel's circumplex model of emotion [21]. The eight musical tracks selected for our current study were chosen on the basis of their positioning in the MDS solution provided by Bigand et al. [2].

Our joy selections were high along the arousal and valence dimensions. Sad selections were low along the arousal and valence dimensions. Fearful and angry selections both were high along the arousal dimension and low along the valence dimension, with the fearful selections being lower on the valence and arousal dimensions. Excerpts selected for the current study varied in length from 24 to 46 seconds. A pretest involving members from the respective labs verified that the selections conveyed the intended emotions. A complete list of the selections used and a brief description of the relevant features (as outlined in [2]) can be found in Table 2. For the CM, excerpts were transmitted uniformly across all eight voice coils. For the FM, excerpts were split into four frequency bands, shown in Table 1, with each band presented on a different pair of voice coils. Lower frequency bands were presented toward the bottom of the back of the chair and higher frequency bands were presented toward the top.

6.1 Procedure

Each participant was "deafened" using a pair of earplugs and a set of noise canceling headphones that played white noise. We next ran several practice vibetracks to ensure that the level of white noise was sufficient to block out any external sounds, and to familiarize participants with the vibrations. Practice vibetracks were also based on excerpts from the list provided by [2]; however, different tracks than those used in the experiment were presented in the practice trials. We used five blocks of experimental trials for each participant, covering the three presentation conditions: FM, CM, and Auditory Model (AM). The AM was run last to prevent familiarity with the pieces; FM and CM conditions were each run twice, in randomized order. Individual excerpts were randomized within blocks. Participants rated each stimulus using the dimensional and discrete scales of emotion described above. The experiment concluded with a poststudy questionnaire, a debriefing, and an open ended interview with participants.

6.2 Variables

Quantitative measures were obtained through participant ratings of each vibetrack using a Likert scale for valence (type of emotion) as negative (1) to positive (7), and arousal (intensity of emotion) as weak (1) to strong (7). In addition, participants rated their enjoyment of the tracks as low (1) to high (7), and provided comments about any additional emotions they felt were expressed by the vibetracks and the audio tracks.

7 Results

We first ran a multivariate analysis of variance (MANOVA) to investigate the differences in ratings between the two models (FM and CM) and determine if participants could discriminate different emotions in the FM more clearly than in the CM. Independent variables used in the MANOVA were the two vibetrack models (FM, CM), four track emotion conditions ($\mathrm{joy} = \mathrm{J},\ \mathrm{sadness} = \mathrm{S},\ \mathrm{anger} = \mathrm{A},$ and $\mathrm{fear} = \mathrm{F}$), and two sample audio tracks for each emotion condition ($\mathrm{track}\ 1 = \mathrm{T1}$ and $\mathrm{track}\ 2 = \mathrm{T2}$).

Dependent variables were the participants' responses to valence, arousal, and enjoyment for each experimental trial. A graph of the mean values for valence, arousal, and enjoyment is shown in Fig. 6. We include mean responses for the AM in Fig. 6 to provide a baseline for comparing vibetrack responses to the audio versions of the excerpts. There were no significant interaction effects found between the independent variables for any of the dependent variables in this analysis.

Results do show a main effect of emotion on valence ($\mathrm{F}_{(3,592)} = 7.236,\ \mathrm{p} = .000$), arousal ($\mathrm{F}_{(3,592)} = 3.851,\ \mathrm{p} = .01$), and enjoyment responses ($\mathrm{F}_{(3,592)} = 5.743,\ \mathrm{p} = .001$), suggesting that characteristics of each emotional condition could

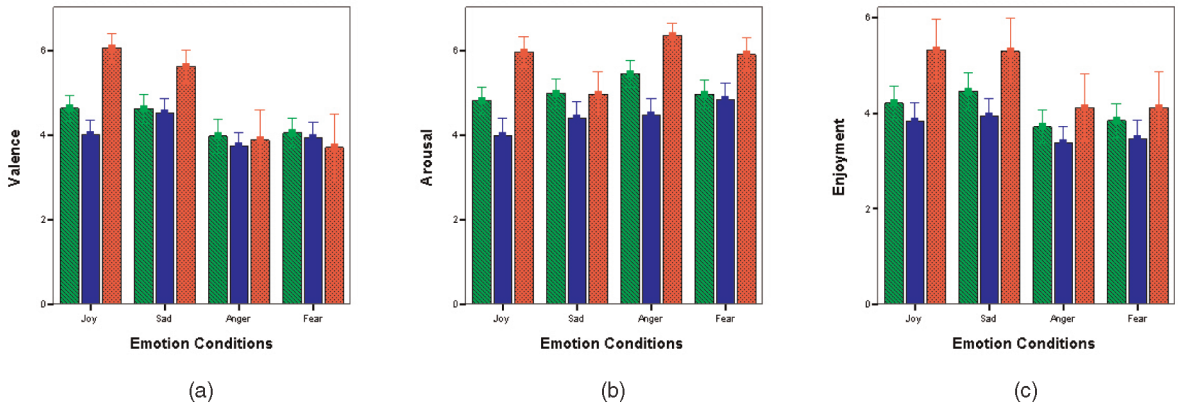

Fig. 6. Mean values for the three models with respect to emotion conditions. *Error bars show 95.0 CI of the mean. (a) Valence, (b) arousal, and

be detected across the different vibetrack models. Main effects were also shown for the two vibetrack models on valence ($F_{(1,592)} = 4.908$, $p = .027$), arousal ($F_{(1,592)} = 23.363$, $p = .000$), and enjoyment ($F_{(1,592)} = 9.610$, $p = .002$), suggesting that each model expresses different levels of information.

7.1 Post Hoc Analysis

We next conducted a post hoc analysis using the least significant difference test (LSD) to further explore the significant results. Results show that valence responses for the joy ($\text{mean} = 4.65$) and sad ($\text{mean} = 4.77$) emotion conditions do not differ significantly from one another, which is not an expected finding. Results also show that valence responses for anger ($\text{mean} = 3.86$) and fear ($\text{mean} = 3.94$) emotion conditions do not differ significantly, which is an expected finding. We find that valence for the sad emotion condition ($\text{mean} = 4.77$) is significantly higher than the anger ($\text{mean} = 3.86$) and fear ($\text{mean} = 3.94$) emotion conditions. Joy ($\text{mean} = 4.65$) was shown to be significantly higher on valence than anger ($\text{mean} = 3.86$). This is in line with our expectations, which places the joy condition higher on the valence scale than both anger and fear.

For the arousal responses, joy ($\text{mean} = 4.70$) was found to be significantly lower than anger ($\text{mean} = 5.23$) and fear ($\text{mean} = 5.10$). The sad excerpts ($\text{mean} = 4.74$) were not significantly different from any of the other emotional conditions, which suggest that responses were generally neutral on the arousal scale. For enjoyment, the joy track responses ($\text{mean} = 4.27$) were significantly higher than anger ($\text{mean} = 3.65$) and fear ($\text{mean} = 3.74$), which follows a similar trend as the AM responses (see Fig. 7c). For the sad tracks ($\text{mean} = 4.40$), we see a higher response for enjoyment overall, which is also similar to the responses for the AM versions (AM-Sad $\text{mean} = 5.29$, AM-Joy $\text{mean} = 5.32$).

Looking at the two models (FM, CM), we find significant differences on ratings for each dependent variable. Fig. 7 presents a line graph of the variables for visual comparison. Valence responses for the FM ($\text{mean} = 4.32$) are significantly higher than the CM ($\text{mean} = 4.05$), showing that overall, the FM provided vibrations that were interpreted as being more positive than the CM (see Fig. 7a). Also, the arousal responses for the FM ($\text{mean} = 5.06$) were significantly higher than the CM ($\text{mean} = 4.43$), which could be because of the overall distribution of the vibrations in the FM, which stimulate more areas on the body than the CM does (see Fig. 7b). This could also be a factor influencing

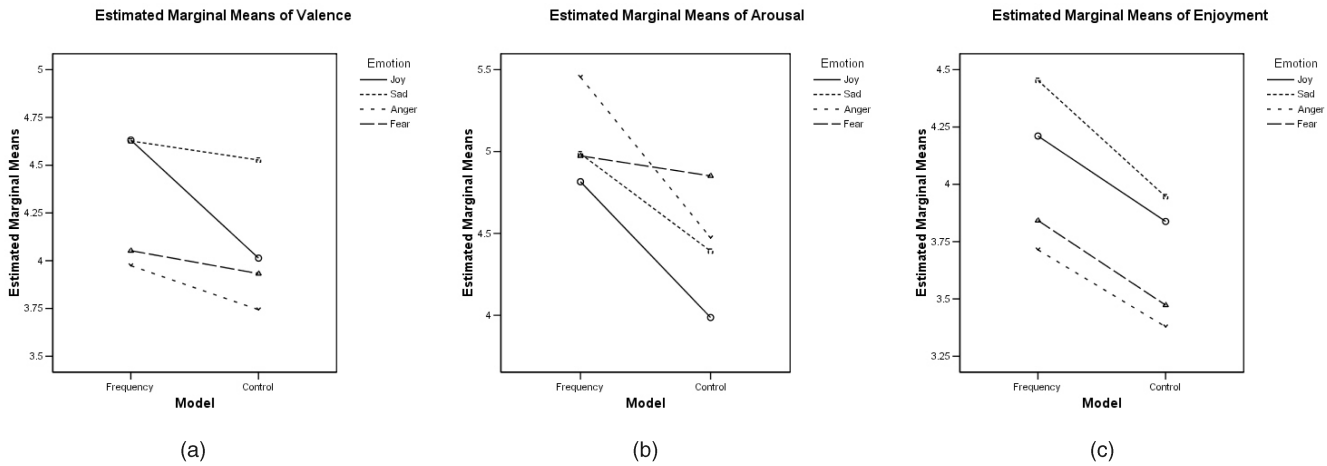

Fig. 7. Line graph representing the responses to the three dependent variables for the two models and the four emotional categories. (a) Valence, (b) arousal, and (c) enjoyment.

the enjoyment responses, which also favored the FM ($\mathrm{mean} = 4.06$) over the CM (mean 3.66). Mean values for enjoyment are presented in Fig. 7c.

These results are further explored by correlational analyses. There was a positive correlation between enjoyment and arousal ($\mathrm{r} = .318,\ \mathrm{p} < 0.01$ level). There was also a positive correlation between enjoyment and valence ($\mathrm{r} = .473,\ \mathrm{p} < 0.01$). Because high-arousal music tends to be characterized by high intensity and fast tempo, while positively valenced music tends to be characterized by variation in pitch, these correlations imply that participants enjoy intense and varied vibrations. Overall, joy ($\mathrm{mean} = 4.27$) received the highest enjoyment responses, while fear ($\mathrm{mean} = 3.74$) received the lowest. Critically, the FM ($\mathrm{mean} = 4.55$) was significantly more enjoyable than the CM ($\mathrm{mean} = 3.66$).

7.2 Qualitative Results

Observations showed that participants tended to group the emotional categories of joy with sadness, and fear with anger in many of the responses, primarily for the valence and arousal variables. This suggests that the arousal or intensity of the music was distinguishable through the vibrations. Interesting results are also found in the comments about the emotions they could detect in both the audio and vibetrack models for each track. For example, one participant made the following comment about track FM-Sad-T1, saying "it felt happy and light. I'm enjoying the tracks with wide range much better than the ones that stay within a tight range."

Another participant said they felt nothing from the CM-Sad-T1, but described the audio version of that track as being wishful and wistful. Further comments about the CM-Sad-T1 vibrations included "I hardly felt anything" and "it was like a bee flying on my back." Other words used to describe the CM-Sad-T1 included weak, serenity, and nothing. These are very interesting results, further suggesting that the FM can invoke a richer sense of emotion than the CM.

The level of imagery that the FM invoked in participants was similar to that which was expressed for the audio versions of the music, which is very exciting, suggesting that the FM may well be an effective way of communicating more information about music to the skin than what is possible without separating the music into multiple channels.

We note that the comments for the FM suggest that more emotional information is being expressed in this model than in the CM, which supports our hypotheses. For example, comments for FM-Sad-T1 included "it felt like someone forcing another person to do something—move along, go that way, etc., the other person was fighting back—stalling, being passive aggressive, etc.," and descriptions of the track as brooding, calmness, and relaxing. FM-Joy-T1 was described as expressing joy, complexity, anticipation, and a summer storm. FM-Joy-T2 was described as excited and carefree. These comments were in contrast to those made for CM-Joy-T1, which ranged from "too low and too weak to feel much of anything," to "epic, energetic, and proud."

Some participants thought the FM was more expressive for the joy tracks, and we found that the overall intensity and variation of the vibrations in the FM did not translate to the CM. Since the CM does not utilize the multiple channels of the MHC, it would not have been able to present as complex and varied a set of vibrations as the FM could, which one could potentially learn to identify, using their body as an effective input modality for a tactile display. This is one hypothesis we will be exploring in future studies.

AM-Anger-T1 was described as being suspenseful, while AM-Anger-T2 was described as "war." Words used to describe FM-Anger-T1 included military, urgency, and impatient, while FM-Fear-T1 was described as suspenseful, fearful, and terrorizing. We note that the general sense of the emotions expressed in the fear and anger tracks do correspond to the terms participants used to describe the vibrations. Although CM-Anger-T1 was described as being uncomfortable, suspense, and violent, CM-Fear-T2 was described as "Boring! didn't say anything." We found, however, that for most of the CM tracks, there were fewer participants who could as clearly identify characteristics of emotion in them as they could in the FM, or the AM.

While we found that participants could more accurately classify elements of the track emotions as an audio signal from the AM than from the tactile signals of the FM or CM, results show that the FM was better at expressing emotion than the CM, suggesting that an increase in the audio-tactile resolution may lead to improved comprehension of musical content through vibrations.

8 Discussion

Results show that the auditory and vibrational modalities do not convey emotion (arousal and valence) and enjoyment in the same manner. However, as expected, the FM responses aligned more closely with the AM responses than the CM responses (see Fig. 6). Thus, it appears that emotional communication in vibetracks may be improved by increasing the number of available channels. The most likely explanation for the relative disadvantage of the CM is the masking that occurs when the audio signal is presented to the body using only one vibrotactile channel.

A general conclusion about the participants' experience with vibetracks is that they seemed to enjoy the sensations of FM vibetracks over the CM versions, showing a preference for the more varied and intense signals of joyful tracks in particular. It is likely that the transient characteristics of notes (attack and decay) influenced the emotional expression of the vibetracks. For example, our anger excerpts contained many notes with rapid attack and rapid decay, which seem to evoke images of violence when presented as vibration. In contrast, our joy excerpts contained many notes with a similarly rapid attack but slow decay. The specific role of transients in conveying emotion through vibration will be the subject of a future study involving the MHC.

8.1 Conclusions and Future Work

One of the main contributions of this research is the MHC prototype, which was evaluated in several studies that support our hypotheses that by increasing the audio-tactile resolution, more of the emotional expression that music provides can be detected through vibration. From the two experiments we discussed in this paper, we have obtained some evidence to suggest that the TM and the FM are more effective than the CM in emotional expression.

Fig. 8. The current MHC prototype uses 16 voice coils in a 2×8 configuration. Voice coils are attached to the back of the chair, with four upholstery pins securing each in place.

In our second experiment, we looked more closely at the difference between the FM and the CM, with results suggesting that the FM could be more effective at communicating information from music than the combined signal presented in the CM. Results from the first experiment suggest that the TM was significantly better than the FM in communicating information, which provides further evidence to suggest that in addition to being able to process more vibrotactile channels, the body can potentially benefit from vibrations that more closely reflect the individual sounds that make up the music.

Our next experiment will provide a closer look at the differences between the FM and the TM for expressing emotional content. One advantage of the TM is that the technology parses the signal with respect to the independent sound sources. While this kind of parsing is not necessary in audition (i.e., we have no trouble understanding a mono recording), it may serve to greatly facilitate perceptual understanding of music presented as vibration. While we did find that emotional responses in FM tended to align with emotional responses in AM, people have considerably less experience interpreting emotion conveyed through vibration. Increasing exposure to vibration should also increase sensitivity.

For example, after limited training with the tactile vocoder (under 100 hours), participants ability to recognize words increased dramatically [4]. Similarly, increasing exposure to vibetracks should increase an individual's ability to discriminate emotion. This is a claim we are planning to investigate in a series of longitudinal studies using the MHC.

We now have a new prototype in the form of a reclining style chair that allows the user to make full contact with all voice coils in the eight-channel MHC, leveraging their body weight to support stronger contact points with the voice coils (see Fig. 8). This version of the MHC has since been used to conduct experiments that explore the use of vibrations to convey specific auditory dimensions such as pitch and timbre. Additional improvements have also been made in our control software, which allows us to easily adjust the frequency bands to suit different excerpts and genres. Future research will also determine whether remapping of higher frequency content (less than 1 kHz) to midfrequencies within the human tactile range can improve overall sensitivity to musical information.

To date, the MHC has been experienced by many different user groups, including Deaf, HoH, and hearing people, who have expressed excitement about the sensory substitution of music as vibration. There are currently many facets of the MHC that have yet to be explored, including the use of a higher resolution configuration, which will present eight separate vibrotactile channels, and the use of biometric measures to assess the emotional responses that people experience when feeling the music on the MHC.

Recently, we conducted experiments at a local center for the Deaf where the chair was experienced by a larger user group [12]. Their reaction to the MHC was very positive, and we have obtained a wealth of comments and suggestions from the Deaf community with regards to different applications of the MHC, and ways to make it more comfortable and accessible universally. Finally, the prototype we have developed will serve as a valuable tool for assisting with the exploration of sensory substitution, crossmodal displays, and research on multimodal integration.

ACKNOWLEDGMENTS

Funding for this project and study was generously provided by the Canadian Natural Sciences and Engineering Council and the Canada Council for the Arts. The authors would like to thank all those who participated in the study, Gabe Nespoli for running the study, and Robert Indrigo for "Vibetracks" and chair prototypes. They also acknowledge Emmanuel Bigand and Stephen McAdams for the musical recordings used in this research.

REFERENCES

[1] G.V. Bekesy, *Experiments in Hearing*. McGraw Hill, 1960.
[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*, vol. 19, no. 8, pp. 1113-1139, 2005.
[3] P. Brooks and B. Frost, "Evaluation of a Tactile Vocoder Device for Word Recognition," *J. Acoustical Soc. of Am.*, vol. 74, pp. 34-40, 1983.
[4] P. Brooks, B. Frost, J. Mason, and D. Gibson, "Continuing Evaluation of the Queen's University Tactile Vocoder i: Identification of Open-Set Sentences and Tracking Narrative," *J. Rehabilitation Research and Development*, vol. 23, no. 1, pp. 129-138, 1986.
[5] E. Chew and A.R.J. Francois, "Interactive Multi-Scale Visualizations of Tonal Evolution in Musa.rt Opus 2," *Computers in Entertainment*, vol. 3, no. 4, pp. 1-16, 2005.
[6] R.W. Cholewiak, A.A. Collins, and J.C. Brill, "Spatial Factors in Vibrotactile Pattern Perception," *Proc. Eurohaptics Conf.*, 2001.
[7] R.W. Cholewiak and C.M. McGrath, "Vibrotactile Targeting in Multimodal Systems: Accuracy and Interaction," *Proc. Virtual Reality 2006 Conf. in Conjunction with the IEEE Conf.*, pp. 22-23, 2006.
[8] J.C. Craig and P. Evans, "Vibrotactile Masking and the Persistence of Tactual Features," *Perception and Psychophysics*, vol. 42, pp. 309-317, 1987.

[9] J.J. Gibson, *The Ecological Approach to Visual Perception.* Lawrence Erlbaum Assoc., 1979.
[10] E. Gunther, G. Davenport, and S. O'Modhrain, "Cutaneous Grooves: Composing for the Sense of Touch," *Proc. 2002 Conf. New Interfaces for Musical Expression (NIME '02),* pp. 1-6, 2002.
[11] M. Karam and D.I. Fels, "Designing a Model Human Cochlea: Issues and Challenges in Crossmodal Audio-Haptic Displays," *Proc. 2008 Ambi-Sys Workshop Haptic User Interfaces in Ambient Media Systems (HAS '08),* pp. 1-9, 2008.
[12] M. Karam, G. Nespoli, F. Russo, and D.I. Fels, "Modelling Perceptual Elements of Music in a Vibrotactile Display for Deaf Users: A Field Study," *Proc. Int'l Conf. Advances in Computer-Human Interaction,* pp. 249-254, 2009.
[13] C. Kayser, C. Petkov, M. Augath, and N. Logothetis, "Integration of Touch and Sound in Auditory Cortex," *Neuron,* vol. 48, no. 2, pp. 373-384, 2005.
[14] J.H. Kirman, "Vibrotactile Frequency Recognition: Forward and Backward Masking Effects," *J. General Psychology,* vol. 113, no. 2, pp. 47-58, 1986.
[15] V. Lévesque, J. Pasquero, V. Hayward, and M. Legault, "Display of Virtual Braille Dots by Lateral Skin Deformation: Feasibility Study," *ACM Trans. Applied Perception,* vol. 2, no. 2, pp. 132-149, 2005.
[16] J. Loomis and S. Lederman, *Tactual Perception,* pp. 31-41. Wiley, 1986.
[17] D.A. Mahns, N.M. Perkins, V. Sahai, L. Robinson, and M.J. Rowe, "Vibrotactile Frequency Discrimination in Human Hairy Skin," *J. Neurophysiology,* vol. 95, pp. 1442-1450, 2006.
[18] M.T. Marshall and M.M. Wanderley, "Vibrotactile Feedback in Digital Musical Instruments," *Proc. 2006 Conf. New Interfaces for Musical Expression (NIME '06),* pp. 226-229, 2006.
[19] J.B. Mitroo, N. Herman, and N.I. Badler, "Movies from Music: Visualizing Musical Compositions," *Proc. ACM SIGGRAPH '79,* pp. 218-225, 1979.
[20] E. Rusconi, B. Kwan, B.L. Giordano, C. Umilita, and B. Butterworth, "Spatial Representation of Pitch Height: The Smarc Effect," *Cognition,* vol. 99, pp. 113-129, 2006.
[21] J.A. Russell, "Evidence of Convergent Validity on the Dimensions of Affect," *J. Personality and Social Psychology,* vol. 36, pp. 1152-1168, 1978.
[22] F.A. Russo, L.L. Cuddy, A. Galembo, and W.F. Thompson, "Sensitivity to Tonality across the Pitch Range," *Perception,* vol. 36, pp. 781-790, 2007.
[23] D.C. Sinclair, *Mechanisms of Cutaneous Sensation,* chapter xi, p. 363. Oxford Univ. Press, 1981.
[24] C. Strumf, *Tonpsychologie,* vol. 1. S. Hirzel, 1883.
[25] Tadoma, The Tadoma Method, Web Resource, 2007.
[26] W. Thompson, F. Russo, and D. Sinclair, "Effects of Underscoring on the Perception of Closure in Film Excerpts," *Psychomusicology,* vol. 13, pp. 9-27, 1994.
[27] T. Tommerdahl, K. Hester, E. Felix, M. Hollins, O. Favorov, P. Quibrera, and B. Whitsel, "Human Vibrotactile Frequency Discriminative Capacity After Adaptation to 25 Hz or 200 Hz Stimulation," *Brain Research,* vol. 1057, pp. 1-9, 2005.
[28] 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. Applied Perception,* vol. 2, no. 2, pp. 106-117, 2005.
[29] P.B.Y. Rita, "Tactile Vision Substitution: Past and Future," *Int'l J. Neuroscience,* vol. 19, pp. 29-36, 1983.

Maria Karam is a postdoctoral fellow at the Centre for Learning Technologies (CLT). Her research focus is on human computer interactions that move users off the desktop with alternative interaction techniques. Music is one of the primary areas of focus for her interaction research, which ranges from using gestures for controlling music players to substituting sound with touch. A lead researcher on the Alternative Sensory Information Display (ASID) project, she has been developing the Model Human Cochlea (MHC) as a tool that can assist research into understanding the effects of translating sound into a tactile interface. She is also a director of a public usability lab situated within a coffee shop in downtown Toronto.

Frank A. Russo received the PhD degree from the Queen's University at Kingston (2002) followed by postdoctoral fellowships in music cognition and hearing science. He is a cognitive scientist, musician, and armchair engineer. His current position is of an assistant professor of psychology and director of the Science of Music, Auditory Research, and Technology (SMART) lab at Ryerson University. He is on the editorial board of the *Psychomusicology: Music, Mind, and Brain* and the board of directors of the Canadian Acoustical Association. His research stands at the intersection of music, mind, and technology.

Deborah I. Fels is a professor at Ryerson University, and the head of the Centre for Learning Technologies (CLT). She is an expert in the field of assistive and adaptive technology and has led several highly successful projects including the ASID, an alternative sensory information display that provides music in the form of tactile stimuli to the deaf and hard of hearing. She and her research staff and students at the CLT have been exploring accessible media, specifically enhanced and emotive captioning, and descriptive video since 2001.