Crossmodal Representations of Music:
A Sensory Supplementation Approach

John Matthew Tennant
University of Toronto

August 17, 2015

ABSTRACT: Research on crossmodal processing suggests that pairing music with corresponding visual and tactile sensations may improve music perception. I define an approach to effective crossmodal representations of music based on the methodology of sensory substitution and supplementation research. I argue first that crossmodal music technologies should emulate sensory substitution technologies by using *deterministic* and *crossmodally congruent* parameter mapping to present information through *high-bandwidth sensory channels*. I then propose additional music-specific considerations, arguing that crossmodal music technologies should present *rich stimuli* which capture the many-layered structure of music while *maintaining the emergence relationships between musical parameters*. Finally, I review existing approaches to representing music crossmodally and evaluate their potential to serve as musical sensory supplementation technologies.

Crossmodal Representations of Music:

A Sensory Supplementation Approach

Music has historically been experienced within a multisensory environment often involving watching performers singing, playing an instrument, or dancing. In the age of digital music however, much of musical experience is now limited to the auditory domain. Nonetheless, many software and hardware technologies exist for representing music using visual or tactile media. Little psychological research has been directly applied to the crossmodal representation of music. However research on sensory substitution and crossmodal correspondence suggests that these technologies may have the potential to both facilitate the perceptual processing of music, and to improve music appreciation for hearing and hearing impaired individuals. I will argue that an audio sensory-supplementation approach tailored to the parameters of music most effectively harnesses the benefits of musical crossmodal representation.

How Should Music Be Represented Crossmodally?

Many approaches to crossmodal representation of music exist, however most lack scientific rigour. Psychological research on sensory substitution and supplementation provides a logical starting point for this investigation into crossmodal representation of music.

Intro to Sensory Substitution

Sensory substitution research investigates methods for improving a person’s awareness of aspects of reality which might otherwise remain inaccessible to them (Lenay, Gapenne, Hanneton, Marque, & Genouëlle, 2003). This can be achieved through behavioral techniques, such as human echolocation (Thaler, Arnott, & Goodale, 20011), or via technology. The Tongue Display Unit (Bach-y-Rita, Kaczmarek, Tyler, Garcia-Lara, 1998; Kaczmarek, 2011) is an example of a tactile-vision substitution system. It converts optical data from a body-mounted camera into a spatially organized pattern of electrical stimulation on the user’s tongue. Once users of a sensory substitution technology successfully learn a crossmodal mapping, they experience improved performance in the relevant domain (Lenay et al. 2003), activation of other sensory processing areas (Poirier, De Volder, & Scheiber, 2007), and qualitative experience of the represented modality (Lenay et al. 2003). For example, experienced users of tactile-vision substitution systems demonstrate improved visual-spatiotemporal awareness(Sampaio, Maris, & Bach-y-Rita, 2001), show activation of the visual cortex(Kupres, Sampaio, Moesgaard, Gjedde, & Ptito, 2003), and report the subjective experience of (albeit imprecise) visual perception (O’Regan, & Noë, 2001).

“Sensory substitution” versus “sensory supplementation”. Lenay et al. (2003) argue that the term “sensory supplementation” is a better descriptor than “sensory substitution” for two main reasons. First, target information can be independent of any existing sensory modality; and second, sensory substitution devices can be effectively used by individuals with normal sensory functioning. These arguments equally apply to crossmodal representations of music, so I will refer to “musical sensory *supplementation*” rather than “*substitution*”.

Auditory Sensory Substitution. Researchers have created a number of devices intended for auditory sensory substitution which show promise for musical sensory supplementation. Most devices present auditory data to either the visual or tactile sensory modalities. The Model Human Cochlea (Karam, Russo, & Fels, 2009) conveys frequency content and loudness information using spatially arranged vibrotactile stimulators. The Voice Visualizer (Pronovost, Yenkin, Anderson, & Lerner 1968) converts auditory sounds into visual shapes and patterns. I will review both of these devices later in terms of their potential for effectively representing music crossmodally.

Principles of Sensory Substitution. Organisms most readily learn and use stable, informative, and relevant stimulus pairings (Rumbaugh, King, Beran, Washburn, & Gould, 2007). Research on sensory substitution has proposed crossmodal mappings based on three main principles. First, using *deterministic mappings* facilitates the process of the brain learning, and adapting to the new “sensory organ”. Second, *presenting information in a manner accessible to the modality* allows for large amounts of information to be precisely conveyed. Finally, *using existing crossmodal correspondences where possible* may improve the effectiveness of the device and shorten the adjustment period. Each of these principles will be considered in turn.

Deterministic Mapping

Sensory substitution relies on the brain’s ability to learn a fixed relationship between a set of sensations and an underlying reality (Lenay et al. 2003). For example, users of tactile-visual substitution systems learn to use patterns of tactile stimulation on their skin or tongue to determine the position and movement of objects in the environment. At first, users of sensory substitution technology report that they perceive the new sensations focally, and use them to deduce features of the underlying reality. After training, users often find that the awareness of the sensations becomes subsidiary, and awareness of the underlying reality becomes focal (O’Regan & Noë, 2001). At this point perception through the new “sensory organ” is said to be implicit; users no longer pay direct attention to the sensations which provide the information. This process of integration relies on the sensations having a stable coupling to an underlying signal (Bach-y-Rita & Kercel, 2003). Whitelaw (2008) argues that the goal of crossmodal representation of music is to achieve this kind of implicit and integrated awareness of an underlying, modality-independent “musical signal”. If so, then the necessity of deterministic correspondence for integration suggests a similar need in a successful crossmodal mapping of music.

Presenting Information Accessibly

Sensory substitution devices provide sensory modalities with access to types of information to which they would otherwise be insensitive. This involves mapping information about inaccessible target stimuli to a set of easily perceived sensations in the new modality (Lenay et al., 2003). Optimal parameter mappings allow for many different stimuli to be quickly and accurately perceived and discriminated. This requires precise knowledge of the sensory capabilities and limitations (psychophysics). Consideration of the strengths and limitations of audition, vision, and touch will inform the present approach to crossmodal representation of music.

Audition. In music, the parameters of rhythm, dynamics, pitch, harmony, and timbre are central. Audition is very sensitive to changes in sound pressure levels. This is involved the perception of dynamics in music. Rhythm perception involves temporal discrimination of events experienced at a rate slower than 20 per second. Above 20Hz, repetitive events become a perceived tone, so around 20Hz the sensation of rhythm gives way to that of pitch. Musical pitch is determined by the fundamental frequency of a musical tone. Humans can detect and discriminate auditory frequencies between 20Hz and 20KHz. Pitch discrimination and contour perception is involved in the perception of melody.

Melody, harmony, and timbre are primarily perceived in terms of the relationships between musical frequencies, rather than in terms of absolute pitch. The phenomenon of timbre arises from the perception of the overall spectral contour of a sound across time. Humans distinguish musical instruments, speakers, and everyday noises by primarily by their timbre (and pitch to a lesser extent).

Vision. Musical rhythms are slow enough (<20Hz) that they can be accurately perceived when represented as visuo-temporal rhythms. However, pitch, harmony, and timbre are constituted by vibrations which are too fast for vision, so they must be mapped to other visual parameters. The visual system can effectively process enormous amounts of information presented as spatiotemporal patterns of color and brightness. Multisensory music technology can make use of vision's sensitivity to space, shape, motion, colour, and brightness information to convey musical parameters across time.

Tactile. Unlike vision, touch is sensitive to vibration. Using vibrotactile perception, humans can discriminate frequencies between 5Hz and 1000Hz with ±2 semitone accuracy (Karam et al., 2009). Humans can also identify musical instruments based on timbre (Russo, Ammirante, & Fels, 2012), perceive consonance and dissonance[1](Yoo, Hwang, Choi, 2014; Okazaki, 2013), and discriminate musical rhythm (Ezawa, 1988) and meter (Huang, Gamble, Sarnlertsophon, Wang, & Hsiao, 2012). These findings suggest that music can be effectively presented as mechanical vibration of the skin.

Vibrotactile perception is limited however, in its ability to parse the signal into auditory objects, and perceive subtleties required for speech perception. This is likely due to the fact that the vibrotactile sensory apparatus does not effectively separate the neural encoding of different

[1] Okazaki et al. (2013) showed that humans can also judge consonance and dissonance between tones when one is heard and one is felt vibrotactilely.

frequency bands. Spatial coding of musical information may be able to compensate for the lack of precision at single loci. Researchers have noted that the skin provides thousands of discriminable points, each of which could potentially be used as an input channel for sensory supplementation (Bach-y-Rita et al., 1998).

Electrotactile Perception. Electrotactile sensitivity has not been thoroughly studied. It is presumed to have temporal acuity similar to touch, however the spatial acuity and frequency discrimination may be better. An electrode can deliver a signal to a single neuron without deforming all the skin around it. This gives electrotactile stimulation an advantage in terms of spatial localization. There are two important differences between frequency perception in physical vibration and electrotactile stimuli. (1) Physical vibration detection relies on the skin transmitting a vibration to the nerve. The skin transmits differentially based on frequency, meaning that some vibrations never make it to the receptor. Electrotactile stimulation avoids this problem by delivering the rhythmic stimulation directly to the nerve. (2) It has also been hypothesized that the alternating polarity of an alternating-current signal may facilitate recovery in sensory neurons, allowing them to fire at a higher frequency than possible when responding to physical vibration (McConnell, 1989).

Psychophysics Conclusions. Other than rhythm, many of the most important elements of music such as melody, harmony, and timbre are inaccessible to, or only vaguely discriminable by vision and touch. Given the different sensitivities of the modalities to parameters such as space, time, and vibration; effective multisensory music technology must take into account the senses' strengths when mapping musical parameters.

Aptness of Crossmodal Correspondence

Sensory substitution researchers often choose crossmodal mappings that seem intuitive (e.g. high visual field with high pitch). Often they do so without explicitly citing crossmodal correspondence literature. However research on crossmodal correspondences has shown that simultaneous presentation of crossmodally-congruent stimuli can enhance information processing, so detailed consideration should be given to the choice of parameters and polarity for information-stimulus mappings.

Crossmodal correspondences are reliable and systematic psychological associations between stimulus parameters in different modalities. Crossmodal correspondences often affect speed and accuracy of information processing. Though arbitrary crossmodal-mappings may be learned, some benefits of crossmodal processing are available without training due to the brain's biological preparedness to crossmodally process certain types of stimuli. Therefore, choosing an apt mapping may reduce the amount of exposure time required to receive the benefits of a crossmodal representation of music. Research on auditory-visual and auditory-tactile crossmodal correspondences are beginning to indicate which pairings may be most effective.

Psychology researchers have used two main methods to investigate crossmodal correspondence. One method investigates explicit subjective associations. In this method participants perform tasks such as matching two stimuli from different modalities (e.g., "Which picture is Bouba and which is Kiki?"(Ramachandran & Hubbard, 2001)), or use crossmodal metaphorical descriptors for a single stimulus (e.g. "how wet/bright/high is this sound?" (Eitan & Rothschild, 2010)). The second way that researchers have investigated crossmodal correspondence is by examining the effect of stimulus pairings on information processing. This is usually achieved via "speeded classification" (e.g., Marks 1975, Evans & Treisman 2010) or “implicit association” tasks (Parise & Spence, 2012). These tasks test for “crossmodal congruence effects”, evidenced by an increase in accuracy or response time when paired stimuli are congruent on a relevant parameter rather than incongruent.

Examples of correspondence. Many auditory-visual and auditory-tactile correspondences have been found using the two research paradigms. An effective musical sensory supplementation technology will make use of these correspondences to map auditory stimuli to visual and tactile parameters.

Auditory-visual correspondences. Congruence effects have been found between the auditory pitch and a number of visual parameters. Higher pitch has been shown to be congruent with higher spatial elevation (Evans & Treisman, 2010), increased brightness (Ludwig, Adachi, & Matzuzawa, 2011), increased lightness (Mondloch & Maurer, 2004), smaller size (Evans & Treisman, 2010), increased angularity of shape (Parise & Spence, 2012), and increased spatial frequency (Evans & Treisman, 2010). See Spence (2011) for a review. Louder sounds have shown congruence effects with increased visual brightness (Marks 1987), and size (Kitagawa & Ichihara, 2002). Auditory rhythms have shown subjective correspondence with spatial frequency (Guzman-Martinez et al., 2012) and congruence effects with visual rhythms expressed as flashes (Frings & Spence, 2010). Auditory timbre has shown subjective correspondence to visual texture (Grill & Flexer, 2012). To my knowledge, no studies have investigated crossmodal correspondences between auditory harmony and visual parameters.

Auditory-tactile correspondences. Tactile, and electrotactile rhythms expressed as pulsed vibrations or electrical current have shown congruence effects with auditory rhythm (Frings & Spence, 2010). Auditory pitch has shown congruence effects with tactile elevation (Occelli, Spence, & Zampini, 2009) and tactile size (Walker & Smith 1985). Tactile frequency has shown congruence effects with auditory frequency (Ro, Hsu, Yasar, Elmore, & Beauchamp, 2009). Vibrotactile texture has shown congruence effects with auditory timbre (McGee, Gray, & Brewster, 2002). Vibrotactile amplitude has shown congruence effects with auditory loudness (Okazaki, Kajimoto, & Hayward, 2012).

Effects on perception. The information processing approach provides evidence that some of the correspondences influence performance and decision making in an automatic and unconscious way. However the effects of crossmodal associations on conscious perceptual experience remains largely unexplored. Some preliminary research suggests that multisensory cues are used to resolve the perception of ambiguous stimuli in favor of crossmodal congruence (Maeda, Kanai, & Shimojo, 2004; Kitagawa & Ichihara, 2002; Smith, Grabowecky, & Suzuki, 2007).

Origins of crossmodal correspondences. Spence (2011) proposes three (possibly interacting) mechanisms that contribute to the development of a crossmodal association. 1) *Structure-based associations* due to similarity in the way neurons communicate two parameters (e.g. brightness and loudness have a similar neural coding scheme). 2) *Statistically-based associations* formed through learning the common pairing of two types of sensations. 3) *Semantically-mediated associations* acquired by using the same symbol or word to refer to phenomena in two different modalities[2]. Spence (2011) concedes that most of the strong

[2] Of course, two parameters in different modalities using the same descriptive language does not preclude the possibility of a statistical or structure-based relationship. It may be that the language developed due to a pre-existing cognitive synergy. And/or prehistoric language may have linked the development and processing certain parameters together long in the evolutionary past, producing a biological predisposition to view those qualities as related.

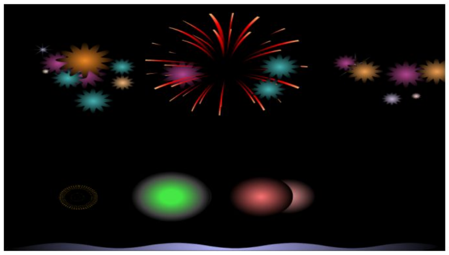
correspondences studied by psychologists are likely due to statistical or structural factors. However, this leaves unanswered questions about the source of the musical auditory-visual and auditory-tactile correspondences. If they are structural, why was the brain wired to process two disparate modalities similarly? If they are statistical, what kind of stimuli led to the development of these correspondences?

Resonating physical objects produce multisensory sensations which correspond in many psychologically congruent ways. For example, higher pitched sounds create smaller and more intricate spatial patterns of resonance which can be made accessible to touch and vision. This suggests that resonance phenomena might provide the statistical basis for certain crossmodal correspondences. Cymatic researchers study methods for visualizing the spatial patterns of vibratory phenomena (Jenny, 2001). I will review cymatic methods of music visualization in a later section.

Beyond the Existing Approach to Sensory Supplementation

Thus far, it appears that an effective crossmodal representation of music will 1) use a deterministic parameter mapping, 2) present information in a manner well-suited to the sensory modality, and 3) make use of existing crossmodal correspondences where possible.

The Object-Based Music Visualizer. Nanayakkara, Taylor, Wyse and Ong (2007) employed a similar sensory supplementation approach to develop an object-based method of music visualization. In their visualizer, each note of a piece generates an on-screen object whose visual parameters are crossmodally determined by the pitch, amplitude, and timbre of the sound. Musical parameters are mapped using psychologically validated crossmodal correspondences: pitch to object size, amplitude to brightness, and timbre to texture. Nanayakkara et al.'s method



thus yields the experience of a dark screen where objects appear and disappear simultaneously with congruent sounds (pictured). Their method successfully employs the previously outlined sensory substitution approach: it uses empirically validated crossmodal correspondences and deterministically maps relevant musical parameters to high-bandwidth sensory channels. However, because it uses explicit mapping of each parameter, it only allows for crossmodal correspondences which have already been investigated, and contains only the musical parameters which the researchers deem relevant. Furthermore, the visual parameters don't seem to bear any strong relation to each other (there usually exists a continuity between rhythm, pitch, timbre, and harmony). The sterility of Nanayakkara's object based method highlights the incompleteness of the three-principled sensory substitution approach. In addition to the aforementioned criteria, I suggest that crossmodal representation of music should present audio data with a minimal amount of information loss, and maintain the relationships between musical parameters.

Information-rich stimuli. The human body has a tendency to automatically organize its biorhythms into harmony with each other and with repetitive external stimuli in a process called entrainment. Entrainment has beneficial effects on attention, memory, and musical enjoyment. Heart rate (Bernardi, Porta, & Sleight, 2006), breathing (Haas, Distenfeld, & Axen, 1986), brainwaves (Zhuang, Zhao, & Tang, 2009), and motoric behaviour (e.g. dancing, foot tapping) (Trainor, Gao, Lei, Lehtovaara, & Harris, 2009) all show entrainment effects. Music cognition seems to rely heavily on entrainment (Honing, ten Cate, Peretz, & Trehub, 2015). Current research is only scratching the surface in determining which kinds of stimuli can serve as targets for entrainment, and which biological processes entrain. Research has used relatively coarse measures and stimuli for entrainment such as isochronous tones, flashing lights, and repeated vibrotactile stimuli. However, entrainment likely occurs to layers of stimuli at many levels of biological and social processes (see, for example Phillips-Silver, Aktipis, and Bryant (2010)). Likewise, crossmodal correspondences have only been investigated between very limited and simple parameters. In contrast, I suggest that the rich hierarchical temporal patterns in music may serve as a complex stimulus for multilayered entrainment and crossmodal correspondence. I therefore argue that auditory-visual and auditory-tactile mappings which avoid interpreting and compressing the signal have the potential to capitalize on crossmodal correspondences and methods of entrainment which have not yet been identified.

Maintaining relationships between musical parameters. In audition, the phenomena of loudness, rhythm, frequency, timbre, and harmony are all interconnected. They all result from temporal patterns of high and low pressure waves. A rhythmic sound played fast enough becomes a pitched tone (around 20Hz). Many pitches sounded simultaneously can produce harmony or timbre depending on their relationship to each other. In the most successful crossmodal mappings, I believe that these relationships between musical parameters will be maintained. The importance of this criterion will become clear when examining technologies for music visualization.

Summary of Proposed Approach

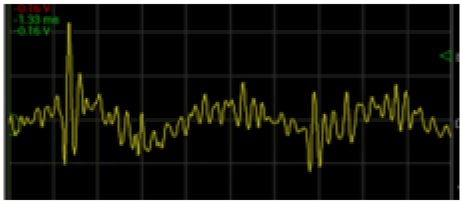
In summary, I have suggested that an ideal multisensory music technology will 1) establish a deterministic relationship between auditory and visual or tactile stimuli, 2) effectively present relevant musical parameters through high-bandwidth sensory channels, 3) aptly map musical parameters to crossmodally congruent stimuli, 4) minimally compress the musical signal in order to preserve nuances which may have unforeseen value, and 5) maintain the relationships of emergence between musical parameters. Using this approach it is now possible to review and evaluate technologies for crossmodal music representation.

Music Visualization Technology

Traditional Methods

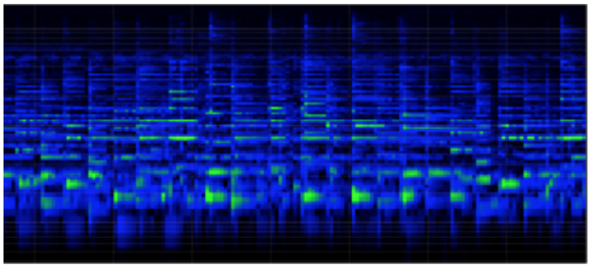
Electrical representation of sound is achieved by converting the pressure over time oscillatory pattern into voltage over time alternating current signal. Electrical audio signals can be converted into sound vibration by driving a speaker with the voltage over time signal. Electrical audio can also be generated independent of any acoustic source and transduced via a speaker in this way. With the advent of electronic representation of audio came the possibility of using the electrical audio signal to drive a music visualizer. The oscilloscope and spectrogram are two ubiquitous styles of music visualizer which have emerged from the electrical signal processing tradition.

Oscilloscope. The oscilloscope generates an *information-rich* cartesian line graph from raw audio data using a single *deterministic mapping*: pressure and time mapped to vertical and horizontal position. The resulting *crossmodally-congruent* mappings for rhythm, dynamics, and pitch are emergent from the pressure-time to vertical-horizontal position relationship. As a result they *maintain the relationships between musical parameters* in the visual domain. However the oscilloscope's inability to produce temporally-stable closed figures severely limits its *visual accessibility*.



The oscilloscope represents rhythm as temporally-congruent changes in visual patterns (Frings & Spence, 2010), and dynamics as shapes with congruent vertical size (Kitagawa & Ichihara, 2002). It visualizes pitch and timbre congruently as horizontal size (Evans & Treisman, 2010) and texture (Grill & Flexer, 2012) respectively. Harmony manifests as visually coherent waveform patterns (no crossmodal congruence data exists for musical harmony). Unfortunately, because the patterns on the oscilloscope do not form closed figures, and appear at different places on screen upwards of 30 times per second, the shapes representing pitch, harmony, and timbre remain largely indiscriminable.

This visualizer likely taps into so many crossmodal correspondences because it models a physical reality (travelling waves on a string) to which humans are adapted. The richness of the oscilloscope allows many related crossmodal correspondences to manifest simultaneously. These may include correspondences which psychologists have not yet identified. Furthermore, the oscilloscope reveals continuity between musical parameters (e.g. a combination of frequencies producing timbre is visualized as a combination of various shapes producing a visual texture). In sum, the oscilloscope shows promise as a device for musical sensory supplementation, but is unfortunately too erratic for precise visual perception.



Spectral displays. In contrast to the information-rich visual signals presented by an oscilloscope, spectrograms display *highly abstracted,* yet still *deterministic* representation of the frequency content in a musical signal over time. The mappings for pitch, rhythm, and dynamics are *crossmodally congruent* and *visually accessible*, but they *do not* *maintain the relationships between musical parameters.* Harmony and timbre are not congruently represented. Nor are they accessible in the visual domain.

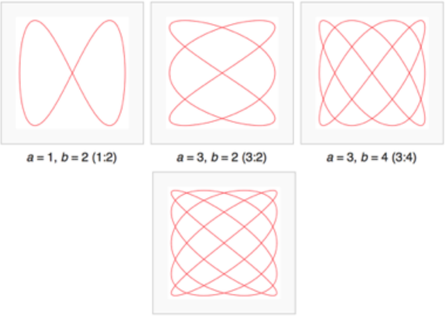
The spectrogram represents rhythm using temporally-congruent changes in visual patterns (Frings & Spence, 2010). It congruently represents dynamics using levels of brightness (Ludwig et al., 2011) and pitch using vertical position (Evans & Treisman, 2010). Timbre appears as brightness contour on the spectrum. Harmony can be deduced using distances between lines, but there is no indication of consonance or dissonance. Unfortunately, these representations do not capture the essence of timbre and harmony.

The data on a spectrogram is highly abstracted, and does not provide much potential for complex crossmodal congruence or entrainment. The relationships between musical parameters is not maintained. In sum, the spectrogram fails to capture important aspects of music, leading to an abstracted feeling contrary to the integration sought by this approach to musical sensory supplementation.

Promising Extensions of the Oscilloscope

Traditional oscilloscopes do not convey harmony effectively. Furthermore, their presentation of data along a single dimension fails to take advantage of the visual system's affinity for two dimensional shapes and patterns. The XY-oscilloscope and my animated software tonoscope transform oscilloscope data into 2-dimensional shapes. In doing so, they potentially address both problems while preserving the advantages of traditional oscilloscopes.

XY Oscilloscope. The XY oscilloscope is a promising variation on the traditional oscilloscope. In this technology, two

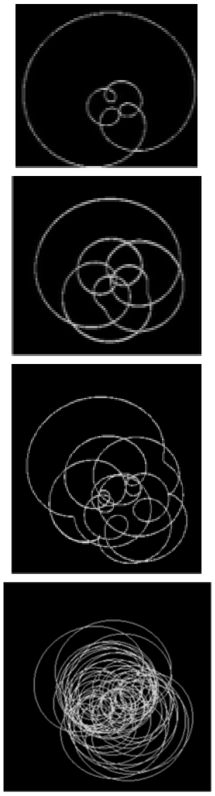

$a = 1$, $b = 2$ (1:2)

$a = 3$, $b = 2$ (3:2)

$a = 3$, $b = 4$ (3:4)

$a = 5$, $b = 4$ (5:4)

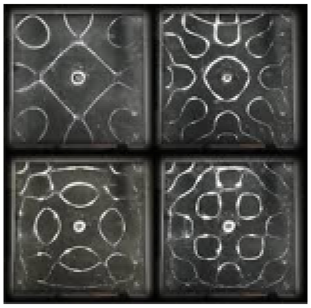
input signals are mapped to the X and Y coordinates, and lines are drawn across time as X and Y vary. This method of visualization has the potential to showcase harmonic relationships between signals. Because an audio signal varies on only one dimension (pressure) across time, some technique must be used to generate a signal with which to compare it. Unfortunately, it is not obvious how a single auditory signal should be divided up into the two channels for comparison. If the same audio signal is used for both channels, the pointer oscillates back and forth, drawing a straight line. When two sine waves of different frequencies are compared via an XY oscilloscope, a Lissajous curve results (pictured). This is a visual representation of harmony.



The Voice Visualizer. The Voice Visualizer (Pronovost et al., 1968) is, to my knowledge, the most visually pleasing and musically effective technique for displaying audio on an XY oscilloscope. Though it was originally developed for audio-visual substitution of speech sounds, the voice visualizer satisfies all the criteria outlined thus far. It is *information-rich*, and uses a *single deterministic formula* to produce *crossmodally congruent mappings*. Furthermore, the emergent mappings *maintain the relationships between musical parameters*, and present information as stable shapes *well-suited to visual perception.*

This XY oscilloscope represents rhythm, dynamics and timbre in ways similar to the traditional oscilloscope. (with visual rhythm, object size, and texture). The important innovation is that pitch is visualized relatively. The fact that the spatial placement of patterns is dependent on the frequency content of the signal means that shapes no longer jump around erratically. A single note of a given a waveform produces the same shape and orientation at any pitch. Similarly, transposing a signal yields does not change the visual pattern. The pattern changes only when the relationships between notes are altered. Generally speaking, when multiple frequencies are present, higher frequencies are congruently visualized by smaller, more detailed patterns than lower frequencies (Evans & Treisman, 2010; Parise & Spence, 2012). In terms of harmony, the frequency ratios between pitches affect the complexity of the patterns drawn (pictured). The simpler the frequency ratio, the simpler the pattern. Interestingly, these types of patterns were identified as a kind of visual harmony by John Whitney[3]. This crossmodal pairing has not yet been investigated, but because of its emergence from other investigated mappings it is plausible that auditory harmony could be crossmodally congruent with visual coherence and proportion.

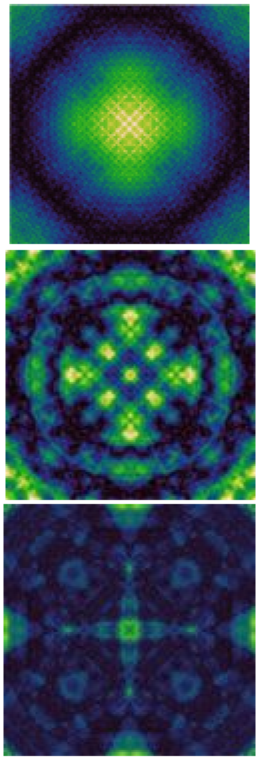
Like the traditional oscilloscope, Pronovost et al.'s XY oscilloscope can be viewed as a model of a physical phenomenon: the (now circular) vibration of a string in three dimensions as viewed cross-sectionally along its length. Again, the visualizer's relationship to a physical reality may explain the richness of crossmodal correspondence. This modified of the XY oscilloscope retains the advantages of the traditional oscilloscope while transforming the erratic line drawings into stable two-dimensional shapes. This the only visualizer to my knowledge which converts musical harmony deterministically into an easily discriminable form of visual harmony. The result is a highly promising device for musical sensory substitution.



Cymatic Visualization. Cymatic technologies show the patterns of resonance formed in vibrating physical media. Most cymatic visualizers use tonoscopes constructed from physical media to achieve

[3] Whitney, J. (1980). *Digital harmony*. Byte Books.

this. For example, sand sprinkled on a Chladni plate is vibrated around until it settles in the nodes of the resonance pattern (pictured). Light can also be shone through pools of water exposed to vibration to make the patterns of crests and troughs visible. I have engineered an animated software tonoscope which visualizes patterns of resonance in an abstracted chladni plate or ripple tank. The water tonoscope and chladni plate have not been subject to much study. Their reliance on physical media makes them impractical and unpredictable. My animated software tonoscope inherits many of the advantages of the physical method, but it produces results which are more consistent and predictable.



Animated Software Tonoscope. This visualizer (pictured thrice) shows raw audio data emanating from a single source, bouncing off the sides of the ripple tank, and forming resonance patterns with itself. The animated software tonoscope satisfies almost all of the criteria outlined thus far. It is *information-rich*, and uses a *single deterministic formula* to produce *crossmodally congruent mappings*. Furthermore, the emergent mappings *maintain the relationships between musical parameters*, and present information as stable shapes *well-suited to visual perception.*

All representations of musical parameters are emergent from the crossmodally-congruent amplitude to brightness mapping. Rhythm appears as temporally-congruent changes in visual patterns (Frings & Spence, 2010). Timbre is visualized congruently as visual texture (Grill & Flexer, 2012). Pitch is mapped congruently to spatial frequency (Evans & Treisman, 2010), which ultimately results in additional congruent mappings to size (Evans & Treisman, 2010) and shape (Parise & Spence, 2012). Pitch is conveyed relative to the resonant frequency physical model. If a frequency related to the tonic of a piece is selected as the resonant frequency, then shapes are indicative of a pitch's relationship to the key. Harmony emerges from the interference between shapes. It forms discriminable patterns, and though intervals and chord types are hard to discriminate, consonance and dissonance are effectively visualized by the degree of temporal stability and spatial coherence. Again, the congruence of these crossmodal mappings for harmony have not been investigated by psychological researchers. However it is highly plausible that congruence effects exist between musical harmony and visual coherence due to this pairings' emergence from other congruent mappings.

This visualizer makes good use of the visual system. Low frequency events like rhythm are displayed in a way which vision can track, and the normally inaccessible high frequency parameters such as pitch and harmony are made visible as shape using resonance. This visualizer makes use of vision's high spatial acuity and sensitivity to colour and brightness contrast to present a very rich and detailed signal in an accessible way.

A final advantage of the animated software tonoscope is its maintenance of the relationships between musical parameters. There is continuity in the representation of rhythm, pitch, timbre, and harmony. It is interesting to note that the relationship of auditory frequency to auditory timbre is analogous to the relationship of visual spatial frequency to visual texture. In this visualizer, that four way relationship is maintained. Auditory frequency is mapped to spatial frequency, and when many frequencies are present, the auditory sensation is timbre and the visual sensation is texture.

Music Visualization Conclusions

The alternative oscilloscopes reviewed here present exciting potential technologies for effective musical sensory supplementation. The effectiveness of physical models in capitalizing on crossmodal correspondence suggests further avenues of psychological research. Statistical crossmodal correspondences reflect real pairings in the world. It has been adaptive for our sensory system to develop crossmodal processing takes advantage of crossmodal correspondences to enhance perception. Physical-modelling visualizations may be particularly effective because they contain a plethora of related audiovisual stimuli which are likely candidates for statistical crossmodal congruences. Psychological research on methods of physical-modelling visualization may help to identify the statistical bases of many crossmodal congruences. It may suggest new crossmodal correspondences, such as the mapping of musical harmony to visual coherence, proportion, and temporal stability.

Tactile Representations of Music

Representation of music in the tactile modality has received much less attention than music visualization. However, because touch is sensitive to vibration, music can be presented without any transformations. The promising result for musical sensory supplementation is that all musical parameters can be mapped in crossmodally congruent ways. The role of the technology, then, is to maximize the accessibility of the various musical parameters.

Musical Vibroacoustic Stimulation

Musical vibrations can be delivered to the body using musical instruments, speakers, and other types of vibrotactile actuators. These methods of presentation congruently map all musical parameters to their vibrotactile correlates. Unfortunately, vibrotactile perception is not very precise due to a number of factors. (1) The skin is not very sensitive to differences in frequency, and (2) cannot separate the frequency bands for a nuanced perception of the signal. (3) Skin filters the signal via damping, thus our tactile system is much less sensitive to some frequencies of vibrotactile stimuli despite the nerves having the capacity to respond to a wide array of vibratory frequencies. Furthermore, traditional vibrotactile perception of music does not take advantage of tactile localization. The Model Human Cochlea (MHC) (Karam et al., 2009) seeks to improve vibrotactile perception of music by incorporating spatial coding of frequency information.

The Model Human Cochlea

The model human cochlea (MHC) uses a number of vibrotactile devices placed along the body (typically along the spine) to deliver music as mechanical vibration. The auditory signal is split into frequency bands, then each band is delivered by vibrotactile actuators. The vibrators are spatially organized such that low auditory frequencies are congruently represented (Occelli et al. 2009) by stimulation at the base of the spine and higher frequencies represented by stimulation towards the head. This technology presents all the audio data to the tactile system in a format which improves its accessibility over vibrotactile stimulation of a single locus. It also adds the crossmodally congruent pairing of pitch with spatial elevation.

Pitch is congruently represented by the spatial location (Occelli et al., 2009) and frequency of the vibrotactile stimulation (Ro et al., 2009). Dynamics are congruently represented by the intensity of vibration (Okazaki et al., 2012). Rhythm is emergent from the amplitude mapping as congruent pulsed tactile rhythms (Frings & Spence 2010). Harmony is conveyed both by the distance between stimuli, and the congruent experience of vibrotactile harmony (Yoo, 2014). Timbre is conveyed congruently by the vibrotactile timbre (McGee, Gray, &

Brewster, 2002), as well as the envelope of the overall set of vibrotactile actuators.

This technology takes advantage of the spatial resolution of touch, as well as its ability to discriminate different levels of vibration. It also makes use of the ability to discriminate frequencies, consonance from dissonance, and textures, all of which have been implicated in audio-tactile crossmodal processing. Overall, the Model Human Cochlea represents a promising development in musical sensory supplementation technology. It satisfies all of the criterion developed here, and extends the traditional method of vibrotactile music representation.

Electrotactile musical stimulation

Electrotactile musical stimulation of a single locus takes advantage of the same crossmodal mappings of vibrotactile musical vibration, but may also benefit from improved frequency discrimination over normal tactile stimulation (McConnell, 1989). Using spatially organized electrodes, a device such as the Model Human Cochlea could be adapted for musical electrotactile stimulation. The improved spatial resolution of touch for electrotactile stimuli would allow the device to be made much smaller. It could, for example fit on the tip of the tongue, or back of the neck. Furthermore, electrotactile music technologies are superior in terms of privacy and portability. Whereas vibrotactile stimulators produce audible sound, require high voltages, and use mechanically vibrating parts, electrotactile stimulation can be delivered silently by stationary electrodes.

Conclusion

From this process of developing and applying an approach to musical sensory supplementation, a number of important conclusions can be drawn. The most effective technologies for representing music crossmodally will likely be based in physical realities. This was equally the case in the reviewed oscilloscopes and tactile representations. These physically-grounded technologies make use of congruent crossmodal mappings, and present information-rich stimuli which maintain the relationships between musical parameters. I suggested that these information-rich stimuli may serve as targets for multilayered entrainment, as well as complex or undiscovered statistical crossmodal correspondences (such as the emergent crossmodal mappings for musical harmony into the visual domain). I therefore argued that future psychological research should investigate the crossmodal relationships inherent in physically-grounded representations of audio to inform future research into crossmodal correspondences.

References

Bach-y-Rita, P., Kaczmarek, K. A., Tyler, M. E., & Garcia-Lara, J. (1998). Form perception with a 49-point electrotactile stimulus array on the tongue: a technical note. *Journal of rehabilitation research and development*, *35*, 427-430.

Bach-y-Rita, P., & Kercel, S. W. (2003). Sensory substitution and the human–machine interface. *Trends in cognitive sciences*, *7*(12), 541-546.

Bernardi, L., Porta, C., & Sleight, P. (2006). Cardiovascular, cerebrovascular, and respiratory changes induced by different types of music in musicians and non-musicians: the importance of silence. *Heart*, *92*(4), 445-452.

Eitan, Z., & Rothschild, I. (2010). How music touches: Musical parameters and listeners' audiotactile metaphorical mappings. *Psychology of Music*, 0305735610377592.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of vision*, *10*(1), 6.

Ezawa, M. (1988). Rhythm perception equipment for skin vibratory stimulation. *Engineering in Medicine and Biology Magazine, IEEE*, *7*(3), 30-34.

Frings, C., & Spence, C. (2010). Crossmodal congruency effects based on stimulus identity. *Brain Research*, *1354*, 113-122.

Haas, F., Distenfeld, S., & Axen, K. (1986). Effects of perceived musical rhythm on respiratory pattern. *Journal of Applied Physiology*, *61*(3), 1185-1191.

Honing, H., ten Cate, C., Peretz, I., & Trehub, S. E. (2015). Without it no music: cognition, biology and evolution of musicality. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *370*(1664), 20140088.

Huang, J., Gamble, D., Sarnlertsophon, K., Wang, X., & Hsiao, S. (2012). Feeling music: integration of auditory and tactile inputs in musical meter perception. PLoS ONE *7*(10): e48496.

Jenny, H. (2001). *Cymatics: a study of wave phenomena and vibration*. Macromedia.

Kaczmarek, K. A. (2011). The tongue display unit (TDU) for electrotactile spatiotemporal pattern presentation. *Scientia Iranica*, *18*(6), 1476-1485.

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

Kitagawa, N., & Ichihara, S. (2002). Hearing visual motion in depth. *Nature*, *416*(6877), 172-174.

Kupers, R., Sampaio, E., Moesgaard, S., Gjedde, A., & Ptito, M. (2003). Activation of visual cortex by electrotactile stimulation of the tongue in early-blind subjects. *Neuroimage*, 19, S65.

Lenay, C., Gapenne, O., Hanneton, S., Marque, C., & Genouëlle, C. (2003). Sensory substitution: limits and perspectives. *Touching for knowing*, 275-292.

Ludwig, V. U., Adachi, I., & Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans. *Proceedings of the National Academy of Sciences*, *108*(51), 20661-20665.

Maeda, F., Kanai, R., & Shimojo, S. (2004). Changing pitch induced visual motion illusion. *Current Biology*, *14*(23), R990-R991.

Marks, L. E. (1975). On colored-hearing synesthesia: cross-modal translations of sensory dimensions. *Psychological bulletin*, *82*(3), 303.

Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(3), 384.

McConnell, J. D. (1989). Method and apparatus for communicating information representative of sound waves to the deaf. *The Journal of the Acoustical Society of America*, *86*(6), 2476-2476.

McGee, M. R., Gray, P., & Brewster, S. (2002). Mixed feelings: Multimodal perception of virtual roughness. *Proceedings of Eurohaptics*.

Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(2), 133-136.

Nanayakkara, S. C., Taylor, E., Wyse, L., & Ong, S. (2007). Towards building an experiential music visualizer. *Information, Communications & Signal Processing, 2007 6th International Conference on*. IEEE.

Occelli, V., Spence, C., & Zampini, M. (2009). Compatibility effects between sound frequency and tactile elevation. *Neuroreport*, *20*(8), 793-797.

Okazaki, R., Kajimoto, H., & Hayward, V. (2012). Vibrotactile stimulation can affect auditory loudness: A pilot study. *Haptics: Perception, Devices, Mobility, and Communication*, 103-108.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, *24*(05), 939-973.

Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: a study using the implicit association test. *Experimental Brain Research*, *220*(3-4), 319-333.

Phillips-Silver, J., Aktipis, C. A., & Bryant, G. A. (2010). The ecology of entrainment: foundations of coordinated rhythmic movement. *Music perception*, *28*(1), 3.

Poirier, C., De Volder, A. G., & Scheiber, C. (2007). What neuroimaging tells us about sensory substitution. *Neuroscience & Biobehavioral Reviews*, *31*(7), 1064-1070.

Pronovost, W., Yenkin, L., Anderson, D., & Lerner, R. (1968). The voice visualizer. *American annals of the deaf*, *113*(2), 230.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia--a window into perception, thought and language. *Journal of consciousness studies*, *8*(12), 3-34.

Ro, T., Hsu, J., Yasar, N. E., Elmore, L. C., & Beauchamp, M. S. (2009). Sound enhances touch perception. *Experimental brain research*, *195*(1), 135-143.

Rumbaugh, D. M., King, J. E., Beran, M. J., Washburn, D. A., & Gould, K. L. (2007). A salience theory of learning and behavior: With perspectives on neurobiology and cognition. *International Journal of Primatology*, *28*(5), 973-996.

Sampaio, E., Maris, S., & Bach-y-Rita, P. (2001). Brain plasticity: 'visual' acuity of blind persons via the tongue. *Brain research*, *908*(2), 204-207.

Smith, E. L., Grabowecky, M., & Suzuki, S. (2007). Auditory-visual crossmodal integration in perception of face gender. *Current Biology*, *17*(19), 1680-1685.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971-995.

Thaler, L., Arnott, S. R., & Goodale, M. A. (2011). Neural correlates of natural human echolocation in early and late blind echolocation experts. *PLoS One*, *6*(5), e20162.

Trainor, L. J., Gao, X., Lei, J., Lehtovaara, K., & Harris, L. R. (2009). The primal role of the vestibular system in determining musical rhythm. *cortex*, *45*(1), 35-43.

Whitelaw, M. (2008). Synesthesia and cross-modality in contemporary audiovisuals. *The Senses and Society*, *3*(3), 259-276.

Whitney, J. (1980). *Digital harmony*. Byte Books.

Yoo, Y., Hwang, I., & Choi, S. (2014). Consonance of vibrotactile chords. *Haptics, IEEE Transactions on*, *7*(1), 3-13

Zhuang, T., Zhao, H., & Tang, Z. (2009). A study of brainwave entrainment based on EEG brain dynamics. *Computer and information science*, *2*(2), p80.