Synaestheatre: Sonification of Coloured Objects in Space

Giles Hamilton-Fletcher[1], Michele Mengucci[2], and Francisco Medeiros[2]

[1] University of Sussex, Brighton, UK
gh71@sussex.ac.uk
[2] Lab.I.O., Lisbon, Portugal
mengucci@gmail.com, excalibas@gmail.com

Abstract. Vision and sound can interact in many ways. For synaesthetes, sounds can automatically and consistently elicit vivid colours (and vice versa). Could a device that converts vision into sound provide not only a window into synaesthetic experiences, but also be a way for the visually-impaired to 'see with synaesthesia?' Sensory substitution devices (SSDs) convert live video into sound so users can experience visual dimensions through changes in auditory feedback. SSDs allow the visually-impaired to 'visually' locate and recognise objects, as well as expand access in education and art. Drawing upon research exploring natural hearing mechanisms and synaesthetic associations, an SSD called the 'Synaestheatre' has been developed that replicates the experience of synaesthesia for use as a visual aid. The Synaestheatre turns 3D spatial and colour information into aesthetically pleasing sounds that vary in real-time. We encourage attendees to experience a new form of synaesthesia through an installation at ICLI 2016.

Keywords: Sensory substitution, seeing, hearing, vision, sound, depth, distance, colour, sonification, synaesthesia.

Introduction

Visual content permeates nearly all aspects of human experience, from picking up a cup of coffee to exploring the natural world. However visual information is not inextricably linked to the eyes. By using sensory substitution devices (SSDs) that convert visual dimensions into auditory ones, it becomes possible to 'hear vision.' Here we explore how sonifying visual information has been previously done to improve access to science, art and the wider visual world for blind and visually-impaired persons (BVIPs). Furthermore we discuss how the use of the psychological associations between vision and sound, as found in synaesthesia and the wider population can not only optimise the design of these devices but also allow access into experiencing a new form of synaesthesia. We present a new SSD called the Synaestheatre that hopes to achieve these aims and will be demonstrated as part of an installation at ICLI 2016.

A History of Sonifying Vision

Representing visual information through sound has a rich history. Galileo Galilei provides the first occurrence of this in 1586, who when needing to count the number of times a fine wire coiled around a bar for a scientific test, found the visual examination too difficult to count. Instead, Galileo ran a dagger down the bar and over the coiled wire, counting the auditory clicks that occurred. This way sonification provided access to the scientific data he required where vision fell short (Anstis, 2015). The sonification of scientific data remains to this day through providing access to graphs for the visually-impaired and in illustrating the temporal aspects of data where hearing provides a finer resolution than sight (Ebert, 2005; Guttman, Gilroy and Blake, 2005). Likewise, the sonification of paintings and performances, have also improved the accessibility of art for BVIPs (Cavaco et al., 2013a; 2013b; Mengucci, Medeiros and Amaral, 2014).

The sonification of sight itself has revealed fundamental processes on how our brains understand and process information as well as construct visual consciousness. In 1992, Peter Meijer designed the 'vOICe' (middle letters sounding out 'Oh I See!'), a device that typically converts one greyscale image into sound every second (Meijer, 1992). A single vertical column sweeps across the image, panning from the left ear to the right, turning the pixels under the column into sound. The vertical location of the pixels determines the pitches played (high locations play high pitched notes) while the brightness of the pixels determines their loudness (bright luminances play louder sounds). For an image such as this:  the user would hear a low pitched tone in the left ear, progress smoothly to a high pitched tone in the right ear. Users of such devices are able to locate and discriminate objects such as books and bottles from sound alone, furthermore, the processing of these auditory shapes takes place in regions of the brain previously thought to be 'sight-specific' for shape discrimination (Amedi et al., 2007). This new way of representing sight shows that the brain is not sense-specific but is able to extract meaningful data from potentially any sense and pass it on to the areas of the brain best suited to the task (Reich, Maidenbaum and Amedi, 2012). For blind users of such devices, the visual cortex becomes increasingly involved and necessary for the effective use of these devices (Merabet et al., 2009). Perhaps most stunningly is that long term users report conscious experiences of sight being generated by the auditory signals (Ward and Meijer, 2010). For those with previous experiences of sight this provides compelling evidence that the visual information provided through sound is enough for the brain to reconstruct a conscious visual world. Furthermore, while the SSDs provide some information, they often lack others, such as smooth motion, colour and depth. However over time, the brain is able to fill-in these gaps to create a smoother, more colourful 3D world (Ward and Meijer, 2010). Practically speaking, the mental construction of an accurate 3D world and identification of upcoming hazards is essential to safe and confident navigation of wider society for most BVIPs.

Designing Sensory Substitution Devices

When designing devices that convert visual information into sound, several factors have to be considered. The first is purely informational capacity; the sheer quantity of information of can be received and resolved by the senses. Overall the eyes can discriminate many more bits of information than the ears (Jacobson, 1950; 1951). As a result visual information has to be simplified in order to fit through the sensory 'bottleneck' of the ears. Which information is best to keep or exclude remains an open question, some devices preserve a high spatial resolution through feeding information in a piecemeal fashion, while others prioritise fast feedback loops between visual and auditory changes over spatial precision. Recently, colour information has also been explored as providing practical benefits to scene and object recognition beyond what basic increases in spatial resolution can provide (Hamilton-Fletcher and Ward, 2013).

Beyond simply transmitting the information, is there an optimal way of representing visual information in sound for the end user or is any consistent pattern comparable to any other? The first optimisation that can be taken is to utilise our natural hearing mechanisms when appropriate, for instance, the horizontal location of an object could be communicated by replicating the experience of hearing a sound emitted from that location in space. This is done by creating a 'head-related transfer function' that describes the relative timing and intensity differences received from a single sound source for the left and right ears. For resolving the vertical location of objects, while the shape of the ear does actually attenuate certain frequencies based on height, this vertical discrimination skill is impaired for BVIPs (Lewald, 2002). Instead the more abstract mapping of pitch-height is a common approach in SSDs. This draws upon a body of psychological literature that finds that humans intuitively relate high pitches to high spatial locations, a tendency that is present from infancy (Walker et al., 2010). These intuitive associations are called 'cross-modal correspondences' and describe a wide variety of mappings between the senses (Spence, 2011). Recently evidence has been provided that sound-colour correspondences in SSDs results in superior colour discrimination and memorisation abilities for users (Hamilton-Fletcher, Wright and Ward, 2015). For those with previous experiences of vision, this may help reduce the difficulty in understanding one sense through another.

Seeing with Synaesthesia?

A related condition to 'seeing with sound' is that of synaesthesia, a perceptual condition that affects approximately 4.4% of the population (Simner et al., 2006). For synaesthetes, stimulation in one sense creates an automatic, consistent and vivid experience in another sense (Simner, 2012), so that listening to music might create a cascade of colours that reflect the quality of sounds being heard. Interestingly, many of these synaesthetic mappings are found to be intuitively linked on an unconscious level in the wider population (Ward, Huckstep and Tsakanikos, 2006). As such, audiovisual media inspired by synaesthesia is rated as more aesthetically appealing by non-synaesthetes (Ward et al., 2008). By using these mappings in SSDs, not only are the intuitiveness and aesthetics of such devices likely to improve, but users could also experience a new synthetic synaesthesia. While there have been multiple attempts to train synaesthesia in non-synaesthetes, only studies with intensive training have reported any perceptual effects (Bor et al., 2014). SSDs are in a unique position to create effort-free synaesthetic experiences in day-to-day life.

Introducing the Synaestheatre: Practical Synaesthetic Sight

In pursuit of a device that can convert vision into sound in a way that is practical, intuitive, aesthetically pleasing and speaks to the wider cross-sensory experience of human psychology, we present a new SSD called the 'Synaestheatre.'

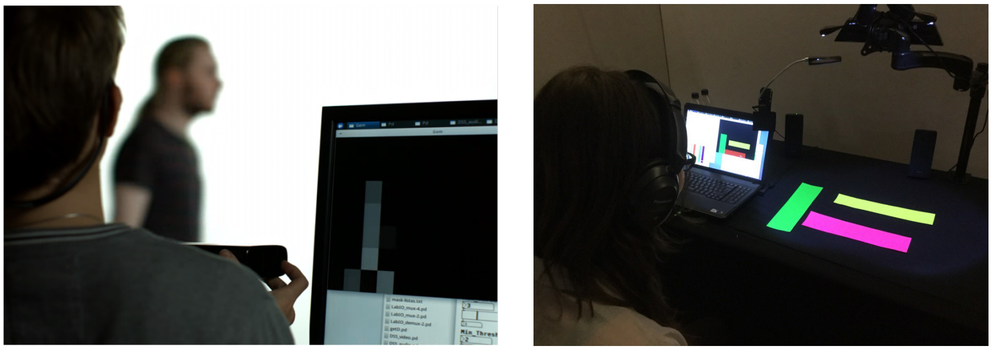

Figure 1. Demonstrations of the Synaestheatre by (left) a blind user detecting the motion of another person and (right) a sighted user listening to various colours. Users can add and move coloured objects while listening to changes in sound. Attendees will also have the option of experiencing both of these at ICLI 2016. See https://youtu.be/7t4NCse6_-w and https://vimeo.com/167031634.

This device turns the 3D location of colours into sounds using both natural hearing mechanisms and synaesthetic sound-colour associations in real time. In particular, a 13 by 7 grid of depth points are selected by a Kinect 3D sensing camera, with colour information in each of the 91 spatial locations categorised into one of seven colour categories (black, grey, white, red, green, blue and yellow). Each of these potential outcomes (91 depth points, 7 colours) has an associated sound that is played when there is the presence of an object in that location. The horizontal position of each pixel gives the sound its associated inter-ear timing and intensity differences. The vertical location specifies the sounds' pitch and the colour provides its timbral quality. The distance from the camera denotes each sound's loudness, with objects beyond the set maximum distance being silent. Furthermore, the device gives the end user a great deal of control over how the image is sonified. For example, pixels can either play independently from one another (mode 1), or can be time-locked together (mode 2). As a result mode 1 sounds similar to an orchestra testing their instruments prior to a concert, while mode 2 sounds like the orchestra playing in rhythmic unison. The timing of sounds can also be offset according to colour categories or spatial position to help users discriminate colour positions.

User Experiences of the Synaestheatre

Experiences with the Synaestheatre by blind users have been documented previously (Hamilton-Fletcher et al., 2016). The fast temporal resolution and precise localisation of sounds was deemed important as it required minimal effort to distinguish the position of objects and allows immediate detection of changes in an object's position in any direction. The instant feedback was appreciated from moving the camera and sensing the immediate change of sound. The vertical pitch mapping also seemed to be well received, with some describing the experience as "turning the room into music." The Synaestheatre was seen as being the most useful in navigating unknown environments in order to detect obstacles like tree branches or scaffolding sets. Experiences of alternative modes by visually-able users found that while mode 1 was deemed to be "pleasing and challenging," mode 2 was seen as more informative. One user suggested that social events could incorporate this through sonifying the location, movement and colour of guests.

Using the Synaestheatre as an Installation

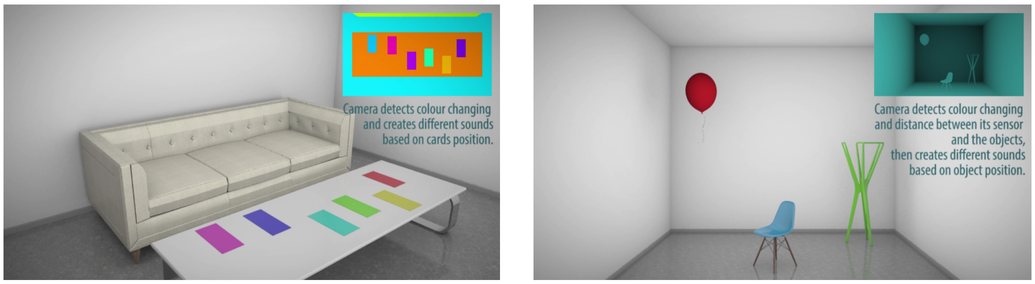

Figure 2. Synaestheatre installation, users can listen to sounds generated by various colours (left) or objects (right).

Attendees will be able to experience the Synaestheatre in a variety of typical use scenarios, from understanding the location of objects to how colour is translated into sound. Through using the 3D sensing camera as a surrogate pair of eyes and listening to the resulting sounds, attendees will be able to discriminate the location and movement of various coloured objects. We hope this installation will help us refine the Synaestheatre, open up discussions on how our senses interact and have attendees experience a unique fusion of sensory substitution and synaesthesia.

References:

Amedi, Amir, William M. Stern, Joan A. Camprodon, Felix Bermpohl, Lotfi Merabet, Stephen Rotman, Christopher Hemond, Peter Meijer, and Alvaro Pascual-Leone. "Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex." *Nature neuroscience* 10, no. 6 (2007): 687-689.

Anstis, Stuart. "Galileo’s dagger." *Perception* (2015): 0301006615594919.

Bor, Daniel, Nicolas Rothen, David J. Schwartzman, Stephanie Clayton, and Anil K. Seth. "Adults can be trained to acquire synesthetic experiences."*Scientific reports* 4 (2014).

Cavaco, Sofia, J. Tomás Henriques, Michele Mengucci, Nuno Correia, and Francisco Medeiros. "Color sonification for the visually impaired." *Procedia Technology* 9 (2013a): 1048-1057.

Cavaco, Sofia, Michele Mengucci, J. Tomas Henriques, Nuno Correia, and Francisco Medeiros. "From pixels to pitches: unveiling the world of color for the blind." In *Serious Games and Applications for Health (SeGAH), 2013 IEEE 2nd International Conference on*, pp. 1-8. IEEE, (2013b).

Ebert, Jessica. "Scientists with disabilities: Access all areas." *Nature* 435, no. 7042 (2005): 552-554.

Guttman, Sharon E., Lee A. Gilroy, and Randolph Blake. "Hearing what the eyes see auditory encoding of visual temporal sequences." *Psychological science* 16, no. 3 (2005): 228-235.

Hamilton-Fletcher, Giles, and Jamie Ward. "Representing colour through hearing and touch in sensory substitution devices." *Multisensory research* 26, no. 6 (2013): 503-532.

Hamilton-Fletcher, Giles, Thomas D. Wright, and Jamie Ward. "Cross-Modal Correspondences Enhance Performance on a Colour-to-Sound Sensory Substitution Device." *Multisensory research* 29, no.4-5 (2015): 337-363.

Hamilton-Fletcher, Giles, Marianna Obrist, Phil Watten, Michele Mengucci, and Jamie Ward. "I Always Wanted to See the Night Sky: Blind User Preferences for Sensory Substitution Devices." In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 2162-2174. ACM, 2016.

Jacobson, Homer. "The informational capacity of the human ear." *Science*112, no. 2901 (1950): 143-144.

Jacobson, Homer. "The informational capacity of the human eye." *Science*113, no. 2933 (1951): 292-293.

Lewald, Jörg. "Vertical sound localization in blind humans."*Neuropsychologia* 40, no. 12 (2002): 1868-1872.

Meijer, Peter BL. "An experimental system for auditory image representations." *Biomedical Engineering, IEEE Transactions on* 39, no. 2 (1992): 112-121.

Mengucci, Michele, Francisco Medeiros, and Miguel Amaral. "Image Sonification Application to Art and Performance." In *proceedings of INTER-FACE International Conference on Live Interfaces, Lisbon* (2014).

Merabet, Lotfi B., Lorella Battelli, Souzana Obretenova, Sara Maguire, Peter Meijer, and Alvaro Pascual-Leone. "Functional recruitment of visual cortex for sound encoded object identification in the blind." *Neuroreport* 20, no. 2 (2009): 132.

Reich, Lior, Shachar Maidenbaum, and Amir Amedi. "The brain as a flexible task machine: implications for visual rehabilitation using noninvasive vs. invasive approaches." *Current opinion in neurology* 25, no. 1 (2012): 86-95.

Simner, Julia. "Defining synaesthesia." *British journal of psychology* 103, no. 1 (2012): 1-15.

Simner, Julia, Catherine Mulvenna, Noam Sagiv, Elias Tsakanikos, Sarah A. Witherby, Christine Fraser, Kirsten Scott, and Jamie Ward. "Synaesthesia: The prevalence of atypical cross-modal experiences." *Perception* 35, no. 8 (2006): 1024-1033.

Walker, Peter, J. Gavin Bremner, Ursula Mason, Joanne Spring, Karen Mattock, Alan Slater, and Scott P. Johnson. "Preverbal Infants' Sensitivity to Synaesthetic Cross-Modality Correspondences." *Psychological Science*21, no. 1 (2010): 21-25.

Ward, Jamie, and Peter Meijer. "Visual experiences in the blind induced by an auditory sensory substitution device." *Consciousness and cognition* 19, no. 1 (2010): 492-500.

Ward, Jamie, Brett Huckstep, and Elias Tsakanikos. "Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all?." *Cortex* 42, no. 2 (2006): 264-280.

Ward, Jamie, Samantha Moore, Daisy Thompson-Lake, Shireen Salih, and Brianna Beck. "The aesthetic appeal of auditory-visual synaesthetic perceptions in people without synaesthesia." *Perception* 37, no. 8 (2008): 1285-1296.