

Sensory Substitution. Visual-to-auditory sensory substitution devices for the blind

Michael J. Proulx and Arne Harder

Introduction

Blind persons rely primarily on auditory and tactile information to recognize and localize objects. The ability to recognize objects and object features with tactile information alone is quite efficient (Lederman & Klatzky, 1993). The main limitation is spatial. How might the blind find distant, silent objects in an unfamiliar environment?

Neuroprosthetic devices are one means of providing visual information about objects (e.g., Loewenstein, Montezuma & Rizzo, 2004), and sensory substitution devices are another. Neuroprosthetic devices substitute for a damaged peripheral or central nervous system or supplement a damaged or healthy one by providing direct neural stimulation. Neuroprosthetic devices have a resolution that is limited by the number and size of the receptors, thus providing the user with just a subset of the information available. Sensory substitution devices for the blind aim to

supply the missing visual input by providing information that another sensory modality can process. One benefit of using a sensory substitution device rather than an invasive neuroprosthetic (besides the obvious non-invasiveness of it) is that the primary information-processing device is not a computer chip, but the brain. As Bach-y-Rita and colleagues (e.g., Bach-y-Rita & Kercel, 2003) have noted, the brain can better perceive the environment with a rich input rather than a paucity of input. Certainly some degree of information is lost as it is translated via sensory substitution, however potentially to a lesser degree than current invasive devices.

Bach-y-Rita and colleagues developed one of the first sensory substitution devices. It converted images from a video camera into tactile stimulation applied to the back of the subject via a modified chair (Bach-y-Rita, Collins, Saunders, White & Scadden, 1969). Black and white images were translated into a mechanical signal such that a pin would press up into the subject's back for a white pixel, but not for a black pixel. In this manner, visual pixels were conveyed directly into a format that another sensory modality could interpret in a visuospatial manner with practice. Since that time, many other devices have been designed to relay visual information to the tactile system via vibrations or electrical stimulation (for a review, see Bach-y-Rita & Kercel, 2003).

Unfortunately there are limitations to such devices. To display complex information with high fidelity, such systems require a sensitive region of the skin with a high density of mechanosensitive peripheral nerves, such as the tongue, which may cause discomfort or irritation (Bach-y-Rita, 1972).

Informatie auteurs:

Dr Michael J. Proulx, Lecturer in Cognitive Psychology, Queen Mary, University of London, United Kingdom

Dr Arne Harder, Mobility Consultant, Cologne, Germany

Correspondentieadres:

Michael J. Proulx, School of Biological & Chemical Sciences, Queen Mary, University of London, Mile End Road, London E1 4NS, United Kingdom, e-mail: m.proulx@qmul.ac.uk, phone: +44 20 7882 7484

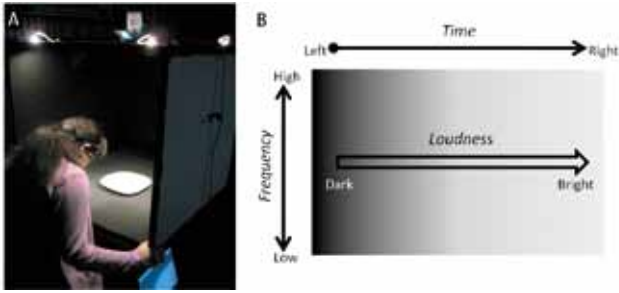


Figure 1.
 A) A blindfolded subject using a camera embedded in 'spy' sunglasses to inspect and attempt to identify a plate.
 B) An illustration of the visual-to-auditory conversion principles for The vOICE.

Meijer (1992) applied the principles developed by Bach-y-Rita to the auditory system with The vOICE. The vOICE is a computer program that requires a video input, a small computer (or even mobile phone) to run the program, and stereo headphones to play the audio output for the user (see Figure 1A). A visual snapshot is taken by the program (every 1 second, for example), scanned from left-to-right, and converted into sound which is then played for the user from left-to-right. Thus, although the visual image is taken of a particular spatial extent provided by the field of view of the camera, the images are also temporally static in that they are only provided once per second (though this rate can be changed). The auditory equivalent is necessarily dynamic and spread out over time to convey the spatial nature of the image. The vOICE maps visual images to sound via three primary dimensions (see Figure 1B): horizontal pixel location is coded by the time provided by the left-to-right scanning transformation of each image and by stereo panning; vertical location is coded by frequency, so that 'up' in the image is represented by high frequencies and 'down' by low frequencies; and pixel brightness is coded by loudness, such that a bright white pixel is heard at maximal volume, and a dark pixel is silent. Figure 2 displays a sample image and a spectrogram of the auditory output provided by The vOICE. Since Meijer's paper, other groups have devel-

oped similar visual-to-auditory substitution devices as well (e.g., Capelle, Trullemans, Arno & Veraart, 1998; Cronly-Dillon, Persaud & Gregory, 1999).

Most research using visual-to-auditory substitution devices has taken a proof-of-concept approach. Now it is time to go beyond proof of concept to describing the underlying principles that make visual-to-auditory substitution useful for particular tasks (see Proulx & Stoerig, 2006, for a review). Independent of the modality chosen to substitute for vision, there are many design considerations necessary to make such a device practical, affordable, ergonomic, and aesthetic; Meijer (1992) noted many of these. The focus of our research has been on the information processing demands of sensory substitution, and in particular on the nature of the attentional demands of using sensory substitution devices.

In normal vision, at any given moment more light is entering the retina from all areas of the visual field than a human can process. Attention is the mechanism used to prioritize the processing of relevant or salient information at the expense of momentarily irrelevant information (Proulx, 2007). The information provided by a sensory substitution device is also more than a human can process. The question is how the information output can be tailored to the attentional system (Parasuraman & Davies, 1984) of the user. This is a critical question because attentional selection in vision works differently than attentional selection in audition (see, e.g., Woods et al, 1998). It cannot be expected that a translation of the visual environment into the auditory domain will result in the same percepts as would normally be available for visual selective attention. Additional research must be carried out on how different aspects of attention (e.g., dividing attention across features or sources of information or selecting some information while ignoring other) may be optimized for using sensory substitution devices.

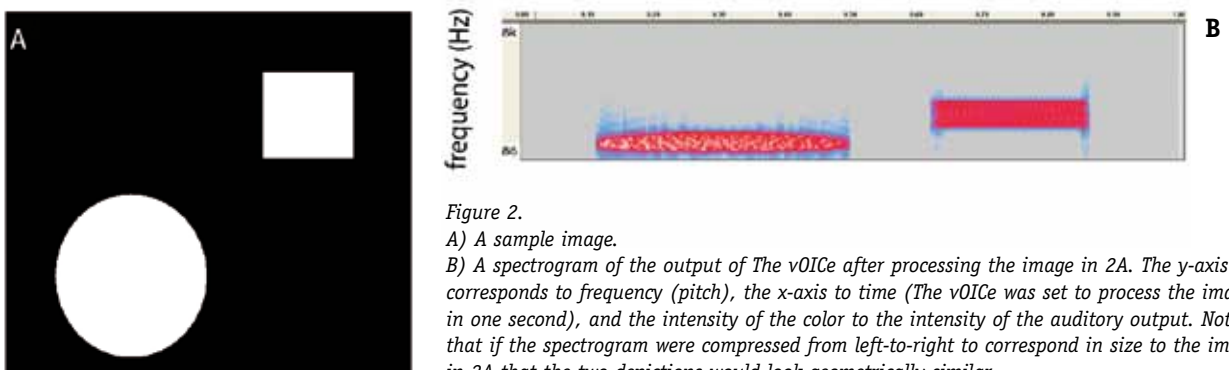


Figure 2.
 A) A sample image.
 B) A spectrogram of the output of The vOICE after processing the image in 2A. The y-axis corresponds to frequency (pitch), the x-axis to time (The vOICE was set to process the image in one second), and the intensity of the color to the intensity of the auditory output. Note that if the spectrogram were compressed from left-to-right to correspond in size to the image in 2A that the two depictions would look geometrically similar.

Divided attention is required for the use of a visual-to-auditory substitution device because information from the device must be attended to simultaneously with information from the environment. Although this is true for tactile devices as well, auditory information is particularly important for navigation by the blind. A blind user would need to monitor simultaneously both the output of a device such as The vOICe and the environmental sounds that would alert one to important objects and events, such as an oncoming car while crossing the street (Barlow, Bentzen, & Bond, 2005). For example, applying work on attentional capture (e.g., Dalton & Lavie, 2007) would be informative to account for such demands by tailoring the output of the device to match features that do not capture attention while allowing environmental auditory onsets to take priority. No research to date has examined the ability of sensory substitution users to attend to both the output of the device and to environmental sounds concurrently, and this is precisely the sort of human factors research that is crucial for the further development and usability of such devices for blind persons.

Selective attention is necessary as well. If a blind user is in an unfamiliar environment and needs to find a silent telephone, then the user of a substitution device would need to sample the environment and pay attention to particular features that would match a mental template of a telephone and guide search to spatial locations which have a higher likelihood of containing the target (see, e.g., Wolfe, 1994). The seeming ease with which a sighted observer can find a telephone in an unfamiliar room belies the difficulty of such a search task. Many phones are of different shapes, sizes, and colors, and no single template would match any given two telephones. Thus, much of visual search research (e.g., Treisman & Gelade, 1980; Wolfe, 1994) could be applied to this issue of performing 'visual' search with a visual-to-auditory substitution device. A better understanding of how a user is able to focus attention on possible targets (and features that would efficiently guide search to targets) in the midst of a distracting background is necessary for improving the design of such devices and for developing the training protocols necessary to learn to use the device effectively.

Sensory substitution devices hold the potential to assist the blind. Research on the attentional demands of sensory substitution will play an important role in improving the design and implementation of auditory substitution devices and may someday allow

blind persons to 'see' and find distant, silent objects through their ears.

References

- Bach-y-Rita, P. *Brain Mechanisms in Sensory Substitution*. New York: Academic Press. 1972.
- Bach-y-Rita, P., & Kercel, S. W. Sensory substitution and the human-machine interface. *Trends in Cognitive Sciences* 2003; 7: 541-546.
- Bach-y-Rita, P., Collins, C. C., Saunders, F. A., White, B. & Scadden, L. Vision substitution by tactile image projection. *Nature* 1969; 221: 963-964.
- Barlow, J. M., Bentzen, B. L., & Bond, T. Blind pedestrians and changing technology and geometry of signalized intersections: Safety, orientation, and independence. *Journal of Visual Impairment & Blindness* 2005; 99: 587-598.
- Dalton, P., & Lavie, N. Overriding auditory attentional capture. *Perception & Psychophysics* 2007; 69: 162-11.
- Capelle, C., Trullemans, C., Arno, P., & Veraart, C. A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution. *IEEE Transaction on Biomedical Engineering* 1998; 45: 1279-1293.
- Cronly-Dillon, J., Persaud, K., & Gregory R. P. F. The perception of visual images encoded in musical form: A study in cross modality information transfer. *Proceedings of the Royal Society of London Series B - Biological Sciences* 1999; 266: 2427-2433.
- Lederman, S. J., & Klatzky, R. L. Extracting object properties through haptic exploration. *Acta Psychologica* 1993; 84: 29-40.
- Loewenstein, O. V., Montezuma, S. R., & Rizzo, J. F., 3rd. Outer retinal degeneration: An electronic retinal prosthesis as a treatment strategy. *Archives of Ophthalmology* 2004; 122: 587-596.
- Meijer, P. B. L. An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering* 1992; 39: 112-121.
- Parasuraman, R. & Davies, D. R. *Varieties of attention*. Orlando: Academic Press. 1984.
- Proulx, M. J. Turning on the spotlight: Do attention and luminance contrast affect neuronal responses in the same way? *The Journal of Neuroscience* 2007; 27: 13043-13044.
- Proulx, M. J. & Stoerig, P. Seeing sounds and tingling tongues: Qualia in synaesthesia and sensory substitution. *Anthropology & Philosophy* 2006; 7: 135-151.
- Treisman, A., & Gelade, G. A feature-integration theory of attention. *Cognitive Psychology* 1980; 12: 97-136.
- Woods, D. L., Alain, C., & Ogawa, K. H. Conjoining auditory and visual features during high-rate serial presentation: Processing and conjoining two features can be faster than processing one. *Perception & Psychophysics* 1998; 60: 239-249.
- Wolfe, J. M. Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review* 1994; 1: 202-238.

Abstract

Sensory substitution devices for the blind aim to supply the missing visual input by providing information that another sensory modality can process. Here we focus specifically on a visual-to-auditory substitution device, but our review applies to the other modalities as well. We conclude that additional research on the roles of attentional processing is needed to improve the usability of such devices. 