Seeing Sounds? Explorations with the "vOICe" Visual-to-Auditory Substitution System

Position Paper by Dr Jamie Ward (jamiew@sussex.ac.uk)
Department of Psychology, University of Sussex, UK

Visual-to-auditory sensory substitution devices (or prostheses) convert visual images, extracted via a live camera rather than the eyes, into an auditory signal. This has the potential to enable blind and visually impaired users to access information about the visual world that would not be readily available to their intact senses of hearing and touch (e.g. objects out of reach). The minimum hardware requirements for a visual-to-auditory sensory substitution device consist of a camera to sample visual images, a portable computer (or similar) to run the software, and headphones to listen to the sounds. Some devices such as the vOICe (www.seeingwithsound.com Meijer, 1992) will run the software on a mobile phone utilising the inbuilt camera and speakers, and thus being portable and affordable (e.g. in the third world). This contrasts with, say, some tactile-to-visual sensory substitution devices (Bach-y-Rita, 2006) in which the hardware requirements can be more costly and cumbersome. The software itself needs to have some way of converting a visual image into a sound. One solution is to divide the visual image into an x-y array of pixels and convert the visual features into auditory ones such that height (y-axis) becomes pitch, brightness becomes loudness, and horizontal position (x-axis) becomes time (i.e. leftmost part of the image is heard first). Research using the vOICe has demonstrated that <u>sighted</u> participants: can learn to match a soundscape to a choice of 4 or 5 pictures with reasonable accuracy (around 70%, Amedi et al., 2007); and can locate objects on a table when blindfolded (Auvray et al., 2007). Although the signal from the device is purely auditory, the use of these devices in sighted people appears to recruit visual as well as auditory brain regions (see Poirier et al., 2007). However, different hardware and software configurations have not been directly compared. For example, the algorithm for converting images to sound could better exploit advances in image processing and knowledge of human auditory perception. There is also a significant lack of research into how blind people adapt to these devices.

My own research to date has been concerned with the following three questions.

<u>(1) What do potential blind users want out of a visual-to-auditory device?</u> In order for such a device to meet the needs of the visually impaired community, it is important to find out how many people would be interested in such a device; who would be interested (e.g. early v. late blind); under what circumstances it may be useful; and what particular aspects of the visual world it would be important to represent (e.g. everything, near objects, a selected area). A survey of a sample of 78 potential users has revealed the following…

- There is a significant amount of interest in the use of visual-to-auditory sensory substitution devices amongst the blind and visually impaired, even if this competes (to some degree) with their normal hearing
- This is equally true of people with early and late acquired visual impairments
- The main interest in the use of such a device was in outdoor settings, although there was a generally high interest (around 4 on a 1-7 scale) in all scenarios that we asked about
- There was particular interest in having the device under some form of manual control

<u>(2) How do different operating characteristics of the device affect objective performance?</u>
Following Auvray et al. (2007) we developed a number of ecologically valid tests that have been carried out on blindfolded participants to assess their ability to use different versions of the vOICe. Our results were as follows…

- Different sensori-motor modes of using the device were found to be effective in different situations. Specifically, a hand-held device was better than a head-mounted device for identifying an object on the table. The reverse was true for locating an object on the table.
- Performance was above chance without any training and was maintained at the same level throughout. However, time taken was halved.
- Performance tends to be higher in participants with higher self-reported visual imagery

(3) What are the subjective experiences of expert users of the device?

We have conducted detailed interviews with two blind expert users of the VOICe device, and both report subjective visual sensations. The following characteristics are noteworthy…

- Both participants report that several months of immersive training is necessary to gain mastery
- Different subjective experiences of vision come into effect at different times. For example, edge and shape detection emerges before depth perception; colour perception is typically absent but is now reported by one of our participants after several years of immersion
- There is evidence that active exploration is important: the camera needs to be anchored on a body part, and sonified still images are virtually impossible for them to recognize
- Some other sounds (e.g. a reversing lorry) may trigger visual experiences when the device is not worn; i.e. the rules for mapping between sound and vision have been internalized as a result of using the device and are applied when the device is not worn (even though the mappings are functionally redundant here)

References

Amedi, A., Stern, W., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., Hemond, C., Meijer, P., & Pascual-Leone, A. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nature Neuroscience, 10*, 687-689.

Auvray, M., Hanneton, S., & O'Regan, J. K. (2007). Learning to perceive with a visuo-auditory substitution system: Localisation and object recognition with 'The vOICe'. *Perception, 36*, 416-430.

Bach-y-Rita, P. (2006). Tactile sensory substitution studies. *Annals of the New York Academy of Sciences, 1013*, 83-91.

Meijer, P. B. L. (1992). An Experimental System for Auditory Image Representations. *IEEE Transactions on Biomedical Engineering, 39*, 112-121.

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