

## Perceiving occlusion through auditory–visual substitution

Alessandra Jacomuzzi • Nicola Bruno

**Keywords** Sensory substitution • Prosthetic systems • Blindness • Occlusions

### Introduction

The development of visual prostheses for the blind has a long history. Presumably, the use of tools such as a long cane for haptic exploration beyond peripersonal space was already common in antiquity. Seeing-eye dogs may also be considered as a special kind of (animate) prosthesis. Modern prosthetic visual devices aim at conveying useful information about the environment to a blind individual. They do not try to induce the conscious, qualitative experience of seeing (a goal that may well prove impossible) or to provide the full complexity of visual information. Rather, they aim at offering sensory support to enhance a blind individual's mobility in non-familiar environments. We find that modern prostheses divide into two broad categories: invasive (artificial retinas or direct cortical stimulation systems), and non-invasive (sensory substitution systems). According to preliminary observations, blind individuals can learn to use invasive prostheses to perform visual tasks surprisingly fast. However, this type of prosthetic device also has obvious disadvantages. It requires expensive neurosurgery and hardware, and there are strong suspects that it may not work at all with early blind patients.

Non-invasive prostheses, conversely, do not require surgery. In addition, there are reasons to think that the sensory substitution approach may work with early blind patients. Prosthetic systems based on sensory substitution have been studied for almost 40 years. They are based on the idea of substituting visual stimulation with stimulation from another, intact sensory system such as the tactile system (Bach-y-Rita 1968) or the auditory system (Cronlly-Dillon and Persaud 2000;

Mejier 1992). Although such systems have been studied for such a long time, they are still poorly understood. In this study we investigated a visual-auditory substitution system developed by Meijer. The vOICE is an inexpensive auditory–visual substitution system that can in principle convert any image into sounds. Due to the specific limitations of the employed mapping of visual to auditory structure, it is currently unknown under what conditions the system could provide useful information and what kind of training would be required to achieve proficiency in specific domains. The vOICE runs on a standard PC computer (running Windows 98, 2000 or XP at 1 GHz) and uses an ordinary webcam for inputting images to the program, which converts them according to a straightforward rule: images are scanned from left to right, horizontal positions are converted to temporal position within 1s sound. Vertical pixel position is converted to frequency (lower frequencies meaning lower positions). Pixel intensity is converted to loudness. Thus each scan generates a complex sound signal having a regular mapping of sound features to image features.

In this study we tested learning spatial task using The vOICE. We made four experiments. In the first experiment we tested learning localization, in the second experiment we tested learning color discrimination, in the third experiment we tested learning orientation discrimination and finally in the fourth experiment we tested learning interpretation of occlusion.

### Methods

Three students of the University of Trieste took part in the experiment. One of them was the first author of this paper. Before beginning the experiment, we explained to subjects the rules used by The vOICE to convert images in sounds. Subjects wore a blindfold, and they used sounds produced by The vOICE to guide pointing responses based on various spatial properties.



*Experiment 1* In each trial, participants heard sounds corresponding to one of nine visual rectangles at different positions on the screen. Blindfolded subjects pointed on the screen where he thinks the rectangle is. Each session consisted of 100 trials. A touch monitor recorded each answer, giving feedback to the subject. Each subject made one or more session every day. The experiment ended when the subject gave more than 90% correct answers in at least three sessions.

*Experiment 2* In each trial, participants heard sounds corresponding to two visual rectangles at different position, one black and one white. Each of blindfolded subjects pointed on the screen where she/he thinks the white rectangle is. Each session consisted of 100 trails. A touch monitor recorded each answer, giving feedback to the subject. Each subject made one or more session every day. The experiment ended when the subject gave more than 90% correct answers in at least three sessions.

*Experiment 3* In each trial, participant heard sounds corresponding to a horizontal or a vertical bar. Each of blindfolded subjects pointed twice on the screen in the starting and final position of the bar making a horizontal or a vertical movement with his/her finger. Each session consists of 100 trails. A touch monitor recorded each answer, giving feedback to the subject. Each subject made one or more session every day. The experiment ended when subject gave more than 90% correct answers in at least three sessions.

*Experiment 4* In each trial, participant heard sounds corresponding to one of eight types of occlusions at different positions on the screen. These occlusions have a horizontal and a vertical bar with different colors partially occluded. Each of blindfolded subjects pointed twice on the screen in the starting and final position of the occluded bar making a horizontal or a vertical movement with his/her finger. Each session consists of 100 trails. A touch monitor recorded each answer, giving feedback to the subject. Each subject makes one or more session every day. The experiment ended when subject gave more than 90% correct answers in at least three sessions.

## Results

*Experiment 1* All subjects were able, even if at different times (AJ in 25 days, MG in 21 days and AM in 31 days), to reach the requested level of learning. All subjects presented the same trend, which was roughly correspondent to a negatively accelerated curve. In the first 12 days learning was very fast. After this period it became slower, and it levelled off near 100% accuracy.

*Experiment 2* All subjects were able, even if at different times (AJ in 22 days, MG in 14 days and AM in 31 days), to reach the requested level of learning. Moreover, subjects presented different trend. Error analysis reveals that two subjects accomplish localization and errors discrimination tasks, but one subject rarely points the right position of the black rectangle. He does only localization errors.

*Experiment 3* The task is very simple for all subjects. In the first session they respond correctly at many trials: AJ responds correctly at 85 trials, MG responds correctly at 93 trials and AM responds correctly at 69 trials. Moreover, all subjects ended the experiment giving the requested level of learning very quickly: AJ in 4 days, MG in 2 days and AM in 4 days.

*Experiment 4* Subjects show a similar trend. In the first 10 days accuracy increases very fast but after this period subjects performances cannot improve. After 40 sessions each subject is able to answer correctly to more than 75 trials. Errors analysis show that subjects have difficulties in the interpretation of two types of occlusions: occlusions where the occluding bar is the vertical bar in right side of figure and occlusions where the occluding bar is the horizontal bar in the bottom part of figure and the occluded bar is the vertical bar in the right side.

## Discussion

Results from the first three experiments suggest that it is possible to learn simple spatial tasks using auditory–visual sensory substitution. Nevertheless, it seems that the system had difficulty to convert visual occlusions in sounds. Even after extensive sessions, none of the subject proved capable of reaching accuracy levels above 75%. It is not clear, at present, why our subjects failed to learn occlusions in our studies. Given that the problems were limited to three specific patterns out the eight studied, independently of pattern position, it seems unlikely that the problem reflects a specific difficulty in learning or interpreting occlusions as rendered by auditory substitution. We speculate that the problem may be related to the mechanics of the conversion rules used by The vOICE to convert images in sounds. The vOICE uses frequencies from 500 to 5,000 Hz. The scan of images goes on for 1,000 ms and in our experiment the scan of occlusions goes on for 660 ms. We think that the problems in the interpretation of these two types of occlusion are due to an effect of masking. For the first type of occlusion, probably, there is a temporal masking effect. When the intensity of the horizontal bar is very strong it masks for 200 ms the following sound. In this case the conversion produces a very confused sound. In



the second type of occlusion, probably, there are both temporary and frequency masking effects that causes the component sounds corresponding to the occlusion patterns to become hard to discern. If this hypothesis is correct, this suggests that occlusion patterns can in principle be understood through auditory substitution even if greater attention needs to be paid to psychoacoustical phenomena arising during temporal interactions at certain frequencies.

### References

- Arno P, De Volder AG et al (2001) Occipital activation by pattern recognition in the early blind using auditory substitution for vision. *Neuroimage* 13(4):632–45
- Bach-y-Rita P (1968) *Brain mechanisms in sensory substitution*. Academic, New York
- Bach-y-Rita P, Tyler ME, Kaczmarek KA (2003) Seeing with the brain. *Int J Hum Comput Interact* 15(2):285–295
- Chow AY, Chow VY, Packo KH, Pollack MD, Peyman GA, Schulard R (2004) The Artificial silicon retina microchip for the treatment of vision loss from retinitis pigmentosa. *Arch Ophthalmol* 122(4):460–469
- Cronly-Dillon J, Persaud KC et al (2000) Blind subjects construct conscious mental images of visual scenes encoded in musical form. *Proc R Soc Lond B Biol Sci* 267(1458):2231–2238
- Dobelle WH (2000) Artificial vision for the blind by connecting a television camera to the visual cortex. *ASAIO J* 46(1):3–9
- Humayun MS, Weiland JD et al (2003) Visual perception in a blind subject with a chronic microelectronic retinal prosthesis. *Vision Res* 43(24):2573–2581
- Meijer P (1992) An experimental system for auditory image representations. *IEEE Trans Biomed Eng* 39:112–121

