

The Vibe: a versatile vision-to-audition sensory substitution device

Sylvain Hanneton^{a*}, Malika Auvray^b and Barthélemy Durette^c

^aLaboratoire de Neurophysique et Physiologie UMR CNRS 8119 Université Paris Descartes, Paris, France; ^bLIMSI-CNRS (CPU team), Bâtiments 508 et 502bis, Université de Paris-Sud, 91403 ORSAY (France); ^cJustus-Liebig-Universität Gießen, Fachbereich 06, Psychologie und Sportwissenschaft, Abteilung Allgemeine Psychologie, Otto-Behaghel-Strasse 10F, 35394 Giessen

(Received 29 January 2010; final version received 29 July 2010)

We describe a sensory substitution scheme that converts a video stream into an audio stream in real-time. It was initially developed as a research tool for studying human ability to learn new ways of perceiving the world: the Vibe can give us the ability to learn a kind of ‘vision’ by audition. It converts a video stream into a continuous stereophonic audio signal that conveys information coded from the video stream. The conversion from the video stream to the audio stream uses a kind of retina with receptive fields. Each receptive field controls a sound source and the user listens to a sound that is a mixture of all these sound sources. Compared to other existing vision-to-audition sensory substitution devices, the Vibe is highly versatile in particular because it uses a set of configurable units working in parallel. In order to demonstrate the validity and interest of this method of vision to audition conversion, we give the results of an experiment involving a pointing task to targets memorised through visual perception or through their auditory conversion by the Vibe. This article is also an opportunity to precisely draw the general specifications of this scheme in order to prepare its implementation on an autonomous/mobile hardware.

Keywords: perception; learning; sensory substitution; audition; pointing

1. Introduction

Sensory substitution devices aim at assisting one or several functions of a deficient sensory modality through another sensory modality. To do so, these systems convert the stimuli normally accessed through one sensory modality (e.g., light when the aim of the device is to compensate for vision) into stimuli accessible to another sensory modality (e.g., sounds or tactile stimuli). The first aim of these sensory substitution devices is to compensate for a lost sense. They are thus part of a rehabilitation context whose aim is to help blind persons to ambulate, localise themselves and recognise objects. In addition, sensory substitution devices provide a unique opportunity to investigate fundamental questions concerning spatial cognition such as how blind persons reconstruct a perceptual space via a new sensory modality or, more generally, how perceptual processes are modified by the use of new tools.

Since their inception in the 60s, various kinds of devices have been developed and tested. Several studies demonstrated that their users behave to some degree as if they possessed some visual abilities (see Auvray et al. 2009 for a review). In particular, two main categories of systems have been designed to compensate for the loss of vision: vision-to-touch substitution devices that convert pictures into tactile stimuli and vision-to-audition substitution devices that convert images into sounds.

1.1. Vision-to-touch substitution devices

In most vision-to-touch substitution devices, pictures captured by a camera are translated into electrical or vibratory stimulation applied to the skin of a part of the body, such as the abdomen, the back, the fingertip, the forehead or the tongue. Research has shown that users can make use of vision-to-touch substitution systems in tasks that would normally require vision, provided that they are granted active control over the camera (e.g., Bach-y-Rita 2002). In particular, experimental studies reported the ability of participants to succeed in localisation tasks (Jansson 1983), simple form recognition (Sampaio et al. 2001; Kaczmarek and Haase 2003) and reading (Loomis 1974; Craig 1981). Some studies have also shown that users of vision-to-touch substitution devices can perform perceptual judgements using perspective, parallax, looming and zooming, as well as depth estimates (Bach-y-Rita 1969). However, it should be mentioned that although various perceptual tasks can be accomplished with vision-to-touch substitution devices, these systems are faced with certain limitations as they depend on the stimulation of a highly sensitive skin surface (such as the tongue for instance) leading to problems of skin irritation or pain. Furthermore, the autonomy of portable versions of these devices is limited due to the substantive energy consumption of tactile stimulators (Lenay et al. 2003).

*Corresponding author. Email: sylvain.hanneton@parisdescartes.fr

1.2. Vision-to-audition substitution devices

The other important field of research that aims at finding compensations for the loss of vision is auditory substitution. The use of audition offers many advantages. First, the auditory system is able to deal with complex and rapidly changing sound patterns such as speech even in a noisy environment (Hirsh 1998). In addition, portable versions of vision-to-audition devices require only simple interfaces: headphones, a webcam and a computer (Capelle et al. 1998). Finally, the technology of digital sound processing is very common and the generation of auditory stimuli requires little energy. In vision-to-audition substitution devices, images captured by a camera are converted into sound and transmitted to users via headphones. The three main vision-to-audition conversion systems that have been designed so far are the Voice developed by Meijer (1992), the PSVA (Prosthesis for Substitution of Vision by Audition) developed by Capelle and his colleagues (Capelle et al. 1998), and the device developed by Cronly-Dillon and his colleagues (Cronly-Dillon et al. 1999, 2000). These three systems convert the vertical position of the registered pixels in the video image into different audio frequencies, with high-pitched sounds corresponding to upper locations and low-pitched sounds corresponding to lower locations in the image. These systems convert the luminosity of the pixels into sound amplitude. With regard to the encoding of the horizontal position of objects, the Voice and the device developed by Cronly-Dillon use left-to-right time scanning to encode horizontal position and the PSVA uses a frequency mapping in addition to the one used for vertical position. More specifically, the frequency associated to each pixel increases from left to right and from bottom to top. Furthermore, in order to enhance the similarity with the human visual system, the receptor field of the PSVA has a higher resolution in the centre of the picture. Studies that have been carried out with vision-to-audition devices have shown that it is possible to localise (Renier et al. 2005; Auvray et al. 2007) and recognise objects (Arno et al. 1999; Cronly-Dillon et al. 1999, 2000; Arno et al. 2001; Pollok et al. 2005; Auvray et al. 2007) with these systems. Interestingly, studies have also demonstrated the possibility of recreating visual illusions with the PSVA (Renier et al. 2005).

1.3. The need for versatility

All the previously described devices have their own qualities and faults and the choice between auditory or tactile systems can be debated. Tactile systems are for instance a good choice for the design of prostheses for blind people since they already use audition to compensate for the absence of vision. However, if the aim is to develop a research tool in order to optimise the characteristics of a prosthesis and the conditions of learning then the auditory system is the best alternative. It requires less power, is cheaper and

can be easily miniaturised to be proposed to their users. But, above all, a great number of video-to-sound coding methods can be tested without the need to physically change or modify the stimulators. The Vibe intends to make the testing of a large panel of methods to obtain sensory substitution easier in a research purpose. There is an infinite number of ways to conceive a video-to-sound mapping. Our proposal does not cover all the conceivable solutions but allows a range from very simple to reasonably complex solutions to be tested. The Vibe allows the testing of mappings that involve something like a set of dynamic resonators continuously fed with a kind of non-regular spatio-temporal convolution of the video input.

2. Methods

2.1. General principle

The Vibe device converts a video stream into a stereophonic sound stream. Basically, it only requires a standard webcam and a standard computer (see Auvray et al. 2005 for an example) but a higher level of performance could be reached with VLSI autonomous implementation. The Vibe uses a kind of virtual retina that has two levels of ‘cells’: sensors and receptors. Each sensor corresponds to a particular pixel. The activity of a sensor is a function (a weighted combination, for instance the mean) of the coded components of the captured pixel (usually RGB or YUV components). A receptor has a receptive field that is a set of sensors (usually a set of neighbouring sensors). Consequently, a receptor is concerned by a particular area of the captured video frames. These areas can overlap. The activity of the receptor is a dynamic function of the activities of the sensors of its receptive field. The characteristics of the transfer function of each receptor are parameterised by the position of its sensors in the two-dimensional (2D) plane of the captured picture. Each receptor produces a signal that can be interpreted as a sound. The signals of all the receptors are mixed together to produce a stereo audio output and the transfer function of the receptor can be adapted to spatialise the audio output adequately for human perception (see figure 1).

2.2. From receptor activity to sound: receptor transfer function

Each pixel of the captured frame corresponds to a vector of measures: one single real value for a grey-level picture or usually three real values for a colour picture (RGB or YUV components for instance). If the retina has n receptors, we define $S_i^j(t)$ the i th sensor ($i \in [1, k_j]$) of the j th receptor ($j \in [1, n]$) as a vector of time-varying real values the dimension d of which depends on the coding of the captured picture. The set of sensors of a receptor is called its ‘receptive field’ (RF). The inputs of a receptor $R_j(t)$ can be

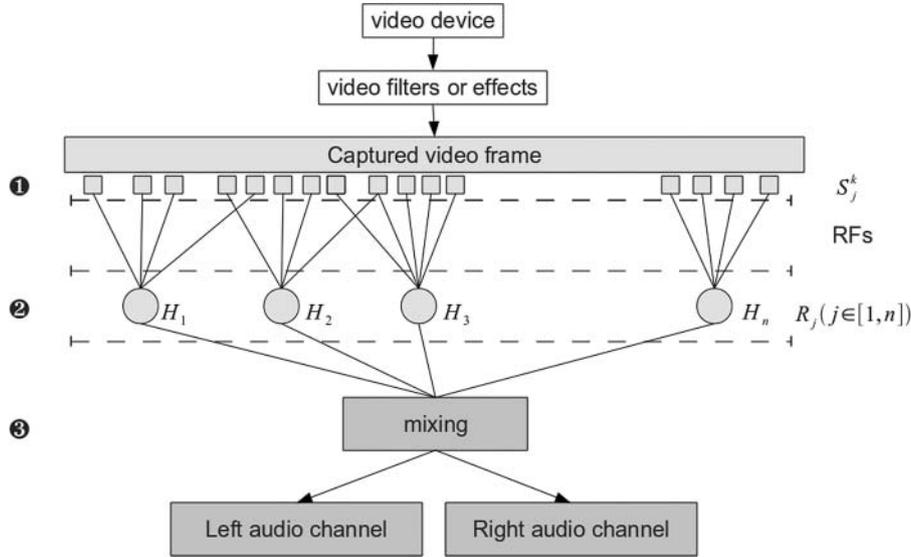


Figure 1. General architecture of the Vibe. Sensors shown in the first layer are grouped into receptive fields. The outputs of sensors are integrated by the receptor layer (2) and each receptor ($R_j, j \in [1, n]$) output is a stereo signal resulting from the application of a transfer function (H_j). All the stereo signals are then mixed to obtain the final stereo output.

written as a time-varying $[k_j, d]$ matrix $M_j(t)$ corresponding to the concatenation of the $S_i^j(t) (i \in [1, k_j], j \in [1, n])$. For instance, if $X_j = [R_j, \dot{R}_j]$ is the state-space representation of the 2D receptor output R_j , then the dynamics of the receptor can be defined in the linear case by the following formula:

$$\dot{X}_j = A_j \cdot X_j + F_j(M_j(t)),$$

where A_j is a matrix of scalar values and F_j is a function that reduces the dimensionality of the inputs from the sensors (for instance, a simple weighted sum of the outputs of sensors). If F_j is linear, this formula can be rewritten in the Laplace formulation $X_j(s) = H_j(s) \cdot M_j(s)$, where $H_j(s)$ is the transfer function of the receptor. The output of the receptor is a two-dimensional (stereo) time-varying signal corresponding to the application of the receptor-specific transfer function H_j to the input $M_j(t)$ (see Figure 2). More concretely, since the receptors are supposed to produce audio signals, the H_j function has to act as a kind of sound source or resonator, parametrised by the position of the sensors of the receptive field in the picture. The outputs of all the receptors are simply summed to produce a global stereo output $W(t)$:

$$W(t) = [W_R(t), W_L(t)] = \frac{1}{n} \sum_{j=1}^{j=n} R_j(t).$$

The role of the transfer function is to produce an audible waveform driven by the state of the receptive field. Each receptor can be considered as a stereophonic instrument played by the sensors of the RF. There are a great num-

ber of eligible transfer functions. For instance, the transfer function can make the receptor act as a damped oscillator (second-order harmonic equation):

$$\ddot{R}_j = F(M_j) - \beta_j \cdot \dot{R}_j,$$

where β_j is a damping factor, F is a $[R^{k_j \cdot d} \rightarrow R^2]$ function applied to the outputs of the sensors (the dot and double dot superscripts indicate single and double time derivatives, respectively). In this case, F and β_j have to be carefully chosen in order to obtain an audible signal. But the simplest case and the only one we have really implemented is the case where the receptors produce periodic signals with an amplitude controlled by the states of the sensors of each RF. We called this application the sinusoidal static receptors scheme.

2.3. Application: sinusoidal static receptors

The more straightforward application of the principle described in the preceding section is to have receptors that produce sinusoidal static outputs. Suppose that the frames of the video stream are coded in terms of levels of grey. The outputs $S_i^j(t) (i \in [1, k_j], j \in [1, n])$ of the sensors are mono-dimensional signals. A simple way to make the receptors to produce audio signals is to associate to each receptor $R_j(t)$ a simple sinusoidal waveform defined by its pulsation ω_j , phase ϕ_j and panoramic¹ $p_j (p_j \in [0, 1])$.

¹The panoramic is an audio parameter that controls the amount of a monophonic waveform attributed to the left and right channels to produce a stereophonic effect.

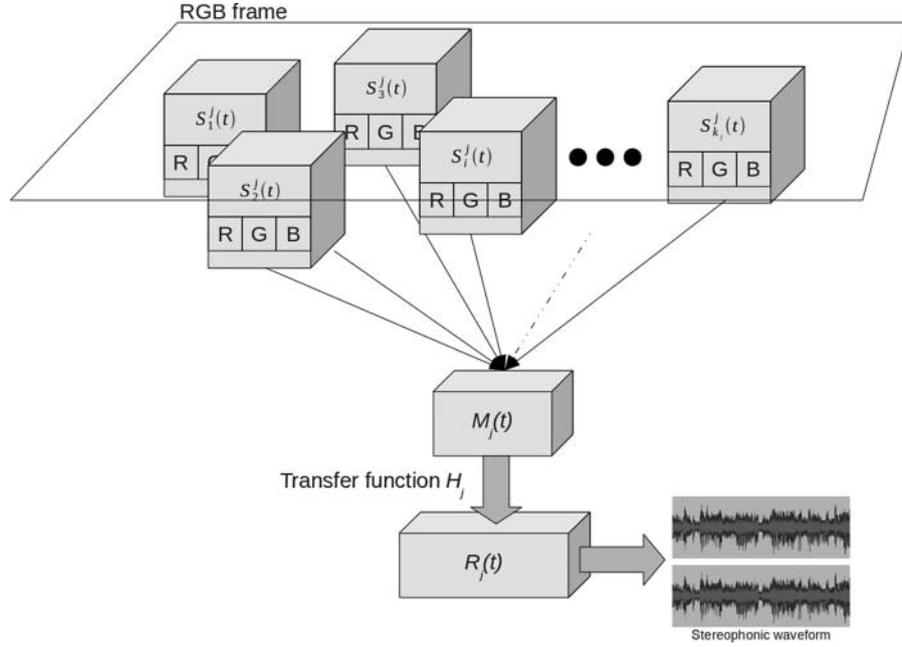


Figure 2. Description of the j th receptor of the Vibe for a RGB-encoded video frame. The group of sensors of one receptor constitutes its receptive field. The RGB outputs $S_i^j(t)$ ($i \in [1, k_j]$, $j \in [1, n]$) of the sensors are grouped in a matrix $M_j(t)$ and then integrated by the transfer function H_j of the receptor R_j to produce a stereophonic signal.

These characteristics of the waveform are constant for a particular receptor R_j but the amplitude of the output is driven by the mean level of grey $\bar{S}_j(t) = \frac{1}{k_j} \sum_{i=1}^{k_j} S_i^j(t)$ over the sensors associated with the considered receptor. The waveforms $R_j^R(t)$ (right ear) and $R_j^L(t)$ (left ear) produced by the j th receptive field are

$$R_j^R(t) = p_j \bar{S}_j(t) \sin(\omega_j t + \phi_j^R),$$

$$R_j^L(t) = (1 - p_j) \bar{S}_j(t) \sin(\omega_j t + \phi_j^L).$$

The sounds $W_R(t)$ and left $W_L(t)$ stimulating the right and left ears, respectively, are the sum of all the sounds produced by the n receptors:

$$W_R(t) = \frac{1}{n} \sum_{j=1}^{j=n} R_j^R(t),$$

$$W_L(t) = \frac{1}{n} \sum_{j=1}^{j=n} R_j^L(t).$$

In order to create a simple mapping between the 2D space of the captured pictures and the sound perceived by the listener, the parameters of the sine waveform can be computed from the positions of the sensors of the receptive field of the corresponding receptors. For instance, the pulsation ω_j (pitch) and the panoramic p_j can be linearly

related to the coordinates of the barycentre of the coordinates of sensors. The phases ϕ_j^R and ϕ_j^L are generally the same (except for the addition of inter-aural cues, see below) and can be randomly chosen between 0 and 2π in order to enhance the capacity of the listener to discriminate between sound sources.

The ‘sinusoidal static receptors’ scheme makes the device produce a continuous sound that conveys a great number of information related to the structure of the visual environment. However, this scheme is only a specific application of the general scheme and one can imagine developing other solutions for instance in order to enhance the ergonomics of the device, its efficiency in certain conditions, or to make it more adaptable to the characteristics of human auditory perception. In the framework of fundamental research on perceptual plasticity, it is essential to have a versatile device that allows the testing of various protocols.

3. A validation experiment: pointing with the Vibe

3.1. Accuracy of the device and influence of handedness

The main objective of this simple experiment is to show that human participants can use our device for the localisation of objects and to have an estimate of the accuracy of the localisation. However, we also propose to address the question

of the influence of the handedness on the localisation of the perceived targets. Right-handed participants hold the device in one hand in order to explore the environment and to perceive the target. They point at the estimated location of the target with the other hand. This localisation can be considered as enactive knowledge since it is gained through perception–action interactions with the environment. Handedness can influence the accuracy of pointing: should the device be held in the right or left hand? There are two alternative hypotheses. (1) Participants are more accurate with the device in the left hand because pointing movements are more skilful with the dominant hand. (2) Participants are more accurate with the device in the right hand because exploratory movements (perceptive movements) are more precisely controlled with the right hand.

According to the second assumption, the dominant hand has to be used for a fine control of perceptive movements rather than for pointing movements. Consequently, we expect better performance when the camera is held in the dominant hand. In an other context, right-handed rifle shooters with a dominant left eye were shown to be more accurate if they held the rifle in the left arm (Porac and Coren 1981). If the dominant eye has more influence than the dominant hand, then better performance should be expected when the camera is held in the non-dominant hand.

3.2. Methods

We implemented the scheme with a standard laptop and a standard webcam (Phillips TouCam Pro) in grey-level 320×240 capture mode. The software was developed in C with the SDL open source API for the control of the sound. It was produced by 20 receptors with receptive fields that used 10 sensors. The sampling frequency of the sound was 44.1 kHz.

Results presented here concerned three young right-handed females. Participants were instructed to point at targets perceived and memorised either visually (vision experimental condition, VEC) or via the Vibe (prosthesis experimental condition, PEC). In the VEC condition, participants were asked to observe the target for three seconds, to close the eyes and to point immediately at the estimated position of the target with either the left or right index finger. In the PEC condition, participants were blindfolded,

wore closed headphones, and held the webcam in the right or left hand with an imposed grasping posture. The elbow of the arm that held the webcam had to be kept at a specific location on the table. Participants had 15 seconds to explore the environment and then pointed at the estimated target location. Each participant performed 45 trials (3 target positions × 15 repetitions) in four experimental conditions (VEC or PEC and pointing with the left or the right hand).

We investigated the influence of experimental conditions on distance to target, on constant and variables pointing errors and on confidence ellipses for each target and each experimental condition. Considering confidence ellipses (a linear local estimates of error distributions) can give access to perception distortion induced by defaults of perceptive organs, memory or actuators and to the structure of the spatial representation of targets (McIntyre et al. 1998).

3.3. Results

The mean distances from the tip of the index to the target with the device were less than two times higher than in the condition involving vision. This fact demonstrates that participants were able, with a short period of practice, to learn to localise the targets quite correctly with the device. The mean distance to target was obviously lower in the VEC condition than in the PEC condition (see Table 1). In the PEC condition, the mean distance to target was lower when pointing with the left hand (webcam in the right hand) than in the opposite condition. The variance of the distance to target was systematically lower in this condition. Characteristics of the confidence ellipses were very variable across participants, targets and experimental conditions. However we have to emphasise that in the PEC condition, participants were more precise in seven cases out of nine (3 targets × 3 participants) when holding the webcam in the right hand. The inversion of the handedness is particularly clear for the second participant (see Figure 3b). However, because of the small number of participants, we have to emphasise that these results are very preliminary and have to be verified with complementary studies involving a larger number of participants.

Table 1. Mean pointing errors (and standard deviations) with either the right (RH) or left (LH) hand at visually (VEC) memorised targets or targets localised with the device (PE).

Condition (pointing with)	RH VEC	LH VEC	RH PEC	LH PEC
Distance to target (cm)	3.87 ± 1.81	3.99 ± 2.17	6.48 ± 5.00	6.06 ± 3.1
Absolute distance error (cm)	1.86 ± 1.41	2.41 ± 1.78	3.83 ± 3.88	4.06 ± 2.7
Absolute direction error (deg)	3.51 ± 2.25	3.25 ± 2.35	5.07 ± 4.47	4.22 ± 3.1

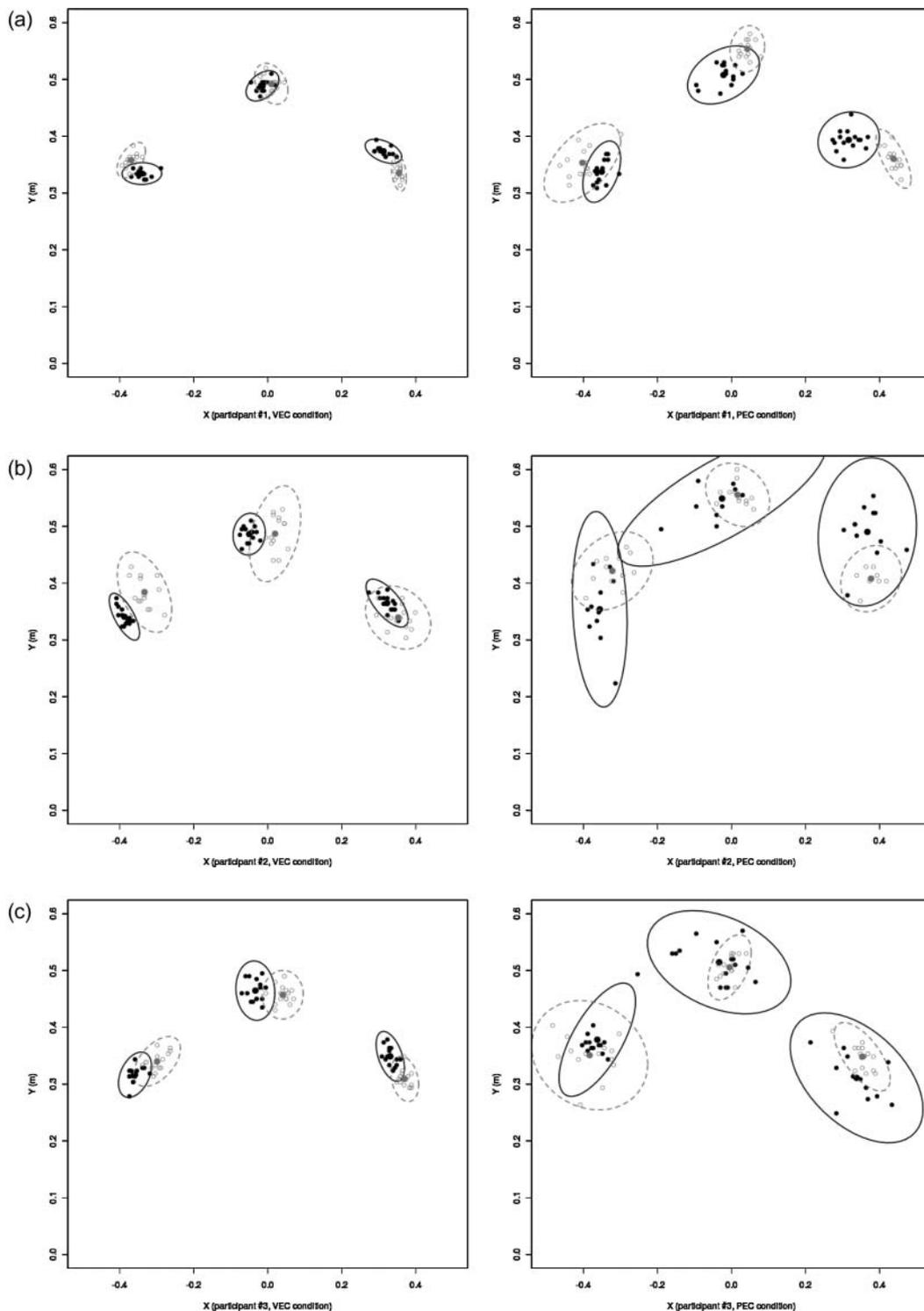


Figure 3. Locations of index finger tip and confidence ellipse of three participants (a–c rows) for visually memorised targets (left) and targets localised with the sensory-substitution device (right). Solid ellipses correspond to pointings with the right (dominant) hand.

4. Conclusion

4.1. Experimental validation

The Vibe has previously been used and tested in experimental studies in order to investigate the emergence of

the feeling of an auditory distal perception (Auvray et al. 2005) and navigation with a vision-to-audition sensory substitution device (Durette 2009). But the experimental part of this study also allows us to demonstrate that the

proposed retina-like scheme for the Vibe can also effectively elicit perceptual abilities for pointing in participants who learned to use the device over a short period of practice. Furthermore, the results support our hypothesis that pointing is more accurate when the device is held in the right dominant hand. Dexterity has to be attributed to the active part of the perceptive system.

4.2. Adaptations and enhancement of the scheme

We specified previously that one way to modify the device is to act on the distribution and characteristics of the receptive fields. The second source of versatility is the possibility to vary the transfer function of the receptors. But there are other changes at various levels that allow designing devices using the same general scheme but that exhibit a different behaviour. We propose in this discussion to describe several examples of adaptations to illustrate the great versatility of the proposed video to sound conversion.

4.2.1. Pre-filtering of the video stream

The video stream can be modified before being processed by the Vibe. A very useful example of pre-filtering is the application of a threshold to the captured pictures. The threshold can make a great number of receptors silent and consequently can make the resulting audio signal less complicated and more comfortable for the user. A second interesting example of pre-filtering is the use of a filter that computes as input to the Vibe a time difference of successive captured frames. With this kind of filtering, instead of producing a continuous and complex audio stream as in the case of sinusoidal static receptors, the Vibe generates sounds only if there are changes in the video stream. This functioning can also be more comfortable and less tiresome for the user than the static solution that continuously produces a complex sound. The comfort level can be measured by a simple questionnaire or by more subtle and indirect ways such as the measure of attention, stress or fatigue. One can imagine to build filters that combine thresholds and time differences to obtain a phasic/tonic behaviour similar to the biological receptors from the Vibe's receptors. There are two main ways to add pre-filtering to the Vibe. The first way is to develop the Vibe as a specific filter with one video input and two audio outputs that can be included in the graphs of media-handling components (such as those proposed by GStreamer™ and DirectX™). The other way, which allows saving of computing power, is to program the pre-filtering at the level of the Vibe's receptive fields (at the level of the transfer function).

4.2.2. Sensor distribution

A very important point is the choice of the method used to initialise the coordinates of the pixels addressed by each

sensor of each receptive field. Receptive fields of receptors can be large or small, identical or different, overlapping or non-overlapping. The distribution of the sensors on the 2D plane of captured pictures may seriously affect the perceptive abilities of the device user. In order to obtain a quasi-random uniform distribution of sensors one can for instance use advanced methods based on Kohonen adaptive maps (Durette, 2009). Non-uniform distribution of sensors can also be used to obtain a space-variant sampling of the video stream (Durette et al. 2007).

4.2.3. Spatialisation

The Vibe also offers the possibility to enhance the binaural perception and differentiation by the listener of the sound sources in the plane of the picture by adding inter-aural disparity cues to the sounds like ITD (inter-aural time difference) or ILD (inter-aural level differences). This can be done by adding delays and specific attenuation to the transfer functions of the receptors. In the case of sinusoidal static receptors, ITD can be obtained by choosing the phases ϕ_j^R and ϕ_j^L for the left and right ears for each receptor j in order to obtain a time difference between the two channels that corresponds to an inter-aural time difference correctly related to the horizontal position of the j th RF in the plane of the picture.

4.2.4. Implementation of the versatility

The map of sensors can be easily described in a text file that can be an input of a main application. The second main source of versatility (the transfer functions of the receptors) is more difficult to explicit in a text file. However, the transfer functions can be implemented with DLLs or lib files and the name of the chosen DLL (or lib) can also be in the text file. Consequently, we can imagine the Vibe to be implemented as a single application that is configured by a simple text file. In our opinion, it is important to provide the Vibe as a single and easily configurable application. The different configurations of the Vibe are intended to be tested by people such as experimental psychologists or ergonomics who are not always familiar with programming. In the preliminary version of the Vibe that exists and is freely available for testing (<http://sourceforge.net/projects/thevibe/files>), the receptive fields and their properties are configured with such a text file. This limited version of the Vibe only implements the static scheme described previously.

The experimental part of this study was achieved in the highly controlled environment of our laboratory. It is obvious that our participants would not exhibit the same level of performance in a more natural environment. In particular, the static sinusoidal scheme that was tested here is certainly not the best solution. A great challenge for

the Vibe and other vision-to-audition or vision-to-touch sensory substitution devices is to evolve to be efficient for perception in more ecological contexts. The versatile Vibe is a good candidate to take up this challenge.

References

- Arno P, Capelle C, Wanet Defalque MC, Catalan Ahumada M, Veraart C. 1999. Auditory coding of visual patterns for the blind. *Perception*. 28:1013–1029.
- Arno P, Vanlierde A, Streel E, Wanet Defalque MC, Sanabria Bohorquez S, Veraart C. 2001. Auditory substitution of vision: pattern recognition by the blind. *Appl Cognit Psychol*. 15:509–519.
- Auvray M, Hanneton S, Lenay C, O'Regan K. 2005. There is something out there: distal attribution in sensory substitution, twenty years later. *J Integr Neurosci*. 4(4):505–421.
- Auvray M, Hanneton S, O'Regan JK. 2007. Learning to perceive with a visuo-auditory substitution system: localization and object recognition with the voice. *Perception*. 36:416–430.
- Auvray M, Myin E. 2009. Perception with compensatory devices: from sensory substitution to sensorimotor extension. *Cognit Sci*. 33:1036–1058.
- Azemar G. 2003. *L'homme asymétrique: gauchers et droitiers face à face*. CNRS Editions, Paris.
- Bach-y-Rita P. 2002. Sensory substitution and qualia. In No A, Thompson E, editors. *Vision and mind*. Cambridge (MA): MIT Press, p. 497–514.
- Bach-y-Rita P, Collins CC, Saunders FA, White B, Scadden L. 1969. Vision substitution by tactile image projection. *Nature*. 221:963–964.
- Capelle C, Trullemans C, Arno P, Veraart C. 1998. A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution. *IEEE Trans Biomed Eng*. 45:1279–1293.
- Craig JC. 1981. Tactile letter recognition: pattern duration and modes of pattern generation. *Percept Psychophys*. 30:540–546.
- Cronly-Dillon J, Persaud K, Blore F. 2000. Blind subjects construct conscious mental images of visual scenes encoded in musical form. *Proc R Soc Lond. B* 267:2231–2238.
- Cronly-Dillon J, Persaud K, Gregory RPF. 1999. The perception of visual images encoded in musical form: a study in cross modality information transfer. *Proc R Soc Lond*. 266:2427–2433.
- Durette B. 2009. *Système de substitution sensorielle pour le handicap visuel: modèle biologique et approche neurocognitive pour la génération des signaux de vision pertinents, application la suppléance perceptive* [PhD thesis]. Université de Grenoble, France.
- Durette B, Gamond L, Hanneton S, Alleysson D, Hérault J. 2007. Biomimetic space-variant sampling in a vision prosthesis improves the user's skill in a localization task. *Proceedings of the 5th Conference and Workshop on Assistive Technology for People with Vision and Hearing Impairments*, Spain.
- Hirsh IJ. 1988. Auditory perception and speech. In: Atkinson RC, Hernstein RJ, Lindzey G, Luce RD, editors. *Handbook of experimental psychology*. New York: John Wiley, p. 377–408.
- Jansson G. 1983. Tactile guidance of movement. *Int J Neurosci*. 19:37–46.
- Kaczmarek KA, Haase SJ. 2003. Pattern identification and perceived stimulus quality as a function of stimulation current on a fingertip-scanned electrotactile display. *IEEE Trans Neural Syst Rehabil Eng*. 11:9–16.
- Lenay C, Gapenne O, Hanneton S, Marque C, Genoulle C. 2003. Sensory substitution: limits and perspectives. In: Hatwell Y, Streri, A, Gentaz, E, editors. *Touching for knowing*. Amsterdam: John Benjamins, p. 275–292.
- Loomis JM. 1974. Tactile letter recognition under different modes of stimulus presentation. *Percept Psychophys*. 16: 401–408.
- McIntyre J, Stratta F, Lacquaniti F. 1998. Short-term memory for reaching to visual targets: psychophysical evidence for body-centered reference frame. *J Neurosci*. 16(20):8423–8435.
- Meijer PBL. 1992. An experimental system for auditory image representations. *IEEE Transact Biomed Eng*. 39: 112–121.
- Pollok B, Schnitzler I, Mierdorf T, Stoerig P, Schnitzler A. 2005. Image-to-sound conversion: experience-induced plasticity in auditory cortex of blindfolded adults. *Exp Brain Res*. 167:287–291.
- Porac C, Coren S. 1981. *Lateral preferences and human behaviour*. New York: Springer.
- Renier L, Laloyaux C, Collignon O, Tranduy D, Vanlierde A, Bruyer R, De Volder AG. 2005. The Ponzo illusion using auditory substitution of vision in sighted and early blind subjects. *Perception*. 34:857–867.
- Sampaio E, Maris S, Bach-y-Rita P. 2001. Brain plasticity: visual acuity of blind persons via the tongue. *Brain Res*. 908:204–207.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

