

Auditory Substitution of Vision: Pattern Recognition by the Blind

P. ARNO, A. VANLIERDE, E. STREEL, M.-C. WANET-DEFALQUE,
S. SANABRIA-BOHORQUEZ and C. VERAART*

Neural Rehabilitation Engineering Laboratory, Université catholique de Louvain, Belgium

SUMMARY

Pattern recognition in a computer environment was investigated in 6 early blind and 6 blindfolded sighted subjects using auditory substitution of vision. Subjects had to scan visual patterns displayed on a PC screen by moving the pen of a graphics tablet, which lead to corresponding displacements of the cursor on the screen. A small screen area centered on the pointer was then translated into sounds according to a visual-auditory transcription code. Subjects were trained to learn this code during 12 one-hour sessions. Performance of both groups significantly increased with practice. This indicates that mental representations of visual patterns can be acquired through the auditory channel, even in the absence of visual experience. Moreover, blind subjects performed significantly better than sighted subjects did. This could be interpreted as a result of partial compensation for their loss of vision. Pattern recognition in a computer environment is thus possible using a fairly natural vision-to-audition coding scheme. Copyright © 2001 John Wiley & Sons, Ltd.

INTRODUCTION

One of the most important developments in information technology over the past ten years has been the graphical user interface (GUI). Although it has been a positive innovation for most sighted people, the GUI has represented a new challenge for blind people given its inherent visual nature, with icons, multi-windows and mouse-based command structure. Similarly, in the current world of communication, graphics play an increasingly important role. Numerous graphs, charts, diagrams and other forms of visual communication are included in documents intended to be read by sighted persons. Even if blind people can now access textual information to a large extent using a standard computer equipped with a Braille keyboard input, or using modern Optical Character Recognition (OCR) and speech synthesizer technology, they remain at a great disadvantage for graphical information access when compared to sighted subjects.

*Correspondence to: Claude Veraart, Neural Rehabilitation Engineering Laboratory, Université catholique de Louvain, Ave. Hippocrate, 54, UCL-54.46, B-1200 Brussels, Belgium. E-mail: veraart@gren.ucl.ac.be

Contract/grant sponsor: Van Goethem-Brichant Foundation.

Contract/grant number: 2191.

Contract/grant sponsor: SSTC.

Contract/grant number: 95/00-189.

In a previous experiment (Capelle *et al.*, 1998; Arno *et al.*, 1999), we have shown that trained normal blindfolded subjects can recognize simple visual patterns using a Prosthesis Substituting Vision by Audition (PSVA). The user carried on the head a miniature TV camera and a pair of headphones, both connected to a PC that simultaneously coded into sounds the images captured by the TV camera. Auditory translation of images was based on a pixel-to-sinusoidal audible frequency relationship. The implemented vision-to-audition coding scheme was fairly natural and allowed trained subjects to recognize any simple visual patterns formed with lines (vertical, horizontal or oblique bars). The recognition was possible thanks to sensory-motor interactions allowed by the prosthesis. As the information contained in some graphics, such as charts, can be decomposed into simple lines, we wondered if the same code as the one implemented in the PSVA, could be useful for providing blind people access to such computer graphics.

Beyond the practical interest of developing a tool that could be useful for blind people, some theoretical considerations have to be taken into account. Indeed, several cognitive components are at stake when users try to recognize visual patterns coded into sounds. On the one hand, the task primarily addresses the auditory ability to discriminate sounds. Indeed, subjects have to detect characteristic sounds (auditory invariants) corresponding to vertical or horizontal bars (visual invariants) (Arno *et al.*, 1999). If blind persons develop capacities of their remaining senses that would exceed those of sighted subjects (Ashmead *et al.*, 1998), one may wonder whether they will perform better than sighted subjects. On the other hand, when scanning a pattern with the PSVA, inexperienced users first try to identify its different components (i.e. the invariants), one by one, on the basis of the acoustic feedback. Users only focus their attention to parts of the stimulation at a time, the total image having to be mentally synthesized over time. This processing stage is closely related to visuo-spatial imagery and recruits short-term memory. If blind people actually suffer from some capacity limitations with regard to imagery processes (Cornoldi *et al.*, 1993), then they should perform worse than sighted subjects.

In order to address these questions, we performed the present study by adapting the original PSVA. In our previous experiment, subjects used head movements to scan the visual environment (Capelle *et al.*, 1998; Arno *et al.*, 1999). In the present study, patterns were displayed on a PC screen. Subjects, carrying a pair of headphones connected to the PC, had to move the optical pen of a graphics tablet, which shifted the pointer on the screen. A small screen area centered on the pointer was then translated into sounds according to the same code as in PSVA. Scanning hand movements were thus related on line to sound modifications. Early blind and sighted control subjects were involved in the experiment. Subjects of both groups learned the vision-to-audition translating code during training sessions. Their performance was assessed before, during and after training. By comparing the two groups, our aim was to investigate the effect of early visual deprivation on performance of recognizing visual patterns using auditory substitution.

METHOD

Subjects

Twelve male volunteers took part in the experiment. Subjects were divided in two groups of 6, the early blind and the control groups. They were paired for age (Av.: 41; SD: 16.5; range: 20–63). Blind subjects were recruited on a voluntary basis with the help of blind

associations. They were completely blind (absence of light perception) and otherwise neurologically normal (see Table 1). All blind volunteers were well integrated. They travelled and lived in an independent way and had a professional or social activity. The Biomedical Ethics Committee of the School of Medicine of the Université catholique De Louvain had approved the protocol of the study.

Subjects were given a pure tone audiometry test using a GSI 17 audiometer (Grason-Stadler, USA) in order to assess their auditory perception. When a slight decrease in auditory perception appeared (never more than 40 dB), the amplitude level of the sound produced by the device was increased accordingly.

Experimental set-up

The instrumentation used in this study was based on the one of Capelle *et al.* (1998) and Arno *et al.* (1999). The original PSVA set-up consists of a TV camera connected to a PC including a frame grabber and a sound generation printed circuit board, itself connected to an audio amplifier and headphones (Capelle *et al.*, 1998). With the original PSVA, a visual image was captured by the head-worn TV camera and its resolution was degraded to produce a two-resolution artificial retina (Veraart, 1989). This artificial retina consisted of a square matrix composed of 8×8 pixels, with the four central ones replaced by 8×8 smaller pixels. The 64 central pixels had thus a 4 times higher resolution than the 60 peripheral pixels. Then, a code translated the processed image into sounds. Each pixel of the processed image was assigned a sinusoidal tone according to its position in the artificial retina. Along each horizontal line, pixel frequencies increased slowly from left to right. Along each vertical line, pixel frequencies were in harmonic relationship, a given pixel frequency being equal to the frequency of the pixel just below, times the square root of 2. A weighted summation of the sine waves built up a complex auditory signal, which was converted into sounds by the headphones. Amplitude of each sinusoid was modulated by the gray level of the corresponding pixel, so that the visual experience of brightness was translated into the auditory experience of loudness (see Capelle *et al.*, 1998 for details).

Table 1. Characteristics of the early blind (EB) and sighted control (C) subjects

Group	Subject	Age at study beginning	Age at onset of blindness (years)	Aetiology	Handedness
EB	R.V.	20	< 5*	Bilateral extensive corneal injury	R
	PM.P.	24	0	Retrolental fibroplasia	L
	L.V.	28	< 3	Congenital glaucoma	R
	M.H.	46	0	Retinoblastoma	R
	R.H.	49	0	Maternal toxoplasmosis during pregnancy	R
	J.P.L.	59	2	Retinoblastoma	R
C	S.C.	24	—	—	R
	B.V.	28	—	—	R
	C.D.	32	—	—	L
	G.D.	56	—	—	R
	J.P.L.	62	—	—	R
	F.M.	63	—	—	R

*This subject was classified as early blind, as he suffered from low vision since birth.

According to the code, frequencies corresponding to adjacent pixels along a same horizontal line lay very close, so that the summation of such frequencies gave a characteristic sound that was perceived as beats. Given the harmonic relationship between the different frequencies corresponding to adjacent pixels along a same vertical line, the perceived complex sound was almost a chord. The code thus generated different acoustic sounds (auditory invariants) related to lines of different orientation (visual invariants).

In the present experiment, a graphics tablet was adapted to the usual PSVA setup. When the graphics tablet was used, a given image was displayed within an area of 64×64 pixels at the center of the PC screen. The resolution of the PC screen used was 1024×768 pixels and the exploration area was restricted to the central 300×300 pixels. A square of 64×64 pixels was centered on the pointer whose position depended on the pen location on the graphics pad. In this area, the artificial retina was implemented. During graphics tablet exploration, a part of the artificial retina might intercept a portion of the displayed pattern. In that case, the related pixels in the artificial retina were translated into sounds in real time, according to the transcription code described above.

In both cases, computer processing performance allowed a refreshing rate of 25 images/s. Subjects could therefore experiment relatively normal sensory-motor interactions with the visual environment.

Training and assessment procedure

A training program with 4 sets of 3 sessions was used. During the first three learning sessions, subjects were familiarized with the use of PSVA. Indeed sensory-motor interactions from head movements proved useful to recognize simple patterns even during the first training sessions (Arno *et al.*, 1999). Therefore, subjects started training by using the head-worn TV camera. During the last nine training sessions, the patterns were displayed on the PC screen and pattern recognition was made by exploring the graphics tablet with the pen. Thus, subjects heard sounds related to their hand movements, similarly to the previous sounds related to their head movements.

Subjects were trained individually. The experimenter was also equipped with headphones in order to receive the same auditory feedback as the subjects and to assist them during training. A total of twenty-five two-dimensional patterns of various complexity levels were employed (see Figure 1). The complexity depended on both the number of the components and their spatial organization within the pattern. All figures were presented once per set of three training sessions. For the first set, the stimuli were presented according to an increasing level of complexity (beginning with pixels, then simple bars, etc.). The order was at random for the three remaining sets. During training, patterns were presented one at a time for a maximum of 3 minutes each. At the beginning, learning focused on developing perceptual-motor control. Subjects had to learn how to move the head, in order to locate and stabilize the studied pattern. They were encouraged to follow their own strategy and to make comments about what they heard. When ready, subjects had to construct their response by arranging aluminum dots (8.5×8.5 mm) and strips (68×8.5 mm) in a frame with the borders corresponding to the vertical and horizontal axes. Dots and strips corresponded to pixels and bars, respectively. A bar was equivalent to eight adjacent collinear pixels (see details in Arno *et al.*, 1999). The role of the experimenter was to provide appropriate guidance. When the response was wrong, the subject was corrected either verbally or by rearranging dots and bars, and was then allowed to haptically explore the correct

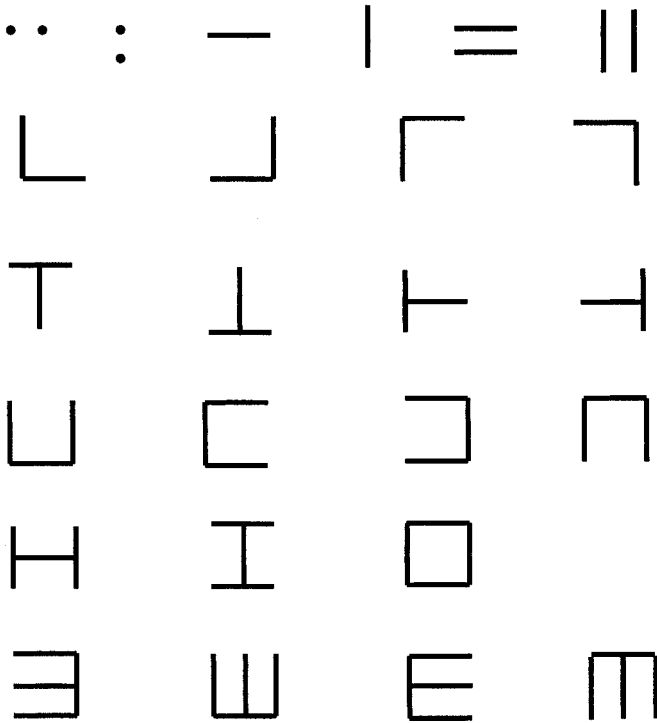


Figure 1. The patterns. Illustration of the twenty-five patterns used in this study

response. At the end of each training session, learning was briefly tested, using four of the patterns presented during the session. Again, subjects had to give their responses within 3 minutes and were given tactile feedback if necessary. The aim of these tests was to reinforce the current learning and to increase subjects' motivation by giving them an easy task. The level of learning reached by each subject was assessed three times, before, during and after training. The three evaluations included all the 25 patterns of the experiment. Before the first evaluation session E1, each subject was given a short explanation about the functioning of the PSVA. Subjects used the head-worn TV camera for the first evaluation, whereas they used the graphics tablet for the second and third evaluations.

Data processing

Performance was assessed on the basis of both response accuracy and processing time needed to answer. The assessment procedure for scoring subjects' responses was the same as in the previous study (Arno *et al.*, 1999). Briefly, each stimulus pattern *S* was superimposed on a representation of the subject's response *R*, in order to get the best correspondence, without applying any rotation. The number *X* of common points (pixels) between *S* and *R* was determined and the score (rated from 0 to 1) was given by the ratio $X/(S + R - X)$.

RESULTS

Qualitative observations

The transition from the head-scanning method to the hand-held approach did not disturb the subjects. Subjects even reported that the task seemed easier using the graphics tablet. Indeed, they found it less difficult to orient the graphics pen with respect to axial references than to move the head in the space, especially when movements were in the horizontal plane.

During pattern exploration, subject strategies appeared very systematic. Using large horizontal and vertical movements, subjects first tried to globally recognize the various sounds corresponding to different pattern elements, i.e. lines inside the pattern. They could deduce the number of elements and their orientation (either vertical or horizontal). Then they chose a starting line. Doing slight and slower movements along its sketching, subjects could deduce how this starting line was spatially organized in relation with the other lines. For example, if an L-shaped pattern lies in the upper left field of view of the retina, frequencies associated with the related pixels are very high. Moreover, the sounds are essentially heard by the left ear, given the implementation of the binaural intensity balance. The sonorous characteristics are sufficient to help the user to orient the head (or the pen) in direction of the pattern, in order to put it in the center of the artificial retina and to allow its recognition. In this particular example, the subject has to move the pen, or the

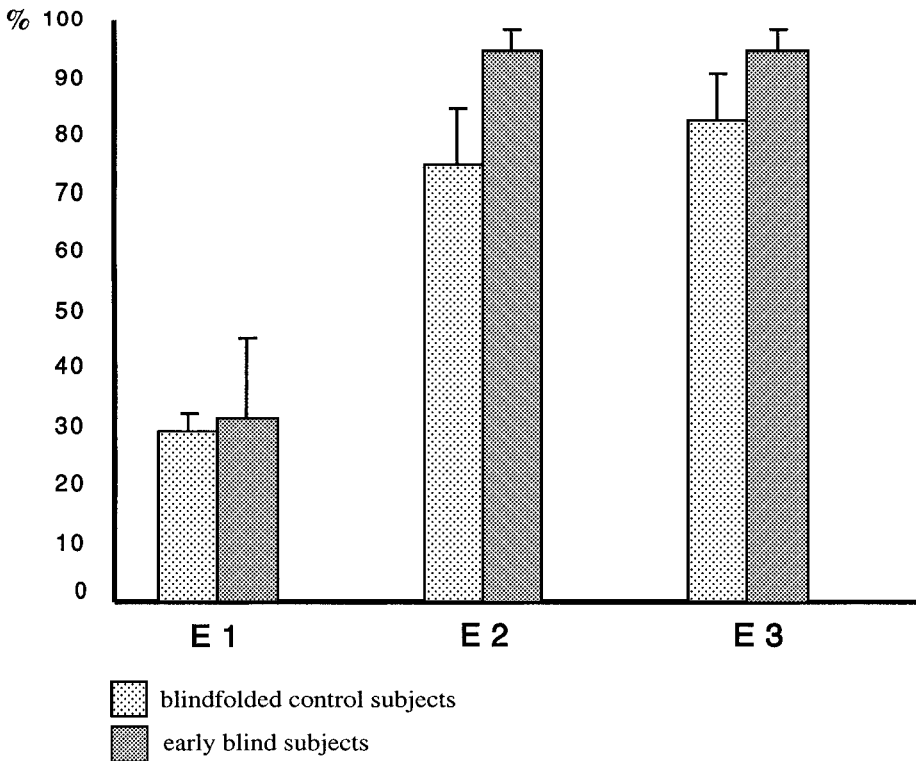


Figure 2. Scores. Mean scores and standard deviations of both groups during the three evaluation sessions. Scores are expressed in percents of the maximum possible score (25)

head, up and to the left. On the basis of the acoustic feedback, the subject can deduce when the artificial fovea is centered on the pattern, because in this case, sounds change from high frequencies to mid-frequencies. Then, doing vertical movements, the subject will hear changing beats and can deduce that an horizontal line is part of the pattern. Moreover, by horizontal movements, he will hear changing harmonious sounds in the left ear. He can thus deduce that a left vertical line is included in the pattern. In the case of an L-shaped pattern positioned below with respect to the artificial retina, the subject will hear very low frequencies. Thus, a same pattern can produce different sounds according to its relative position in the artificial retina.

Quantitative results

An analysis of variance (ANOVA 2×3) was performed separately on the score and the processing time, with the group as between-subject factor and the evaluation session as within-subject factor.

Scores

Scores of both groups as a function of the session are shown in Figure 2. There was a significant main effect of the group ($F(1, 298) = 51.49, p < 0.001$). Blind subjects performed significantly better than sighted subjects. There was also a significant main effect of the session ($F(2, 596) = 731.85, p < 0.001$), showing that performance increased over sessions. The interaction between these two factors was significant ($F(2, 596) = 13.46; p < 0.001$).

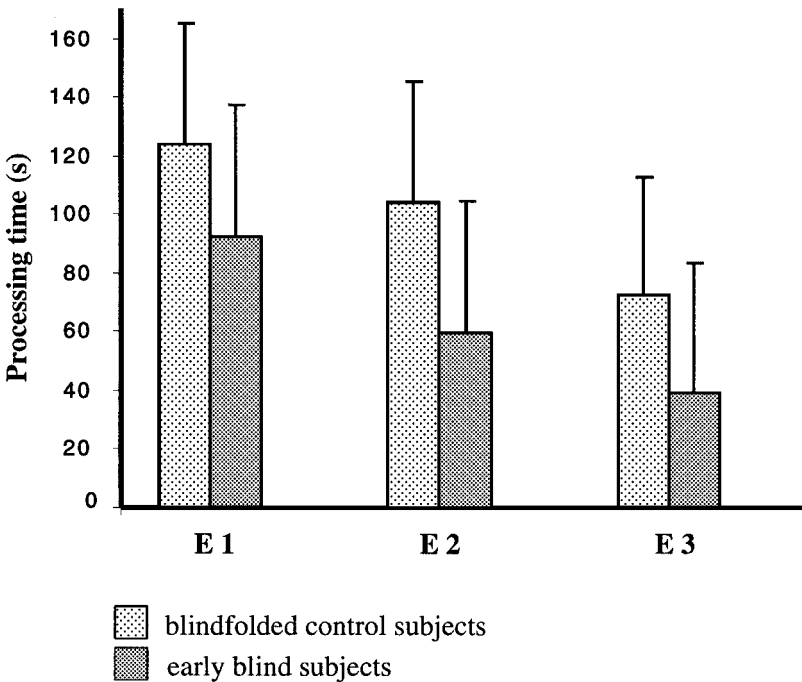


Figure 3. Processing time. Mean processing time and standard deviations, in seconds, of both groups for each evaluation session

A test for the group effect showed that both groups did not differ for evaluation E1, but did significantly for evaluations E2 ($F(1, 298) = 53.78, p < 0.001$) and E3 ($F(1, 298) = 24.96, p < 0.001$).

A test for evaluation session effect showed that the increase between E1 and E2 was significant ($F(2, 298) = 447.27; p < 0.001$), as well as between E2 and E3 ($F(2, 298) = 6.48; p < 0.001$).

Processing time

Processing time of both groups as a function of the evaluation session is shown in Figure 3. There was a significant main group effect ($F(1, 298) = 134.37, p < 0.001$). Blind subjects were significantly faster than sighted subjects. There was also a significant main effect of the evaluation session ($F(2, 596) = 122.48, p < 0.001$), showing that processing time decreased over sessions. The interaction between the two factors was not significant.

A test for the evaluation session effect showed that the decrease between E1 and E2 was significant ($F(2, 298) = 24.994, p < 0.001$) as well as the decrease between E2 and E3 ($F(2, 298) = 68.198, p < 0.001$).

DISCUSSION

The results of this experiment show that early blind and blindfolded sighted subjects are able to recognize auditory coded visual patterns on the basis of a real-time acoustic feedback related to hand movements. These results confirm and extend our previous observations with sighted subjects using PSVA and in which auditory feedback was provided by head movements (Arno *et al.*, 1999). Indeed, sighted subjects' performance in both score and processing time for these two studies were not significantly different ($F(1, 16) = 0.058, p = 0.813$ and $F(1, 16) = 0.541, p = 0.473$ for score and processing time, respectively). In this comparison, we only considered data related to patterns with identical training schedule.

Taken together, the results of the present and the previous experiments underline the potential of sensory substitution using audition in visual pattern recognition. For both groups of the present experiment, the handling of the graphics pen after having used the miniature head-worn TV camera did not interfere with performance level or with processing time. Accordingly, self-induced movements are required for sensory substitution whatever the motor system involved, since plasticity mechanisms occur at a rather central level (Bach-y-Rita, 1972).

Subjects did not merely memorize simple associations between sounds and patterns; they did learn the relationship between auditory code and spatial attributes of the patterns. Subjects actually learned a way to explore any pattern, improving their ability to extract the auditory invariants related to horizontal and vertical bars within a pattern from the auditory information provided by the device. It was the self-induced movements, which allowed extracting the auditory invariants and the way they are organized. The process of sensory substitution would not be possible without the real-time sensory-motor feedback, as it has also been demonstrated in case of tactile substitution by Bach-y-Rita (1972). Besides, we had previously demonstrated that learning with PSVA generalized to new stimuli (see Arno *et al.*, 1999).

Recognition of visual patterns with PSVA, whether the scanning is performed with the head-worn TV camera or the graphics pen, implies both to identify its bars of various

orientation and to understand how they are spatially assembled (Arno *et al.*, 1999). This is consistent with claims of e.g. Marr (1982) and Biederman (1987) that visual recognition requires identification of both object parts and their spatial relationship. In the present study, as in the previous one (Arno *et al.*, 1999), component identification requires to recognize the auditory invariants corresponding to visual invariants. On the other hand, ascertainment of spatial relationships is provided by sensory-motor interactions. Reaching a global unified representation requires having a mental representation of the pattern, especially for the more complex ones. Of course, subjects can use verbal strategies to code the different elements in the pattern, by naming them ('there is an horizontal line, with a vertical one on the left, etc.'). But verbal strategies can probably not entirely explain the results. Indeed, such strategies are time consuming, and the short recognition time observed for some patterns is not in favor of an explanation exclusively based on verbal strategies. Moreover, with exclusively verbal strategies, recognition of patterns composed with more elements, like an E-shaped pattern, could take more time to be described with words than simpler L- or T-shaped patterns. With PSVA, it was not the case (see also Arno *et al.*, 1999). Thus, the present and previous results suggest that subjects can get to and acquire a mental unified representation of visual patterns through the auditory channel. Moreover, the high performance level reached by the early blind group in the present study suggests that vision is not a prerequisite for the development of these representations, in agreement with previous studies on visuo-spatial imagery (Kerr, 1983; De Beni and Cornoldi, 1988) and mental rotation by blind subjects (Marmor and Zaback, 1976).

The successful recognition process we demonstrated here requires both good auditory abilities allowing auditory invariant recognition, and good working memory capacities to maintain and process information during the scanning. As the set of patterns used in this study was relatively simple, the working memory load likely remained limited. One may wonder whether the performance of the early blind group would be maintained in a task using more complex stimuli. Indeed, blind people seem to be impaired in imagery tasks requiring a high cognitive load (De Beni and Cornoldi, 1988; Cornoldi *et al.*, 1993). Accordingly, we could thus expect that increasing the stimuli complexity, and hence the working memory load, should lessen the blinds' performance more than that of the controls. However, if training makes the recognition process more automatic, the working memory load should be reduced. By providing additional training with complex stimuli, blind people might thus obtain a satisfying performance level as well.

The present study also shows that the early blind subjects performed better than the control group, both in accuracy and processing time. According to the compensation hypothesis (see Ashmead *et al.*, 1998 for review), this could be accounted for by superior auditory abilities of early blind subjects. Indeed, they are compelled to frequently deal with auditory cues in daily life while sighted subjects tend to favor visual cues among information coming from different sensory modalities, as suggested by studies on intermodal conflicts (for example, Craske, 1966). However, some studies do not support this view (e.g. Wanet and Veraart, 1985) and when increased auditory capacities are observed in case of compensation to visual disabilities, it would not exceed the normal range of human hearing (Ashmead *et al.*, 1998). Moreover, one could not totally exclude the possibility of an increased motivation from the blind group to succeed in the task. As our subjects sample was small and included well-integrated blind people, maybe those blind volunteers were particularly efficient from a cognitive point of view.

The results presented here are encouraging from a rehabilitation point of view. As graphical user interfaces and other forms of visual communication (graphs, charts, etc.)

are not generalized and incorporated into education, employment and daily life, it is a necessity for blind and visually impaired computer users to have reliable access to graphical information. Some kind of information, like numbers in a graph, is not inherently visual in nature. They can thus be easily coded in an alternative form (using speech synthesizer or Braille dots). However, for information which is pictorial in nature, like the outline of a curve, verbal descriptions are very lengthy, or not exhaustive. In that case, blind computer users need mechanisms allowing them to explore and directly interpret the images. Accordingly, nonverbal coding of visual information deserves consideration. Tactile coding of visual images seems straightforward, as both signal and sensory organs appear as two-dimensional. However, transmission of a large flow of information requires together high spatial resolution and high data rate, from sensory organ up to perception. As the auditory sense exhibits a better absolute range compared to the tactile sense, auditory coding could be preferable, provided that a natural coding scheme is used. In the study, we show that a rather short training period is sufficient to blind individuals to be able to explore and interpret simple visual patterns on the basis of sensory-motor interactions between the hand movements and the related sound modifications. For some kind of graphics in which information is mainly linear, such as bar graphs, well-trained blind users might access them when displayed on a PC screen, without a sighted person at hand.

ACKNOWLEDGEMENTS

This work was partly supported by grant #2191 from the Walloon Region of Belgium, by the Van Goethem-Brichant Foundation from Belgium and also by SSTC grant #95/00-189. The authors are very indebted to Professor Raymond Bruyer for his helpful comments on a previous version of this manuscript. We also very gratefully acknowledge Dr Christian Capelle, MM. Benoit Gérard and Jean-Christophe Popeler for their technical assistance. Thanks are also due to the Oeuvre Nationale des Aveugles and to the Institut Royal pour Sourds et Aveugles of Belgium and to our volunteers for their helpful collaboration.

REFERENCES

- Arno P, Wanet-Defalque M-C, Capelle C, Catalan-Ahumada M, Veraart C. 1999. Auditory coding of visual patterns for the blind. *Perception* **28**: 1013–1029.
- Ashmead DH, Wall RS, Ebinger KA, Eaton SB, Snook-Hill M-M, Yang X. 1998. Spatial hearing in children with visual disabilities. *Perception* **27**: 105–122.
- Bach-y-Rita P. 1972. *Brain Mechanisms in Sensory Substitution*. Academic Press: San Diego.
- Biederman I. 1987. Recognition-by-components: A theory of human image understanding. *Psychological Review* **94**(2): 115–147.
- Capelle C, Trullemans C, Arno P, Veraart C. 1998. A real time experimental prototype for enhancement of vision rehabilitation using auditory substitution. *Institute of Electrical and Electronics Engineers Transactions on Biomedical Engineering* **45**(10): 1279–1293.
- Cornoldi C, Betucelli B, Rocchi P, Sbrana B. 1993. Processing capacity limitations in pictorial and spatial representations in the totally congenitally blind. *Cortex* **29**: 675–689.
- Craske B. 1966. Intermodal transfer of adaptation to displacement. *Nature* **210**: 765.
- De Beni R, Cornoldi C. 1988. Imagery limitations in totally congenitally blind subjects. *Journal of Experimental Psychology: Learning, Memory and Cognition* **14**: 650–655.
- Kerr NH. 1983. The role of vision in 'visual imagery' experiments: Evidence from the congenitally blind. *Journal of Experimental Psychology, Human Learning and Memory* **5**: 424–426.

- Marmor GS, Zaback LA. 1976. Mental rotation by the blind: does mental rotation depend on visual imagery? *Journal of Experimental Psychology: Human Perception and Performance* **2**(4): 515–521.
- Marr D. 1982. *Vision*. Freeman: San Francisco.
- Veraart C. 1989. Neurophysiological approach to the design of visual prostheses: a theoretical discussion. *Journal of Medical Engineering and Technology* **13**: 57–62.
- Wanet M-C, Veraart C. 1985. Processing of auditory information by the blind in spatial localization tasks. *Perception and Psychophysics* **38**: 91–96.