THE ROYAL SOCIETY

The perception of visual images encoded in musical form: a study in cross-modality information transfer

John Cronly-Dillon[1*], **Krishna Persaud**[2] **and R. P. F. Gregory**[1]

[1]*Department of Optometry and Neuroscience and* [2]*Department of Instrumentation and Analytical Science, UMIST, Manchester M60 1QD, UK*

This study demonstrates the ability of blind (previously sighted) and blindfolded (sighted) subjects in reconstructing and identifying a number of visual targets transformed into equivalent musical representations. Visual images are deconstructed through a process which selectively segregates different features of the image into separate packages. These are then encoded in sound and presented as a polyphonic musical melody which resembles a Baroque fugue with many voices, allowing subjects to analyse the component voices selectively in combination, or separately in sequence, in a manner which allows a subject to patch together and bind the different features of the object mentally into a mental percept of a single recognizable entity. The visual targets used in this study included a variety of geometrical figures, simple high-contrast line drawings of man-made objects, natural and urban scenes, etc., translated into sound and presented to the subject in polyphonic musical form.

Keywords: cross-modality transfer; vision through sound; blindness

1. INTRODUCTION AND OUTLINE OF THE CONCEPT

This paper addresses some of the problems associated with the cross-modality transfer of sensory information of static, high-contrast line figures translated into musical form. In common with some other devices based on sense substitution, the approach explored here translates visual images into sound (Meijer 1992). However, it differs from other attempts in that it incorporates a system for feature extraction which is designed to enable a blind subject to deconstruct a complex optical image selectively into a set of simpler representations that make it easier for the subject to analyse. Each representation selectively isolates one or more sets of features in the visual display which are each then translated into an equivalent sound pattern and presented to the subject as the several 'voices or submelodies of a complex polyphonic musical composition' which encodes the whole image. The blind subject listens and analyses these representations one at a time or in combination and uses the combination of submelodies to reconstruct a representation of the visual target mentally. The process resembles the tactile scanning of a solid object carried out by a blind subject or the normal scan-path behaviour of the human eye as it pursues a search pattern of foveations in exploring a complex novel object such as a strange face. In the course of carrying out a scan-path trajectory, the visual system selects and gathers information about the salient features in a patch-like manner. These serially acquired patches of information contribute collectively to form a representation which characterizes the object and aids future recognition (Stark & Ellis 1981).

(a) *Mapping visual images into musical form*

A musical score is a graph which displays notes as a spatial distribution. This fundamental observation formed the basis of the first automatic player piano or 'pianola'. Here, music notated in perforations on a paper roll is passed over a *y*-axis tracker bar so that different notes are produced according to where the different perforations intercept the tracker bar. The pattern of perforations need not be confined to representing a musical composition, but may equally be used to represent the locations of a set of points which define the outline features of a visual scene or object in sound. Exploiting this notion and simulating the process on a PC has enabled us to represent many of the attributes and contents of simple high-contrast, black and white, visual images in a musical form which can be analysed and reconstructed in the mind by a blind or blindfolded person (Cronly-Dillon & Persaud 1998).

(b) *The perceptual analysis of complex visual forms translated into sound requires segmentation of the image*

While several devices based on an image-scanning principle are capable of encoding visual images and generating sound impressions that may characterize even complex natural scenes (N. B. Fournier d'Albe exhibited the 'optophone' at the 1912 Optical Exhibition at South Kensington (Wagg 1932; Mansur *et al*. 1985; Meijer 1992), problems arise because the brain is then faced with the task of extracting meaningful percepts from what may appear to be a confusing array of sounds within which are embedded

*Author for correspondence (mjcjcd@fsl.op.umist.ac.uk).

Received 26 July 1999 *Accepted* 13 September 1999

elements of structure. The task is non-trivial and precisely the same problems were faced by the visual system, which had to evolve diverse means for segmenting various attributes of the visual image for separate processing and then recombine these into meaningful perceptual units before it could develop into a system capable of conveying the diversity and detail of visual experience that we enjoy today.

(c) *Segmentation and stimulus binding*

The human visual pathway provides a neural basis for recognizing external objects and for judging their different locations, speeds and direction of motion in the outside world. To perform these functions, information relating to different features in a retinal image is first partitioned into discrete packages, where each package corresponds to a selected feature or group of features within the image of a natural scene or object (Regan 1991; Zeki 1993). Then the brain somehow combines the different relevant packages to generate a single percept of the object or scene. In modern parlance this is sometimes referred to as 'the binding problem' and relates to the fact that the different sensory features of an object, although processed in different brain areas, are mentally grouped together to generate a Gestalt-like impression which groups the separate features together and unites them into a global percept of a single entity.

(d) *Basic icons used in the analysis of simple shapes*

The visual system of mammals and primates contains neurons sensitive to line orientation, which are an important subset of the feature-selective elements in the brain which play a key role in the analysis and perceptual representation of visual form. Indeed, the mosaic of orientation-specific columns found in the mammalian striate cortex (Livingston & Hubel 1984; Zeki 1993) have all the necessary characteristics for allowing line contours of different orientation in a patterned retinal image to be segregated and displayed selectively as a set of simpler representations, each containing only those line contours which correspond to the selected orientation. In the immature brain, these and other feature detectors represent the functional building blocks (basic icons) for later assembly through learning of more elaborate and sophisticated feature detectors which may encompass the whole or substantial portions of the whole object, i.e. gnostic circuits (Konorski 1969; Barlow 1972; Rolls 1997; Rolls & Treves 1998). Meanwhile, studies by Mountcastle (1978) have suggested that the mosaic columnar arrangement of some of these high-order modules within area IT, similar to that displayed by the orientation columns in area V1, may serve as a higher order set of pattern primitives, which in combination might represent the appearance of a particular object (Stryker 1992).

(e) *Feature extraction can segment the image before it is translated into sound*

Segmentation appears to be the key to analysing the structure of complex visual images translated into sound. To regain the original figure some process of 'orchestration', not unlike that which takes place in a chorale work of music, is required within the brain to generate a Gestalt-like grouping which leads to the unification of these separate 'voices' to form a single perceptual entity.

2. METHODS

(a) *The basic module*

(i) *Transforming visual images into sound*

This was done with a personal computer programmed to simulate a pianola which is used to generate the sound patterns associated with different high-contrast line contours of visual shapes. The computer screen on which the image was displayed was divided into 'vocal' pixels of uniform size so that all notes had the same duration. Any pixel that is occupied by a portion of the figure displayed on the monitor represents a single musical note which remains silent until it is intercepted by a moving y-axis tracker bar. The latter is subdivided and programmed along its length to represent a scale consisting initially of 50 white musical notes arranged symmetrically about middle C on a piano keyboard, with the treble end situated at the top and bass at the bottom of the screen. As the tracker bar intercepts and traverses the boundary of a figure displayed on the screen, it generates a temporal mosaic of sounds which can be analysed by subjects who may then reconstruct the shape it has traversed. Provision was made to allow the subjects to test for bilateral symmetries along the vertical and horizontal axes of the figure to assist in their reconstruction and identification of the visual target.

(b) *Procedures used when testing subjects*

(i) *Masking*

Masking is an optional facility which enables subjects to block out areas of the image selectively (e.g. top or right half, etc.) to allow them to carry out a patch-by-patch exploration and analysis of the target.

(ii) *Image simplification by feature extraction*

A further pre-processing stage was introduced which allowed the subjects to extract and sequester vertical, horizontal or oblique line components within the target image selectively to produce a set of simplified representations which they could analyse. Thus, one pre-processed image may select for priming all sets of occupied pixels that lie on well-defined vertical lines in the image, while another may do the same for horizontal lines, etc. This enabled the subjects to select which image feature or combination of features they wished to have included (or excluded) in a particular presentation. Effectively, it allowed the subjects to isolate and examine separately and selectively the contents of several segmented packages in which different features of the image had been sequestered.

(iii) *Analysis of complex segmented images*

The testing protocols employed for analysing the musical signatures of complex figures were as follows.

(i) The subject first listened to the sound signatures generated by scanning along the two principal axes (vertical and horizontal) of the whole non-segmented figure.
(ii) Next the target figure was segmented or deconstructed into a number of simpler representations, the auditory signatures of which were studied and analysed either separately or in combination by the subject.
(iii) The subject listened to the sound signature of the whole non-segmented figure and was asked to use all the information they had gathered to attempt a mental reconstruction of the target. The subject reported their conclusion either verbally or by sketching it.

(iv) *Subjects*

The majority of the 42 subjects used in these experiments had normal vision which they were prevented from using during testing. Among these we included four totally blind but previously sighted subjects. B.K. is a 52-year-old male who lost all light sense at the age of 27 years (through retinitis pigmentosa). M.S., aged 49 years and another former retinitis pigmentosa patient, lost her sight 20 years ago but has retained some light sense. M.P., aged 45 years, lost all vision and light sense 34 years ago through a laboratory accident. P.L., aged 31 years, lost all vision and light sense ten years ago through a car accident, which resulted in pressure to the optic nerves. None of the subjects tested suffered from any hearing defect. Some subjects, including blind subject P.L., did not carry out the full range of tests reported in this paper.

(v) *Testing*

This took place once a week in a session that lasted *ca.* 90 min. The sweep speed of the tracker bar was set so that it took 25 ms to traverse a vocal pixel of 2 mm width. This we found was adequate in allowing for the initial analysis of most stationary figures. Once a figure had been correctly analysed we later noted that the subjects were able to recognize the target using a much faster scan rate, but we have not yet investigated this systematically.

Before testing, the subjects were given a verbal description of the musical signatures associated with the following basic shape and line segments.

(i) Horizontal line: a rapid temporal repetition of the same note.
(ii) Vertical line: a set of different musical notes sounding simultaneously.
(iii) Oblique line: an ascending or descending scale.
(iv) Two lines forming an acute angle ('<' and '>'): overall sound of 'opening out' or 'closing in'.

Other than in circumstances specifically mentioned (see experiment 2b and the few cases of blind subjects retested for recall several months later), all subjects were presented with the sound signatures associated with a set of visual shapes they had not been presented with before. They were encouraged to analyse these in terms of the basic line elements, the characteristic sound signatures of which had been described to them previously. Their task was to explore and analyse the target figure by listening to the associated sound pattern. If necessary, the target shape was presented to the subjects *ad libitum*. Using information gathered during these explorations, the subjects attempted to reconstruct the whole figure mentally from the set of melodic sequences presented to them. They were also instructed to describe verbally or, alternatively, to represent their reconstruction graphically in a drawing.

3. RESULTS

(a) *Experiment 1: recognition of basic line icons and simple figure combinations of two lines*

Experiment 1a (cross-hatched bars in figure 1) shows the average performance of 30 naive (sighted but blindfolded) subjects who were restricted to scanning the visual target in one direction, i.e. tracker scanning only from left to right across the screen.

All the sound signatures associated with the test figures presented were encountered by the subjects for the first time. Before giving their response, each was allowed to scan the target several times. The subjects were scored on the basis of their first attempt at identifying the target from its musical signature. The uniformly high scores obtained by most subjects confirmed their ability to identify the basic line elements correctly from the verbal description given before the experiment. This contrasted with the greater variability in performance in correctly identifying some simple figures consisting of two line combinations.

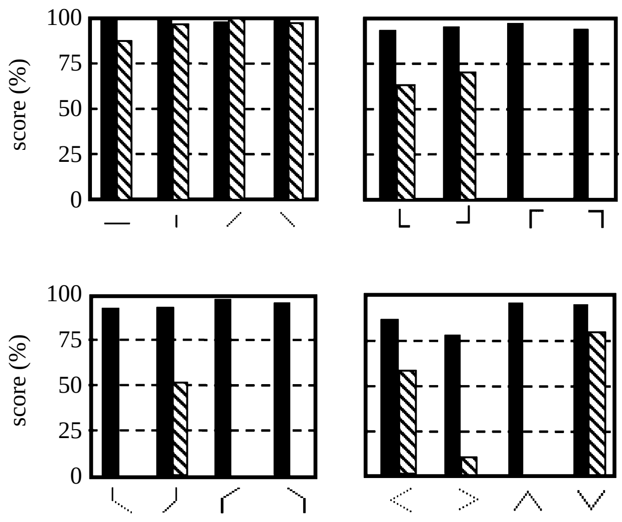


Figure 1. Recognition of simple combinations of basic line elements. Experiment 1a: mean accuracy of the first response of subjects given no prior training other than verbal instructions (cross-hatched bars). Experiment 1b: mean performance of subjects with prior training (black bars).

Experiment 1b (black bars in figure 1) involved training the subjects in icon identification. This was a variation of experiment 1a except that a new set of 12 subjects (eight sighted but blindfolded and four blind) was used and (unlike experiment 1a) if they identified icons incorrectly they were informed of the correct answer and allowed to continue until they had achieved an uninterrupted score equal to or better than seven out of ten for each item in a randomly selected sequence of visual targets comprising a set of 16 icons. This was repeated with different random permutations among the 16 members of the set to generate a subject's average performance.

(b) *Experiment 2: analysis and identification of geometrical shapes*

The principal difference between the two experiments described below (experiments 2a and 2b) lies in the number of different ways a subject could interrogate the target to ascertain its form. In experiment 2a the subject was allowed to scan the target in only one direction along the horizontal axis (figure 2, cross-hatched bars), while in experiment 2b the subject could scan the target in both directions along the vertical and/or horizontal axis of the figure to test for symmetry (figure 2, black bars) and/or selectively mask the top or bottom half of the target figure (figure 2, horizontally hatched bars).

Experiment 2a (figure 2, cross-hatched bars) was first carried out with the same group of 30 subjects that were

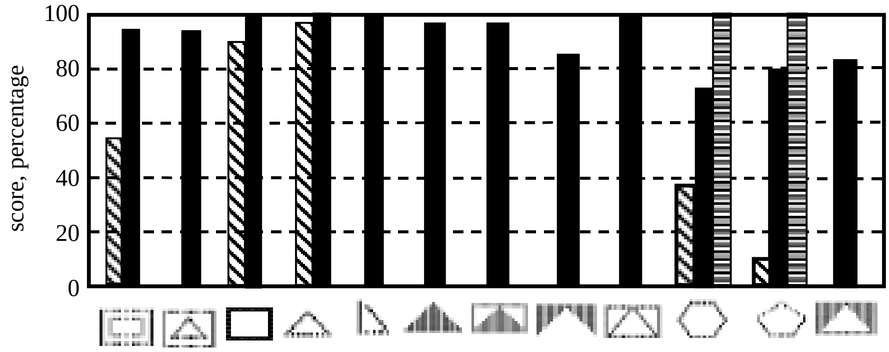
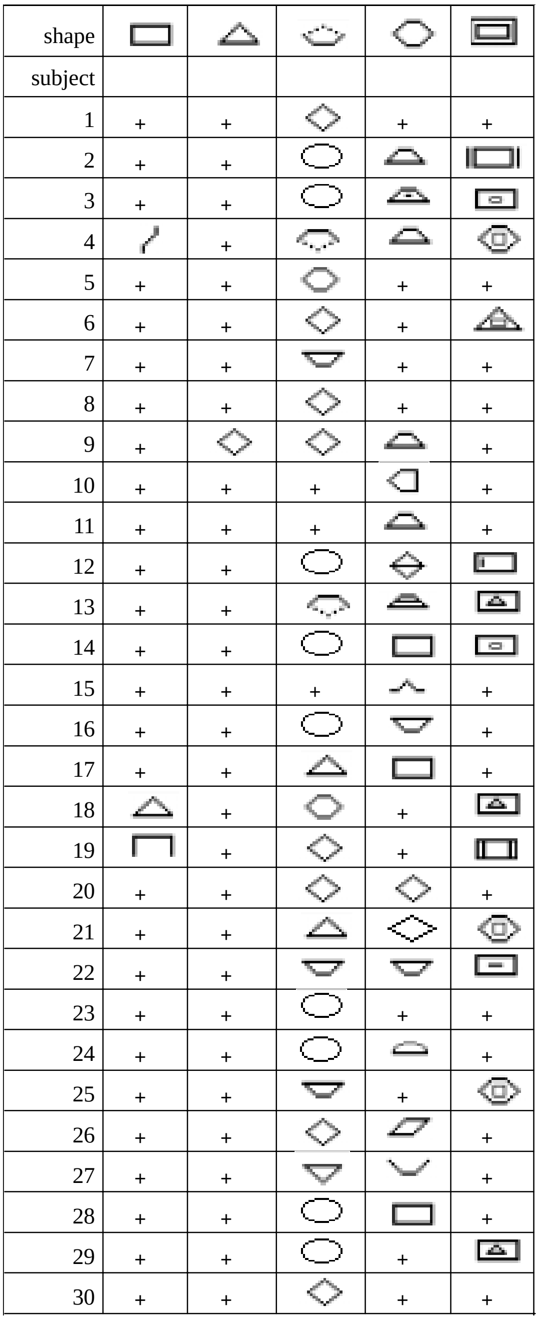


Figure 2. Geometrical figures. Cross-hatched bars: mean accuracy of the first response of 30 subjects in experiment 2a who were allowed to scan the target in one direction only. Black bars: seven of the 12 subjects in experiment 2b chose to scan the shapes in both directions along both the vertical and horizontal axis. Horizontal-hatched bars: five of the 12 subjects in experiment 2b chose to segment the pentagon and hexagon by masking portions of the figure. This enabled these subjects to identify these figures correctly as easily as a square.



employed in experiment 1a. The aim was to investigate their ability to analyse, name or draw simple geometrical shapes from their auditory signatures when exploration of the target was limited to scanning in one direction only. No prior information was provided about the visual targets other than that they were high-contrast line representations of geometrical forms. The subjects were scored on the basis of the first response to the target they produced. The results depicted in figure 2 (black bars) give the performance averaged over the 30 subjects for each of the geometrical forms presented. Under the constraints imposed in this experiment, the two geometrical forms which presented subjects with the most difficulty in correct identification were the pentagon and hexagon. Despite this, the drawings produced by individual subjects showed that, even when the subjects failed to identify the target from the sound signature correctly, many were still able to capture some essential features of the figures that were presented to them (figure 3).

Experiment 2b (figure 2, black and horizontal hatched bars) was a variation of experiment 2a carried out with 12 subjects. These consisted of seven sighted and three blind subjects of the original set of 12 subjects employed in experiment 1b who had received training on the basic icons (figure 1) and two new sighted subjects who received an abridged training session on the basic icons. The set of geometrical forms presented in experiment 2b included four of those used in experiment 2a and several additional ones including some enclosed figures. The principal difference between this and experiment 2a, apart from the number of participants, was that the subjects were allowed additional choices in the number of ways they could explore the shape to determine its characteristic features, i.e. they were permitted to scan the target in both directions along the vertical and horizontal axes of the target (figure 2, black bars) or, in addition, they could elect to segment the figure (figure 2, horizontally

Figure 3. (*Opposite*) Examples of the erroneous drawings made of the geometrical figures by the subjects in figure 2*a* during testing. Note the correct identification of some essential features in the incorrectly drawn figures. (A plus sign denotes a subjects correct identification of the visual target.)

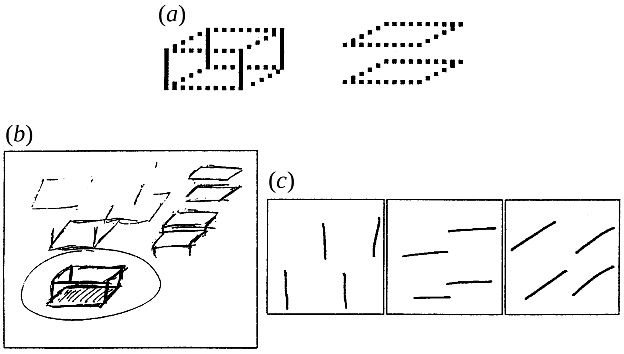


Figure 4. Analysis of a two-dimensional representation of a three-dimensional cube. (*a*) The target figure and segmentation with vertical line components removed. (*b*) Sketches showing the process of reconstruction produced by a sighted subject (P.H.). (*c*) Sketches showing the process of reconstruction produced by a blind subject (B.K.).

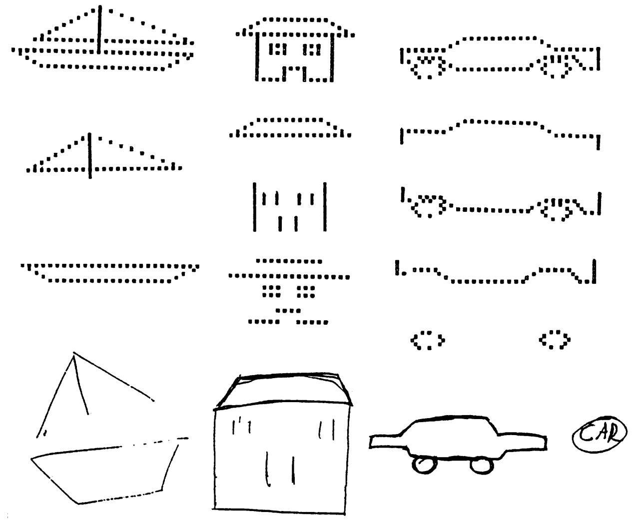


Figure 5. Encoding and segmentation (masking and/or some feature extraction) of cartoon images each presented as a melody with several 'voices'. Bottom row: drawings of a 'house' and 'car' produced on first presentation by sighted subjects M.W. and J.W. with only 30 min prior experience with simple geometrical shapes. The 'yacht' drawing was by a blind subject.

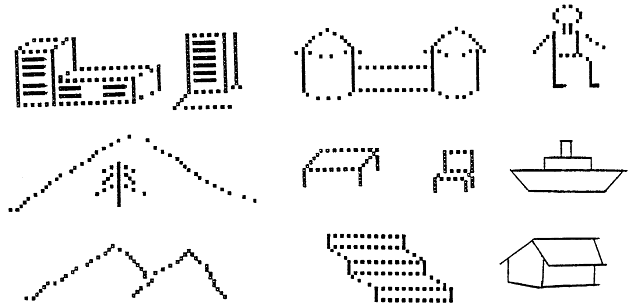


Figure 6. Set of additional cartoon images presented and analysed by four sighted and three blind subjects using a combination of masking and/or feature extraction. The images were identified verbally by the subjects in all cases. On initial presentation, all except the 'mountain range' and 'stairs' required segmentation for their identication.

hatched bars) and analyse the auditory representation of each segmented representation separately, for example, by masking either the top or bottom half of the figure and then comparing this against the auditory signature produced by the whole non-segmented figure. The subjects were scored on the basis of their first response for each set of explorations of the test figure. The responses for each test item are shown in figure 2 (black and horizontally hatched bars). Most of the test items were correctly identified without being deconstructed, while others, including the pentagon and hexagon which most subjects had failed to identify in experiment 2a, required deconstruction in some cases (figure 2, horizontally hatched bars). Our observations on the ability of subjects in identifying polygons strongly suggested that segmenting the image greatly facilitates the task of correctly analysing and identifying the visual target from its sound signature. Later observations also revealed that the time taken for recognition was substantially reduced. Indeed, once the subjects had successfully analysed and identified a visual target many were subsequently able to identify it correctly from the global sound pattern generated by a later presentation of the whole non-segmented figure.

(c) *Experiment 3: analysis of a three-dimensional cube viewed in perspective*

Two subjects were presented the sound signature associated with the classical figure of a Necker cube. Before the presentation each was told that it was a two-dimensional representation of a three-dimensional wire model of some object and they were required to analyse the sound pattern and, from this, reconstruct its form. A sighted subject (P.H.) was given a drawing pad and invited to sketch the shapes based on the inferences he had drawn from listening to the sound pattern during the process of reconstruction (figure 4*b*). Subject P.H. first reported that the 'whole figure was difficult'. The experimenter simplified the figure without informing the subject of the nature of the change (figure 4*a*). P.H. asked for the whole figure to be replayed, then drew it (figure 4*b*). Our blind (previously sighted) subject (B.K.) was able to describe the figure correctly as a 'cube' without it being deconstructed. When asked how he had analysed the shape he produced the diagrams shown in figure 4*c*.

(d) *Experiment 4: analysis and identification of cartoon drawings*

Three blind subjects (B.K., M.S. and M.P.) and four sighted subjects were tested on a set of simple cartoon, high-contrast, line drawings of objects and structures generally similar to objects or scenes that a previously sighted subject would have been familiar with prior to losing their sight. No information about their nature was disclosed other than a general statement that they were 12 simple line drawings (figures 5 and 6) depicting a variety of man-made artefacts or simple representations of urban or natural scenes, etc. to which the subject was to respond by giving a verbal description or a graphical representation. In presenting these figures we followed the testing protocols for complex figures described previously. Not infrequently a subject would ask for the presentation sequence to be interrupted while they mentally tested various hypotheses about the data they had already gathered. The response times taken to analyse and respond to the initial presentation of each of these novel items varied considerably both between items and between subjects, ranging from less than 3 min to over 6 min for some items (including interruptions for reflection). This variation in time was due to the number of occasions the subject requested a segmented portion to be repeated, as well as the number of different segmented representations of the figure he or she required to reconstruct the entire figure. Once the subjects had successfully analysed and identified an item, they were frequently able to do so again many months later with considerable saving in time. On first presentation one subject required 6 min 40 s to describe the line drawing representing a house correctly. When presented with the same item six months later with no intervening practice, he was at first unable to recognize its musical signature, but was almost immediately able to identify it when this was followed by the segmented version depicting only the vertical line components of the image (figure 5). On this occasion the whole process leading to recognition was accomplished within 25 s. The same time compression is observed with simple items. Identification of a rectangle or a triangle on first presentation took *ca*. 30 s and for a rectangle enclosing another rectangle *ca*. 1 min. On subsequent presentation, identification of the triangle and rectangle occurred in less than 2 s and for a rectangle enclosing another *ca*. 4 s or less, even when the presentations were separated by a considerable lapse of time (i.e. several months). Another feature of the response given by the subjects was that they would sometimes produce an accurate description of an item but name it incorrectly. In one notable case a subject was presented the image of a house (figure 5) which he described as 'a bridge-like structure with an open archway, which is topped with a trapezium whose ends are projecting from either side'. The error arose because the subject failed to take note of the windows.

4. DISCUSSION

These experiments demonstrate that blind and sighted subjects prevented from using sight can use sound representations of visual images to analyse the structure of novel, relatively complex visual forms. They are able to achieve this if the image is pre-processed to generate a set of simplified representations which they can analyse individually and then combine mentally to reconstruct the original image through a process which includes mental imagery, selective attention and hypothesis testing (Gregory 1980; Desimone *et al*. 1996; Farah 1996; Kosslyn & Sussman 1996). In some instances, while a subject failed to name an object correctly, they were nevertheless still able to produce an accurate description or sketch of the figure presented. This shows that we were not simply furnishing our subjects with a dictionary of complex sound patterns associated with specific objects but were providing them with tools which enabled them to analyse novel complex objects in terms of simpler building blocks whose associated sound signatures they had either learned previously or appeared to recognize naturally.

(a) *Global versus serial representation of visual forms encoded in musical form*

In figure 1, where the targets consisted of a variety of simple combinations of two lines, the subjects may either (i) have learned to analyse these shapes by identifying the constituent basic line components and their spatial arrangement within the figure or, alternatively, (ii) because they had previously been informed of the nature of the target after each incorrect response, they eventually learned to recognize the global sound pattern associated with each shape, without necessarily deconstructing and identifying the separate line components which formed the basis of its construction. This was sometimes evident for combinations of two oblique lines of different orientation where performance was more variable, particularly for the chevron figures (> and <). In the latter case, difficulties arose due to the presence of an ascending and descending scale of notes which were produced simultaneously as the tracker bar traversed the chevron. The subjects were able to resolve this 'analytically' by shifting the figure upwards towards the treble end or downwards towards the base end of the scale to try and highlight each of the component oblique lines. Alternatively, they were also able to carry out discrimination by resorting to a strategy of global processing in which they listened to the overall sound pattern (sound signature), which either closes in for > as the tracker bar moves from left to right or opens out as it traverses < .

Once the subjects had successfully identified certain critical features of a visual form, some of these appeared to become encoded in newly established detector modules which operate globally to assist in future recognition of the object. In P.H.'s analysis of the Necker cube, his identification of the rhombus components in the segmented figure (figure 4*a*) assisted the reconstruction of the whole figure and also his recognition of the cube in later presentations. Similarly, the simplified representation containing only the vertical line components of a house (figure 5) provided B.K. with a vital clue which facilitated recognition of the house on subsequent occasions. These and other examples suggest that there are two ways in which patterned sensory information may be analysed to yield global percepts. The first involves a laborious elemental process which serially reconstructs the figure from a set of elemental features, while the other operates 'globally' through a set of higher-order feature detectors which operate like gnostic circuits and which the elemental process may sometimes but not necessarily precede and then be supplanted by the higher-order global process. Recent studies on functional brain asymmetry using functional magnetic resonance imaging also point to a distinction between elemental and global forms of processing (Fink *et al*. 1997) as being ubiquitous within the human brain.

We thank the Royal Scottish Asylum and School for the Blind, the Northern Eye Institute and UMIST for financial support. In addition, we thank Krystynne Harvey and Dr Roy Blore for assistance in our experiments and Professors David Foster and Neil Charman for their critical appraisal of our manuscript.

REFERENCES

Barlow, H. B. 1972 Single units and sensation: a neuron doctrine for perceptual psychology? *Perception* **1**, 371–374.

Cronly-Dillon, J. R. & Persaud, K. 1998 Transforming visual images into musical forms. *IOVS Abstr. ARVO 98*.

Desimone, R., Miller, E., Chelazzi, L. & Lueschow, A. 1996 Multiple memory systems in the visual cortex. In *Cognitive neurosciences* (ed. M. Gazzaniga), pp. 475–486. Cambridge, MA: MIT Press.

Farah, M. J. 1996 The neural basis of mental imagery. In *Cognitive neurosciences* (ed. M. Gazzaniga), pp. 963–975. Cambridge, MA: MIT Press.

Fink, G. R., Halligan, P. W., Marshall, J. C., Frith, C. D., Frackowiak, S. J. & Dolan, R. J. 1997 Neural mechanisms involved in the processing of global and local aspects of heirachically organized visual stimuli. *Brain* **120**, 1779–1791.

Gregory, R. L. 1980 Perception as hypotheses. *Phil. Trans. R. Soc. Lond.* B **290**, 181–197.

Konorski, J. 1969 *Integrative action of the brain*. Amsterdam: Elsevier.

Kosslyn, S. M. & Sussman, A. L. 1996 Roles of imagery in perception. In *Cognitive neurosciences* (ed. M. Gazzaniga), pp. 1035–1042. Cambridge, MA: MIT Press.

Livingston, M. S. & Hubel, D. H. 1984 Anatomy & physiology of a colour system in the primate visual cortex. *J. Neurosci*. **4**, 309–356.

Mansur, D. L., Blattner, M. M. & Joy, K. I. 1985 Sound graphs: a numerical data analysis method for the blind. *J. Med. Syst. Jun*. **9**, 163–174.

Meijer, P. 1992 An experimental system for auditory image representations. *IEE Trans. Biomed. Engng* **39**, 112–121.

Mountcastle, V. B. 1978 An organizing principle for cerebral function: the unit module and the distributed system. In *The mindful brain* (ed. V. B. Mountcastle & G. M. Edelman), pp. 7–50. Cambridge, MA: MIT Press.

Regan, D. 1991 Spatial vision for objects defined by colour contrast, binocular disparity and motion parallax. In *Spatial vision* (ed. D. Regan), pp. 135–178. Basingstoke, UK: Macmillan Press Ltd.

Rolls, E. T. 1997 A neurophysiological and computational approach to the functions of the temporal lobe cortical visual areas in invariant object recognition. In *Computational and psychophysical mechanisms of visual coding* (ed. I. L. Harris & M. Jenkins), pp. 184–219. Cambridge University Press.

Rolls, E. T. & Treves, A. 1998 *Neural networks & brain function*. Oxford University Press.

Stark, L. & Ellis, S. R. 1981 Scanpaths revisited: cognitive models direct active looking. In *Eye movements: cognition & visual perception* (ed. D. F. Fisher, R. A Monty & J. Senders), pp 308–320. Hillsdale, NJ: Lawrence Erlbaum Associates.

Stryker, M. P. 1992 Elements of visual perception. *Nature* **360**, 301–302.

Wagg, H. J. 1932 *A chronological survey of work for the blind*. London, UK: National Institute for the Blind.

Zeki, S. (ed.) 1993 *A vision of the brain*. Oxford, UK: Blackwell Science.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.