# An Audiotactile Vision-Substitution System

**David Dewhurst**
HiFiVE
www.HFVE.org

## Abstract
This poster describes "work-in-progress" on "HiFiVE" (Heard & Felt Visual Effects), an experimental vision-substitution system that uses verbally-orientated audiotactile methods to convey visual images to blind and deafblind people.

## Keywords
Blindness, deafblindness, sensory-substitution, vision-substitution, audiotactile, haptic, braille.

## Introduction
There is often the need to convey general visual information to blind people. An existing approach is to use relief images e.g. tactile maps. While these are convenient for conveying unchanging two-dimensional images, the instantaneous production of vision substitution images is much more difficult to achieve. Devices can be devised that present other senses with information that includes aspects of sight, but other senses are not as powerful, or as able to comprehend such information [6].

The HiFiVE system aims to simulate the way that sighted people perceive visual features, rather than conveying raw optical measurements. The system highlights the features of visual images that are normally perceived categorically, and substitutes with coded sound effects and their tactile equivalents. It simulates the instant recognition of properties and objects that occurs in visual perception, by using the near-instantaneous recognition of phoneme sounds that occurs in speech. The user can instantly understand the colours and objects that are present in an image by hearing the coded phonetic sounds (and feeling the corresponding

tactile/braille effects). The system also conveys shape and location, and may in future versions convey texture, distance and "change".

When the system continuously changes the pitch and binaural positioning of the sounds, they appear to move, whether following a systematic path or conveying a specific shape.

It is intended that the project will lead to a practical application. The HiFiVE system is designed to run on a personal computer. It uses standard equipment for sound output, and low-cost force-feedback devices for haptic input and output.

## Coded phonetics

The HiFiVE system uses speech-like sounds, consisting of specific "coded phonetics" that can be rapidly interpreted in a categorical and linguistic way. These sounds convey the categorical properties of an image e.g. the colours, the distribution of those colours, recognised objects etc.

People can easily recognise speech-like sounds and rapidly assign meaning to them. Speech is a natural and efficient method of conveying information, it is perceived in a classified/coded way, and the information content is not greatly effected by distortion. Most people are able to retain several such spoken words in their short-term memory, including "nonsense" words [4]. The effort needed to learn the coded phonetics is low.

Visual properties are presented to the user via groups of CV (Consonant-Vowel) syllables, each of which is assembled from a consonant and a vowel sound selected from a set of 16 consonant (C) and 16 vowel (V) sounds. The user can recognise the sounds instantaneously, in the same way as they can recognise language. Certain visual properties, such as colour, tend to be perceived in a categorical way [1], but properties which are not "naturally" categorical, e.g. distance, can be assigned to bands of values.

The approach is best illustrated by a simple example, shown in figure 1. An image (A) is first reduced to 8 by 8 pixels (B). Then the pixels in each quarter of the image are set to one of two colours (C). Finally the image is presented via audio (D) and tactile (E) methods : for each quarter, one CV syllable conveys the two colours, and two CV syllables convey the layout of those two colours, to the detail shown in the pixelated image (C).

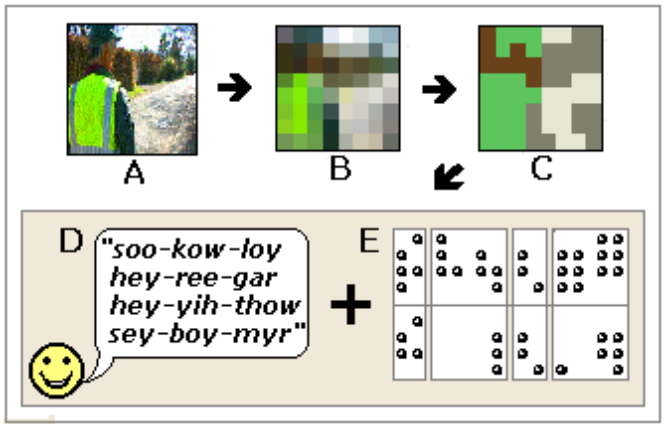


**figure 1.** Diagram illustrating conversion of an image into coded phonetics and braille.

For the top left square of 4x4 pixels in the pixelated image (C), the CV syllable "soo" conveys the two colours, and the two CV syllables "kow-loy" present the layout of the two colours as 4x4 pixels. The whole image is conveyed by the four spoken words "soo-kow-loy, hey-ree-gar, hey-yih-thow, sey-boy-myr", which completely describe the 8x8 pixels shown in (C).

## Tactile effects

The audio effects have tactile equivalents which can be presented using:- standard low-cost force-feedback devices to convey location and shape; and braille or other touch-based methods to convey the categorical properties.

As only 16 consonants and 16 vowel sounds are used, each CV syllable conveys one of 256 possible combinations (16 x 16). This corresponds to the number of different dot-patterns that can be presented on an 8-dot braille cell (programmable 8-dot braille cells are available commercially [5]). Figure 1 (E) shows one way in which the information given by the spoken sounds could also be conveyed via 12 braille cells. (Note that the conventions used in figure 1 are likely to change before the system is completed.)

A force-feedback joystick makes an effective pointing device with which to indicate areas of the image, as it can also be programmed to tend to position itself to one of a number of set positions, so that a "notchy" effect is felt as the joystick is moved, giving a tactile indication of location. A force-feedback joystick can also be moved by the system, pushing and pulling the user's hand and arm both to convey shapes (by tracing them out), and to indicate a location within an image. Standard force feedback devices are currently used to present tactile effects.

Using both audio and tactile modalities allows the user to spread the load of information to suit their needs and abilities, and could be used by deafblind people. Having a degree of "redundancy" of information may result in less tiring usage [3].

Some of the other features of the HiFiVE system are described below:-

### Audiotactile Tracers

The HiFiVE system can produce apparently-moving audio (and/or tactile) effects that trace out the shapes of features and identified objects within an image, by continuously changing the pitch and binaural positioning of the sounds (the "sound space" normally uses a high = high-pitch / low = low-pitch convention). These are known as "audiotactile shape tracers". In the tactile modality, tracer location and movement can be conveyed via a force-feedback device.

The audiotactile tracers can also move systematically through an area while outputting the properties of the corresponding parts of the image.

Moving effects are generally easier to mentally position than stationary ones.

### Combotex

It is planned that the fine detail of an area or entity will be conveyed via small, rapid fluctuations in the volume of the speech-sounds. These are referred to as "combotex" effects, as they combine the effects of small changes in brightness, colour, and distance, to give a single volume-conveyed "texture" effect. This simulates the effect found in vision whereby the overall properties of an area tend to be perceived categorically, and the minor variations in properties across it are perceived as a general texture. The user need not follow the precise detail conveyed by the combotex effects, but gets an impression of the general level of fine change occurring in an area. (Combotex effects have not been implemented at the time of writing.)

### Viewports

Sections of an image can be selected via a pointer, so that only those parts are conveyed, and at a higher resolution. These sections are known as "viewports", and the user can instruct a viewport to "zoom in" to higher levels of detail, as well as "zoom out" to convey a low-resolution representation of the whole image.

Viewports could be rectangular, hexagonal or "rounded" (circular or elliptical) and several viewports could be active at any moment. Viewports could be nested so that a "child" viewport moves within a "parent" viewport. One possible configuration would be to use nested viewports to simulate an eye's macula, fovea and/or areas of focal attention. (Only rectangular viewports are currently implemented.)

## Audiotactile Entities

As well as conveying general visual features, the system attempts to simulate the way in which features and objects are perceived in vision. Conveying basic properties does not do much to identify "entities", separate "figures" from the background, or assist with the other processes that occur naturally when people see things.

The simplest features are conveyed via shape-tracers, which can highlight identified shapes and features within a scene, emphasising the shape and layout of the feature or area.

## Audiotactile Objects

"Audiotactile objects" are items in an image that have been identified to the extent that they can be presented as specific entities rather than being described in terms of their properties, shapes and features. They are signified by special CV syllables and braille patterns.

Initially audiotactile objects will mainly be used from within pre-processed images, but in the future automatic recognition of certain objects may be possible.

A shape-tracer can present the shape of an object as found in an image. However it may be better to convey the distinctive "classic" shapes of objects, rather than the outline that happens to be formed by the object at its current distance and orientation, allowing "shape constancy" and "size constancy" to be simulated [2].

## Image pre-processing

When completed, the system will be able to convey prepared programmes of material. Pre-processing allows a sighted designer to select features and areas of an image, and specify the most appropriate methods of conveying them.

Pre-processed information could be embedded in the image pixels using "steganography", so that the images can also be viewed by sighted people using standard equipment and software. Images and movie sequences prepared in this way could be transmitted through currently available media, e.g. via DVDs, the Internet or broadcasts, enabling pre-processed sequences to be embedded in otherwise standard video material.

## Summary

When fully implemented, it is intended that the HiFiVE system will allow a continuum of visual features, from basic visual properties, to fully-identified objects, to be conveyed to blind and deafblind users. At the time of writing several of the features described above can be demonstrated, as well as "work-in-progress" on some of the others.

## References

[1] Berlin, B., and Kay, P. Basic color terms: Their universality and evolution. University of California Press, Berkeley, CA, USA, (1969).

[2] Coren, S., Ward, L.M., and Enns, J.T. Sensation and Perception (Fourth Edition). Harcourt Brace & Company (1994), 487-501.

[3] Jansson, G. Implications of perceptual theory for the development of non-visual travel aids for the visually impaired. Electronic Spatial Sensing for the Blind. (Eds. Warren, D.H., and Strelow, E.R). Matinus Nijhoff (1985).

[4] Miller, G.A. The magic number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63 (1956), 81-93.

[5] Programmable braille displays. See, for example, website http://www.kgs-america.com.

[6] Zahl, P.A. (Ed.) Blindness : Modern approaches to the unseen environment. Hafner Publishing (1963 – originally published 1949).