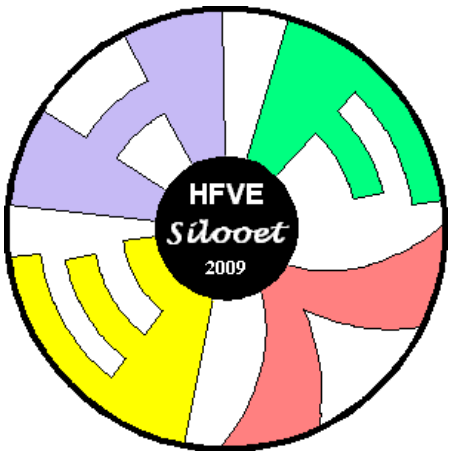


Sonic
Interaction
Design Workshop at the University of York

# HFVE

# Silooet

## Audiotactile Vision-Substitution Software


### ABSTRACT

This paper reports the latest developments of the HFVE (pronounced “HiFiVE”) vision substitution system; and the “Silooet” software, which implements some features of the system. The system uses verbally-orientated audiotactile methods to present features of visual images to blind and deafblind people. Included are details of presenting objects from prepared and live images; object-related moving effects and layouts; and embedding predetermined features in multimedia files.



### Keywords

HFVE, HiFiVE, Silooet, Silooets, blindness, deafblindness, sensory-substitution, vision-substitution, audiotactile, haptic, braille, Morse code.


### INTRODUCTION

The HFVE (Heard & Felt Vision Effects - pronounced “HiFiVE”) vision-substitution system has in the past been shown exhibiting areas of images via speech and tactile methods, with demonstration shapes also shown; and the “Silooet” (Sensory Image Layout and Object Outline Effect Transmitter) software implementation has been shown presenting predetermined object outlines and corners of items present in a sequence of images [1 & 2]. See the Appendices at the end of this paper for a fuller description of 1) Key Features; 2) the Silooet Software; and 3) Colour-to-Speech Mapping.

In this paper details of the latest developments are given, some of which were demonstrated at the SID (Sound Interaction Design) Workshop at the University of York. They include:- exhibiting objects found in live images and non-prepared media Figure 1; presenting object-related layouts and moving effect paths (including symbolic paths); and embedding predetermined features in multimedia files (both visual media, and “audio-only” formats e.g. “.MP3” files).

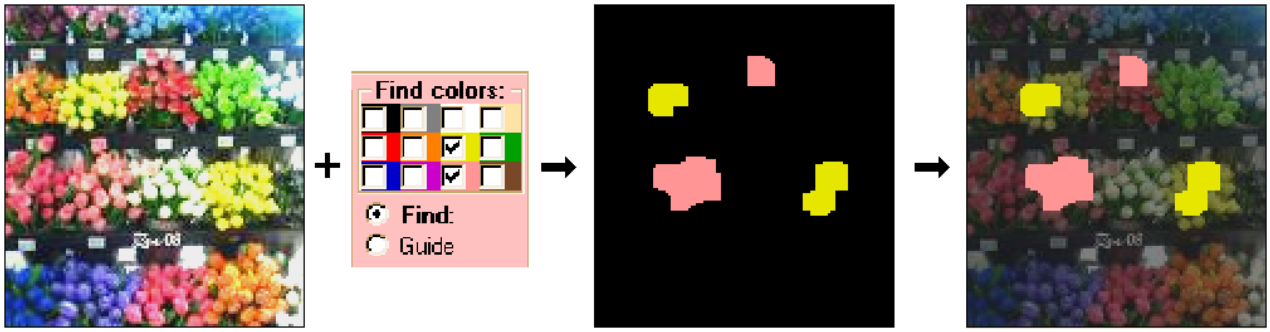


**Figure 1. Finding four yellow and pink objects. (From Silooet screenshots.)**

### Overview

The HFVE system aims to simulate the way that sighted people perceive visual features, by highlighting the features of visual images that are normally perceived categorically, and substituting 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 coded phonetic sounds (and feeling the corresponding tactile/braille effects). By smoothly changing the pitch and binaural positioning of the sounds, they can be made to appear to "move", whether following a systematic path or describing a specific shape. Such moving effects are referred to as “tracers”, and can be "shape-tracers", whose paths convey the shapes of items in an image; or "area-tracers", which systematically present the layout of properties of parts of an image. In the tactile modality, tracer location and movement are presented

via force-feedback devices; and categorically-perceived features via braille, or via Morse code-like “tapping” effects. The system also conveys corners and location, and may in future versions convey texture, distance and "change".

The approach differs from other haptic methods which allow people to explore a shape by moving around it under their own control, for example where a force-feedback device is used to model a groove on a surface. Instead, the HFVE system generally "conducts" a user around a shape, under the control of the system (albeit with user-controlled parameters), which might be less tiring and require less attention of the user than when requiring them to actively explore a shape. (The system could be used in combination with other systems that use different approaches.)

As the system outputs both audio and tactile effects, users can choose which modality to use; or both modalities can be used simultaneously, allowing more information to be presented during a certain period of time.

The software implementation of the HFVE system is called "Silooet", partly as an acronym, and partly because the system is effective when presenting an object as a silhouette i.e. conveying the the outline of the object while transmitting the colours of which it (and optionally the background) is composed. Silooet runs on a personal computer, and uses low-cost force-feedback devices for haptic input and output.

Other work in the field includes tone-sound scanning methods that have been devised for presenting text [4], and for general images [5]; and software for presenting audiotactile descriptions of pixels in computer images [6]. Audio description is used to supplement television etc. (The merits of other approaches are not discussed in this paper.)

### Potential applications

The HFVE project is not focused on a specific application, but is trying various methods for presenting sequences of visual images via touch and sound. More straightforward material (such as simple shapes, maps, and diagrams) can also be presented.

Possible applications include:- presenting shapes and lines for instructional purposes; adding shape, colour and texture data to diagrams; providing ad-hoc information to users wishing to know the colour and shape of an item; and for specific tasks such as seeking distinctively-coloured items (for which corresponding sets of parameters are provided).

## RECENT DEVELOPMENTS

(See the appendices for previously-reported details of the HFVE system [3].)

### Embedding audiotactile features in prepared images

(The term “object” is used to refer to a specific entity that is being presented, for example a person (face or figure), background, part of a diagram, etc.)

HFVE Silooet can present prepared media, in which predetermined shapes etc. are embedded in common multimedia formats (e.g. BMP, GIF or JPEG still images; AVI movie files; and WAV or MP3 audio files). These are produced via a straightforward procedure, and they can also be played on standard media players (without audiotactile effects).

The predetermined sequences are presented as a series of one or more “Views” that present a scene, the set of Views being referred to as a “Guide”. It has been found that narrative movies typically require about one View every 10-12 seconds or so, but this will of course vary considerably depending on the material being presented.

For each View, one or more objects can be presented. These are marked-up in bitmap images, each containing groups of non-overlapping objects. Generally one group of objects will present the background, and one or more further groups of objects will present the foreground and details Figure 2.

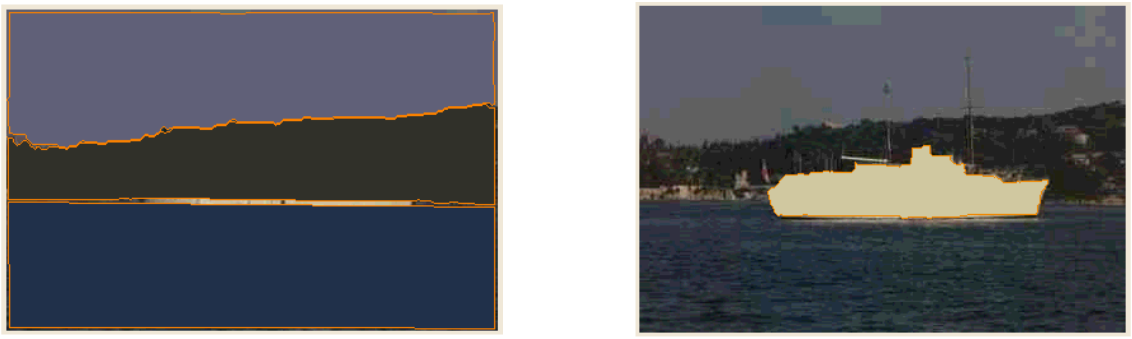

**Figure 2. An image marked-up with “objects”. This example has two non-overlapping groups, one for the distant background, and one for a single “object” (the ship) in the foreground.**

In creating a Guide, a sighted designer can select features and areas of an image, and specify the importance etc. of the entities within an image. “Paths” can be included to illustrate (a) the shape of objects and/or (b) the paths that objects move along in the scene being portrayed. For example, for a bouncing ball, the shape of the ball, and the path that it follows, can be presented. Paths generally refer to a particular entity, usually an object, though sometimes they refer to the whole View, for example when indicating the path of a panning or zoom shot.

Although the perimeter of an object is often the best path to present, this is not always the case, and it is useful to be able to include “diversion” paths, which take priority when the system is presenting the shape tracers representing the objects Figure 3. (See below for details of other types of object path tracer.)

**Figure 3. The speedboat has a “diversion path” marked on it (shown in purple), so that the tracer presenting the outline of the boat does not include the outlines of the crew.**

Once completed, a Guide can be bound to an MP3 or WAV audio file soundtrack. In a test, a sequence of approximately 150 seconds was presented via a Guide file (with embedded JPEG images) bound to a corresponding MP3 file of acceptable sound quality. The combined file was about 500 kilobytes in size.

Figure 10 shows the GUI for handling guided pre-determined sequences. It is used both to facilitate the creation of Guides, and for presenting them to the user.

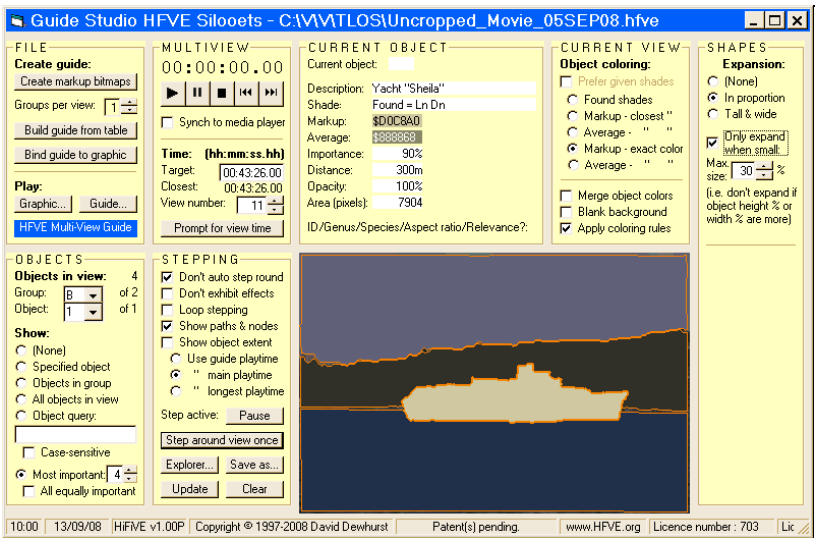


**Figure 10. Silooet's GUI for building and presenting prepared sequences.**

### Presenting the predetermined effects

The system can present the predetermined effects within Views to the user, for example by presenting the most important objects/features; a selected object/feature; etc. Alternatively the user can specify a keyword included in the descriptions of the objects/features, so that only items containing that keyword in their description are presented. For each object/feature, the system moves the “tracer” to describe the shape (or other path – see below) for the object/feature, as well as presenting related categorical information (e.g. colours etc.). The tracers can be sized to correspond to the item size and shape; or be expanded; or expanded only when an item is smaller than a certain size.

It is found to be very effective to “step” around the qualifying objects in a View, showing the most important objects and features, in order of importance.

### Finding objects

For non-prepared media (e.g. “live” images) the system has to “find” any objects to exhibit. The user can control the object selection : for example the checkboxes Figure 10 (15) provide a simple method of telling the system to look for particular colours. Figure 1 shows this process in action. Object detection is not a main focus of the project as it is a major separate area of research, but simple “blob”-detection methods are included, and in future standard face-detection facilities etc. [7] may be included.

Any found objects can then be presented in the same way as if they had been marked-up in a prepared image.

The GUI controls Figure 10 (14) allow the user to specify whether a Guide (if available) is used, or whether objects should be searched for. An “automatic” mode can be used where for example the system uses Guide information if available, then attempts to find appropriate objects, and if none are found then outputs simple area layouts.

### Object path tracers

Objects can be presented via audiotactile path tracers, for example presenting the outline of the object, as previously described. However they can follow one of several routes as described below and illustrated in Figure 5.

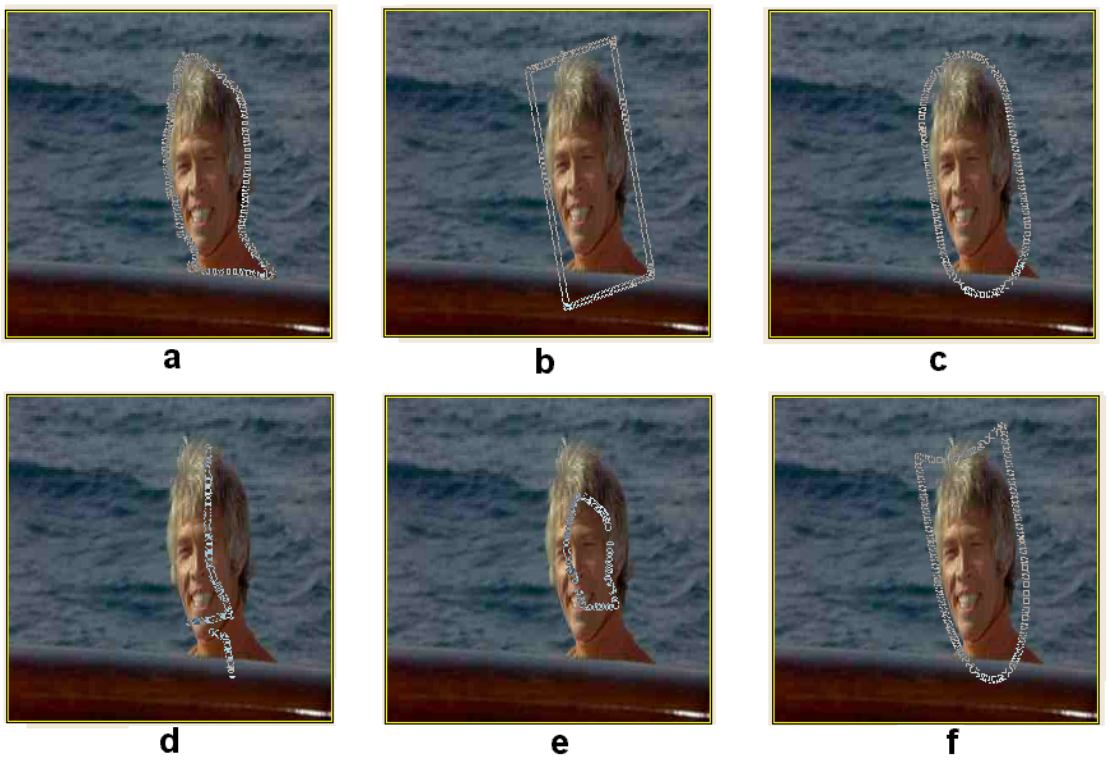


**Figure 5. Object path tracers.**

**Object Outlines**. The outline (a) of the object can be presented, or other "keylines".

**Object Frames**. The audiotactile path tracer can follow a path that “frames” the extent of the object. The frame can be rectangular (b), or be rounded at the corners (c), and sloped to match the angle of the object. The corners of the rectangular frames are not normally emphasised via effects, as such corners do not convey useful extra information.

**Object Medial Paths**. The tracer can follow the “centre-line” of an identified object (d). This is most effective for elongated objects, where the path travels in the general direction of the longest edge, but is not as effective for objects with no clear elongation : for them, a "circuit medial" (e) can be used, where the path travels in a loop centred on the middle of the object, and is positioned at any point along its route at the middle of the content found between the centre and the edge of the object.

**Symbolic Object Paths**. For identified objects, the system can present a series of lines and corners that symbolise the classification of those objects, rather than attempting to present the shape that the object currently forms in the scene. Human figures and people's faces (f) are examples of entities that can be effectively presented via symbolic object paths.

Currently, these would mainly be presented for prepared material. However image processing software can at the present state of development can perform some object identification, for example by using face-detection methods [7]. In such cases a standard symbolic shape Figure 6 (a) can be presented when the corresponding item is being output. Symbolic object paths are generally angled and stretched to match the angle and aspect ratio of the object being presented (b).

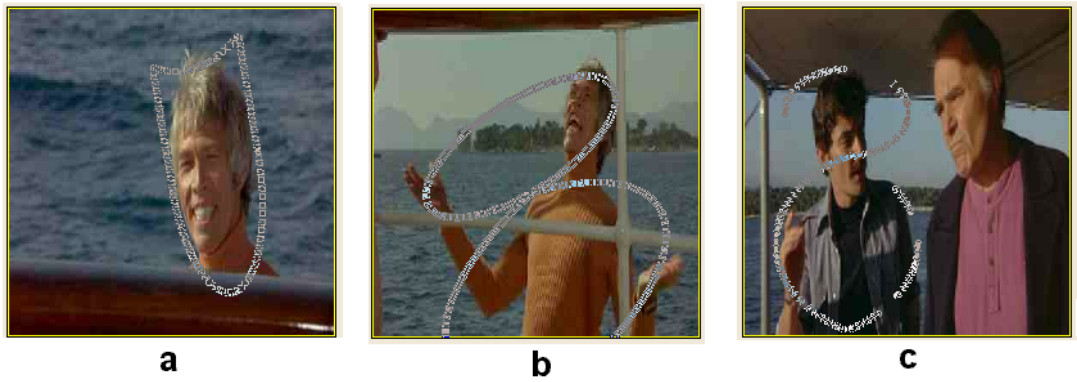


**Figure 6. Symbolic object paths.**

Basic symbolic shapes can be assigned to particular classifications/types, and embellishments can be added to represent sub-classifications e.g. a shape representing a face can be embellished to include features representing a pair of glasses, left-profile, right-profile etc. by having additional effects added. By using this approach, basic symbolic shapes of common object classifications can be easily recognised by beginners, with sub-classifications recognised by more experienced users. It was found to be useful to have sub-categories of symbolic shapes that show parts of an object. For example it is useful to provide a shape for the the top half of a human figure, head & shoulders, etc., as these are often what is present in a visual image (c).

### Object-related layouts

When presenting objects, a "layout" related to the object can be presented at the same time, for example by using a braille/tactile display, or by using speech codes.

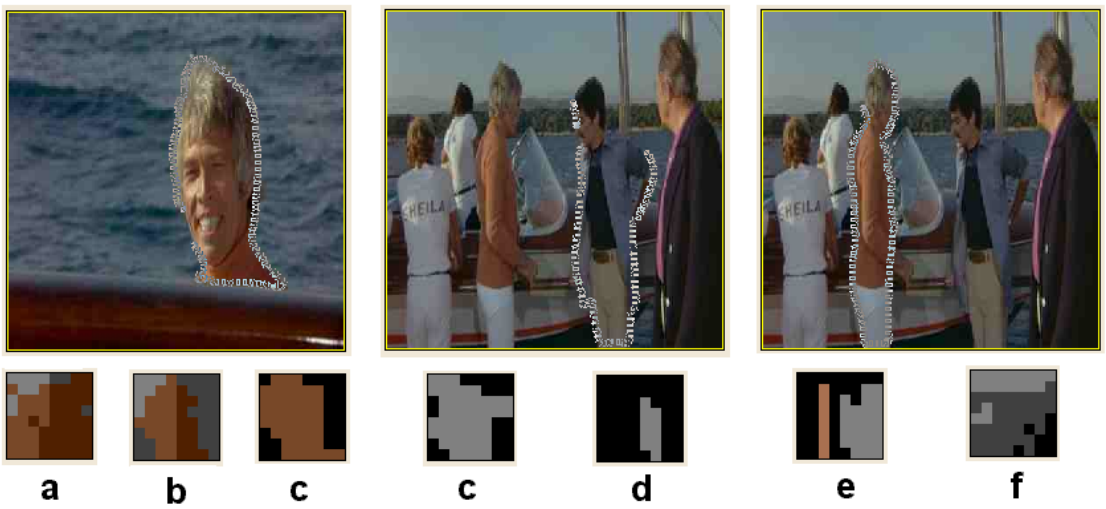


**Figure 7. Object-related layouts.**

The Layout content can comprise material selected from the following options:-

a) **Object content**. Because the shape of the object is known, the image content in only the area covered by the object can be presented Figure 7 (a), stretched to fill the layout area.

b) **Framed content**. The content of the rectangular frame enclosing the object can be presented (b) with the content "stretched" if necessary in one direction to use the full height and width of the frame.

c) **Figure/ground format**. The content of the frame can be presented using an approach which incorporates the perceptual concept of "figure/ground" i.e. the effect whereby objects are perceived as being figures on a background. If one object is being presented then the system can present the layouts as showing:-

i) The regions covered by the object within the "frame" enclosing the object (c) (stretched or un-stretched).

ii) The location of the object ("Figure") within the whole scene ("Ground") can be presented (d). When the system is "stepping" round the scene presenting the selected objects, the object-related layouts will appear and disappear as the corresponding objects are presented, giving the user additional/alternative information about their location, size and colour.

iii) All of the objects being presented within the whole scene can be presented (e). The objects are in the same location as for ii), however they are all presented at the same time. This method works best when not too many objects are being presented/"stepped" round, so that separate objects can be clearly perceived against the background.

The "figure/ground" object layout methods are generally most effective when a few foreground objects are being presented, rather than backgrounds (as the latter tend to occupy large areas).

(When particular colours are being sought and presented, and “framed” layouts are being presented i.e. not the whole scene, the frame can be set wider than the exact extent of the frame enclosing the object, otherwise the typical effect is for the layout to show mainly the colour found, while if the framing is set wider, then the context in which the found colour was located is also presented.)

d) **Symbolic layout format.** If the object has been identified, then symbolic layouts (using a similar concept to the symbolic object tracer paths described previously) can be presented, where the arrangement of dots is constant for particular object types (not shown).

### Compact object layouts

Layouts that are output as speech or morse (tap codes or audio) (i.e. not braille) tend to be long-winded. If object-related Layouts are being presented, a compact format can be used : only the location of the centre of the object can be presented, via a single “CV” syllable, the C & V giving the vertical and horizontal “coordinates” of the centre of the object. A second CV syllable can give the approximate size and/or shape of the object.

### Processing simple images

It is important that the HFVE system effectively handles simple images or visual materials containing a limited number of colour shades, and with clearly defined areas. Examples include certain maps, diagrams, cartoons etc., and these are often encountered, particularly in environments where a computer might be being used (e.g. office or educational environments). Though they can be handled via the standard/general routines that handle any type of image, it was found to be effective to have special processing for simple images.

Simple images can be detected by inspecting pixels and testing if the number of different shades is less than a certain number. An effective way of automatically determining the background colour shade is by finding the most popular colour shade along the perimeter of the image.

Such images do not require special optical filtering, as “objects” are already clearly defined in them Figure 10. The approach works effectively for simple images held in “lossless” file/image formats, e.g. “GIF” and “BMP” formats. For example diagrams drawn using Microsoft's Windows “Paint” program can be effectively presented in this way.

## SUMMARY

The HFVE Silooet software allows blind people to access features of visual images using low-cost equipment. The characteristics of the image (e.g. colour) that are conveyed by the sounds and tactile effects may not be perceived by a blind user in the same way as a sighted person perceives them. However the information may be useful in itself - for example knowing the colour of something can be useful to a blind person, even if they have never perceived and cannot imagine the property of colour.

The HFVE system's aim of presenting the features of successive detailed images to blind people, is challenging. Some users might only use the system for accessing more straightforward material, for example maps and diagrams; or it could be used for instructional purposes, to present certain shapes. Most of the features of the system can now be demonstrated. It remains to be seen which aspects of the system are most effective.

## APPENDIX 1 : KEY FEATURES OF THE SYSTEM

(This section recaps some of the previously-published features of the HFVE system [3])

The Silooet software can present images as arrangements of pixels known as “layouts” Figure 8; and as the outlines of the objects in images Figure 9. In both cases the categorical properties of the areas are exhibited via groups of coded CV (Consonant-Vowel) speech sounds; or as braille dots; or via Morse code-like “tapping” effects.

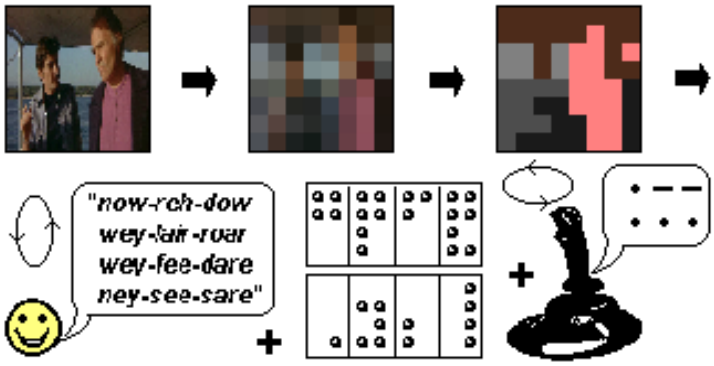


**Figure 8. Diagram illustrating the conversion of the "layout" of an image into coded phonetics, braille, and "tap codes" (while moving audiotactile "tracers" describe key shapes).**

Figure 8 shows an example of an image being reduced to two colour shades in each image quarter / "panel", and some of the corresponding speech sounds, braille cells and tapping codes, which describe the 8 by 8 pixels shown. "Layouts" can be used to present any section of an image.

Alternatively, if the shapes of entities in an image can be determined (Figure 9), then audiotactile "shape-tracers" can exhibit those shapes and their corners, at the same time as the categorical properties are being presented. (Alternatively, if shapes are undetermined, "area-tracers" can be used to show which part of a Layout is currently being presented.)

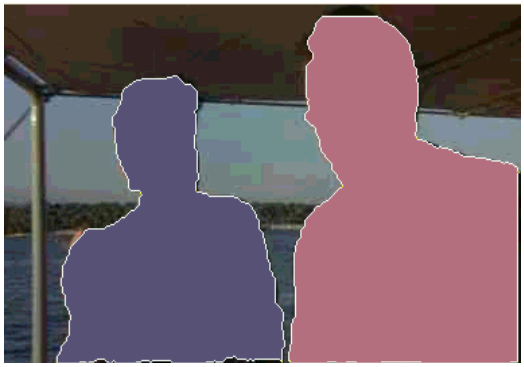

**Figure 9. Identified entity shapes, whose outlines and corners can be presented via audiotactile "shape tracers".**

### Tactile effects and user interaction

Programmable 8-dot braille cells are available commercially, and are an effective way of presenting categorical data to blind people who are able to read braille.

Alternatively, coded pulses can be induced on a standard force-feedback device, presenting "tap codes" to the user. It is found that short, evenly-spaced pulses of two force levels are more straightforward to interpret than conventional Morse-code timings. As with conventional Morse-code, shorter combinations can be assigned to more common data. "Tap codes" are relatively slow when compared to speech or braille, but may be useful for deafblind users who cannot read braille.

A force-feedback joystick makes an effective control and pointing device, as it can also be moved by the system, pushing and pulling the user's hand and arm, tracing out any shapes (and highlighting corners) that are to be presented.

### Highlighting corners

When sighted people see things, corners (such as the vertices of a rectangle) produce a considerable effect in giving the impression of the shape, and it is useful if this can be reproduced in the audio and tactile modalities. It is particularly important that vertices are highlighted when they are essential features of an entity but are not sharp angles, as is the case for an octagon, for example.

In both the audio and tactile modalities, to emphasise a vertex the system can add a tactile effect, and/or momentarily stop the movement of the "shape-tracer" (for example by momentarily stopping a moving force-feedback joystick); and in the audio modality the system can alter the volume of the sound and/or add an audio effect.

## APPENDIX 2 : USING THE HFVE SILOOET SOFTWARE

Using the HFVE system can be summarised by describing the numbered features shown on the screenprint of Silooet's main GUI Figure 10, which is shown exhibiting a simple diagram, one object at a time.

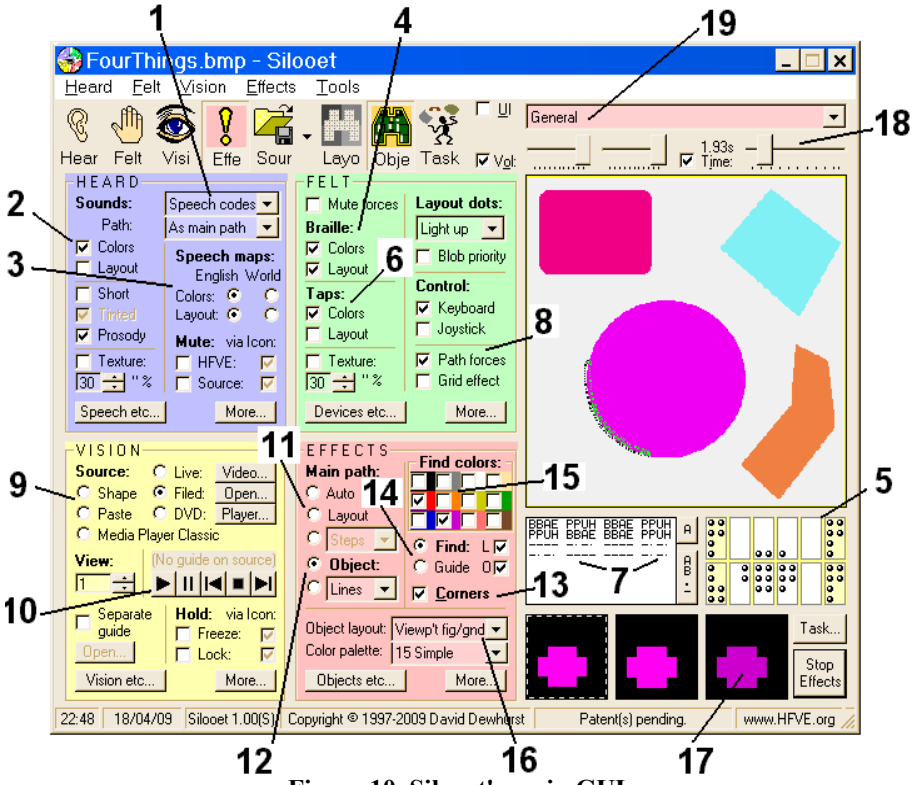


**Figure 10. Silooet's main GUI.**

With reference to the numbered features:-

| | |
|---|---|
| **1** | The primary method of outputting sounds is via speech-like sounds (see Appendix 3 below), as these are very efficient. Alternative sounds include tone-like and morse-like sounds. |
| **2** | The speech (or morse) sounds can present the colours of the entity being exhibited, and/or the "layout" of the colours in the entity. For colours, compact mapping methods are available. |
| **3** | Two main methods of speech mapping are available, and these are described in Appendix 3 below. |
| **4-7** | Tactile information can also be output, namely braille (4 & 5), and tap codes (6 & 7). Braille (5) is very effective for presenting layouts (central area), as it corresponds directly to the layout, and no mapping/coding is required. Braille can also present colours in a compact format (in the right- and left-hand cells, with 2 colours to each cell). |
| **8** | The user will normally want path forces presented on their force-feedback device. These can present, for example, the outline of an object. |
| **9** | The images can be obtained from a variety of sources. |
| **10** | For prepared material, a series of images can be presented using "player"-style controls. |
| **11-12** | The path that the tracer follows can be selected from "Layout" (11) (i.e. area tracers) and "Object" (12). For "Layout", the path can step, or move smoothly around the area being presented. For "Object" (12), the path can present one of several routes, which are described above. (An "automatic" mode is also available.) |
| **13** | It is important to present corners effectively when conveying the shape of an object. |
| **14** | For non-prepared material (e.g. "live" images) the system always has to "Find" any objects. However for prepared material, predetermined details of objects (held in a "Guide") can be embedded in media (or held in a standalone file), giving the exact object shape etc. to be presented. |
| **15** | If the system is in "Find" mode, then the user can control the object selection. For example the checkboxes (15) provide a simple method of telling the system to look for particular colours. |
| **16** | When objects are being presented, the system can also exhibit object-related layouts, the content of which can comprise material selected from one of several options, which are described above. |
| **17** | In Figure 10, the system is showing the figure/ground arrangement of the currently-presented object within the overall scene. The layout is presented in braille format (5) (and could also be presented as speech or tap codes). |
| **18** | The volume and timing of the effects is user-controlled. |
| **19** | Users can rapidly switch between sets of parameters and options by selecting particular activities/tasks. |

## APPENDIX 3 : COLOUR-TO-SPEECH MAPPING

The system normally uses two colour shades (e.g. “blue and red”) when presenting an area or entity, allowing the shades to be given via a modest number of corresponding effects. The two-colour-shade approach can simply present the two associated shades, or have the effect of painting a picture with one colour on a differently-coloured background.

A 15-colour-palette format allows two colours to be easily presented via a single “CV” syllable; via a single 8-dot braille cell; or via “tap codes” of 8 or less pulses.

One approach to mapping colour to speech is to base the speech on English colour names, as shown (for the first fifteen shades of the colour palettes) in Table 1 (phonemes shown in “Arpabet” format).

**Table 1. Colour-to-speech mapping for the first 15 colour shades, using “English” colour coding.**

| Colour Name | Short Name | “Split” Name | Vowel Sound | Phon-emes |
|---|---|---|---|---|
| 1 Red | Reh | S-eh | bed | S-EH |
| 2 Orange | Joh | J-oh | bob | JH-AA |
| 3 Yellow | Yow | Y-ow | boat | Y-OW |
| 4 Green | Gee | N-ee | bee | N-IY |
| 5 Blue | Boo | B-oo | boot | B-UW |
| 6 Purple | Puh | Z-uh | book | Z-UH |
| 7 Pink | Pih | P-ih | bid | P-IH |
| 8 Brown | Bow | R-ow | brow | R-AW |
| 9 Black | Bah | K-ah | back | K-AE |
| 10 White | Wuy | W-uy | buy | W-AY |
| 11 Mid Grey | Mey | M-ey | bay | M-EY |
| 12 Dark Grey | Dey | D-ar | bar | D-AR |
| 13 Light Grey | Ley | L-air | bear | L-XR |
| 14 Turquoise | Toy | T-oy | boy | T-OY |
| 15 Dk.Brown | Dow | V-ore | bore | V-OR |

This “English” coding makes use of the relatively large number of vowel sounds available in English; and the surprisingly varied set of vowel (V) and consonant (C) sounds used in the English names for basic colours. The “Short Name” column shows single syllable “names”, of C-V format, that are similar to the English colour names. Two syllables can be used to present colour pairs, for example “boo-yow” for “blue and yellow”. These sounds can be further contracted : notice how the vowels for the short-format names of the 11 “Basic Colour Categories” are all different. By changing the C or V of certain colours so that every C and V is different (see “Split Name” column), single-syllable colour pairs can be produced by taking the consonant of the first colour and adding the vowel of the second colour, for example “bow” for “blue and yellow”.

The mappings shown in Table 1 are based on the phonemes found in spoken English, which could make comprehension difficult for people whose mother-tongue is not English, especially as they will not be able to use the context to help with understanding, as would occur with normal speech. Many languages use only five different vowel phonemes, centred around the vowels A, E, I, O and U.

The “International” format shown in the table in Table 2 can use just five consonant and five vowel sounds (alternatives shown in brackets). For example the colour pair “blue and yellow” would be presented as “doh-reh”. A disadvantage of using this double-syllable format is that the sounds take longer to speak, though the individual syllables can be spoken faster than for the “English” format, as the consonant and vowel sounds are slightly shorter on average, and more distinctive.

**Table 2. Colour-to-speech mapping using "International" coding.**

| Vowel sound 2nd ↓ | Consonant sound 1st → S (w) | R (L) | K (g) | N (m) | D (b,p,t) |
|---|---|---|---|---|---|
| I | Lt.Purple | Lt.Brown | White | Cream/24 | (Special) |
| E | Pink | Yellow | Light Grey | Lt.Green | Light Blue |
| A | Red | Orange | Mid Grey | GrnYell/23 | Turquoise |
| O | Purple | Brown | Dark Grey | Green | Blue |
| U | Dk.Purple | Dk.Brown | Black | Dk.Green | Dark Blue |

One advantage of the "International" format is that users can get an impression of the lightness or darkness of an area from the vowel sounds alone.

For people who can easily distinguish English phonemes out of context, the "International" format vowels can be "coloured", where the standard five vowel phonemes are replaced by similar "R"- or "Y"-sounding vowels when "warm" or "cool" colours respectively are being presented. For example "blue and yellow" could alternatively be presented as "doy-rare" (instead of "doh-reh"), allowing a user to know that the first colour (blue) is "cool" and the second colour (yellow) is "warm", without fully interpreting the coded sounds.

For 5-level monochrome shades, the two levels can be conveyed via a single syllable even in "International" format, with the consonant sound conveying the first level and the vowel sound conveying the second level.

### Cultural factors

The given mappings can be changed to allow for cultural issues (for example to avoid phoneme combinations that happen to produce unacceptable words). The speech sounds can also be adjusted to match the phonemes found in the users' languages.

## REFERENCES


1. Dewhurst, D.: An Audiotactile Vision-Substitution System. In: Proc. of First International Workshop on Haptic and Audio Interaction Design Vol.2, pp. 17-20 (2006)
2. Dewhurst, D.: "Silooets" - Audiotactile Vision-Substitution Software. In: Proc. of Third International Workshop on Haptic and Audio Interaction Design Vol.2 (2008)
3. U.S. Patent Appl. No. US 2008/0058894 A1
4. Fournier d'Albe, E. E.: On a Type-Reading Optophone. In: Proc. of the Royal Society of London. Series A, Vol. 90, No. 619 pp. 373-375 (Jul. 1, 1914)
5. See website http://www.seeingwithsound.com.
6. See website http://www.ifeelpixel.com.
7. Viola, P., Jones, M.: Robust real-time object detection. In: IEEE ICCV Workshop on Statistical and Computation Theories of Vision, Vancouver, Canada (2001)



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


**23rd April 2009**