Accessing Audiotactile Images with HFVE Silooet

David Dewhurst

www.HFVE.org
daviddewhurst@hfve.org

Abstract. In this paper, recent developments of the HFVE vision-substitution system are described; and the initial results of a trial of the "Silooet" software are reported. The system uses audiotactile methods to present features of visual images to blind people. Included are details of presenting objects found in prepared media and live images; object-related layouts and moving effects (including symbolic paths); and minor enhancements that make the system more practical to use. Initial results are reported from a pilot study that tests the system with untrained users.

Keywords: Vision-substitution, sensory-substitution, HFVE, Silooet, blindness, deafblindness, audiotactile, haptics, braille, Morse code.

1 Introduction and Background

HFVE Silooet software allows blind people to access features of visual images using low-cost equipment. This paper will first summarise the system, then report on the latest developments, and finally describe initial results of a pilot study that tests the system.

At the 1st HAID Workshop, the HFVE (Heard & Felt Vision Effects - pronounced "HiFiVE") vision-substitution system was shown exhibiting areas of images via speech and tactile methods, with demonstration shapes also shown [1]. At the 3rd HAID Workshop the "Silooet" (Sensory Image Layout and Object Outline Effect Translator) software implementation was shown presenting predetermined object outlines and corners of items present in a sequence of images [2]. Recent developments include:- presenting found or predetermined objects; symbolic moving effects and layouts; and minor enhancements such as an adapted joystick, and methods for rapidly creating and presenting simple images and diagrams in audiotactile format. A pilot study of the system has recently commenced.

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. The main approach used differs from other methods which, for example, allow people to explore a shape or image by moving around it under their own control. Instead, the HFVE system generally "conducts" a user around an image, 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 an image. (The system could be used in combination with other approaches.)

M.E. Altinsoy, U. Jekosch, and S. Brewster (Eds.): HAID 2009, LNCS 5763, pp. 61–70, 2009.

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, theatrical performances etc. (The merits of other approaches are not discussed in this paper.)

2 System Features

The HFVE system aims to simulate the way that sighted people perceive visual features, and the approach is illustrated in Fig. 1:-

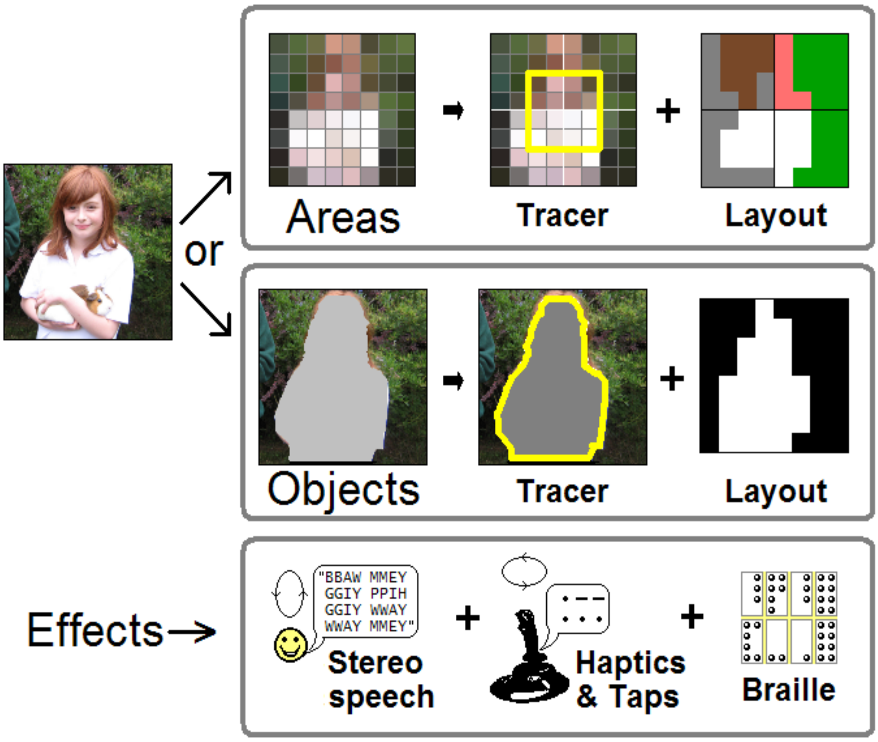

Fig. 1. Presenting features of a visual image as Area or Object "tracers" and "layouts"

For any image, or section of an image, the property content (colour, texture etc.) of "Areas" can be presented; or the properties of identified "Objects". (The term "object" is used to refer to a specific entity that is being presented, for example a person, a person's face, part of a diagram, a found coloured "blob" etc., whether found by the system, or highlighted in material prepared by a sighted designer.) For both "Areas" and "Objects", the information is presented via moving audiotactile effects referred to as "Tracers" - for Areas, the tracer path shows the location of the areas, and for Objects the path shows the shape, size, location and (if known) the identity of the objects. "Layouts" present the arrangement of (usually two) properties within an Area or Object, and normally use a regular grid-like format Fig. 1.

The paths of the tracers are presented:- via apparently-moving sounds, that are positioned in "sound space" according to location, and pitched according to height; and via a moving force-feedback device that moves/pulls the user's hand and arm – in both modalities the path describes the shape, size and location (and possibly identity) of the Areas or Objects. As the system outputs both audio and tactile effects, users can choose which modality to use; or both modalities can be used simultaneously.

The properties (colours, textures, types etc.) of the Areas or Objects are either presented within the audiotactile tracers, or separately. In the audio modality, speech-like sounds generally present categorical properties (e.g. "boo-wuy" or "b-uy" for "blue and white"). In the tactile modality, Morse-code like "taps" can be presented on a force-feedback device, or alternatively a separate braille display can be used Fig. 1. The "layout" of properties is best presented on a braille display, though, as will be described later, there are practical ways of presenting certain object layouts via speech or taps. (Appropriate mappings for speech etc. have previously been reported [1,2,3]). Until recently, layouts were used for presenting the arrangements of properties in rectangular Areas. However the content of objects can be also presented via layouts.

A key feature of the system is the presenting of corners/vertices within shapes, which initial tests show to be very importing in conveying the shape of an object. Corners are highlighted via audiotactile effects that are included at appropriate points in the shape-conveying tracers.

Although one possible tracer path for presenting an object's shape is the object's outline Fig. 1, other paths such as medial lines and frames can be used Fig. 5. "Symbolic Object Paths" are found to be effective, as they present the location, size, orientation and type of object via a single tracer path.

3 Recent Developments

3.1 Presenting Predetermined and Found Objects

HFVE Silooet can present both objects found in images "on the fly", and predetermined objects from prepared media. Fig. 2 illustrates the process:- for non-prepared media (e.g. "live" images) the system attempts to Find (a) objects according to the user's requirements, and builds a "Guide" (b) of the found objects. Alternatively a previously-prepared "Guide" (b) can be used to give the objects and features that are present. Finally, the corresponding Effects (c) are presented to the user.

The system can use predetermined Guide information if available, otherwise it attempts to find appropriate objects, and if none are found, it outputs area layouts.

Fig. 2. The HFVE system processing approach

For prepared media, a sighted designer can highlight the entities present in images, and identify what they are, their importance etc. Such predetermined entity information can be embedded in files of common multimedia formats (e.g. MP3 or AVI). The combined files 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 comprise a scene, the set of Views being referred to as a “Guide”. For each View, one or more objects can be defined. These can be marked-up on bitmap images, each bitmap containing groups of non-overlapping objects. Usually one bitmap will contain the background objects, and one or more further bitmaps will handle the foreground and details Fig. 3.

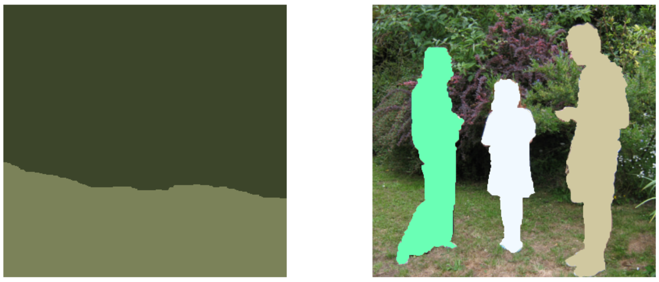

Fig. 3. An image marked-up with “objects”. This example has two groups of non-overlapping objects : one for the background, and one for the “objects” (the figures) in the foreground.

Extra “Paths” can be included to illustrate the route 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.

A Guide can be bound to an audio file soundtrack (e.g. MP3 or WAV file). In a test, a sequence lasting approximately 150 seconds was presented via a Guide file bound to a corresponding MP3 file of acceptable sound quality. The combined file was about 500 kilobytes in size.

The system can present the most important objects and features. Alternatively the user can specify a keyword, so that only items that include the keyword in their description are presented. For each item, the system moves the “tracer” to describe the shape for the item (for example via an outline tracer, or via a symbolic tracer etc.), as well as presenting related categorical information (e.g. colours, texture etc.). The tracers can be sized to correspond to the item's size and shape; or be expanded; or expanded only when an item is smaller than a certain size.

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

For non-prepared media, the system has to look for objects to exhibit. The user can control the object selection. For example the check-boxes Fig. 4 provide a simple method of telling the system to look for particular colours. More precise parameters (e.g. specifying size, shape etc.) can be given elsewhere. Fig. 4 also shows the check-boxes for requesting that certain types of object (faces, figures, or moving objects) are looked for. (Advanced object recognition is not currently implemented for live images, but the controls could be used to select particular objects types from prepared media – for example people's faces could be requested to be presented.) Object detection and identification is not a main focus of the project as it is a major separate area of research, but simple blob-detection methods are currently implemented, and in future standard face-detection facilities etc. [7,8] may be included.

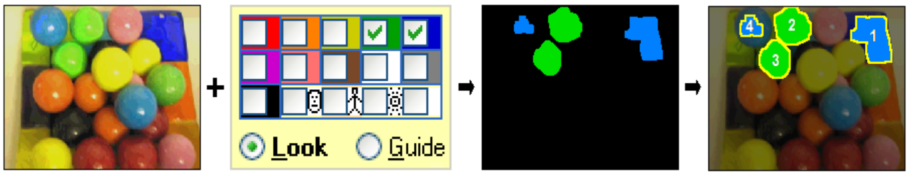

Fig. 4. Finding four blue or green objects, and presenting them in size order

Any found objects can be presented as audiotactile effects in the same way as if they had been marked-up in a prepared image – though the system has to decide which of the found objects (if any) are presented (i.e. which objects best match the user-controlled parameters), and their order of importance (e.g. by order of size).

3.2 Object Tracer Paths

The object tracer paths can follow several different types of route, and these are described below and illustrated in Fig. 5.

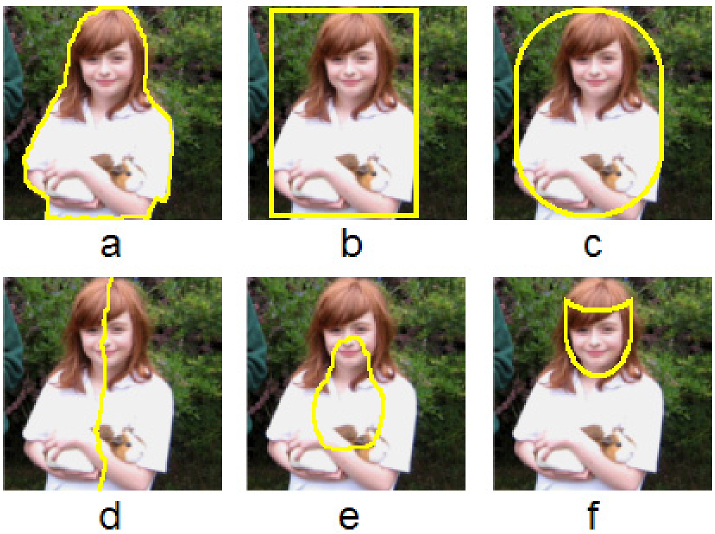

Fig. 5. "Object" tracer path types

The outline (a) of an “object” can be presented, as previously described. Alternatively the audiotactile 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 tracer can follow the “centre-line” of an 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 Tracer Paths. For identified objects, the system can present a series of lines and corners that symbolise the classification of those objects, rather than presenting the shapes that the objects currently form in the scene. Fig. 6 shows example symbolic paths. Human figures (a) and people's faces (b) are examples of entities that can be effectively presented via symbolic object paths. It is best if the paths are such that they would not normally be produced by the outline of objects, for example by causing the paths to travel in the reverse direction to normal.

Currently, symbolic object tracer paths would mainly be displayed for prepared material. However image processing software can at the present state of development perform some object identification, for example by using face-detection methods [7,8]. In such cases a standard symbolic shape Fig. 6 (b) can be presented when the corresponding item is being output. An “X”-shaped symbolic object path representing “unknown” is provided Fig. 6 (c), allowing “unidentified” objects to be handled in the same way. (Alternatively the system could revert to presenting the outline or other shape when an unidentified object is processed.) Symbolic object paths are generally angled and stretched to match the angle and aspect ratio of the object being presented.

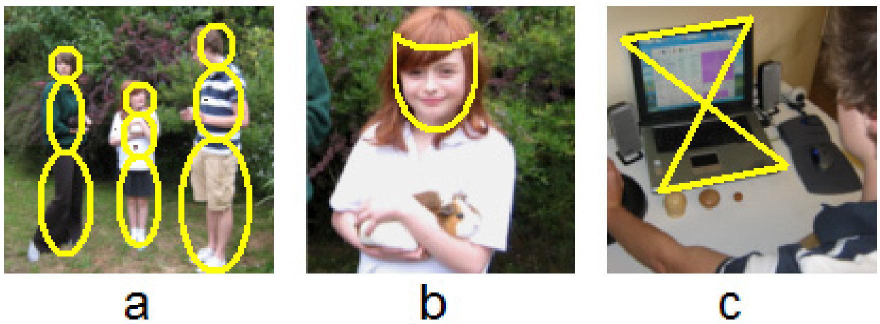

Fig. 6. Symbolic object tracer paths

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 and shoulders etc., as these are often what is present in a visual image.

3.3 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 display.

Fig. 7. Object-related layout types

Because the shape of the object is known, the image content in only the area covered by the object can be presented, spread to fill the layout area. Alternatively the content of the rectangular frame enclosing the object can be presented, with the content stretched if necessary in one direction to use the full height and width of the frame Fig. 7 (a & b).

Alternatively, 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 the regions covered by the object within the "frame" enclosing the object (optionally stretched to match the layout dimensions) Fig. 7 (c & d); or the location of the object ("Figure") within the whole scene ("Ground") can be presented (e) - when the system is "stepping" round the image, presenting the selected objects, the highlighted objects within the layout will appear and disappear as the corresponding objects are presented, giving the user information about their location, size and colour etc. (Alternatively all of the objects being presented within the whole scene can be displayed simultaneously Fig. 7 (f).)

If object types can be identified, then "Symbolic Layouts" can be presented (using a similar approach to that used for Symbolic Object Paths), wherein the arrangement of dots is constant for particular object types (as previously reported [9]).

When objects of particular colours are being looked for and presented, and "framed" layouts are being used (i.e. not the whole image), the frame can be set wider than the exact extent of the frame enclosing the found object, otherwise the typical effect is for the layout to show mainly the found colour. By setting the framing wider, the context in which the found colour is located is also presented.

Layouts that are output as speech or Morse (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(onsonant) and V(owel) giving the vertical and horizontal "coordinates" of the centre of the object. Additional coded "CV" syllables can give the approximate size and/or shape, colour etc. of the object if required.

3.4 Processing Simple Images

It is important that the HFVE system effectively handles simple images or other visual materials containing a limited number of colour shades, and with clearly defined coloured regions. 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 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 value. An effective way of automatically determining the "background" colour shade is by finding which colour shade predominates along the perimeter of the image.

Such images do not require special optical filtering, as "objects" are already clearly defined in them, and these can be presented. The approach works well for simple images held in "lossless" image file formats, e.g. "GIF" and "BMP" formats. For example diagrams drawn using Microsoft's Windows "Paint" program can be effectively presented in this way, or a special facility can be provided to allow shapes etc. to be rapidly drawn and then immediately presented in audiotactile format.

4 Pilot Study

A pilot study/trial with untrained users has recently commenced. The prototype Silooet software was installed on an ordinary laptop computer, and separate speakers and two types of low-cost force-feedback devices were used, namely a Logitech Force Feedback Mouse, and an adapted Microsoft SideWinder Force Feedback 2 joystick Fig. 8.

a b c

Fig. 8. The pilot study equipment (a); the force feedback mouse and adapted force feedback joystick, with alternative handles (b); and the main GUI for the trial Silooet software (c)

The standard joystick vertical handle configuration is designed for computer games, flight simulators etc. The handle was detached and some wires de-soldered, so that four control buttons, the slider, and potential twist action remained (the original handle could easily be re-fitted if required). Three alternative wooden handles (roughly the size of an apple, a door-knob, and a large grape) Fig. 8 were tested.

The two force-feedback devices tested in the study are low cost but not current (though relatively easily obtainable) - testing the system with current force-feedback hardware may be worthwhile. Braille output took the form of simulated braille cells shown on the main GUI, which is obviously not suitable for blind testers (a programmable braille display has not been implemented at the time of writing).

The system was presented to several sighted participants, of different ages, in informal trial/interview sessions, and their impressions were noted and performances assessed. Standard shapes were presented at various speeds and sizes, in audio, tactile, and combined audiotactile format. Sequences of filed and live images were tested.

The initial findings are:-

- After a few minutes practice, participants could typically recognise more that 90% of standard demonstration shapes. Shapes with many features, representing more complex objects, were more difficult to interpret directly, but could be recognised after several repetitions, when interspersed with the standard shapes.
- Emphasised corners are essential for recognising shapes.
- Standard shapes could be recognised at Full-, 1/2-, and 1/4- image diameters with no difficulty. Recognition became more problematic at 1/8 diameter sizes. This finding suggests that small shapes should be automatically enlarged, perhaps with audiotactile cues given to indicate that this has been done.
- Of the two haptic devices tested, their was no clear preference. The force feedback mouse had a more suitable handle configuration for the HFVE Silooet application and gave very accurate positioning (but users needed to hold it lightly by their fingertips), while the joystick gave more powerful forces. All participants preferred one of the replacement joystick handles to the standard vertical handle : the door-knob handle was preferred by a child, while older testers preferred the apple-sized handle. A tennis ball cover was added to this to provide a softer surface, and this was the most preferred joystick configuration. (The standard joystick handle was usable, but not as effective for presenting shapes.)
- Audiotactile output (i.e. both modalities together) generally worked best. Audio (speech) was most effective for categorical information, and tactile was most effective for comprehending shapes.
- None of the testers liked the Morse-code like effects (either audio or tactile "taps"), but this could be due to their lack of familiarity with Morse. The speech-based categorical effects and braille layouts are more immediately accessible.
- A "novice" mode was requested, wherein the colours (and recognised objects) are not coded, but spoken in full (this was what was originally planned [9]).
- Sudden moves of the joystick, when it was repositioning to present new objects, were found distracting by certain testers, but others felt it gave a clear indication that a new object was being presented. Some clarification via audiotactile effects is needed, perhaps with several styles of repositioning being made available.

- The most liked features were recognising standard shapes; corners; symbolic tracers; using the system to find things in live images; and the "quick draw" / simple image feature.
- The least liked features were Morse-style output, and very small shapes.

The trial/interview sessions lasted about an hour each. Participants reported feeling tired after this period, though that may have been due to the intensity of the sessions and their unfamiliarity with the system. The effects of longer-term use of the system has not yet been assessed.

At the time of writing testing has only recently commenced, and all of the testers have been sighted. It is hoped that fuller results, and the feedback of blind testers, can be reported at the HAID workshop.

5 Conclusions and Future Development

The HFVE system has now been developed to the point where the prototype Silooet software is being tested in a pilot study. The system's aim of effectively presenting the features of successive detailed images to blind people, is challenging. Some users might only use the system for accessing more straightforward material.

Future developments can build on the trial results, and attempt to create a useful application. The test done so far show that most people are able to easily recognise standard shapes. The positive response to recognising shapes, to symbolic object shapes, and to live images suggests that a future development could be to incorporate automatic face-recognition and other object-recognition facilities.

Possible applications include:- presenting shapes, lines, maps and diagrams for instructional purposes; providing information to users wishing to know the colour and shape of an item; and for specific tasks such as seeking distinctively-coloured items. The recently-commenced pilot study should help to clarify which aspects of the system are likely to be the most useful.

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, pp. 14–15 (2008)
3. U.S. Patent Appl. No. US 2008/0058894 A1
4. Fournier d'Albe, E.E.: On a Type-Reading Optophone. Proc. of the Royal Society of London. Series A 90(619), 373–375 (1914)
5. Vision Technology for the Totally Blind, `http://www.seeingwithsound.com`
6. iFeelPixel, `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)
8. Yang, M., Kriegman, D., Ahuja, N.: Detecting Faces in Images: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(1), 34–58 (2002)
9. The HiFiVE System, `http://www.hfve.org`