CREATING AND ACCESSING AUDIOTACTILE IMAGES WITH *"HFVE"* VISION SUBSTITUTION SOFTWARE

David Dewhurst

www.HFVE.org
daviddewhurst@HFVE.org

ABSTRACT

The HFVE (Heard and Felt Vision Effects) vision substitution system uses moving speech-like sounds and tactile effects to present aspects of visual images. This paper describes several audio- and interaction-related improvements. A separate "buzz track" allows more accurate perception of shape, and additional sound cues can be added to this new track, instead of distorting the speech. Details are given of improved ways of presenting image "layout", and the HFVE approach is compared to other audio vision substitution systems. Blind users can create or add to images using a standard computer mouse (or joystick), by hearing similar sound cues. Finally, a facility for defining and capturing material visible on a computer screen is described.

1. INTRODUCTION

HFVE (Heard and Felt Vision Effects - pronounced "HiFiVE") is an experimental audiotactile vision substitution system which presents aspects of visual images, with the user interacting to control what is presented [1]. Apparently-moving speech-like sounds (and corresponding tactile effects) known as "tracers" follow the paths of key shapes (with corners being emphasised), or convey the layout of areas, the speech-like sounds describing features (e.g. colour, layout etc.) of the images being presented.

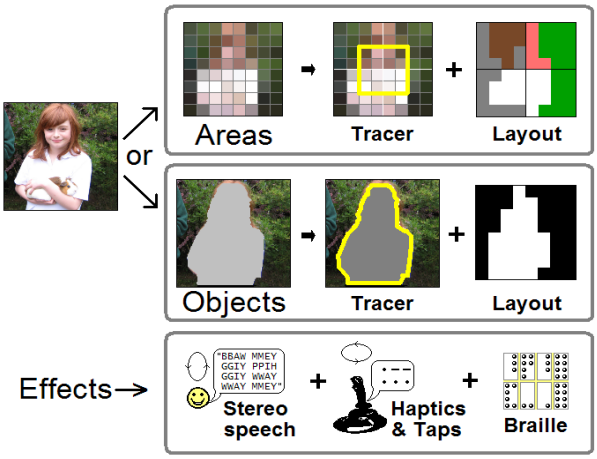

Figure 1. *Presenting an image via audiotactile effects.*

The apparent speed of travel of an audiotactile "tracer" is generally constant when presenting a particular entity, but can vary from entity to entity so that there is time to present speech-conveyed information.

This paper describes several new features of the system.

2. IMPROVING THE PERCEPTION OF "TRACERS"

In earlier versions of the HFVE system, the shapes perceived by users were not always clearly defined if presented using only apparently-moving speech-like sounds. Highlighting corners greatly improved matters, particularly in the tactile modality, but extra cues are needed in the audio modality. The speech-like sounds were presenting additional information via volume changes - relatively slow changes to present size, change, etc., and a more rapid "flutter" to convey the "texture" of an area - and this could make the speech more difficult to understand.

"Optophone"-like systems [2,3,4] typically use a systematic left-to-right "scanning" action, which gives "time-after-start" cues to the horizontal location of material within images Fig 3.

However such cues are not generally present with the HFVE system, as the "tracer" can move in any direction when presenting the path of a lineal feature (e.g. the perimeter or medial line of an item).

In the tactile modality, a moving force-feedback device can give clear "proprioceptive" cues to horizontal location, but the tactile modality has the disadvantage that it requires users to hold or touch a tactile display of some kind.

In the audio modality, users had to rely on stereophonic effects to obtain the horizontal position of the tracer, and these cues can be weak. (Vertical positioning is mapped to pitch).

These issues have been addressed by using a separate sound track (referred to as a "buzz track") that is played at the same time as the speech-like sounds. The buzz track is easier to "mentally position" in "soundspace", and allows more accurate perception of shape, than when speech-like sounds alone are presented. Additional location and direction cues, and timbre-conveyed information, can be loaded onto the buzz track.

2.1. Adding a "buzz track"

In conveying a particular entity (e.g. object or abstract shape), the speech tracer presents categorically-perceived properties of the entity, for example colour and object type; while the buzz track tracer, played at the same time as the speech tracer, presents other properties of the same entity, for example volume-conveyed properties, as well as presenting the shape and position more clearly than the speech tracer.

Any volume-altering effects (conveying information such as texture, width, change, etc.) can be applied to the buzz track, rather than distorting the speech.

Both the speech tracer and buzz track can follow the same apparent path at the same time. However small objects Fig 2 (C) can be enlarged (B) to better convey their shape, and optionally only the buzz track tracer can be enlarged to present the shape more effectively, while the speech tracer gives the location of the small shape within the image.

The buzz track sounds can be system-generated, or can be recordings of sampled sounds e.g. musical instruments; voices; natural sounds; etc. One of the more effective sounds was a "buzzy" sound, but with a clearly defined pitch. (Similar sounds are often used to demonstrate "3D sound" environments, indicating that such sounds are effective for conveying location in "soundspace".)

2.2. Adding timbre to the "buzz track"

If a "buzz track" is being presented, changes to its timbre can be made in order to convey additional information in a non-linguistic way. For example, the horizontal positioning can be further enhanced by gradually changing from a "buzzy" sound to a square-wave sound as the tracer sounds move from left to right. Other visual data, such as the characteristics of the edge of an object, can also be conveyed via the buzz track's timbre.

(Timbre may be used to convey colour, although this does not emulate the categorical manner in which people perceive colour. Instead, the timbre can be gradually changed to convey the "colour temperature" of the region being presented.)

2.3. "Pillar" and "stratum" effects

If buzz tracks and timbre effects are used, it is still sometimes difficult to interpret the shape of the lines described by a moving tracer from the audio effects alone. Furthermore, for a tracer moving in a mainly upwards direction, it is difficult to determine the direction of the slope (i.e. whether to the left or right) from the slowly-changing timbre.

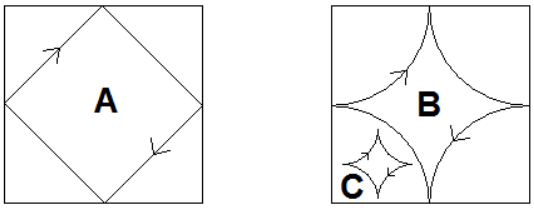

Figure 2. S*imilar shapes, and enlarged small shape.*

Consider shapes A and B (Fig 2) - from the buzz track alone it is not always clear whether the edges are straight or curved.

Additional effects can clarify the shape of sloping edges.

One approach is to divide the image to be presented into several equal-width columns and/or rows. Then effects can be triggered whenever the tracer moves from one column to another (referred to as "pillar effects"), or from one row to another (referred to as "stratum effects").

Using pillar and/or stratum effects allows the shape of lines to be perceived more clearly : as the tracer travels at a constant speed, the rate at which the effects are presented will correspond to the angle of slope. For example the diamond shape (A) will produce an even rate of pillar effects, while the "concave diamond" shape (B) will produce a changing rate of effects as the slope becomes more horizontal or more vertical.

Different effects are presented when the tracer moves from left to right, and right to left, so that the direction of travel is clear. One approach is to apply a sawtooth-shaped volume profile. If applied as pillar effects, as the buzz track moves horizontally it presents effects sounding like "bing-bing" as the it moves left to right, similar to the "attack-decay" effect heard when a percussion instrument is struck; and presents effects sounding like "nyib-nyib" as it moves right to left, similar to the sounds heard when a soundtrack is played backwards. The rate at which such effects are heard indicates the slope of the line described by the tracer. (Other directional effects can be used, for example distinct sounds.)

As an option, the pillar and stratum spacing can vary dynamically from entity to entity, so that a consistent effect frequency is given for a particular angle of slope.

(Further clarity can be given to the horizontal definition of shapes by starting the tracer at, say, the leftmost point of the shape, so that the user knows that any initial horizontal movement will be rightwards.)

3. IMPROVING THE PERCEPTION OF "LAYOUT"

The system can present the arrangement of properties within a defined rectangular area Fig 1, or within an object being presented. Until recently, development has mainly focused on using speech-like sounds to describe such "layouts", although it was previously planned [5] that texture etc. would be conveyed using multiple speech tracers with fluctuating speech volumes.

The use of supporting effects, similar to those presented by "optophone"-like systems, has now been further investigated, and the two approaches have to some extent been integrated.

Fournier d'Albe's 1914 Reading Optophone [2] presented the shapes of printed characters (or other material) by scanning across lines of type with a column of five spots of light, with each spot controlling the volume of a different musical note, producing characteristic sets of notes for each letter Fig 3.

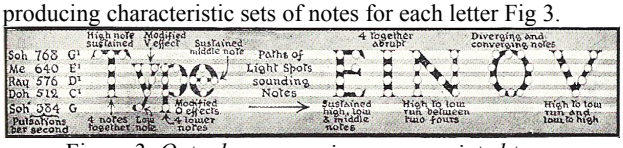

Figure 3. *Optophone scanning across printed type.*

Other systems have been independently invented which use similar conventions to present images and image features [3 & 4], or to sonify the lines on a 2-dimensional line graph [6]. Typically height is mapped to pitch, intensity to volume (either dark- or light- sounding), with a left-to-right column scan normally used. Horizontal lines produce a constant pitch, vertical lines produce a short blast of many frequencies, and the pitch of the sounds representing a sloping line will change at a rate that indicates the angle of slope. For example a "V"-shape would be presented as a series of notes reducing, and then rising, in pitch. A combination of recognition of familiar shapes, and analysis of new sounds, allows users to interpret shapes. Simple images such as printed characters and diagrams, containing horizontal and diagonal elements, produce clear effects. The mapping for such systems is straightforward, and "time-after-start" cues give the left-right positioning clearly. However complex scenes can be confusing.

3.1. Improved layout coding

Previously, HFVE used somewhat arbitrary coded speech-like sounds to convey layout. A single consonant-vowel ("CV") syllable could present the arrangement of 4 or 8 "blobs" of content [1]. The colours (or other properties) of the areas were also presented in a coded, but less arbitrary, manner, for example "boo-yow" or "bow" for "blue and yellow". However when tested in a small trial, real-name (non-coded) colours were greatly preferred by participants [1], and it made the system more accessible to untrained users. The real-name colours could be spoken more quickly by the system, as the user was expecting a colour name, and could "fill in" parts of the speech that they heard less clearly - this effect is not available with the theoretically more efficient coded phonemes. Even long colour names such as "DarkPurple" could be spoken rapidly (in about a third of a second) and still be understood.

Unlike for colour, where common names are available, there are few standard terms for particular arrangements of "blobs". However it was straightforward to give reasonably sensible (and easily distinguishable) "real-word" names to 16 layout arrangements, allowing a 8-by-4 layout matrix Fig 4 (A) to be presented to beginners via 8 "real" words in a "column-by-column" arrangement (B). (Such an arrangement also maps well to a 4-dot-high refreshable computer braille display Fig 4 (C).)

Figure 4. *8-by-4 layout conveyed as speech and braille.*

A comfortable limit of about 4 to 6 short words per second is practical. This gives a limit to how much layout information can practically be presented via words. Furthermore, a well-known psychological effect [7] states that about 6 to 8 unrelated "chunks" of information can be comfortably handled in people's short term memory, indicating that a limit of about 4 to 6 "words" are available for presenting layout information for any particular area, if other property information is also given.

A modification made to the coded "CV" syllables was to strictly match the consonant to the first half of a layout, and the vowel to the second half Fig 4 (B).

Blob arrangements presented "column-by-column" or "row-by-row" (e.g. 1-by-4, or 1-by-6) Fig 4 (B) (whether coded or real-word) may be easier for users to follow than 2-dimensional arrangements (e.g. mapping to 2-blobs-by-2 or 2-blobs-by-4).

It remains to be seen whether coded or "real-word" colour and layout presentation is preferred longer term : using real-words may be more distracting to ambient sounds, whereas the coded sounds may be more easily ignored when required. Furthermore, the codings are not difficult to learn.

With practice, users may become familiar with groups of sounds representing several columns, so that, say, a 4-blobs-by-4 arrangement is immediately understood as a single entity "chunk", rather than having to be "assembled" from the component sounds. (This has not yet been tested.)

3.2. Layouts supported with multiple tracers ("polytracer")

Just as apparently-moving speech-like sounds can be supported by using a buzz track to clarify the shape, so speech-like layouts can be supported using optophone-like multiple-tracer effects (referred to as a "polytracer"), which may allow more accurate perception of the distribution of material within entities.

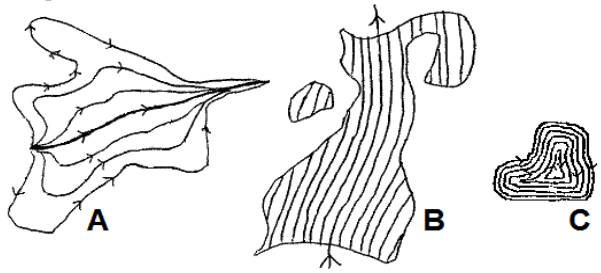

Figure 5. *"Contoured" and "parallel" "polytracers".*

A "polytracer" can present non-speech-like "tone sound" tracers in a similar manner to existing optophone-like systems; or the extra tracers can also be speech-like, presenting the same speech phonemes as the main tracer, but moving in "soundspace" so that their pitch and binaural location at any moment corresponds to the location of the image matter that they are representing. The latter approach produces a "choir" of voices that "chant" the speech sounds (this effect is referred to as a "chorus"). The "tone-like" multiple tracers may allow better positioning accuracy than the "chorus" approach.

The paths that the polytracers follow can either be straight parallel lines, as used in previous optophone-like systems, or if a shaped entity is being presented then the tracers can follow paths that give the overall shape of the entity.

The system can combine a coded speech-like medial-tracer with a several simultaneously-conveyed tracers that travel in approximately the same direction as the medial-tracer, but vary in the width that they represent Fig 5 (A), so that the shape of the entity is conveyed quickly, and more of the detail and texture is also conveyed. The "contoured" polytracer method works best when the general direction of movement of the tracers is horizontal (A), as the spread of frequencies used helps to convey the width of the entity.

Equal-width non-medial-tracers ("parallel") polytracers Fig 5 (B) travel quasi-parallel to the main (i.e. medial) tracer. The number of tracers conveyed at any moment will vary, with the outer-edge tracers being activated and de-activated according to the width of the entity at any point (B). Hence the changing number of tracers active at any time gives an indication of the shape and width of the entity at different points.

The medial line tracer is an effective main tracer on which to base the polytracers, for both contoured Fig 5 (A) and parallel (B) polytracers. However a "circuit" medial path can also be used, where the path follows a loop centred on the middle of the object (C).

As an alternative to shaping the tracers' paths, an optophone-like "rectangular" polytracer arrangement can be used, where the tracers are straight, parallel, and of equal length. This approach is effective when a polytracer is presenting the special layouts that are used for highlighting objects and entities. For example, "object-related" and "symbolic" layouts have previously been described [1] which highlight the shape and location of entities within an image Fig 6, so that a perceptual "figure/ground" effect is produced, either emphasising the shape of the object (A) or the location of the object(s) within the scene (B & C).

Figure 6. *"Figure/ground" and object-related layouts.*

Such silhouette-like images are particularly effective when presented via polytracers, either as additional optophone-like tone sounds, or as a "chorus" effect.

Rectangular polytracer arrangements can effectively present the information presented by the braille display area Fig 4 (C).

The pitch range used for the polytracer effects can match the pitching used elsewhere by the system, or a polytracer-specific musical pitch range can be used.

The polytracers can be set to be "light-sounding", "dark-sounding", or "least-sounding", the latter setting being used to emphasise either dark or light effects, whichever is least present, in order to minimise the confusion of sounds.

The tone sounds can be similar to those used for "buzz tracks". For example the timbre of the tracers can change to indicate the left-right positioning.

To summarise, polytracer effects are generally used to support the layout effects, by giving greater clarity to the shapes being presented, and to the distribution of material within those shapes.

4. INTERACTING WITH AUDIOTACTILE IMAGES

HFVE can present both objects found in images “on the fly”, and objects from prepared media. For non-prepared media (e.g. “live” images), the system attempts to find objects according to the user's requirements, and builds a “guide table” of the found objects. Alternatively a previously-prepared “guide table” can be used to specify the objects and features that are present : a sighted designer can highlight the entities present in images, and identify what they are, their importance, etc. For each image, one or more objects can be defined, and these can be “marked-up” on the image. “Paths” can be included to give the route that moving objects follow.

4.1. Drawing and “marking-up” images

The HFVE system provides a facility tailored to the process of creating a guide table, and marking up images with objects to be presented, which are then linked to objects in the table. It can additionally be used to draw shapes that can be immediately presented to a blind person. The user can edit the guide table and the system can automatically adjust selected colours so that the system can match the drawn objects to those in the guide table. If no guide table is active, then a simple default guide table is used, to which references to the drawn objects can be added.

The user can draw lines (for example via a computer mouse) which can then be presented as tracer paths; or “closed” lines can be “filled” with colour. These are then presented using the current system settings. An aim is to make this facility accessible : blind people can “draw” onto the image (or blank background) via a joystick; or a mouse with constrained movement (e.g. the Logitech Force Feedback Mouse). An “unconstrained” mouse (i.e. standard mouse) can also be used, as described below. For blind users, stereophonic tone-like sounds, using conventions similar to those used for “buzz tracks”, give continuous feedback to the user about the location of the mouse pointer (e.g. drawing “pen”) at any time. Timbre, "pillar" and "stratum" effects can be exhibited, and a "dwell" action can be used to mark specific corners.

When users move the mouse (or joystick) in a certain path, the sounds they hear will be similar to those produced when a tracer moves in the same path, and they will hear similar sounds when the "buzz track" of the same shape is replayed.

4.2. Using a standard computer mouse to draw images

An “unconstrained” computer mouse is normally considered to be of little use to a totally blind person, as they are unable to visually follow the mouse pointer on the screen.

Although location-conveying audio feedback can be used to give the approximate mouse pointer location and the shape of the path in which it is moving, for a “drawing” application the user has to locate the mouse pointer in the drawing area/"canvas", which is difficult to do even with audio feedback.

An effective solution is to allow the mouse/pointer to be moved anywhere over the computer's screen/“desktop” area, but with the location processed to map to the application's drawing area. However as the mouse may move over other applications, the standard main mouse button cannot be used in this mode. To do a "mouse down" action (to draw a line etc.), users can press a particular keyboard key, such as "M"(ouse). Alternative the middle button of a 3-button mouse can be used, as it normally produces no change when clicked over most applications, although this is a slightly less safe approach.

"Alternative" input devices that simulate the action of a mouse may be also used, e.g. a graphics pad, an interactive whiteboard, or an interactive touch-screen.

4.3. Using a "viewfinder" to capture images

The images presented by HFVE can be gathered from various sources, such as media files, or live video images. However by using a sizeable and moveable "viewfinder" frame Fig 7 (A) that can “hover” over any part of the computer screen/“desktop”, the screen content framed by the viewfinder can be captured, and then presented in the same manner as other images.

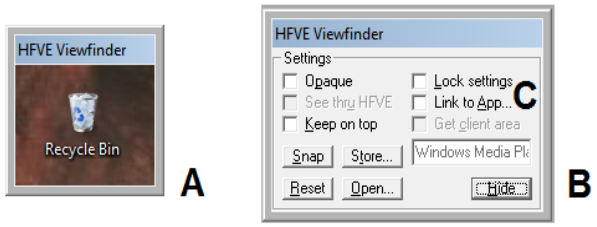

Figure 7. *The HFVE “viewfinder” (A) and controls (B).*

The "viewfinder" can be sized and moved via the keyboard, or via a mouse used with audio feedback. The mouse can define the region to be presented e.g. via a diagonal movement. The HFVE system then "frames" the defined region with the viewfinder, and the enclosed content is presented.

The viewfinder can be linked to another application (C), so that it "follows" it if the other application is moved.

5. SUMMARY AND CONCLUSION

This paper has described several new techniques for the interactive sonification of images using the HFVE system. Although an earlier version was assessed in a small pilot study [1], the latest features, being incomplete, have not yet been tested by users, and this is a necessary next step. The system's current state of development will be demonstrated at ISon 2010.

The HFVE project's aim, of effectively presenting aspects of successive images to blind people, is challenging. It remains to be seen which features of the system are the most effective.

6. REFERENCES

[1] D. Dewhurst, “Accessing Audiotactile Images with HFVE Silooet” in *Proc. Fourth Int. Workshop on Haptic and Audio Interaction Design*, Springer-Verlag, 2009.

[2] E. E. Fournier d'Albe, “On a Type-Reading Optophone” in *Proc. Royal Society of London. Series A*, Vol. 90, No. 619 (Jul. 1, 1914), pp. 373-375.

[3] P.B.L. Meijer, “An Experimental System for Auditory Image Representations” in *IEEE Trans on Biomedical Engineering*, Vol. 39, No. 2, pp. 112-121, 1992.

[4] U.S. Patent No. US 6,963,656 B1.

[5] The HiFiVE System, http://www.hfve.org.

[6] D.L. Mansur, M.M. Blattner and K.I. Joy, ``Sound Graphs, A Numerical Data Analysis Method for the Blind," in *Journal of Medical Systems*, Vol. 9, pp. 163-174, 1985.

[7] G.A. Miller, “The magic number seven, plus or minus two: Some limits on our capacity for processing information” in *Psychological Review*, 63, pp. 81-93, 1956.