THE DESIGN AND EXPLORATION OF AUDITORY DISPLAY EFFECTS FOR GENERAL VISION SUBSTITUTION

David Dewhurst

www.HFVE.org

david.dewhurst@HFVE.org

ABSTRACT

This work forms a part of a wider project, in which the author is developing a system to present visual images, and other material, via sets of auditory (and tactile) effects.

The main contribution of this Extended abstract is to describe relevant developments that have occurred since the author last reported at ICAD 2019 (June 2019), and which will be demonstrated at ICAD 2025. The author also describes possible future developments of the system.

The key new developments described in this Extended abstract include details of:- a new system architecture, and techniques; new effect types (including what are termed "Pan Effects"); new user interaction and interfaces; online installation on a virtual private server (accessible via open source cross-platform browser-based Myrtille Remote Access Gateway, or via Microsoft's Remote Desktop Connection); and other features and possible future work.

(This Extended abstract also summarises auditory display effects and techniques that have previously been reported.)

1. INTRODUCTION

Several attempts have previously been made to present aspects of vision to blind people via other senses, particularly hearing and touch. The approach is termed “sensory substitution” or “vision substitution”.

To convey general visual information to blind people, apparatus or software can be devised that presents other senses with information that includes aspects of sight, but other senses may not be as powerful, or as able to comprehend such information [1].

Early work in the field includes Fournier d'Albe's 1914 Reading Optophone [2], which presented the shapes of characters by scanning across lines of type with a column of five spots of light, each spot controlling the volume of a different musical note, producing characteristic sequences of notes for each letter of the alphabet Fig. 1.

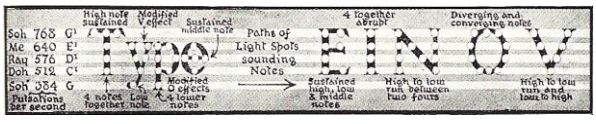

Figure 1: Optophone scanning across a line of type.

Other systems have been invented which use similar conventions to present images and image features [3], [4], or to sonify the lines on a 2-dimensional line graph [5]. Typically height is mapped to pitch, brightness 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 frequency at a rate that indicates the angle of slope.

Some previous work in the field is summarised in [6], [7].

1.1. The HFVE system.

The author's HFVE (Heard & Felt Vision Effects) system attempts to present aspects of visual images to blind people, via a rich set of audio (and tactile) effects, conveying images as a series of items, with the user deciding what is presented.

A major feature of the system is presenting modified speech i.e. spoken word sounds that are changed, multiplied, and moved, in order to intuitively convey the location, size, shape, and other properties, of the items they are presenting. "Tracers" are moving audio and tactile effects that trace out shapes, with corners emphasised (A) Fig. 2. Many of the described auditory effects can be combined.

The user can control which items are being presented, via a mouse pointer, joystick, or via touch; or the system can automatically sequentially step around or list the most important items found within a user-controlled rectangular "Frame" (optionally the whole image).

Another feature allows a blind person to navigate between levels of view within visual or non-visual representations, rising up levels for an overview, and drilling down levels for more detail, via, for example, a mouse wheel or dial device.

(Note that most of these features have been reported previously [9], [10], [11], [12], [13].)

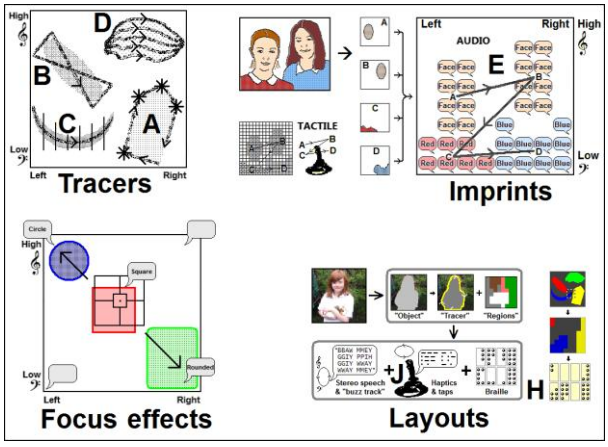

Figure 2: Some HFVE effect types.

Particular item types, such as faces, text, persons, etc., can be presented using preferred per-item-type effect settings held by the system. For example faces and people could be presented as symbolic tracers (B) Fig 2, while blobs / areas of particular colour could be presented via imprints (E) Fig 2.

The nature and aesthetics of the auditory display effects can be experienced by visiting the author's website, which includes demonstration videos, and provides access to online HFVE sessions [8].

1.2. Tactile display effects.

Though not the primary concern of this Extended abstract, many of the auditory effects have tactile equivalents.

A force-feedback joystick makes an effective pointing device with which to indicate areas of an image, as it can also be programmed to position itself to one of a number of set positions, so that notch-like effects are felt when the joystick is moved, giving a tactile indication of location. Suitable devices include Microsoft's Sidewinder® Force Feedback 2 joystick, and Logitech's Wingman® Force Feedback Mouse Fig. 3, controlled via Microsoft's DirectInput software.

A force-feedback joystick can also be programmed to allow free movement only within a restricted range, or along a lineal route. It can also be moved by the system, by it doing successive DirectInput "Spring" effects, so pushing and pulling the user's hand and arm to trace out shapes and highlight corners. It can also be used to command the system via button and twist actions, and can output tapping effects to present morse-like information to deafblind users (J) Fig. 2.

Figure 3: Microsoft's Sidewinder® Force Feedback 2 joystick; and Logitech's Wingman® Force Feedback Mouse.

The method of user interaction can be "exploring" in style, using a moving pointer to inspect a scene; or alternatively the system can announce items, and the user can then select one of those items for further inspection (the latter approach perhaps requiring less mental effort).

Users can tap commands onto a touch- screen or pad, and touch or drag over them to indicate parts of an image [12].

2. NEW DEVELOPMENTS

2.1. New system architecture and approaches.

When the system was last presented (at ICAD 2019), the software that was demonstrated had been developed using a somewhat haphazard approach, being experimental in nature.

After that point, the earlier approaches were frozen, and prior features (now termed "Legacy" type or "Special" type effects) were locked / not developed further.

One major feature of the Legacy / Special type effects was presenting the content of a rectangular "Lens" (which can be updated at any point from the current moveable and sizeable "Frame" by pressing the "X" key). The Lens content (colours, layouts, item types, etc.) can then be presented as braille, or coded speech, or tactile taps (H & J) Fig. 2 [9].

Legacy / Special type effects include for example:-

RegionRectangle	Rectangular Region tracer (presenting Lens content).
Region Circle	Circular Region tracer effect path (showing Lens extent).
Region Middle	Presents stationary (middle of Lens) effects.
Object Outline	Tracer follows perimeter of the object(s) (A) Fig2.
Object Symbolic	Object tracer follows a path that symbolises the nature of the object(s), and shows their extent (B) Fig 2.
Object Medial	Object tracer follows lineal medial path though middle of the object(s), showing shape and extent (C) Fig2.
Object Circuit	Object tracer follows circuit medial path of the object(s).
Object Framed	Object tracer follows sloping and object-enclosing rectangular path, showing extent of the object(s).
Imprint Standard	Groups of voices, speaking in unison, that rapidly convey the properties and approximate extent of Item(s).
Imprint Bubbles	Imprint effects with bubbly / dynamic energetic effervescent effects (E) Fig 2.
Polytracer Both	Tone- and / or speech- based multi-tracer Polytracer effects that trace out shapes (D) Fig 2.
Complex 1	Polytracer effects, plus Imprint effects.
Complex 4	Default Object tracer, and Imprint effects with bubbly / dynamic energetic effervescent effects.

With the new "Standard" type effects (i.e. not "Legacy"), commands (messages, requests, etc.) are passed between "Instances" i.e. HFVE executables, which can run on separate Machines e.g. VMs. Such commands can, for example, be used for triggering effects, or for requesting resources.

To provide flexibility and to allow for future expansion, the system can produce several concurrent effects, for example to present several Items simultaneously. In order to spread the processing load, the system can run across several Machines (virtual or physical). Several Instances of the application (i.e. HFVE executables) can run on each Machine. On each HFVE Instance, one or more "Effect Kits" can run — these are separate Visual Basic forms (visible or not) that operate as classes, and most effects run from such Effect Kits.

The new version of the software uses what is termed a system "Inbox", into which are added requests to present. E.g. the folder "…\Inbox" could have files "Visual_Items.bmp" and "Visual_Items.txt" added for presentation. These comprise:- a standard 24-bit colour bitmap file that includes the basic item blobs (termed the "ItemMap" file); and a standard text file that describes how those blobs are marked via particular bit settings on the bitmap, and specifies how those blobs are consolidated to higher-level items (termed the "ItemKey" file). This pair of files is described further in [12].

Visual outputs of AI processing may be further processed into a pair of such files, and so be presented by the system.

2.2. New effect types.

Some of the new Standard type effects are described below:-

- **"Basic" effects** : Items to be presented are simply spoken pitched and stereophonically placed, as the system steps round them. They can include new Pan effects (see below).

- **New "Rich" / "Regular" effects** include for example:-

Solo Outline	Lone key tracer(s) traces out perimeter outlines of the item's blobs, showing shapes and extents. (A) Fig 2.
Solo Symbolic	Lone key tracer(s) presents path that shows the nature of the item, and also the extent of the item. (B) Fig 2.
Solo Medial	Tracer(s) follow lineal medial paths through middle of the item's blobs, which shows extent of item. (C) Fig 2.
Imprint Speech	Imprint effects via many stationary voices, speaking in unison, that convey item info. and extent. (E) Fig2.
Polytracer Both	Tone- and / or speech- based multi-tracer Polytracer effects that trace out shapes (D) Fig2.
Object Framed	Object tracer follows sloping and object-enclosing rectangular path, showing extent of the object(s).
Imprint & polytracer	Default Imprint effects with multi-tracer Polytracer effects.

2.3. Pan Effects.

A recently-developed effect type is termed "Pan Effects" (and is related to the Audio Preview effects — now renamed "Spanner" effects — described in [13]).

Pan Effects are intended to give an instantaneous impression of horizontal / lateral i.e. Left<>Right location, analogous to the mapping of height to pitch. Existing stereophonic pan effects, though clear for stationary or slow-moving sounds, can be less easy to interpret for rapidly-changing horizontal positioning — for example when freely exploring an image with a pointer. Furthermore, stereo sound presentation is not always available, and some hearing-impaired users may be unable to perceive stereo positioning.

Unlike for height-to-pitch mapping, there is no obvious intuitive property to use, so several experimental mappings are provided, from which users can choose preferred settings.

The effects are described in relation to the initial experimental controls of the Pan Effects settings UI screen-shot Fig. 4 (which is likely to change following testing).

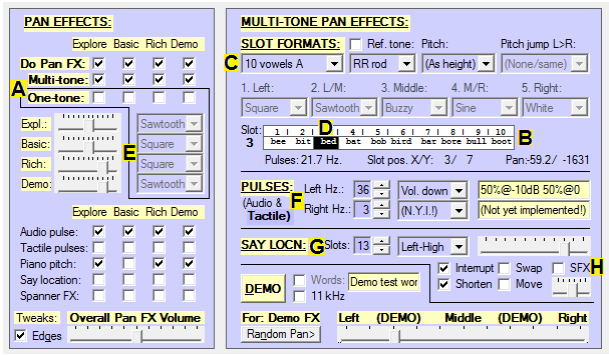

Figure 4: User interface for experimental Pan Effects.

Pan Effects currently provide:-

- Basic stereophonic audio effects (and tactile proprioceptive effects). The system already used an improved panning method, which is described in [11].
- "Pan tones", which can be tone-like e.g. square or sine waves; or speech-like sounds that can be sounded continuously (e.g. non-diphthong vowel, or continuous consonant, speech sounds). Pan tones can be multi-tone / multi-timbre; or one-tone / timbre (A) Fig. 4.

When "multi-tone", different tones or speech sounds are assigned to particular horizontal locations (B) Fig. 4. In the example screenshot, 10 vowel sounds have been selected (C), and their assignment is shown in the lateral slot display (B), via English words containing such vowel sounds (D). 3- or 5-slot formats can overlap sounds covering intermediate areas.

The motivation for providing speech sounds is that they can be instantly recognised and associated with a horizontal position. (Previous attempts at using a second tone sound for horizontal location mapping were disappointing.).

When "one-tone", a single vowel-like sound e.g. EE or OO, or a standard waveform such as Sine, Square, or Sawtooth (E) Fig. 4, is used for all lateral positions.

Up to 13 distinct tones can be provided in the X / horizontal direction (B) Fig. 4 (analogous to the 13 semitones, spanning one octave, that are typically used in the Y / vertical direction).

- Varying pulses can be applied to the pan tones (F), the frequency corresponding to the Left<>Right position.
- Speech "Say location" effects (G) Fig. 4 can directly announce the X,Y location, for example "Left-High" or "1-A". Effects such as echo or reverb (H) can help distinguish such words from other speech sounds.

Feedback from an initial informal assessment included:-

- Certain sounds were perceived as harsh-sounding, particularly sine and white noise, and fricative consonants. This may depend on voice synthesiser.
- The multi-tone approach, assigning specific sounds to particular horizontal slots, seemed promising.
- Overlapping multi-tone sounds (available if only e.g. 3 or 5 slots are used) were perceived as harsh.
- Varying pulses were thought effective.
- Directly announcing the location proved popular.

2.4. New user interaction and interfaces.

The Main user interface (UI) (A) Fig. 5 was too complex for practical use, due to the experimental nature of development.

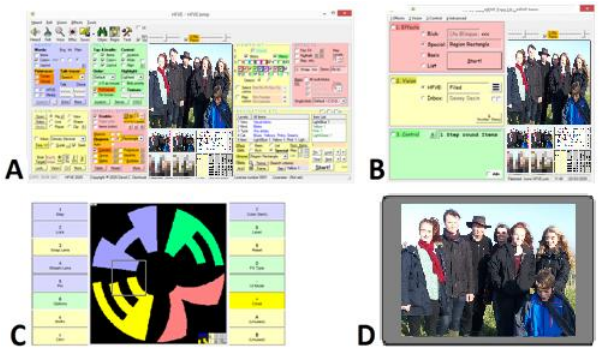

Figure 5: Original Main UI (A), "Easy" UI (B), "Touch" UI (C), and "Full" UI (D) formats.

In order to simplify usage, three new user interfaces / UI formats have been developed, as shown in Fig. 5 (B to D).

- **"Easy" UI (B)** : This uses a traditional desktop UI, but hides content behind checkboxes and drop-down combos.

The Easy UI format is the default UI for desktop use.

- "**Touch" UI (C)** : This format uses a simplified set of navigable commands, grouped into sets of 12. Commands can be triggered via touch, tap codes, speech, keys (if available), or via a grid of commands [12].

The Touch UI format is the default UI for online use.

- "**Full" UI (D)** : The image fills the screen, effectively like the Touch UI format but without the command buttons. Commands can be triggered via touch, tap codes [12], speech, or via keyboard keys (if available).

An assessment of the UIs is described in section 3 below.

The Full UI format allows an "explore" mode to use the system pointer to present Item information; or the system pointer can act as an exploring colour meter. If a touch-screen tablet is being used with the Full UI format, then the user can tap commands, drag over the monitor area, and touch the tablet screen to indicate parts of the image.

A touch pad, as often found on laptop computers, may be used to control the system via taps and drags / swipes in the same manner. A Synaptics® TouchPad® set to "absolute" mode can be used to indicate locations within the monitor (D) Fig. 5, and to trigger touch screen tap and gesture commands.

2.5. Online use via "Myrtille"; and via Remote Desktop.

Since the system was last reported on, and presented, an online version of HFVE has been developed [8].

Having investigated many possible technical solutions (e.g. a video-conferencing approach), the method chosen was that of simply running the existing Windows-based HFVE desktop software on virtual Windows servers, the servers hosted on Amazon Web Services (AWS) "Lightsail" [14], accessible via open source cross-platform browser-based "Myrtille" Remote Access Gateway [15]; or via Microsoft's Remote Desktop client [16] (or Windows App).

Three servers are provided, located in Europe (London), West (Oregon), and East (Tokyo) (using a relatively nearby server improves the application's responsiveness). See https://www.hfve.com/TryHfve.htm.

The current setup runs using minimal resources on each server (1 CPU, 512MB RAM, and 32GB storage).

Using the above approach allows a single codebase to run both on Windows Desktop, and online, with minimal flags needed to adjust processing.

Useful features found when installing online included:-

- Myrtille supports most HTML5-compatible browsers, and no plug-in or extension is needed.
- Microsoft provides Remote Desktop clients for Windows, Android/Chrome OS, iOS, and macOS.
- Using AWS "burstable" EC2 instances allows minimal resources to be used most of the time, with extra CPU "burst" performance being temporarily available when the HFVE software is running.

Both Myrtille, and Remote Desktop, ran smoothly via standard Windows computers, and via an Android Samsung Galaxy Tab S, and via an iOS iPad Pro. An add-on keyboard and/or mouse assists with using HFVE on a tablet computer.

3. INFORMAL ASSESSMENTS

Informal assessments of the new effect types, and user interaction and interfaces, have commenced.

With testing done so far, the feedback on the UIs is that the desktop-style Easy UI seems more user friendly (though this could be due to familiarity with Windows controls), and that the Touch UI format would be more suitable for blind users and for touch-based operation (B & C) Fig. 5.

Some feedback from an initial informal assessment of Pan effects is included in section 2.3 above.

It is important to also conduct tests with blind participants (only sighted testers have been used so far).

More results and features will be reported at ICAD 2025.

4. CONCLUSIONS AND FUTURE WORK

In this Extended abstract, new developments of the HFVE system have been described.

Some tidying work is needed, for example with keyboard shortcut commands, which can clash with browser commands. For instance keys Alt+F4 will immediately fully close the whole browser — rather than close the HFVE session. Additionally, adjustments to the volume of the Pan effects in certain circumstances may prove worthwhile.

Future developments may include enhancements to the Imprint effects, including a new effect type termed "Sliders". A Slider is a moving and / or zooming / changing Imprint of an item, typically used to convey change and movement.

Currently, the online version of HFVE only performs audio stereophonic output i.e. does not provide tactile output. However, the author is currently assessing the software "VirtualHere" [17], which allows local USB devices (e.g. force-feedback devices) to be used. Other approaches include using audio, or visual, signals from the online application to convey the required location and other actions to a joystick.

Furthermore, future work should include detailed evaluations, with an examination of specific tasks and approaches, detailed statistical analysis of results, and a qualitative analysis of post task interview data.

The system will be demonstrated at ICAD 2025.

5. REFERENCES

[1] Zahl, P.A. (Ed.) *Blindness : Modern approaches to the unseen environment.* Hafner Publishing, 1963.

[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] 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.

[6] A. Edwards, "Auditory Display in Assistive Technology" in *The Sonification Handbook*, T. Hermann, A. Hunt, J.G. Neuhoff (Eds.), 2011.

[7] T. Pun et al., "Image and Video Processing for Visually Handicapped People" in *EURASIP Journal on Image and Video Processing*, vol. 2007, Article ID 25214, 2007.

[8] *The HFVE system*, https://www.hfve.com.

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

[10] D. Dewhurst, "Creating and Accessing Audiotactile Images With "HFVE" Vision Substitution Software" in *Proc. of ISon 2010, 3rd Interactive Sonification Workshop*, KTH, Stockholm, Sweden, 2010.

[11] D. Dewhurst, "Using "Imprints" to Summarise Accessible Images" in *Proc. of ISon 2013, 4th Interactive Sonification Workshop*, Fraunhofer IIS, Erlangen, Germany, 2013.

[12] David Dewhurst and Tony Stockman, "The Design and Exploration of Interaction Techniques for the Presentation of Foreground and Background Items in Auditory Displays" in *Proc. of ISon 2016, 5th Interactive Sonification Workshop*, CITEC, Bielefeld University, Germany, 2016.

[13] D. Dewhurst, "The Design and Exploration of Auditory Display Effects for Blind Drivers in Autonomous Vehicles" in *Proc. of 25th International Conference on Auditory Display*, Northumbria University, 2019.

[14] *What is Amazon Lightsail?*,https://docs.aws.amazon.com /lightsail/latest/userguide/what-is-amazon-lightsail.html.

[15] *Myrtille*, https://www.myrtille.io.

[16] *Remote Desktop client*, https://learn.microsoft.com/en-us/previous-versions/remote-desktop-client/overview.

[17] *VirtualHere*, https://www.virtualhere.com.