# 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 report on relevant recent developments which can be demonstrated at ICAD 2026. The author also describes possible future developments of the system.

The key new developments described in this Extended abstract include details of:- new effect types (including what are termed "Pan Effects" and “Span Effects”); online installation on a virtual private server (accessible via open source cross-platform browser-based Myrtille Remote Access Gateway [15], or via Microsoft's Remote Desktop Connection [16]); means of using USB devices with the online version via “VirtualHere” software [17]; and other features and possible future work.

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

## 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 normally termed “sensory substitution” or “vision substitution” [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.

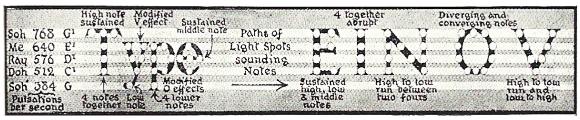


Figure 1: Fournier d'Albe's Optophone presenting shapes of letters by scanning lines of type.

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 typically 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 that they are presenting.

“Tracers” (A) Fig. 2 convey the shapes and corners of items, and “Imprints” (E) rapidly summarize the content of a scene via multiple stationary effects.

Most of these audio effect types have been reported previously [9], [10], [11], [12], [13], and several of them can be combined.

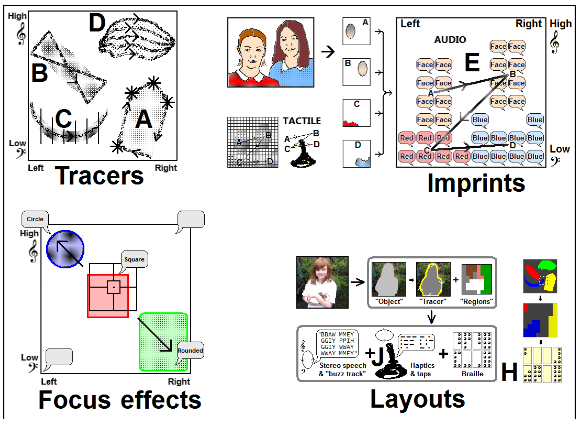


Figure 2: Some HFVE effect types.

Particular common item types, such as faces, text, persons, etc., can be presented using preferred per-item-type effect settings held by the system.

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. Using HFVE, and tactile effects.

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

A blind user can 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.

Though not the primary concern of this Extended abstract, many of the auditory effects have tactile equivalents. A force-feedback joystick can be moved by the system, and 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.

Suitable devices include Microsoft's Sidewinder® Force Feedback 2 joystick, and Logitech's Wingman® Force Feedback Mouse, Fig. 3, both of which use Microsoft's DirectInput software. They can be programmed to allow free movement only within a restricted range, or along a lineal route. They 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 (A) Fig. 2.

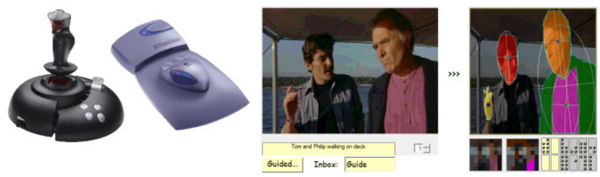

Figure 3: Microsoft's Sidewinder® Force Feedback 2 joystick, and Logitech's Wingman® Force Feedback Mouse; and presenting items in a marked-up "guided" movie.

## 2. RECENT DEVELOPMENTS

### 2.1. System architecture and approaches.

When the system was 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 "Special" (or "Legacy") 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].

**Special / Legacy type effects** include, for example:-

| | |
|---|---|
| RegionRectangle | Rectangular Region tracer (presenting Lens extent and content). |
| Region Circle | Circular Region tracer effect path (presenting Lens extent and content). |
| 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 a circuit medial through object(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. |

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 now run from such Kits.

The software uses a system "...\Inbox" folder, into which are added requests [12]. For example files "Visual_Items.bmp" and "Visual_Items.txt" could be added, comprising, respectively:- a standard 24-bit colour bitmap file that includes the basic item "blobs"; and a standard text file that describes how those blobs are marked via particular low-significance bit settings on the bitmap, and specifies how those blobs are consolidated into higher-level items Fig. 4.

In the visual domain, the system can produce higher-level consolidations of image content. While HFVE knows how to consolidate general visual images (A) Fig. 4, it does not know about other domains such as, for example, spreadsheets (B). Such entities can be submitted to HFVE as client entities, for HFVE to present them as a spreadsheet-domain view (B), consolidating cells to columns and rows, then to blocks, etc.

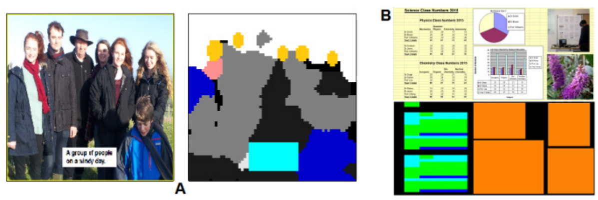


Figure 4: HFVE computer vision processed text, faces, and colour areas (A); and a spreadsheet and corresponding externally-generated "ItemMap" (B).

Visual outputs of, for instance, Artificial Intelligence may be further processed into a pair of such files ("ItemMap" and "ItemKey"), 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 can simply be spoken pitched and stereophonically placed, as the system steps round them. (They can include the new Pan, or Span, 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. |
| Polytracer Both | Tone- and / or speech- based multi-tracer Polytracer effects that trace out shapes (D) Fig. 2. |
| Imprint & polytracer | Default Imprint effects (E) Fig. 2, with multi-tracer Polytracer effects (D). |

### 2.3. "Pan" Effects.

Recently-developed "Pan" Effects (short for "panorama") 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 controls of the experimental Pan Effects settings UI Fig. 5 (which will be updated following testing, and demonstrated at ICAD 2026).

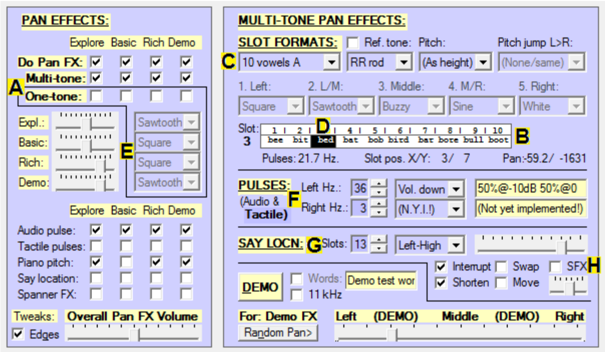


Figure 5: User interface for experimental "Pan" Effects.

Effects currently provided:-

- 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. 5.

When "multi-tone", different tones or speech sounds are assigned to particular horizontal locations (B) Fig. 5. 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. 5, is used for all lateral positions.

Up to 13 distinct tones can be provided in the X / horizontal direction (B) Fig. 5 (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-to-Right position.
- Speech "Say location" effects (G) Fig. 5 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 could be dependent on the particular voice synthesiser used).
- 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.

At AMICAD 2025, 12 anonymous random testers, who were asked to give their opinions and preferences, showed a mild preference for "Left OO <—> EE Right" vowel sound mapping (but with several "no preferences"); and showed a clear preference for "Left 2Hz <—> 32Hz Right" pulse frequency mapping (but with one tester stating a strong preference for using pulse frequency mapped to the Y axis).

Further details will be given at ICAD 2026.

## 2.4. "Span" Effects.

Recent development work involves presenting the span/ (range/extent) of an Item–enclosing rectangle, via its location, aspect ratio, and size, within the larger extent of the full presentation area. Such "Span" effects are related to Pan effects, in that both address the problem of rapidly presenting the horizontal location and extent of items. However Span effects also convey the size and aspect ratio of said rectangle.

Span effects Fig. 6 can optionally be used to quickly convey the location and extent (size and aspect ratio) of an item immediately prior to its details and exact shape etc. being presented via the other methods previously described, especially when only audio methods (i.e. no tactile) are available. Standard height to pitch mapping, and left to right panning, is used, with stereo sound placement Fig. 6.

The non-speech methods tested included using:- fixed height or sloping/diagonal lead-in and trail-out tracer phases, to convey the top and base heights of the rectangle; different sound timbres for each phase; musical step changes with height change; 2-level oscillating pitches; click sounds between phases; a second sloping tracer to mirror the slope of the first; following an "L"-shaped path (i.e. not the default diagonal path) to represent the extent of the rectangle; extra presentation areas to the left and right of the presentation square (for when the rectangle is located fully at the right or left edge); and variable speed, volume, and pitch range.

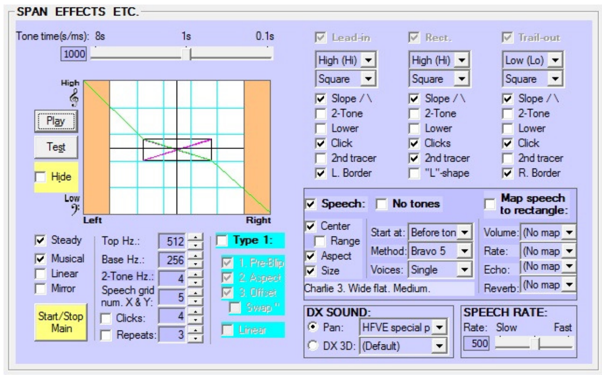


Figure 6: User interface for experimental "Span" Effects.

Alternatively, or additionally, speech output can directly announce the rectangle extent e.g. "Charlie 3, Wide flat, Medium", the speech being pitched and panned to correspond to the location of the rectangle. Optionally two voices can speak in unison or in sequence, being audio-located at opposite corners of the rectangle.

The location can be presented via cartesian coordinates (e.g. chessboard-style "B5", or phonetic alphabet-style "Bravo 5", or "Top right" style); or via polar coordinates (e.g. "8 o'clock edge").

The volume, speech rate, optional echo and reverb effects, and other speech properties, can be mapped to particular properties such as "Left-ness" or "Large-ness" etc.

The test rectangle can be drawn in a particular position with a mouse, or randomly positioned and dimensioned by the system, and can be hidden for testing purposes. The sounds can be repeated a number of times.

In initial informal testing, participants generally felt that speech was easier to use and gave immediate information, but some thought that they may be able to more quickly and intuitively interpret the tone sounds with further practice.

Further details will be given at ICAD 2026.

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

An online version of the system has been developed [8]. The approach used is that of simply running the existing Windows-based HFVE desktop software on a virtual Windows server, the server 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). (See https://www.hfve.com/TryHfve.htm.)

Three servers were initially provided, located in Europe (London), West (Oregon), and East (Tokyo) (using "nearby" servers improved the application's responsiveness). However these are now reduced to a single London server, which uses minimal resources: 1 CPU, 512MB RAM, and 30GB storage.

Using the above approach allows a single codebase to run both on Windows Desktop, and online, with minimal flags needed to adjust processing. It costs about $12 US dollars per month to run (including a backup copy of the system disk).

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 on most standard Windows computers tested, and on an Android Samsung Galaxy Tab S, and on an iOS iPad Pro. (An add-on keyboard and/or mouse assists with using a tablet computer.)

## 3. ASSESSMENTS

Assessments of the new effect types, and of "VirtualHere" software, are in progress. Some feedback from initial informal assessments of Pan, and Span, effects, is included 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 2026.

## 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 only close the online HFVE session.

Future developments may include enhancements to the Imprint effects, including a new effect type termed "Sliders", which are moving and / or zooming / changing Imprints of items, 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 should allow local USB devices (e.g. force-feedback devices) to be used. Other approaches are using audio or visual signals from the online application to convey required locations, 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 2026.

## 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.