USING "IMPRINTS" TO SUMMARISE ACCESSIBLE IMAGES

David Dewhurst

www.HFVE.org
daviddewhurst@HFVE.org

ABSTRACT

"Imprints" are a new feature of HFVE (Heard and Felt Vision Effects), an experimental audiotactile vision substitution system being developed by the author. Imprints comprise groups of simultaneously-presented apparently-stationary audio and tactile effects, which have apparent spatial locations that correspond to the spatial locations of the content of the items that they represent. Imprints convey the approximate extent of items in a scene.

When the Imprint effects are speech-like sounds, they may give the impression of a group of people, each at a different location, speaking in unison. Imprints can produce the effect of successive visual items being "stamped out" or "printed", and can be used in conjunction with existing features.

The intention is to rapidly summarise the content of a scene, according to the task or activity being performed.

This paper describes several types of Imprint effects, and methods of producing them. Interaction methods are considered, and blind people's use of computer mouse-like devices to interact with the system is described. Possible applications are suggested, and the results of an informal assessment session with a totally blind person are reported.

1. INTRODUCTION

It is estimated that there are about 39 million blind people in the world [1]. Several attempts have previously been made to present aspects of vision to blind people via other senses, particularly hearing and touch. The approach is known as "sensory substitution" or "vision substitution".

1.1. Previous work

Work in the field dates back to Fournier d'Albe's 1914 Reading Optophone [2], which presented the shapes of printed characters 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.

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 at a rate that indicates the angle of slope.

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

Several tactile image-presentation systems have been developed that allow visual features to be presented via touch, usually via a matrix of tactile actuators (described later).

González-Mora et al. [8] have developed an experimental device which produces stereophonic "clicks", with a randomised order of emission, corresponding to the calculated 3-D coordinates of objects.

Many audio description methods have been devised, and blind people can use speaking colour identifiers to determine an item's colour.

Previous approaches allow users to actively explore an image, using both audio and tactile methods [9, 10]. The GATE (Graphics Accessible To Everyone) project allows blind users to explore pictures via a grid approach, with verbal and non-verbal sound feedback provided for both high-level items (e.g. objects) and low-level visual information (e.g. colours) [11].

The author has previously reported other features of HFVE, notably using audio and tactile effects ("tracers") to trace out the shapes of items in a scene; using distinct effects to emphasise the corners within an item's traced-out shape; and using buzzing sounds and other effects to clarify the shapes of items [13, 14]. These methods are effective for presenting item features that can be summarised via single or multiple lineal effects (e.g. the outlines of items). However in order to convey the two-dimensional arrangement of the content of an item the system previously used coded "Layout" descriptions. These presented the locations of content, via categorical coded speech sounds, braille, or Morse code-like taps.

1.2. HFVE "Imprints"

"Imprints" rapidly summarise the content of a scene via multiple stationary audio and tactile effects Fig 1, using mappings similar to those used for "tracer" effects. They are a new feature of HFVE and are believed to be novel, though having similarities to other approaches, which are described in [9, 10, 11, 12].

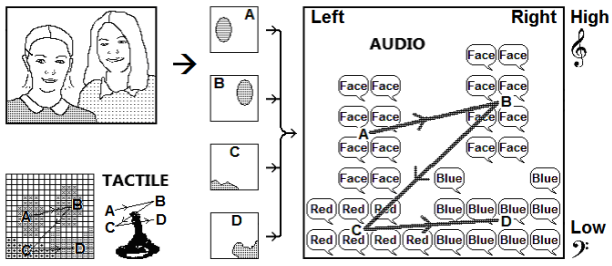

Figure 1. *Presenting image items via "Imprints".*

HFVE 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 interacting to control what is presented. ("Items" can be objects within a scene; regular regions of an image; abstract shapes; etc.) If we take the definition of interactive sonification as "the discipline of data exploration by interactively manipulating the data's transformation into sound" [15], then for Imprints "the data" is the content of a visual scene, with the user interacting via a variety of methods to control what is presented.

Imprint effects present the spatial distribution of items by using groups of simultaneously-presented effects to convey the arrangement of the items' content, which may be found to be a speedy and intuitive approach. The clusters of speech and other effects can instantaneously present the extent of the items being represented. The system can "step round" a scene or part of a scene, sequentially presenting Imprints of the items in the scene Fig 1.

The intention is that each item presented is perceived as a single "unified whole" or "gestalt", in a similar manner to how sighted people perceive successive visual features in a scene.

The HFVE approach, although allowing exploration (as often used in previous work), attempts to allow the system to decide what is presented, based on the user's current task or activity, so that the system is less tiring to use.

When Imprints are presented using speech sounds, they could be regarded as a form of augmented audio description, in which the speech describing the items is "spread" in "soundspace" to convey the location and extent of items Fig 1.

Blind users may not need to know the exact size, shape and location of each item – the approximate size and extent presented by an Imprint is often sufficient. However users can command the system to "lock on" to an item when it is presented, in order to obtain the exact shape etc. of the item.

Imprints can be presented in conjunction with other effects, such as shape-conveying buzz-track tracers, and optophone-like multiple tracer "polytracer" effects [14].

The nature and aesthetics of the sonification effects can be experienced by visiting the author's website [16], which includes demonstration videos.

2. "IMPRINT" TYPES, AND THEIR PRODUCTION

An "Imprint" consists of a group of simultaneously-presented apparently-stationary audio and/or tactile effects, which have apparent spatial locations that correspond to the spatial locations of the content of the item that they represent. In the audio modality, horizontal position is mapped to left-right stereophonic positioning, and vertical position is mapped to frequency (i.e. similar mappings to those used for "tracers").

Tactile Imprints have also been investigated, but at the time of writing have not been implemented in a practical manner, and are not covered in detail in this paper. Tactile Imprints could be displayed on a braille-like array; or on a matrix of tactile actuators, for example Telesensory's "Optacon" finger-read vibro-tactile array; Wicab's "Brainport" tongue-placed electro-tactile display; or EyePlusPlus's "Forehead Sensory Recognition System" electro-tactile display.

The author's website [16] contains demonstration videos showing Imprints summarising the colours within images, and shows braille-like tactile equivalents.

2.1. Types of Imprint effects

Audio Imprint effects can be speech-like, or use non-speech-like sounds, or present combinations of both.

If speech effects are used, then all effects at any moment usually "speak" the same words or encoded sounds (although in theory different parts of an audio display could output different speech, and the user could focus on one part at any time, using the "cocktail party effect").

Imprints produce a combined effect that may rapidly and intuitively convey the approximate extent of the item being presented. Wide-ranging items produce a "dispersed" effect of a wide range of pitches and apparent stereophonic locations. Compact items produce a more "constricted" effect of fewer, or closer, voices and of narrower pitch range.

An array ("lattice") of Imprint effects can comprise effects arranged at regular fixed points in a scene Fig 1. Alternatively, the effects in the regular lattice of effects Fig 2 (A) can be arranged to cover the presented item. If, for example, a smaller regular region is being presented, the several voices can be arranged to be apparently closer together (B), and the distinct reduced range of frequencies and stereophonic positioning may be easily and intuitively interpreted by the user. The lattice of effects can be varied in the vertical and horizontal direction, so that they match an object's area, "framing" the object (C).

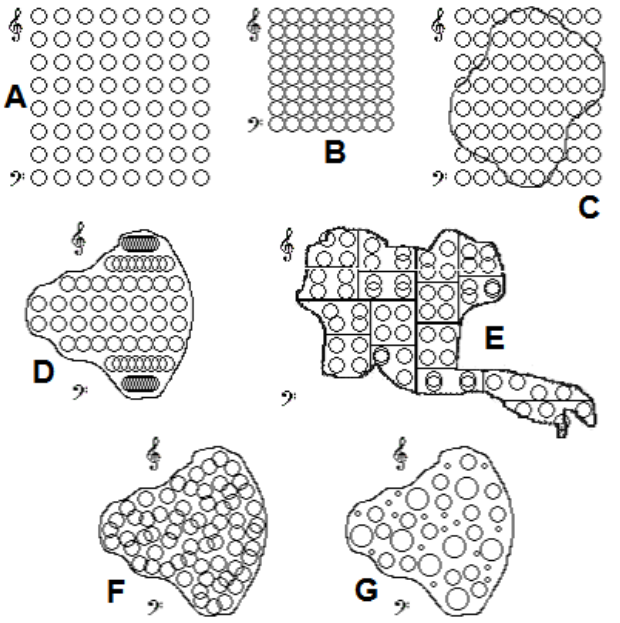

Figure 2. *"Imprint" effect arrangements.*

The lattice of effects may be arranged so that the same number of active individual effects are presented for each item (i.e. none are "switched off" as will normally be the case if a fixed (A) or rectangular (C) lattice of effects is used). The effects can be aligned vertically or horizontally (D). The lattice can be arranged in both directions so that the effects are evenly distributed according to the shape of the item (E), or a randomly scattered arrangement of effects can be used (F).

As well as speech, the effects can comprise non-categorical effects such as certain varying tone-sounds or buzzing effects, with certain continuously-changing properties used to present continuously changing quantities such as brightness.

Furthermore the energy (e.g. volume) of individual effects can be rapidly varied (G). The frequency and amount of variations in energies may be perceived as "bubbling" effects. The effects, though remaining in the same approximate location, can be rapidly "moved" in their apparent location. The movements can be regular (e.g. back and forth, in small circles or rectangles, spirals etc.) or irregularly. These "dynamic Imprints" produce a "bubbling" / effervescent effect. The "dynamic Imprint" effects can be mapped to visual properties e.g. brightness or texture. The frequency, evenness or unevenness of frequency, and amplitude of changes, can rapidly convey the texture of an item.

The items presented can be objects within a scene or section of a scene, and can be "stepped round" in sequence, as already described. Alternatively the content of an image, or a section of an image, can be continuously "streamed" via the several effects (i.e. simultaneously presented with no "stepping round" effect), with the categorical content and/or smoothly changing properties of each effect corresponding to the content of the location that they each represent, as it changes with time. In such cases the "spread" of the Imprint effects will correspond to the size, shape and location of the section of the scene being presented.

Both categorically-perceived speech Imprints and non-speech Imprints can be presented, either in succession, or simultaneously, with the balance controllable by the user.

Categorically-perceived speech sounds or sounds of distinct timbre can exhibit non-categorical continuously-varying intensity properties such as volume.

Optionally the volume and/or length of time of presentation of each Imprint can be varied to correspond to e.g. the size of the item that they represent.

2.2. Using Imprints with other effects

Differently-shaped items may sound similar if presented only as Imprint effects – the spread of pitches and stereophonic positioning may give a clear general impression of the extent of an item, but the exact form/shape, vertices, etc. of the item will not be clear from the Imprint effects alone. Consequently Imprints can be presented in conjunction with other effects, such as shape-conveying buzz-track tracers Fig 3 (C), or optophone-like "rectangular polytracer" effects (D).

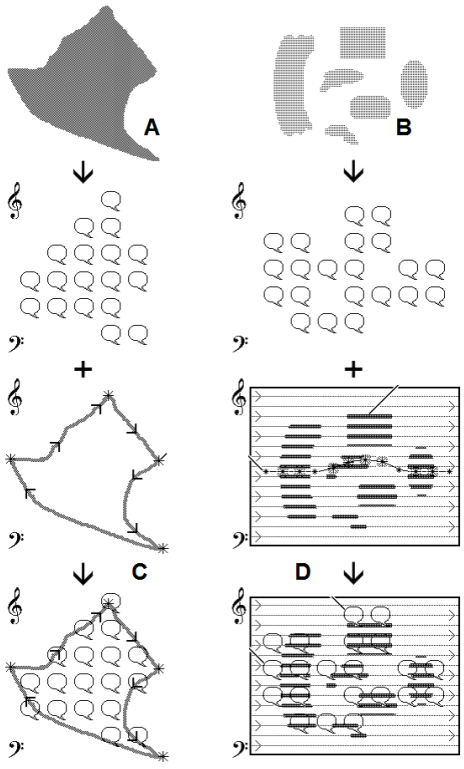

Figure 3. *Using "Imprints" with other effects.*

One effective approach is to present a buzzing outline "tracer" (C) if the item being presented is a single contiguous non-fragmented item (A), and optophone-like "rectangular polytracer" effects (D) if the item is fragmented (B).

One issue that needed addressing was how to integrate the short time-period Imprint effects with corresponding tracer and polytracer effects, which by definition require a certain period of time to trace out the required shape.

One approach is to "play" the Imprints at the same time as the tracers Fig 3, but this may cause confusion for the user, as well as requiring equal periods of time to be assigned to both processes, whereas one of the main motivations of using Imprints is to rapidly summarise the items in a scene.

An alternative approach is to allow the user to control when the detailed tracers or polytracers are presented. For example, if the system is "stepping round" a scene, sequentially presenting Imprints of the items Fig 1, a blind user does not generally need to know the exact size, shape and location of each item – the approximate size and extent presented by the Imprint may be sufficient. When a particular item is presented about which the user wishes to discover more, they can command the system to "lock on" to that item, then, for example, obtain the shape of the item via "tracers". The user does not have to seek such items – instead they can wait for the required item to be presented before issuing a "lock on" command.

In this way the user can get the benefit of the rapidly-presented Imprints, as well as the detail presented by tracers (or other effects).

2.3. Producing speech-like "Imprint" effects

It is usually necessary to produce stretched versions of the speech sounds used to produce Imprints, so that when the sounds are presented at differing pitches, the several speech sounds will still be synchronized (although pitching "on the fly" can alternatively be performed). If "panning" is used to achieve the stereophonic positioning (i.e. the same sounds are played on the left and right channels, but the volume of each channel is altered to give a horizontal positioning effect), then only one sample of stretched speech is required for each row of effects Fig 2 (D), because the same sample can be used for each position within a row of effects.

The algorithm to produce the speech-like Imprint effects is:-
a) Produce appropriately stretched or shortened monophonic waveforms of the same pre-recorded speech sample, with the frequency unchanged, using standard techniques (one stretched sample is required for each different pitch to be presented); then
b) Play the stretched samples simultaneously, with the pitch shifted and the sound location set appropriately for each point represented.

It was found to be beneficial to use a musical / logarithmic pitch relationship between rows of effects : if a linear relationship is used then it can produce harsh-sounding harmonic effects (this particularly effects tone-like sounds).

The system uses Microsoft's DirectSound "SetVolume", "SetFrequency", "SetPosition" and "SetPan" methods to set the volume, height-conveying pitch, and stereophonic or panned sound position respectively of the replayed samples.

Panned sounds generally use less resources than 3-D sounds, and produce effective Imprint effects if the pan parameter-setting technique described later is used. By using these methods it was practical to use 64 panned sound buffers in an 8 by 8 arrangement Fig 2.

The system is currently implemented on a standard Windows PC, using standard sound facilities, with force feedback effects presented on consumer devices (described later), moved via Microsoft's DirectInput "Spring" effects.

2.4. Producing non-speech Imprint effects

For non-speech sounds and other sounds that do not need to be synchronised on replay, step a) of the algorithm is not required, and the sounds can be replayed in a continuous loop, pitched appropriately.

Non-speech Imprint effects can be produced by processing samples of tone-like, bubble-like, "raindrop"-like, tapping, buzzing, and humming sounds etc. for outputting as non-speech Imprints. For non-continuous sounds such as "raindrop"-like sounds, the start, length, and intensity of the component sounds of the effects can be randomised around average values. This produces a "fluttering" or "rain on roof" effect, and the frequency, the length, the intensity, and the amount of randomisation, can be user-controlled and mapped to visual properties e.g. brightness or texture.

2.5. Deciding what to present

The effects presented can convey the nature of an item if identified (the "whatness" – e.g. face, blob, area of movement etc.), and its colours or other properties (e.g. for faces, the property could be a facial expression). They can be presented via speech sounds or via non-speech effects.

The speech sounds can be coded for brevity. However when tested in a small trial, for the case of colours, real-name (non-coded) colours were greatly preferred by participants [14]. 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. Even long colour names such as "DarkPurple" could be spoken rapidly (in about a third of a second) and still be understood. It has also been suggested that very short "Spearcons" [17] could be used, which would have the advantage of even greater brevity.

However comprehension is an issue with Imprints, as the multiple voices, of different pitches and locations, though synchronised, give the impression of a small crowd of people speaking in unison, which may reduce comprehension when compared to a single voice.

An informal assessment session with a totally blind person (described later) suggested that both speech and non-speech effects should be available, and user-controllable.

For example one very effective combination was to use a fixed "lattice" of "raindrop"-like Imprint effects Fig 2 (A) (i.e. not adjusted to frame or cover objects), with a speech tracer (with subdued buzz) conveying information. As each item was presented, the tap frequency and intensity of the "raindrops" was proportional to the area occupied by the item, with larger items causing more Imprint effects to be activated.

2.6. Improved stereophonic positioning

Whatever method is used to achieve the stereophonic positioning of sound effects (e.g. sound "panning" or "3-D" sound), it is important that the location conveyed by the sounds accurately reflects the location that is being represented.

One method of improving the stereophonic positioning effects is to allow the user to specify the 3-D or pan sound parameters for several locations along the horizontal axis that produce the most accurate impression of that horizontal location. The system can then interpolate positioning parameters to use for intermediate locations. In this way the user-perceived stereophonic locations may better match the locations being presented.

Such improved left-right stereophonic positioning can be used for any of the audio effects, and in the author's opinion produces a considerable improvement in the clarity of the presented locations. A similar approach can be used for the vertical axis if 3-D sound is used.

(Subtle psychoacoustic effects sometimes seem to influence the overall pitch of Imprints perceived by users – if several items are presented, of differing sizes, but centred on the same "height", then for some users, for speech sounds, the overall pitch appears higher with larger, more spread items; and lower with more constricted items. However for other sounds the effect can be reversed. This effect does not occur for all users. A possible explanation could be that a combination of masking effects and other psychoacoustic effects are occurring. The sounds could be adjusted to compensate for the effect, but this needs further investigation.)

2.7. System design

The sonification of a visual image into Imprints (and other effects) can be considered as a two-stage process, a "Vision" stage and an "Effects" stage Fig 4.

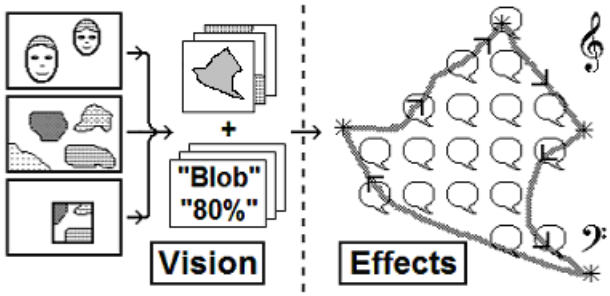

Figure 4. *Simplified system architecture.*

The "Vision" stage gathers "visual items" from images and decides which to present. This can vary according to the current task or activity, and this is a significant feature of the system : in vision, the importance of items depends on the task or activity being undertaken. For example, if you are looking for a red item of clothing, then only red items are of interest, while in a social situation people's faces would be of more interest. It is important that users can rapidly switch task/activity so that the system selects the most appropriate items.

The sources of visual items can be pictures, live images, media, shapes, data that can be presented visually, computer "desktop" or "clipboard" contents, etc.; and can be provided by external systems.

The process of identifying items can be complex, can involve "computer vision", and will not be covered in detail in this paper. The output of the vision processing is a set of "visual items", which can be "Regions" i.e. regular rectangular regions; or "Objects" i.e. entities (identified via computer vision processing or highlighted manually) including:- "blobs" of colour or other basic visual properties, recognised objects (e.g. people's faces), areas of movement, abstract shapes, components of graphs and charts, etc. (For prepared media, a sighted person can directly identify items, properties, etc.)

A convenient system architecture might be as shown in Fig 4 – items can be submitted to the "Effects" stage as bitmaps showing the item isolated against a neutral background, with accompanying data giving the nature of the item if identified (the "whatness" – e.g. face, blob, area of movement etc.); and

its “importance” (which will be task-dependent). External systems could also submit items in this manner.

The “Effect” stage can then prioritise and present the most important items (appropriate to the task) as audiotactile effects in the time available, according to the currently-selected options (which can also be task-dependent). The audiotactile effects can be "tracers" (including "symbolic tracers" and "polytracers"); "categorical” information e.g. layouts; and "Imprints" (which are the main subject of this paper).

3. INTERACTION

In considering interaction, there are a number of issues that need to be addressed. Most aspects of the system can be controlled by the user, including what is presented, and how. However this may be difficult for a blind person to do interactively. In any case, it may be beneficial to have a relatively low amount of user interaction during use, so that the system is less tiring to use.

3.1. Task/activity control

Instead it may be desirable for users to be able to command the system to set the system for particular tasks/activities as described above, and for the system to then set the several settings accordingly, so that for a given task or activity the user can be presented with suitable items without having to continuously control the content and presentation methods – instead these can be defined for each user-selected task or activity. One way of achieving this is to allow the user to record the settings that they change during a particular period of time, and link them to an activity. Later, on selecting the activity, only those controls that were changed during the recording period will be updated.

3.2. Control actions

An approach to interaction has been devised which is intended to be straightforward for a blind person to use. The approach makes volume and speed of effect presentation easily adjustable during use, and also provides three basic control actions for commanding the system. These can be triggered via keyboard keys, mouse or joystick buttons, or via specialist switches, and can be extended via “modifiers” (having a similar effect to pressing a keyboard “control” or “shift” key).

The first control action is a toggle action, causing the system to start or stop presenting effects. The second control action triggers task selection, allowing scrolling (e.g. via a mouse scroll-wheel) though a list of tasks/activities that are spoken by the system, until the desired task/activity-linked set of settings is reached. The third control action is a “lock on” command, selecting items for further activity e.g. presenting an item in more detail, or causing a selected region to follow the mouse or be selected for tracking.

Such control actions can be also implemented using other standard computer control methods, such as speech recognition with a “command and control” approach (i.e. using a limited number of recognised commands, making misinterpretation less likely), as well as via keyboard, touch pad or stylus input.

3.3. Using mouse-like devices to interact with the system

As blind people do not generally use a mouse or joystick when interacting with a computer, having a separate mouse or joystick purely for use with this application might be beneficial. The author has previously described using a mouse to “draw” with audio feedback, and using a mouse to navigate around areas of an image [14].

An interaction device should ideally be able to act as a tactile display, providing haptic / force feedback effects (e.g. presenting tactile tracers); and allow the user to indicate an absolute location within an image, and command the system via buttons etc.

Although most joysticks provide at least three buttons, their vertical handle orientation is designed for computer games etc., and is not ideal for presenting and receiving location information [13]. Logitech’s Wingman “Force Feedback Mouse” Fig 5 (A) overcomes these issues, and can be programmed to move like a powered joystick in order to trace out key features etc., yet can be moved and clicked by the user to perform mouse-like actions. Its constrained area of movement makes it straightforward for blind people to indicate an absolute location. It has previously been used and adapted to provide a larger area of movement for assistive technology applications [18]. Even in its unmodified state it can be an effective controller, as multiple clicks can be used to act as modifiers, so that, if desired, just one or two buttons can control the system, retaining at least one button for standard mouse click actions.

Figure 5. *Logitech’s Wingman Force Feedback Mouse (A), and an “MMO” mouse (B).*

The approach of "force joystick plus mouse" is an effective one, and the computer mouse has developed several useful features in recent years, notably wireless control, scroll-wheel, and extra buttons. The same functionality as a force feedback mouse can be achieved by attaching a mouse to a powered joystick – for example the Microsoft Sidewinder Force Feedback 2 joystick Fig 8 (A) can have a standard 5-button wireless mouse fitted to its handle, so providing scroll-wheel and multi-click facilities, as well as a hand-grip orientation more suited to this application.

For blind people, commanding via mouse clicks can be a problem as they may trigger unwanted actions. A solution is to lock the mouse pointer to a “controlled” part of the computer desktop, where any such mouse clicks can be correctly handled as control action commands.

Using the “three control actions plus modifiers” approach allows the system to be controlled by standard mice that have additional buttons that can be programmed to act as modifiers. “MMO” mice Fig 5 (B) typically have more than 12 separate programmable buttons, and these can be mapped to common actions, allowing full control without the use of modifiers.

Wireless “air mouse” devices such as Logitech’s “MX Air” or Gyration’s “Air Mouse” Fig 8 (B) allow mouse-like actions

without requiring a surface to work on, and so may be suitable to use as portable controllers. Gyration's Air Mouse has 3 extra programmable buttons and 8 programmable gestures, allowing 11 programmable actions to be performed. Gesture actions may be more intuitive, not requiring finding particular buttons, although when a blind tester was asked to try using an Air Mouse, he had no difficulty finding and operating the extra buttons.

Severely disabled people can use special switches to control the system, in a similar manner to button control, for example via a switch-adapted mouse. Single-switch control is possible, for example by using single-, double-, and triple-clicks to trigger the three control actions, with proceeding or following long-period clicks acting as modifiers.

Recently-developed touch devices, such as touch pads, can convey mouse-like signals to the system, and enable a blind person to easily give absolute location information to the system (unlike for standard unconstrained mice, which require e.g. audio feedback to indicate mouse pointer location). For example a small touch-controlled Windows "tablet" computer might be a suitable platform for the system, being very portable, and allowing the user to easily indicate locations within images via touch, for example indicating a section of an image for which they wish to receive an Imprint-presented summary.

In order that a totally blind person can control a tablet computer, large button areas can be arranged around the edge of the tablet screen so that the user can straightforwardly touch the intended command area.

4. USING "IMPRINTS"

There are many possible applications of the HFVE system, but two applications of Imprint effects used alone will now be described.

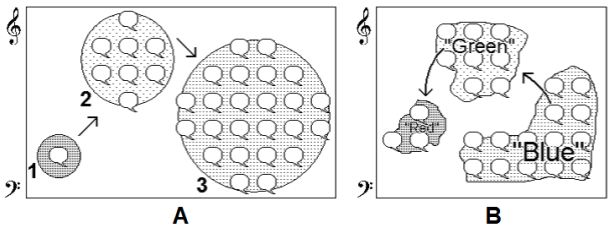

Figure 6. *A bubble chart, and an enhanced colour identifier, presented using "Imprints".*

4.1. Application : A bubble chart

"Bubble charts" Fig 6 (A) are effective when presented via Imprints, as the "spread" of the effects (and optionally the variation in intensity/volume and length of presentation time) may rapidly and intuitively convey the relative sizes of the "bubbles" in the bubble chart. The bubbles can be presented sequentially in, say, order of size (or any other appropriate order), but if the bubble chart is presented in the audio modality only, then it may be worthwhile to present the bubbles in the order in which they occur along the horizontal axis Fig 6 (A), so that their order along that axis is clear, as the horizontal audio location effects will generally be weaker than the pitch-based vertical axis effects. The intensity and length of the effects can correspond to the size of the bubbles. The "locking" facility described elsewhere could allow any particular bubble to be temporarily "locked on" to, so that the location and relative size of the bubble can be more clearly perceived – the system can then switch to presenting shape tracers or giving other details.

4.2. Application : An enhanced colour identifier

Certain visual properties, such as colour, tend to be perceived in a categorical way [19].

Imprint effects can be used to present the distribution of colours, or other visual properties, within an image, so as to produce, for example, an enhanced colour identifier Fig 6 (B).

If the several colours of an area and their distribution are to be presented (rather than the precise colour of a single point or the single average colour of an area) then one issue is how to decide on a limited number of colour shades which effectively describe the colours of the area. For a simple image or diagram comprising a limited number of colours, each colour can be presented in succession, via Imprint effects. However for an image containing many shades, for example a colour photograph, a different approach is needed.

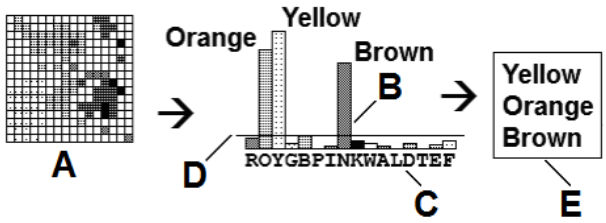

Figure 7. *Identifying prominent colour shades in an image.*

One approach is to identify a "sub-gamut" of colour categories, comprising the colour shades/categories that are found in more than a certain proportion of samples of an image, using a "histogram" approach Fig 7. A set of samples of smoothed colour (or other property) values (A) are obtained from an image, and each is categorised as one of the colour categories (B) in the full gamut (C) of colour categories. Those colour categories that have more than a certain proportion (D) of pixels samples assigned to them can be deemed a "predominant colour" and added to the "sub-gamut" (E). (Clusters of colour shades that straddle two or more colour categories should be assigned to one or other of the colour categories, and not divided into several colour categories.)

The same general approach can be extended to select and present categories of other property types, for example categories of textures. In this way many samples of visual properties can be represented via a limited number of appropriately-selected visual categories.

Once the "best" colour categories (or other properties) are determined, each part of the content of the image can be assigned to the nearest best colour (or to none).

The distributions of the colours can then be presented via Imprints Fig 6 (B), optionally with additional effects as previously described Fig 3.

Computer vision processing can be used to segment the image into larger blobs, for example by doing "moving average" filtering or other "blob extraction" techniques, so that larger non-fragmented regions of common colour can be presented.

Other applications for Imprints include presenting the results of "computer vision"-related techniques such as "blob extraction", motion detection, and object detection and tracking.

4.3. Using Imprints : An informal assessment session

It was important to obtain an independent assessment of the approaches described in this paper. "AB" (not his real initials), who has been totally blind since birth, kindly agreed to help assess the system, especially the new sonification and interaction methods. AB has considerable prior knowledge and experience of computer access for blind people, and was able to make many helpful points and constructive criticisms. In a free-format discussion session, the approaches were demonstrated, and the pros and cons considered.

The author first recapped the existing system, including:- using audio and tactile effects to trace out the shapes of items in a scene; using distinct effects to emphasise the corners within an item's traced-out shape; and using buzzing sounds and other effects to clarify the shapes of items. AB could recognize straightforward shapes using an unmodified powered joystick (a Microsoft Sidewinder Force Feedback 2 Fig 8) as a tactile display, and could also recognize them when they were presented via moving "buzz track" tracers and audio corner effects alone (i.e. with no tactile cues). AB found the buzzing effects and corner effects helpful in clarifying the shapes – without these features recognition was difficult.

An unexpected observation was that AB found the horizontal/left-right audio positioning clearer than the vertical/up-down positioning, despite having only the stereophonic cues (whereas vertical positioning is conveyed via pitch). This would tend to indicate that the new "panning" methods may have some benefit to users, although a counterargument could be that panning, unlike "3-D" sound, contains no inherent vertical cues (this could benefit from further investigation).

Moving to the new "Imprints" feature, AB was generally positive about these, and felt that they were an effective way to summarise a scene. The demonstrations were limited to test images containing solid items each of distinct colour, and, when speech was output, spoke only the colour of the item. However AB liked the feature of the system "stepping" sequentially round the items, and particularly liked the facility for the user to "lock on" to a particular item, so that it can be inspected more closely, then released and the Imprint stepping then continue (at the time of the demonstration, the locked-on item was repeatedly presented as an Imprint until unlocked, rather than allowing immediate presentation of the selected item via outline tracers etc.).

When items were presented via Imprints, AB could tell the difference between large items, containing a spread of pitches and horizontal positioning cues, and small items, with a more constricted range of cues, when both were centred on the same point (i.e. with the same "average" pitch and horizontal cues), and could also distinguish such items when they were offset from each other.

AB was unsure whether he preferred the technique of using a fixed 8 by 8 grid of effect points Fig 2 (A), where the number of effects presented indicated the area presented (with a reduced number for smaller items, giving a "sparse" effect); or preferred the "richer" sounds produced when the effects in the 8 by 8 grid are relocated with each item presented to either "frame" the item Fig 2 (C), or cover the item more precisely (D).

Interestingly, AB found it helpful having the buzz track included, even if the Imprints were using speech alone (the buzz track sound in such cases was a single buzzing effect centred on the middle of the item being presented, and moved from item to item as the system sequentially presented them). Similarly, when the Imprints are non-speech sounds, a single effect presenting the corresponding speech – for example the colour name – can be centred on the current item. AB said that it generally helped to have a speech component of some kind active. These observations tend to indicate that both the speech and non-speech effects should be available and user-controllable, with the user able to alter the relative amount of each.

The author briefly demonstrated the types of non-speech sounds that could be used for Imprints, including tone-like, buzzing, humming, tapping, bubbling and "rain on roof"-like effects (i.e. randomised "tapping" sounds). Of these AB preferred the latter (or speech).

Figure 8. *Microsoft's Sidewinder "Force Feedback 2" joystick, and Gyration's "Air Mouse".*

We concluded by considering interaction methods. As already mentioned, AB particularly liked being able to "lock on" a particular item when the system is presenting successive items via Imprints. Users can interact with the system (for example giving commands and indicating locations in an image) by using touch, speech, or a conventional keyboard and mouse. The author demonstrated using a Gyration "Air Mouse" Fig 8 (B) to start and stop effects, lock on an item, and set groups of controls via task-linked commands, as well as moving the mouse (on a tabletop or "in the air") to indicate locations, draw shapes, and perform recognised gestures to control the speed, volume, and zoom the area of interest. AB was able to use the Gyration Air Mouse in this manner after a few minutes practice.

5. CONCLUSIONS AND FUTURE WORK

"Imprints" are a new way of summarising the visual content of a scene, and, when combined with the previously-reported methods, allow a blind person to access several aspects of visual images. The initial results and feedback are encouraging, and indicate that the approach is worth progressing. Future work should include obtaining more feedback from blind users, and if possible should include a systematic evaluation with multiple users, and a statistical analysis of the results.

The system's current state of development will be demonstrated at ISon 2013.

6. REFERENCES

[1] World Health Organization, "Visual impairment and blindness" in Fact Sheet No. 282, Updated October 2013, http://www.who.int/mediacentre/factsheets/fs282/en/.

[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, pp. 431-453.
[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] González-Mora, J.L., et al., Seeing the world by hearing: Virtual acoustic space (VAS) a new space perception system for blind people, in *Touch Blindness and Neuroscience*, S. Ballesteros and M.A. Heller, Editors. 2004, UNED: Madrid, Spain. p. 371-383.
[9] Roth P, Richoz D, Petrucci L, Pun T., "An audio-haptic tool for non-visual image representation" in *Proceedings of the Sixth International Symposium on Signal Processing and its Applications* 2001 (Cat.No.01EX467) : 64-7.
[10] Patrick Roth, Thierry Pun: "Design and Evaluation of Multimodal System for the Non-visual Exploration of Digital Pictures". In *Proceedings of INTERACT 2003*
[11] Kopeček, I and Ošlejšek, R. "GATE to Accessibility of Computer Graphics" in *Computers Helping People with Special Needs: 11th International Conference*, ICCHP 2008. Berlin: Springer-Verlag, pp. 295-302, 2008.
[12] Kopeček, I and Ošlejšek, R. "Hybrid Approach to Sonification of Color Images" in Proceedings of the 2008 International Conference on Convergence and Hybrid Information Technologies. Los Alamitos: IEEE Computer Society, 2008. p. 722-727
[13] D. Dewhurst, "Accessing Audiotactile Images with HFVE Silooet" in *Proc. Fourth Int. Workshop on Haptic and Audio Interaction Design,* Springer-Verlag, 2009.
[14] 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.
[15] A. Hunt and T. Hermann, "Interactive Sonification" in *The Sonification Handbook*, T. Hermann, A. Hunt, J.G. Neuhoff (Eds.) 2011, p. 274.
[16] *The HFVE System*, http://www.hfve.com.
[17] Palladino, D., & Walker, B. N. (2007). Learning rates for auditory menus enhanced with spearcons versus earcons. *Proc. of the International Conference on Auditory Display* (ICAD2007) (pp. 274–279), Montreal, Canada.
[18] Roth, P., Giess, C., Petrucci, L., & Pun, T. (2001). Adapting Haptic Game Devices for non-visual Graph Rendering. Paper presented at the *HCI 2001*, International Symposium on Signal Processing and its Applications, New Orleans, LA, USA.
[19] Berlin, B., and Kay, P. Basic color terms: Their universality and evolution. University of California Press, Berkeley, CA, USA, (1969).