THE DESIGN AND EXPLORATION OF AUDITORY DISPLAY EFFECTS FOR BLIND DRIVERS IN AUTONOMOUS VEHICLES

David Dewhurst

www.HFVE.org

david.dewhurst@HFVE.org

Hyundai Motor Company Design Challenge :

“Auditory User eXperience Design for Autonomous Vehicles and Future Mobility”

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) display effects. The main contribution of this paper is to describe the design, and examine the effectiveness, of these effects in an automotive context, specifically in the context of blind drivers travelling in autonomous (self-driving) vehicles.

This paper also brings together and summarizes auditory display effects and techniques that have previously been reported by the author, and describes several new features. The effects are termed tracers; polytracers; drone and matrix effects; imprints; and multi-level multi-talker “focus” effects.

The paper describes the potential automotive application of such auditory display effects in:- command and control; route presentation; maps/cartography; and enhancing the journey experience of blind travelers.

Methods of presenting rectangular areas within a scene, (termed “audio previews”) are described and discussed, as is the concept of a small set of effects termed a “glimpse”.

The results of informal assessment sessions with a totally blind person, and two sighted people, are described.

1. INTRODUCTION

One source estimates that there are about 39 million blind people in the world; and another estimates that there are nearly 253 million people who are blind or visually impaired [1], [2]. 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”.

The coming introduction of autonomous vehicles presents many opportunities for blind people. One estimate foresees fully autonomous cars accounting for up to 15 percent of passenger vehicles sold worldwide in 2030 [3].

This paper focuses on the use of auditory displays as part of the user experience of autonomous cars, mainly for blind travelers, but with application to sighted users whose visual attention is elsewhere.

1.1. Other previous work

There is often the need to convey general visual information to blind people. An existing approach is to use relief images e.g. tactile maps. While these are convenient for conveying unchanging two-dimensional images, the instantaneous production of vision substitution images is more difficult to achieve. Devices can be devised that present other senses with information that includes aspects of sight, but other senses are not as powerful, or as able to comprehend such information [4].

Work in the field dates back to Fournier d'Albe's 1914 Reading Optophone [5], 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 Fig. 1.

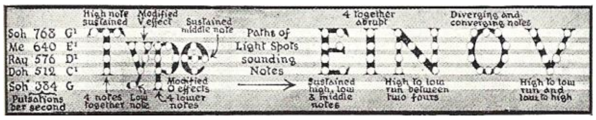


Figure 1: Optophone scanning across printed type.

Other systems have been invented which use similar conventions to present images and image features [6], [7], or to sonify the lines on a 2-dimensional line graph [8]. 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.

Previous work in the field is summarized in [9], [10]. Previous approaches have allowed users to actively explore an image, using both audio and tactile methods [11], [12]. BATS (Blind Audio Tactile Mapping System) presents maps via speech synthesis, auditory icons, and tactile feedback [13]. The GATE (Graphics Accessible To Everyone) project allows blind users to explore pictures via a grid approach, with verbal and nonverbal sound feedback provided for both high-level items (e.g. objects) and low-level visual information (e.g. colors) [14], [15]. An approach used by the US Navy for attending to two or more voices is to accelerate each voice, and then serialize them [16].

The Discrete REconfigurable Aural Matrix (DREAM) is a multi-speaker array technology, and [23] describes an approach to presenting multiple geometric shapes, including vertex highlighting; and producing “aural paintings”.

Google's Lookout software allows blind users to identify information about their surroundings [2]. Microsoft's Seeing AI software allows users to touch an image on a touch-screen to hear a description of objects within an image and the spatial relationship between them [17].

(The merits of these other approaches are not discussed further in this paper.)

1.2. 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 controlling 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.

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 several of the features described in this paper have been reported previously [19], [20], [21], [22].)

2. SUMMARY OF AUDITORY DISPLAY EFFECTS, AND USER EXPERIENCE

In this section the auditory display effects produced by the HFVE system are summarized. Tactile equivalents will also be briefly described.

Most of the auditory display effects described below can be combined – for example imprints and tracer effects can present the same item simultaneously.

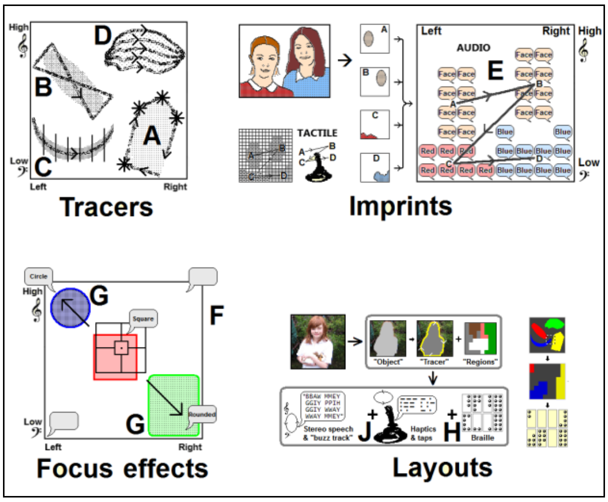


Figure 2: HFVE effect types (see text for details).

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

The nature and aesthetics of the auditory display effects can be experienced by visiting the author’s website, which includes demonstration videos [18].

The application of the effects to automotive use is described and discussed in Section 3 below.

2.1. Tracers

By smoothly changing the pitch and binaural positioning of particular sounds, speech and other sounds can be made to appear to move, whether following a systematic path, or describing a specific shape. Such moving effects are termed “tracers”, and can be “shape-tracers” (A) Fig 2, whose paths convey the shapes of items in an image [19].

In the tactile modality, tracer location and movement can be presented via a moving force-feedback device Fig. 5 that moves/pulls the user's hand and arm – in both modalities the path can describe the shape, size and location of the items.

As the system outputs both audio and tactile effects, users can choose which modality to use; or both modalities can be used simultaneously.

The system presents corners/vertices within shapes (A) Fig 2, which are found to be very important in conveying the shape [19], [21]. Corners are highlighted via audiotactile effects that are included at appropriate points in the shape-conveying tracers, for example by momentarily stopping the tracer, or outputting a short distinct audio or tactile effect.

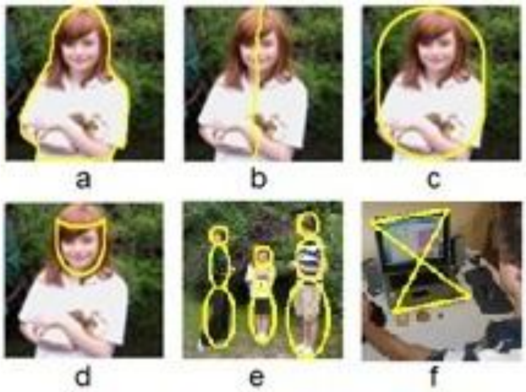


Figure 3: Shape, and symbolic, tracer paths (the yellow lines show the routes travelled by several types of tracer).

Although one possible tracer path for presenting an item's shape is the item's outline (a) Fig. 3, other paths such as medial lines (b), or frames (c), can be used. Symbolic paths e.g. for Face (d), Person (e), and “Unknown” (f) & (B) Fig 2 are found to be effective, as they present the location, size, orientation and type of object via a single tracer path.

A “drone” or “buzz” tracer, played at the same time as the speech tracer, can more clearly convey the size and shape, and present other information [20]. One effective such sound is a buzzing sound, but with a clearly defined pitch. (An additional non-speech tracer, of differing timbre, can convey distance information, if available, via pitch. Alternatively, the pitch of either the standard speech or standard buzzing sound can convey distance information, with the other conveying height. A similar approach can be used for presenting distances for polytracers, imprints, etc. – see below.)

“Matrix” effects (C) Fig 2 are produced by dividing the image into several equal-width columns and/or rows, so that special effects can be triggered whenever the tracer moves from one such column or row to another, allowing the shape of lines to be perceived more clearly – if the tracer travels at a constant speed, the rate at which the effects are presented will correspond to the angle of slope [20].

“Polytracers” (D) Fig. 2 are additional multiple tracers that are used to present the same item [20]. A polytracer can present non-speech tone sound tracers in a similar manner to existing optophone-like systems; or the extra tracers can also be speech, presenting the same speech sounds 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 moving speech sounds that the voices present are each stretched or shortened as required in order that the re-pitched voices together present synchronized speech sounds.

2.2. Imprints

“Imprints” (E) Fig. 2 rapidly summarize the content of a scene via multiple stationary audio and tactile effects, using mappings similar to those used for tracer effects [21]. (In (E) Fig. 2 each speech bubble represents a voice/speech source.)

Audio imprint effects can be speech-like, or non-speech-like (e.g. bubbling/dynamic) sounds, or combinations of both.

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, with a narrower pitch range.

The multiple voices, of different pitches and locations, but synchronized, give the impression of a group of people speaking in unison.

The system can step round the items of a scene, sequentially presenting imprints of the items (E) Fig. 2.

In an assessment session a totally blind person suggested that both speech and non-speech imprint effects should be available, and be user-controllable [21].

2.3. Multi-level multi-talker (“Focus”) effects

Multi-level multi-talker effects (termed “Focus effects”) [F] Fig. 2 allow several properties and items, at different levels of view, to be presented and perceived at the same time [22].

A blind user can rapidly navigate between such levels, e.g. by using a mouse wheel or a dial device, while hearing the focus effects speaking the level of view (e.g. spreadsheet cell, column, row, or block) that is currently emphasized, and at the same time being made aware of the levels above and below the current level of view, which have distinguishing effects applied (e.g. voice character, persona, etc.).

Focus effects can also be used to present property values of non-visual and non-spatial properties, for example levels of categorization and analysis, as found in academic and other fields. For example a car manual, or the Dewey Decimal system [30] could be presented and navigated round using focus effects, as described in section 3.1 below.

The system presents the items that are currently the primary focus of attention via crisp non-modified sounds, for example via speech sounds. At the same time the system presents the speech sounds for items that are not at the focus of attention, but applies a distinct differentiating effect on them, for example by changing the character of the speaker, or by applying echo or reverberation effects.

The system can artificially move the presented items (G) Fig. 2, so that the audio separation is maximized. This helps users to focus their auditory attention on the item emphasized by the system, or switch their attention to another item that is also presented but not emphasized. The user can then cause the system to highlight that other item instead.

The differentiated focus effects, for example echo and reverberation, can be applied to most of the other effect types, such as polytracers or imprints, so allowing such effects to have a faraway, hazy, unfocussed, quality analogous to the way that photographers use depth of field to accentuate focused items, with out-of-focus items also present which the observer is aware of but not directed towards.

In the visual domain, the system can produce higher-level consolidations of image content [22]. While HFVE knows how to consolidate general visual images, it does not know about other domains such as, for example, Excel spreadsheets. Instead such entities can be submitted to HFVE as client entities, for HFVE to present. For example consider the spreadsheet (A) Fig. 4. Although it could be presented as a visual-domain view i.e. as a series of patches of color and perhaps some text recognition, it is more meaningful to be able to inspect it via a spreadsheet-domain view (B), consolidating cells (Level 5) to columns and rows (Level 4), then to individual blocks (and objects such as charts and pictures) (Level 3), then to all blocks (and all objects) (Level 2), then to top level Spreadsheet (Level 1). (Level 0 gives a top-level overview of all available domain views.)

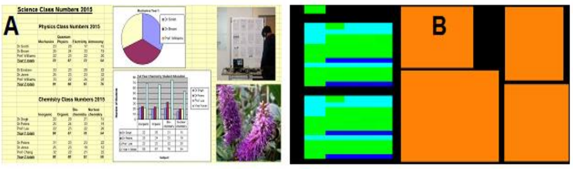

A
Science Class Numbers 2015
Physics Class Numbers 2015
Chemistry Class Numbers 2015
B

Figure 4: A spreadsheet and corresponding ItemMap.

Such higher-level view groupings facilitate obtaining meaningful summaries/overviews of content, and help with navigating around the items of the image/entity.

In order to present externally-produced images and other entity types via HFVE, a straightforward interfacing method has been devised. This comprises submitting a standard 24-bit color bitmap (.BMP) file e.g. (B) Fig. 4 that includes all of the required basic item blobs (referred to as the “ItemMap” file); and a standard text (.TXT) file (referred to as the “ItemKey” file) that describes how the blobs are marked via particular bit settings on the bitmap, and specifies how those blobs are consolidated to higher-level items. This pair of files, that fully describes the blobs of the image/entity, and how they are consolidated, can be created manually using a simple image painting application and a text editor, or can be created via an external application.

In the case of a spreadsheet Fig. 4, the ItemMap bitmap and ItemKey text file can be produced automatically via an Excel Add-In that has been developed. HFVE does not know about Excel, but processes the resultant pair of files like any other, getting item identifier bits from the ItemMap bitmap pixels, then looking up the corresponding item details (e.g. words to speak) from the ItemKey text file.

For certain entities some blobs may overlap (for example detected faces, and areas of color), and the system can reserve a certain number of bits in the 24-bit bitmap for particular sets of non-overlapping blobs. Such content is resolved by the ItemKey text file, which specifies which bits are significant, and their values for particular items.

Demonstration videos of the auditory display effects are available at the website: http://hfve.com. (Note that most of the effects that are described above in this section have been reported previously.)

2.4. Tactile display effects

Though not the primary concern of this paper, many of the auditory effects have corresponding tactile equivalents.

A force-feedback joystick makes an effective pointing device with which to indicate areas of the image, as it can also be programmed to tend to position itself to one of a number of set positions, so that a "notchy" effect is felt as the joystick is moved, giving a tactile indication of location.

A force-feedback joystick can 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 moving successive Spring condition effects, so pushing and pulling the user's hand and arm to trace out shapes (and highlight corners). Additionally it can be used to command the system via button and twist actions, and tap effects can present morse-like information to deafblind users (J) Fig 2.

Microsoft's Sidewinder Force Feedback joystick and Logitech's Wingman Force Feedback Mouse Fig. 5 are suitable devices, and can be controlled via Microsoft's DirectInput methods.

Figure 5: Microsoft's Sidewinder Force Feedback 2 joystick, and Logitech's Wingman Force Feedback Mouse.

Though both of these example force-feedback devices are relatively dated, bespoke new force-feedback devices could be developed for use in new automotive applications.

(Tactile braille effects (H) Fig. 2 and tactile tap effects (J) (known as "Layouts") can also be output to tactile devices by the system, as described in earlier papers [19], [20].)

The method of user interaction can be "exploring" in style, using a moving pointer to inspect a scene; or alternatively allowing the system to announce content, and then selecting areas for further inspection – the latter approach requiring less input from the user, and applicable to information navigation.

The user can tap commands onto a touch-screen or touchpad, and touch or drag over them to indicate parts of the image [22]. An optional pitched and panned buzzing sound can convey the location of the pointer within the image area.

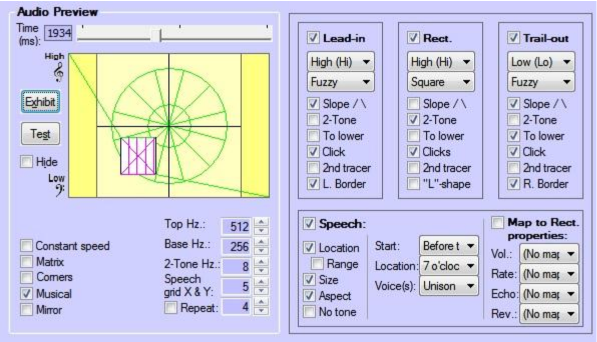
2.5. Audio previews

To summarize, the project uses both tone and speech sounds, suitably modified, in order to convey visual information to blind people. This is exemplified in some very recent (incomplete) development work, which involves presenting the location, size, and aspect ratio, of a smaller rectangle within the larger square extent of the full presentation area Fig 6.

It is intended that such "audio previews" can be used to quickly convey the location and extent (size, aspect) of an item immediately prior to its details and exact shape etc. being presented via the other methods described, especially when only audio methods are available i.e. no tactile feedback. 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 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; "L"-shaped presentation 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.

The audio properties such as high or low pitch start, oscillating pitch frequency, and other sound properties, can be mapped to rectangle (or other) properties.



Audio Preview

Time (ms): 1934

□ Hide

□ Constant speed
□ Matrix
□ Corners
☑ Musical
□ Mirror

Top Hz.: 512
Base Hz.: 256
2-Tone Hz.: 8
Speech grid X & Y: 5
□ Repeat: 4

☑ Lead-in
High (Hi)
Fuzzy
☑ Slope / \
□ 2-Tone
□ To lower
☑ Click
□ 2nd tracer
☑ L. Border

☑ Rect.
High (Hi)
Square
□ Slope / \
☑ 2-Tone
□ To lower
☑ Clicks
□ 2nd tracer
□ "L"-shape

☑ Trail-out
Low (Lo)
Fuzzy
☑ Slope / \
□ 2-Tone
☑ To lower
☑ Click
□ 2nd tracer
☑ R. Border

☑ Speech:
☑ Location
□ Range
☑ Size
☑ Aspect
□ No tone

Start: Before t
Location: 7 o'cloc
Voice(s): Unison

□ Map to Rect. properties:
Vol.: (No maj
Rate: (No maj
Echo: (No maj
Rev.: (No maj

Figure 6: User interface for experimental "audio previews".

Alternatively or additionally, speech output can directly present the rectangle extent e.g. "Eight o'clock, square, medium", the speech 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 using cartesian coordinates (e.g. chessboard-style "B3", or phonetic alphabet-style "Bravo 3", or "Top right" style), or polar (via clock-face positions). The volume, speech rate, optional echo and reverb effects, and other speech properties, can be mapped to particular properties.

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 test purposes. The sounds can be repeated a number of times.

See section 4 below for initial informal test results. 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.

2.6. Glimpses

In developing the system, it was found to be effective to apply the concept of a short list of typically 4 to 8 items that are presented in a burst of a few seconds, for example as imprints stepping through the list and presenting the approximate location and size of each item. These small sets of items are termed "glimpses" Fig. 7. The concept can be used to help indicate an appropriate number of items to present at any point, whether the user is exploring an image ad-hoc, or navigating around the items in the layers of items.

The author speculates that the effectiveness of the number of items, and the timing of such glimpses, may be related to the neuroscientific and psychological concept of a "psychological present", wherein there is a window of about 2 to 3 seconds within which your brain fuses what you are experiencing [24]. It may also relate to the well-known effect that only about 6 to 8 unrelated "chunks" of information can be comfortably handled in people's short term memory [25].

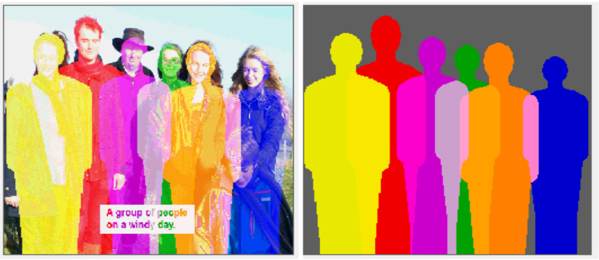


Figure 7: A "glimpse" – a set of a few items, that are presented over an approximately 2 to 3 second period.

Furthermore, from a system development point of view such a concept leads to an effective system interface – it allows a clear separation of, and interfacing between, the item selection and navigation processing; and presenting the auditory (and tactile) effects representing the items.

Another advantage of this approach is that it provides a scalable design and allows effects to be distributed across several instances of the application, and in theory across several machines (virtual or actual). For example the system may produce an impression analogous to that of "covert attention" in vision – several instances can each present the content of separate locations i.e. the user can be simultaneously presented with data about several locations, whereby the effect known as covert attention is simulated.

3. AUDITORY DISPLAY EFFECTS FOR BLIND TRAVELLERS IN AUTONOMOUS VEHICLES

In this section the potential applications of the described auditory display effects for blind travelers in autonomous vehicles are considered and discussed.

It is assumed that any such vehicle will correspond to Level 4 or Level 5 of the SAE's automation level definitions i.e. requiring no driver attention [26].

The following application areas will be considered: command and control; route presentation; maps/cartography; and enhancing the journey experience of blind travelers.

Automatic list to bitmap production, route presentation, and maps, will be demonstrated at the ICAD conference.

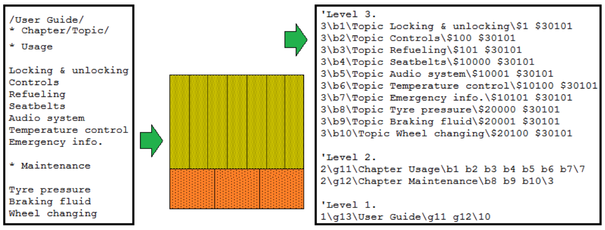
3.1. Command and control

A blind person in charge of an autonomous vehicle will often need to both give instructions to the vehicle; and receive information and feedback from the vehicle.

One way of giving and receiving such information is to use pseudo-visual representations of hierarchical multi-level structures such as menu structures, lists, etc. These are termed "ListMaps", and can be thought of as 3D explorable entities that can be automatically created from text lists.

The example of a simple car user guide Fig. 8 shows a simple text file that is automatically converted by the system into an ItemMap bitmap, and an ItemKey describing the basic items, and how they are consolidated up to group items.

This is performed by initially totaling up the content of whatever quantity is to be expressed by the area shown, for each group item (the quantity can simply be number of basic items). Then starting at the highest level the image area is split into rectangular areas each sized according to e.g. the basic item count for the group items at the highest level, then within each such rectangular area splitting further according to the next level content, until a pattern of similar-area small rectangles representing the basic items is produced, grouped according to their higher-level classifications Figs. 8 & 9.



```
/User Guide/
* Chapter/Topic/
* Usage

Locking & unlocking
Controls
Refueling
Seatbelts
Audio system
Temperature control
Emergency info.

* Maintenance

Tyre pressure
Braking fluid
Wheel changing
```

```
'Level 3.
3\b1\Topic Locking & unlocking\$1 $30101
3\b2\Topic Controls\$100 $30101
3\b3\Topic Refueling\$101 $30101
3\b4\Topic Seatbelts\$10000 $30101
3\b5\Topic Audio system\$10001 $30101
3\b6\Topic Temperature control\$10100 $30101
3\b7\Topic Emergency info.\$10101 $30101
3\b8\Topic Tyre pressure\$20000 $30101
3\b9\Topic Braking fluid\$20001 $30101
3\b10\Topic Wheel changing\$20100 $30101

'Level 2.
2\g11\Chapter Usage\b1 b2 b3 b4 b5 b6 b7\7
2\g12\Chapter Maintenance\b8 b9 b10\3

'Level 1.
1\g13\User Guide\g11 g12\10
```

Figure 8: A simplified car user guide text file section listing, and the corresponding ItemMap and ItemKey.

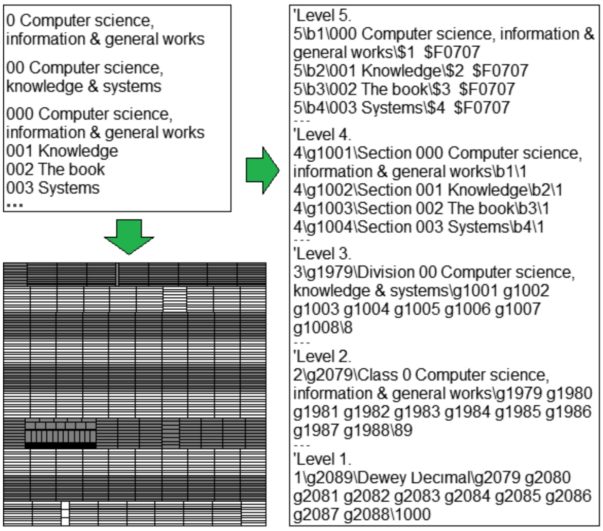
Another example in an automotive context is using the approach to present the legs of a journey route, wherein the route is structured as a set of legs, and each leg is sized to match either its distance, or estimated duration, so giving an impression of the relative distance or duration of each leg.

Complex non-visual multi-level/structured entities may also be presented as pseudo-visual/spatial representations. An example would be a full maintenance manual, that might have a complex hierarchy structure. A non-automotive example of such a structure is the Dewey Decimal classification system [30] Fig. 9. The levels might be Level 2 Class (e.g. 500 / Science & Maths) – Level 3 Division (e.g. 510/Maths) – Level 4 Section (e.g. 516 / Geometry) – Level 5 Sub-section (e.g. 516.3 / Analytic Geometry) (with Level 1 giving the entity/domain view name). The lowest level items i.e. Sub-sections can be automatically marked on a bitmap as block patterns of small rectangles Fig. 9, each of a unique color shade, which can then be consolidated up through the levels to the higher-level group items. Then when presented as audio (and tactile) effects, the user can obtain an impression of the size and distribution of the items at each level of the entity, and navigate around them.

The basic items in the bitmap, and resultant group items, can then be presented via any of the auditory display effect types described above, with the user controlling the level of view, and then either receiving the information at any level via automatic stepping, with the user locking on the item (e.g. a chapter, in the Fig. 8 simple handbook example) that they wish to explore further when it is presented; or directly exploring the items at any level – in use, the user can freely move the pointer to find a higher level group item, lock on it, and then explore the lower level items within that item.

In this way a spatial / dimensional impression of a visual or non-visual structure can be produced.

(When moving the pointer, one option is for a different voice to start when a new item is to be announced (optionally with a different persona), but with the earlier voice continuing on at a reduced volume level, being reduced further with each subsequent item (the previous voices can also be moved to the side to keep them distinct from the new main voice). This may produce a less abrupt effect on change of item, with the previous voices gradually fading away.)



```
0 Computer science,
information & general works

00 Computer science,
knowledge & systems

000 Computer science,
information & general works
001 Knowledge
002 The book
003 Systems
...
```

```
'Level 5.
5\b1\000 Computer science, information &
general works\$1  $F0707
5\b2\001 Knowledge\$2  $F0707
5\b3\002 The book\$3  $F0707
5\b4\003 Systems\$4  $F0707
---
'Level 4.
4\g1001\Section 000 Computer science,
information & general works\b1\1
4\g1002\Section 001 Knowledge\b2\1
4\g1003\Section 002 The book\b3\1
4\g1004\Section 003 Systems\b4\1
---
'Level 3.
3\g1979\Division 00 Computer science,
knowledge & systems\g1001 g1002
g1003 g1004 g1005 g1006 g1007
g1008\8
---
'Level 2.
2\g2079\Class 0 Computer science,
information & general works\g1979 g1980
g1981 g1982 g1983 g1984 g1985 g1986
g1987 g1988\89
---
'Level 1.
1\g2089\Dewey Decimal\g2079 g2080
g2081 g2082 g2083 g2084 g2085 g2086
g2087 g2088\1000
```

Figure 9: Part of the Dewey Decimal classification levels, and the corresponding ItemMap and (part of) the ItemKey.

In an automotive context, the contents of handbooks and troubleshooting guides, and journey details (as described); as well as a car’s technical settings and control and gauge values etc.; could be presented in a similar manner, and this facility could also be used by sighted people.

3.2. Route presentation

It is assumed that any autonomous vehicle will have the planned route available In addition to the approach just described, the route could be presented to a blind traveler as a tracer, whether auditory, or auditory and tactile. If a force-feedback device Fig. 5 is available, then it can be locked to the route within the mapped area, analogous to moving a pencil tip along a shaped groove. The force-feedback device can either be moved by the system to show the route, or can be free to move, but locked to the route, so that the user can move it back and forth along the route to feel its shape/pattern, while the system announces points along the route corresponding to the current location of the device.

If a touch-screen or other touchpad is available, the user can control the position along the route by dragging back and forth across the screen of the device.

The whole route, the fraction of the route travelled so far, and/or the fraction remaining, can be presented to the blind traveler via a tracer, so giving them an auditory (and tactile) impression of the each fraction of the journey. The timbre of the tracer can change between presenting the fraction covered, and the fraction remaining. The speed of the tracer can be constant i.e. presenting the distances involved; or related to the actual expected speed of travel, so presenting the estimated timings. (This presentation approach could also be useful for sighted people.)

Route presentation can provide the following features:-

- Presentation of features for the point in the journey being presented, including : road name, road number, town/village name, predicted traffic, road works, speed limit, landmarks, nearby hotels, restaurants, tourist attractions, landscape (urban, forest, moorland, etc.), fuel/recharging points, etc. These can use the focus effect and selection methods described above. Higher level geographical/political regions such as city, county, or country can also be presented.

- The system can use a bitmap marked with features for possible presentation, in a similar manner to that described for visual and other domains in section 2.3 above. Alternatively the system can note the current coordinates of the location along the route and look up the features from other sources on the fly.

- Ability to zoom in and out of the route.

- If the route doubles back on itself, this can be difficult for a blind person to be aware of if they are pushing a force-feedback device along the route. The system can highlight the issue via audio or tactile means. Additionally the system can automatically drive the force-feedback device though the tight corner and then return control to the user, or the user can instruct the system to do this.

- If a force-feedback device is being used then the device can display damper or friction conditions (i.e. be made harder or easier to move) depending on the speed limit and expected traffic conditions along the route. This gives an intuitive impression of likely speed of progress along the route. Corresponding audio effects can also be presented.

- Alternative routes can be presented.

3.3. Maps / Cartography

The HFVE system clearly has application in presenting maps to blind people. The maps can be geographical, or can be political, for example structured as levels showing State – Country – Region – County – Town etc., allowing the user to explore using the methods already described, and to obtain the shape and extent of any such area.

Alternatively a user can do a simple search of any named area, and, for example, get an impression of its distance from the current location, relative size, etc. via e.g. tracer and imprint effects.

Several places could be presented simultaneously, or stepped round, so giving an auditory impression of their distance separations, and relative sizes. Many similar cartography applications can be devised.

3.4. Enhancing the journey experience of blind travelers

As well as presenting practical information related to the journey, the system can be used to present many auditory display effects related to other aspects of travel (as well as allowing the blind traveler to access non-travel-related media, and such things as spreadsheets and structured data as described elsewhere – such uses will not be discussed here).

The system could be used to allow the blind traveler to be more aware of their surroundings along their journey. For example the system can use standard AI-related methods to present information about signage, people, etc. along the route travelled, by using text recognition and face detection respectively. For the demonstration system, open source Tesseract OCR is used for text recognition, and open source OpenCV is used for face, and motion, detection, though cloud-based services could also be used [27], [28].

Person detection is less straightforward to achieve in arbitrary situations, but the demonstration system can optionally use the simple approach of assuming that any face has a person's body below it, and that the size and distance of the person is related to the size of the detected face, with nearer persons overlapping further-away persons. (If similar-sized faces are detected close together, then the higher-located faces will typically be for persons further away than the lower-located faces. Adjustments can also be made for age, gender, etc.)

Fig. 10 shows this process in action, with 5 faces detected, and with the resultant assumed figures overlapped appropriately (note that no statistical testing has yet been done to assess the applicability of the assumptions of this approach). The resultant figures can then be presented using the effects, for example as imprints, or symbolic tracers.

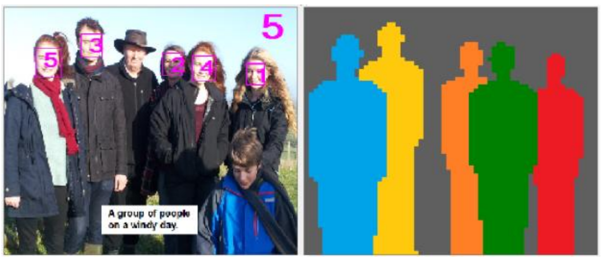


Figure 10: Human figures assumed from detected faces.

Other robust object identification methods can be used where available. Alternatively the system could make use of the large amount of Cloud-based information that is available concerning fixed landmarks etc. that may be encountered along the route of a journey, and these could be presented to the user in a similar manner.

Locked-on items

At any moment the user can lock on the item (e.g. route) being presented. When an item is locked on, and the user moves the pointer within the area of the item, typically the items at lower- (and/or higher-) levels than the locked item can also be presented, so that the user can be aware of items in adjacent levels (or items nearby on the same level), and can switch to being locked on one of them instead. Alternatively the system can step around the lower-level items within the locked-on higher-level item, and the user can at any time lock on the item being presented.

Once an item is locked on, the subsequent interaction depends to some extent on the equipment being used to access the entity:-

Force-feedback : If a force-feedback mouse or joystick Fig. 5 is being used, the system can restrict the free movement to the area(s) of the current item/route – when pushed by the user away from the item, a spring force will attempt to push the mouse or joystick handle back to the centre or nearest part of the selected item (or to the point at which they left the item). When within the area of the item, the mouse or joystick handle will be loose/“floppy” and can be moved freely. The user can feel around the edge of the item, and get audio feedback as well.

If the item is multi-blob, e.g. a group item or fragmented basic item, then the user can command a jump to the next blob, and then explore that shape and content. Alternatively, with a force-feedback device the user can simply push the handle around the image and it will tend to snap to the nearest applicable blob.

Mouse : If a standard computer mouse is being used, an audio cue can signify and warn that the user has attempted to leave the area of the item/route. However the cursor pointer can be locked at the edge of the item (e.g. via a Windows SetCursorPos action), so that the user does not need to find the item again and can simply move their mouse back in the opposite direction.

Touch : If a touch-screen, or an absolute mode touchpad, is being used, then the system cannot easily restrict the physical movement of the user’s finger, so needs to directly tell the user or give non-speech cues to indicate how to move back to the locked item/route area. However users will typically be better able to recall the approximate location of the item (e.g. route) within the physical fixed area of the touch-screen or touchpad, than when using a standard relative mode mouse.

The system could use a virtual reality 360-degree camera or similar to gather images containing the distributed items that surround the blind (or sighted) traveler's vehicle, and corresponding effects then located in 3D soundspace.

Online facilities exist to provide words summarizing the content of images, so providing a top-level summary term for visual images [29].

4. ASSESSMENTS

An informal assessment of both the new (incomplete) audio preview feature, and the application of features for automotive use, was conducted with one blind participant “AB” (not his real initials), who has been totally blind since birth, and two sighted participants (“CD” & “EF”).

AB, who is very familiar with TTS speech synthesis, felt that the monotone voice (that accurately conveys height through pitch mapping) was sometimes hard to comprehend, and suggested that it should be easy to rapidly switch to one of the standard PC voices (which are more prosodic, but of less clear pitch level). This applied to both the speech used for audio preview, and for the automotive applications.

Concerning automotive applications, AB was interested in being able to easily produce a navigable hierarchical structure from a simple list, for several possible applications. He was very interested in the route presentation facility, and thought it could have application beyond automotive use.

Regarding audio preview, AB felt that while speech gave a relatively clear description, the tone sounds gave more of an impression of the rectangle.

CD (sighted) thought that using timbre to distinguish audio preview tracer phases was effective, and thought it should also be used to distinguish the two legs presenting the "L"-shape representation of the rectangle. She preferred the clock-face approach to location coordinates.

CD felt that if too many special properties where applied to the non-speech sounds then they could sound confusing.

She was generally positive about the automotive applications, and liked the feature for handling sharp turns in locked routes (see section 3.2 above).

EF (sighted) was positive about the mapping and route presentation application.

Regarding audio preview, she initially preferred speech to tone presentation, but didn't like using the phonetic alphabet (“Alfa”, “Bravo” etc.) when hearing coordinates, preferring the chessboard-style (e.g. "B2") terminology.

Overall the participants liked using polar coordinates (clock-face directions are easy to rapidly interpret).

Another approach discussed was presenting location by playing a circular tracer from the 12 o'clock position round to the location of the rectangle, then presenting the rectangle.

Other features discussed included controlling the volume of each phase; and presenting location by successively saying two numbers in the range 1 to 9, the first number representing the position of the target location within one of 3x3 squares numbered as the keys on a typical telephone handset, and with the second number representing a smaller square within the area of the first square, in a similar manner.

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. One thought that additional tone sounds might be helpful for exact rectangle positioning.

(All of these points require further investigation.)

5. CONCLUSIONS AND FUTURE WORK

In this paper possible applications of the HFVE system to automotive vehicles have been described and considered, with particular emphasis on applications for blind travelers.

Assessment sessions with a totally blind participant, and two sighted participants, are reported above.

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 2019.

6. REFERENCES

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

[2] *Google Lookout,* https://blog.google/outreach-initiatives/accessibility/ lookout-discover-your-surroundings-help-ai/.

[3] McKinsey, “Rethinking car software and electronics architecture”, 2016. https://www.mckinsey.com/ industries/automotive-and-assembly/our-insights/ rethinking-car-software-and-electronics-architecture.

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

[5] 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.

[6] 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.

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

[8] 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.

[9] A. Edwards, “Auditory Display in Assistive Technology” in *The Sonification Handbook*, T. Hermann, A. Hunt, J.G. Neuhoff (Eds.) 2011, pp. 431-453.

[10] 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.

[11] Roth P, Richoz D, Petrucci L, Pun T, “An audio-haptic tool for non-visual image representation” in *Proceedings of the 6th International Symposium on Signal Processing and its Applications* 2001 (Cat.No.01EX467) : 64-7.

[12] Patrick Roth, Thierry Pun, “Design and Evaluation of Multimodal System for the Non-visual Exploration of Digital Pictures”. In *Proceedings of INTERACT 2003.*

[13] Parente, P. and G. Bishop, BATS: The Blind Audio Tactile Mapping System. ACMSE. Savannah, GA. March 2003.

[14] 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.

[15] 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, pp. 722-727, 2008.

[16] Derek Brock, Christina Wasylyshyn, and Brian McClimens, “Word spotting in a multichannel virtual auditory display at normal and accelerated rates of speech” in *Proc. of 22nd International Conference on Auditory Display(ICAD-2016),*Canberra,Australia, 2016.

[17] *Seeing AI,* http://microsoft.com/en-us/seeing-ai.

[18] *The HFVE system,* http://hfve.com.

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

[20] 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.

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

[22] 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.

[23] Ivica Ico Bukvic, Denis Gracanin, Francis Quek, “Investigating Artistic Potential of the Dream Interface: the Aural Painting”, in *International Computer Music Conference Proceedings,* Volume 2008, August 2008.

[24] Laura Spinney, “How long is now?” in *New Scientist*, vol. 225, no. 3003, pp. 28-31, 10th January 2015.

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

[26] *Self-driving car,* http://en.wikipedia.org/wiki/Self-driving_car.

[27] *Tesseract,* http://github.com/tesseract-ocr/tesseract/wiki.

[28] *OpenCV (Open Source Computer Vision),* http://opencv.org

[29] *IBM Watson Visual Recognition service,* http://ibm.com/watson /developercloud/doc/visual-recognition.

[30] *Dewey Decimal Classification,* http://en.wikipedia.org/ wiki/Dewey_Decimal_Classification.