A Survey on Hardware and Software Solutions for Multimodal Wearable Assistive Devices Targeting the Visually Impaired

Ádám Csapó[1], György Wersényi[2], and Myounghoon Jeon[3]

[1]Széchenyi István Univesity, Department of Informatics, Egyetem tér 1, H-9026 Győr, Hungary (e-mail: csapo.adam@sze.hu)

[2]Széchenyi István Univesity, Department of Telecommunications, Egyetem tér 1, H-9026 Győr, Hungary (e-mail: wersenyi@sze.hu)

[3]Michigan Technological University, Department of Cognitive and Learning Sciences and the Department of Computer Science, Houghton, Harold Meese Center, Houghton, MI 49931 USA (e-mail: mjeon@mtu.edu)

Abstract: The market penetration of user-centric assistive devices has rapidly increased in the past decades. Growth in computational power, accessibility, and "cognitive" device capabilities have been accompanied by significant reductions in weight, size, and price, as a result of which mobile and wearable equipment are becoming part of our everyday life. In this context, a key focus of development has been on rehabilitation engineering and on developing assistive technologies targeting people with various disabilities, including hearing loss, visual impairments and others. Applications range from simple health monitoring such as sport activity trackers, through medical applications including sensory (e.g. hearing) aids and real-time monitoring of life functions, to task-oriented tools such as navigational devices for the blind. This paper provides an overview of recent trends in software and hardware-based signal processing relevant to the development of wearable assistive solutions.

Keywords: assistive technology; blind user; haptics; spatial rendering; sonification

1 Introduction

The first (assistive) wearable devices developed in the 1990s tried to address the basic needs of specific target groups. Particular devices were primarily considered as medical devices incorporating the basic functionalities of sensors and actuators (e.g. microphones, amplifiers, vibro-/electrotactile actuators) in order to complement human sensorimotor capabilities.

In the case of visually impaired users, the ability to travel in both familiar and unfamiliar [1, 2], as well as indoor and outdoor [3, 4] environments is crucial. Such a large variety of contexts creates a myriad of challenges, both in terms of functionality and safety. Key issues include avoidance of obstacles, accessibility to salient points of the environment (e.g. finding doors, obtaining key notifications from signs and other information displays), and even co-presence (participating to the fullest possible extent in social interactions that occur within the environment).

However, given that the processing of visual information requires powerful sensory equipment as well as significant resources for processing incoming data, the earliest electronic travel aids (ETAs) were quite bulky and processed information at relatively low resolutions. Significant changes to this state of affairs only became possible in the 21^{st} Century, in parallel with exponential reduction in the size, weight, and processing capacity of both dedicated electronic aids and mobile devices such as smartphones and tablets. Even today, the widespread adoption of SSDs (sensory substitution devices) outside of academia seems to face steep challenges. As suggested by Elli et al., this may be due to the fact that effectively coordinating issues arising from the different perspectives of ergonomics, neuroscience, and social psychology is a tricky business [5].

Even with these challenges, taking into account that visually impaired users have at best very basic visual capabilities through residual vision, the role of both substitutive and augmentative feedback through the senses of audition and taction is widely seen as essential. While these senses can be used both separately and in parallel, the ways in which information is allocated to them, and the ways in which information is presented (i.e. filtered, processed, and displayed) to users cannot be arbitrary. While it is often said (but sometimes questioned) that healthy individuals process up to 90% of information of the environment through vision [6], even the information obtained through this single modality is heterogeneous. Thus, both 2D and 3D information are acquired in foveal and peripheral parts of the retina and at different degrees of attention; not all events and objects are attended to directly - even if they reach consciousness to some extent - and it is important that the structure of such details be reflected in the structure of substitutive and augmentative feedback.

At the same time, it is important to take into consideration how the modalities of audition and taction are fundamentally different from vision - at least in the ways we use them. Users are able to focus through these modalities less acutely, and the spatial resolution of incoming information is also lower than vision. With audition, perception (or rather, estimation) of distance is based partially on sound pressure level, which is very inaccurate [7, 8]. Presenting multiple (concurrent) audio sources result in increased cognitive load, reduced localization, and reduced identification accuracy. As a result, mapping 3D visual information to pure auditory cues has had limited success. The crucial realization is that while the visual modality would be capable of effectively suppressing irrelevant information, in the case of auditory modality this kind of suppression has to be carried out by design, prior to auditory mapping and rendering through sonification and directional stimulation. With haptics, many of the conclusions are similar to audition. Here, the exclusion of visual feedback also results in a relative strengthening of substance parameters (in this case, hardness and texture) and a weakening of shape salience [9]. This tendency can be countered by appropriate "exploratory procedures", in which subjects move their hands across the given surface in specific ways to obtain relevant information [10, 11]. While such procedures can potentially be richer in the case of haptic as opposed to auditory exploration, enabling users to capitalize on such procedures requires that the structure of the stimulus be assembled in appropriate ways, and whenever the stimuli are temporal, parameters such as burst frequency, number of pulses per burst, pulse repetition frequency, waveform shape and localization also need to be taken into consideration [12]. The general conclusion is that in both cases of audition and haptic/tactile feedback, results are optimal if a selective filtering and feature encoding process precedes the mapping and rendering phases.

The following sections of this paper focus on representations and signal processing techniques associated with the auditory and haptic / tactile modalities, respectively. Finally, an application-oriented section concludes the paper.

2 Audio

Sound is the most important feedback channel for the visually impaired. Aspects of sound such as reverberation (echolocation), pressure, and timbre convey useful information about the environment for localization, navigation, and general expectations about the surrounding space. In fact, in the case of healthy individuals, localization of sound sources is the most important part in identifying hazardous objects, obstacles, or even the “free path” to navigate through [13]. Both ecologically motivated and abstract (artificial) sounds can deliver crucial supplementary information on real life, and increasingly virtual reality situations.

Auditory Localization

For localizing sound sources, the auditory system uses both monaural (using one ear) and interaural (using two ears) cues [14]. Monaural cues are essentially responsible for distance perception and localization in the median plane (elevation perception). For example, distant sounds are usually softer and sounds that gradually grow louder are perceived as approaching. Interaural cues are based on the time, intensity, and phase differences between the two ears in the case of sound sources outside the median plane. Interaural Time Differences (ITD) mean that the arrival times of the sounds to the two ears differ with respect to each other. Interaural Level Differences (ILD) mean that the sound intensities perceived by the two ears differ due to the “head shadow”, i.e., due to the shape of the listener's head blocking certain high-frequency sound components. Interaural Phase Differences (IPD) are also due to the shape of the head, which alters the phases of the sounds reaching the two ears. These mechanisms are better suited towards the identification of horizontal directions. For all of these phenomena, the hearing system uses the filtering effects of the outer ears, head, and torso. These response characteristics are called Head-Related Transfer Functions (HRTFs) and they can be measured, stored, and used for reverse engineering usually in the form of digital IIR and/or FIR filters [15-18].

Actual localization performance of human subjects depends on various other factors as well, such as:

- real-life environment vs. virtual simulation,
- training and familiarity with the sounds,
- type (bandwidth, length) and number of sound sources,
- spatial resolution and overall accuracy of the applied HRTFs (if any),
- head-tracking,
- other spatial cues (reverberation, early reverb ratio, etc.).

General findings show (1) decreased localization performance in the vertical directions in contrast to horizontal plane sources, (2) preference for individually measured HRTFs during binaural playback, and (3) increased error rates when using headphones [19-24].

Methods for Directional Simulation

Several methods exist for the spatialization of sound sources. As distances are generally mapped to sound level only, the localization task using directional information is usually associated with sources that are at a constant distance. Hence, localization has to be tested both in the horizontal and vertical planes.

Sounds can be played back over loudspeakers or headphones. Loudspeaker systems incorporate at least two channels (stereo panning, panorama stereo), but otherwise can range from traditional 5.1 multi-channel systems to ones with up to hundreds of speakers. Headphones are generally two-channel playback systems using electrodynamic transducers for airborne conduction. Traditional open or closed type headphones (especially, if they are individually free-field equalized [25-28]) are the best solutions. However, if they cover the entire ears and block the outside world, blind users will refrain from using them. Therefore, loudspeaker solutions are not applicable for mobile and wearable applications. Other types of headphones exist such as multi-speakers, multi-channel (5.1) solutions, partly covering the ears, or bone conduction phones. People are likely to think that the latter technology may provide lossy information compared to everyday air conduction phones; however, research has shown that virtual threedimensional auditory displays can also be delivered through bone-conduction transducers using digital signal processing, without increased percept variability or decreased lateralization [29, 30].

Signal processing solutions for spatial simulation include the following methods:

- Amplitude panning (panorama). Simple panning between the channels will result in virtual sources at a given direction. Mono is not suitable for delivering directional information. Stereo panning is limited for correct sound source positioning between the two speakers. The classical setup is a 60-degree triangle of listener and loudspeakers. Amplitude panning can be used for two-channel headphone playback, not necessarily limited to ±30 degrees, but it can be up to ±90 degrees.

- HRTF filtering is for binaural playback over headphones. HRTFs are stored in a form of digital filters in a given number of length (taps, filter coefficients) and spatial resolution (number of filters in the horizontal and vertical plane). Real-time filtering or pre-filtered pre-recorded samples are needed, together with some kind of interpolation for missing filters (directions). Although this is a commonly applied method, localization performance is sometimes low and errors, such as front-back-confusions in-the-head localization and others influence the performance [31-33].

- Wave-Field Synthesis (WFS) incorporates a large number of individually driven speakers and it is not designed for wearable applications [34]. The computational load is very high, but the localization does not depend on or change with the listener's position.

- Ambisonics uses a full sphere of loudspeakers, not just in the horizontal plane but also in the vertical plane (above and below the listener). It is not the traditional multi-channel system and the signals are not dedicated speaker signals. They contain speaker-independent representation of a sound field that has to be decoded to the actual speaker setup. Its advantage is that the focus is on source direction instead of loudspeaker positions and it can be applied to various setups. However, using multiple speakers and high signal processing makes it unavailable in wearable devices [35].

Further Aspects of Auditory Representation

Besides directional information, a number of mapping strategies have been applied to wearable assistive devices making it possible to communicate bits of information with semantics that lie outside of the scope of auditory perception.

Traditional auditory cues

Auditory icons were defined by Gaver in the context of 'everyday listening' - meaning that one listens to the information behind the sound as a “caricature" of physical-digital phenomena [36-38]. This was the first generalization of David

Canfield-Smith’s original visual icon concept [39] in modalities other than vision. Around the same time, earcons were defined by Blattner, Sumikawa, and Greenberg as "non-verbal audio messages used in the user interfaces to provide users with information about computer objects, operation, or interaction". Today, the term is used exclusively in the context of 'abstract' (rather than 'representational', see [40]) messages, i.e., as a concept that is complementary to the iconic nature of auditory icons.

Sonification

Whenever a data-oriented perspective is preferred, as in transferring data to audio, the term ‘sonification’ is used, which refers to the “use of non-speech audio to convey information or perceptualize data” [41] (for a more recent definition, the reader is referred to [42]).

Sonification is a widely used term to cover applications of sound as information. However, spatial simulation and directional information is not generally part of sonification: such techniques are generally applied separately. From a visual perspective, sonification focuses on finding an appropriate mapping between visual events and auditory counterparts, that is, how certain visual objects or events should sound like. Several methods have been proposed for finding such mappings based on the conceptual structure of the problem domain (e.g. [43]). For example, Walker used magnitude estimation to identify optimal sonification for diverse tasks with sighted and visually impaired users [44]. Following the specification of an appropriate mapping, the parameters of the individual sound events may also be modified through time, based on real-time changes in the physical attributes of the visual objects or environment that is represented. Whenever the semantic content to be mapped (e.g., an image in the case of auditory substitution of vision) has more parameters than can easily be associated with sound attributes, the following direct kinds of parameter mappings are used most frequently:

- frequency of sound (increasing frequency usually means that an event parameter is increasing, or moving to the right / upward)

- amplitude of sound is often mapped to distance information (increasing loudness means that a parameter is increasing or approaching)

- timing in case of short sound events (decreasing the time interval between sound samples conveys an impression of increasing urgency, or a shortening of a distance)

- timbre, i.e. the "*characteristic quality of sound, independent of pitch and loudness, from which its source or manner of production can be inferred*" [45] can represent iconic features, such as color or texture of the visual counterpart

Besides such "direct" mapping techniques, various analogy-based solutions are also possible. For example, sonification can reflect the spatio-temporal context of events, sometimes in a simplified form as in 'cartoonification' [46-49]. In Model-Based Sonification [50], the data are generally used to configure a sound-capable virtual object that in turn reacts on excitatory interactions with acoustic responses whereby the user can explore the data interactively. This can be extended to interactive sonification where a higher degree of active involvement occurs when the user actively changes and adjusts parameters of the sonification module, or interacts otherwise with the sonification system [51].

Whenever such mappings are selected intuitively, empirical evaluations should be carried out with the target user group (i.e., whatever the designer considers to be an intuitive sonification does not necessarily correspond to the target group’s expectations [52]). In such cases, cognitive load due to a potentially large number of features is just as important as is the aspect of semantic recognizability.

Speech and music

Using speech and music has always been a viable option in electronic aids, and has always been treated somewhat orthogonally to the aspects of sonification described above. Speech is convenient because many steps of sonification can be skipped and all the information can be directly "mapped" to spoken words. This, however, also makes speech-based communication relatively slow, and language-dependent text-to-speech applications have to be implemented. Previously, concatenative synthesis (the recording of human voice) tended to be used for applications requiring small vocabularies of fewer than 200 words, whereas Text-to-Speech (TTS) was used for producing a much larger range of responses [53]. There were two types of TTS synthesis techniques depending on signal processing methods. The technique of diphone synthesis uses diphones (i.e., a pair of phones) extracted from the digitized recordings of a large set of standard utterances. The formant synthesis uses a mathematical model of formant frequencies to produce intelligible speech sounds. Nowadays, these techniques are integrated and used in parallel in electronic products [54].

Recently, several novel solutions have emerged, which try to make use of the advantageous properties of speech (symbolic), and iconic and indexical representations [55]. These speech-like sounds, including spearcons, spemoticons, spindexes, lyricons, etc. are a type of tweaked speech sound, which uses part of the speech or the combinations of speech and other sounds [56-61]. Spindexes are a predesigned prefix set and can be automatically added to speech items. Lyricons are a combination of melodic speech and earcons and thus, require some manual sound design. However, spearcons and spemoticons can be algorithmically made on the fly. Spearcons’ time compression is accomplished by running text-to-speech files through a SOLA (Synchronized Overlap Add Method) algorithm [62, 63], which produces the best-quality speech for a computationally efficient time domain technique. Spemoticons can be made in the interactive development environment by manipulating the intensity, duration and pitch structure of the generated speech [64].

Some of these auditory cues can represent a specific meaning of the item, but others can represent an overall structure of the system. For example, each spearcon can depict a specific item like auditory icons with a focus on “what” an item is and provide a one-on-one mapping between sound and meaning. In contrast, spindex cues can provide contextual information, such as the structure and size of the auditory menus, and the user’s location or status like earcons with a focus on “where” the user is in the system structure. Interestingly, lyricons or the combinations of melodic speech and earcons can represent both the semantics of the item (speech) and the contextual information of the system (earcons).

Music or musical concept can also be used for electronic devices. Music is pleasant for long term, can be learned relatively fast, and can be both suitable for iconic and continuous representation. Chords or chord progression can deliver different events or emotions. Different instruments or timbre can represent unique characteristics of objects or events. By mapping a musical scale to menu items, auditory scrollbars enhanced users’ estimation of menu size and their relative location in a one-dimensional menu [65]. Using a short portion of existing music, musicons have been shown successful application as a reminder for home tasks (e.g., reminder for taking pills) [66]. On the other hand, intuitive mappings between music and meaning are inherently difficult given the subjective characteristics of music. First, individual musical capabilities differ from person to person. Second, the context of the application has to be considered to alleviate the possibility of both misinterpretations (e.g., when other sound sources are present in the environment at the same time) and of "phantom" perceptions, in which the user thinks that she has perceived a signal despite the fact that there is no signal (understanding such effects is as important as guaranteeing that signals that are perceived are understood correctly).

Hybrid solutions As a synthesis of all of the above techniques, increasingly ingenious approaches have appeared which combine high-level (both iconic and abstract, using both music and speech) sounds with sonification-based techniques. Walker and his colleagues tried to come up with hybrids integrating auditory icons and earcons [67]. Speech, spearcons, and spindex cues have also been used together in a serial manner [68]. Jeon and Lee compared subsequent vs. parallel combinations of different auditory cues on smartphones to represent a couple of submenu components (e.g., combining camera shutter and game sound to represent a multimedia menu) [69]. In the same research, they also tried to integrate background music (representing menu depths) and different auditory cues (representing elements in each depth) in a single user interface. Recently Csapo, Baranyi, and their colleagues developed hybrid solutions based on earcons, auditory icons and sonification to convey tactile information as well as feedback on virtual sketching operations using sound [70, 71]. Such combinations make possible the inclusion of temporal patterns into audio feedback as well, e.g. in terms of the ordering and timing between component sounds.

3 Haptics

Haptic perception occurs when objects are explored and recognized through touching, grasping, or pushing / pulling movements. “Haptic” comes from the Greek, “haptikos”, which means “to be able to touch or grasp” [72]. Depending on the mechanoreceptor, haptic perception includes pressure, flutter, stretching, and vibration and involves the sensory and motor systems as well as high-level cognitive capabilities. The terms "haptic perception" and "tactile perception" are often used interchangeably, with the exception that tactile often refers to sensations obtained through the skin, while haptic often refers to sensations obtained through the muscles, tendons, and joints. We will use this same convention in this section.

Haptic/tactile resolution and accuracy

In a way similar to vision and hearing, haptic / tactile perception can also be characterized by measures of accuracy and spatial resolution. However, the values of such measures are different depending on various body parts and depending on the way in which stimuli are generated. From a technological standpoint, sensitivity to vibration and various grating patterns on different areas of the body influences (though it does not entirely determine) how feedback devices can be applied. Relevant studies have been carried out in a variety of contexts (see e.g., [73]). The use of vibration is often restricted to on/off and simple patterns using frequency and amplitude changes. As a result, it is generally seen as a way to add additional information along auditory feedback about e.g. warnings, importance, or displacement.

With respect to the resolution of haptic perception, just-noticeable-differences (JNDs) ranging from 5 to 10% have been reported [74]. The exact value depends on similar factors as in the case of tactile discrimination, but added factors such as the involvement of the kinesthetic sense in perception, and even the temperature of the contact object have been shown to play a significant role [75].

The overall conclusion that can be drawn from these results is that the amount of information that can be provided using tactile and haptic feedback is less than it is through the visual and auditory senses. As we will see later, information obtained through taction and force is also less conceptual in the sense that it seems to be less suitable for creating analogy-based information mapping.

Haptic/tactile representations

As in the case of audio, several basic and more complex types of tactile (and more generally, haptic) representations have been proposed as detailed in this section. One general tendency in these representations that can be contrasted with audio representation types is that the distinction between iconic and abstract representations is less clear-cut (this may be evidence for the fact that people are generally less aware of the conceptual interpretations of haptic/tactile feedback than in the case of auditory feedback).

Hapticons (haptic icons) Haptic icons were defined by MacLean and Enriquez as "brief computer-generated signals, displayed to a user through force or tactile feedback to convey information such as event notification, identity, content, or state" [76]. In a different paper, the same authors write that "Haptic icons, or hapticons, [are] brief programmed forces applied to a user through a haptic interface, with the role of communicating a simple idea in a manner similar to visual or auditory icons" [77].

These two definitions imply that the term haptic icon and hapticon can be used interchangeably. Further, the discussions of MacLean and Enriquez refer both to 'representational' and 'abstract' phenomena (while the definitions themselves reflect a representational point of view, the authors also state that “*our approach shares more philosophically with [earcons], but we also have a long-term aim of adding the intuitive benefits of Gaver’s approach…”[*76]). All of this suggests that the dimensions of representation and meaning are seen as less independent in the case of the haptic modality than in vision or audio. Stated differently, whenever an interface designer decides to employ hapticons/haptic icons, it becomes clear that the feedback signals will provide information primarily 'about' either haptic perception itself, or indirectly about the occurrence of an event that has been linked - through training - to the 'iconic' occurrence of those signals. However, the lack of distinction between haptic icons and hapticons also entails that designers of haptic interfaces have not discussed the need (or do not see a possibility, due perhaps to the limitations of the haptic modality) to create higher-level analogies between the form of a haptic signal and a concept from a different domain.

Tactons (tactile icons) Brewster and Brown define tactons and tactile icons as interchangeable terms, stating that both are “structured, abstract messages that can be used to communicate messages non-visually” [78]. This definition, together with the conceptual link between the 'representational' and 'abstract' creates a strong analogy between tactons and hapticons. In fact, tactons may be seen as special kinds of hapticons which make use of Blattner, Sumikawa and Greenberg's original distinction between signals that carry a certain meaning within their representation, and those which do not (which can be generalized to other modalities) is nowhere visible in these definitions. Once again, this may be due to limitations in the 'conceptual expressiveness' of the haptic/tactile modality. Subsequent work by Brewster, Brown and others has suggested that such limitations may be overcome. For example, it was suggested that by designing patterns of abstract vibrations, personalized cellphone vibrations can be used to provide information on the identity of a caller [79]. If this is the case, one might wonder whether a separation of the terms haptic icon and hapticon, as well as those of tactile icon and tacton would be useful.

4 Applications

When it comes to developing portable and wearable assistive solutions for the visually impaired, power consumption is generally a crucial issue. If several computationally demanding applications are required to run at the same time, the battery life of even the highest quality mobile devices can be reduced to a few hours. Possible solutions include reducing the workload of the application (in terms of amount of information processed, or precision of processing); or using dedicated hardware alongside multi-purpose mobile devices, such as smartphones or tablets. In this section, we provide a broad overview of the past, present, and (potential) future of devices, technologies, and algorithmic solutions used to tackle such challenges.

Devices

State-of-the-art mobile devices offer built-in sensors and enormous computational capacity for application development on the Android and iOS platform [80, 81]. Without dedicated hardware, software-only applications provide information during navigation using the GPS, compass, on-line services and others [82-84]. However, microcontrollers and system-on-a-chip (SoC) solutions – e.g. Arduino or Raspberry Pi - are also gaining popularity, as the costs associated with such systems decrease and as users are able to access increasingly sophisticated services through them (e.g., in terms of accessing both versatile computing platforms, such as Wolfram Mathematica, and dedicated platforms for multimodal solutions, such as the Supercollider, Max/MSP [85] and PureData [86]).

Most of the above devices afford generic methods for the use of complementary peripheral devices for more direct contact with end users. In the blind community, both audio and haptic/tactile peripherals can be useful. In the auditory domain, devices such as microphones, filters / amplifiers and headphones with active noise cancelling (ANC) for the combination and enhancement of sound are of particular interest. In the haptic and tactile domains, devices such as white canes, "haptic" bracelets and other wearables with built-in vibrotactile or eletrotactile displays are often considered. While the inclusion of such peripherals in solutions increases associated costs (in terms of both development time and sales price), it also enables the implementation of improved functionality in terms of number of channels, "stronger" transducers, and the ability to combine sensory channels both from the real world and virtual environments into a single embedded reality.

Assistive Technologies for the Blind

In this section, we provide a brief summary of systems which have been and are still often used for assistive feedback through the auditory and haptic senses.

Assistive technologies using audio

Auditory feedback has increasingly been used in assistive technologies oriented towards the visually impaired. It has been remarked that both the temporal and frequency-based resolution of the auditory sensory system is higher than the resolution of somatosensory receptors along the skin. For several decades, however, this potential advantage of audition over touch was difficult to be taken due to the limitations in processing power [80].

Systems that have been particularly successful include:

- SonicGuide, which uses a wearable ultrasonic echolocation system to provide users with cues on the azimuth and distance of obstacles [87, 88].

- LaserCane, which involves the use of a walking cane and infrared instead of ultrasound signals [87, 89].

- The Nottingham Obstacle Detector, which is a handheld device that provides 8 gradations of distance through a musical scale based on ultrasonic echolocation [90].

- The Real-Time Assistance Prototype (RTAP), which is a camera-based system, equipped with headphones and a portable computer for improved processing power that even performs object categorization and importance-based filtering [91].

- The vOICe [92] and PSVA [87] systems, which provide direct, retinotopic temporal-spectral mappings between reduced-resolution camera-based images and audio signals.

- System for Wearable Audio Navigation (SWAN), which is developed for safe pedestrian navigation, and uses a combination of continuous (abstract) and event-based (conceptual) sounds to provide feedback on geometric features of the street, obstacles, and landmarks [93, 94].

Text-to-speech applications, speech-based command interfaces, and navigational helps are the most popular applications on mobile platforms. Talking Location, Guard my Angel, Intersection Explorer, Straight-line Walking apps, The vOICe, Ariadne GPS, GPS Lookaround, BlindSquare etc. offer several solutions with or without GPS for save guidance. See [80] for a detailed comparison and evaluation of such applications.

Assistive technologies using haptics

Historically speaking, solutions supporting vision using the tactile modality appeared earlier than audio-based solutions. These solutions generally translate camera images into electrical and/or vibrotactile stimuli.

Systems that have been particularly successful include:

- The Optacon device, which transcodes printed letters onto an array of vibrotactile actuators in a 24x6 arrangement [95-97]. While the Optacon was relatively expensive at a price of about 1500 GBP in the 1970s, it allowed for reading speeds of 15-40 words per minute [98] (others have reported an average of about 28 wpm [99], whereas the variability of user success is illustrated by the fact that two users were observed with Optacon reading speeds of over 80 wpm [80]).

- The Mowat sensor (from Wormald International Sensory Aids), which is a hand-held device that uses ultra-sonic detection of obstacles and provides feedback in the form of tactile vibrations inversely proportional to distance.

- Videotact, created by ForeThought Development LLC, which provides navigation cues through 768 titanium electrodes placed on the abdomen [100].

- A recent example of a solution which aims to make use of developments in mobile processing power is a product of a company, “Artificial Vision For the Blind”, which incorporates a pair of glasses from which haptic feedback is transmitted to the palm [101, 102].

Today, assistive solutions making use of generic mobile technologies are increasingly prevalent. Further details on this subject can be found in [80].

Hybrid solutions

Solutions combining the auditory and haptic/tactile modalities are still relatively rare. However, several recent developments are summarized in [80]. Examples include the HiFiVE [103, 104] and SeeColOR [104] systems, which represent a wide range of low to high-level visual features through both audio and tactile representations. With respect to these examples, two observations are made: first, audio and taction are generally treated as a separate primary (i.e., more prominent with a holistic spatio-temporal scope) and secondary (i.e., less prominent in its spatio-temporal scope) modality, respectively; and second, various conceptual levels of are reflected in signals presented to these modalities at the same time.

Algorithmic challenges and solutions

In general, designing multimodal applications shows tradeoffs between storing stimuli beforehand and generating them on the fly. While the latter solution is more flexible in terms of real-time parametrization, it requires more processing.

Signal generation

An important question on any platform is how to generate the desired stimuli. In this section, we summarize key approaches in auditory and haptic/tactile domains.

Auditory signals With the advance of technology, multiple levels of auditory signals can be generated from electronic devices, including assistive technologies. However, auditory signals can be largely classified into two types. First, we can generate auditory signals using a buzzer. The buzzer is a self-sufficient sound-generation device. It is cheap and small. It does not require any additional sound equipment or hardware. The different patterns of auditory signals can be programmed using even lower level programming languages (e.g., C or Assembly), varying the basic low-level signal processing parameters of *attack*, decay, *sustain* and *release*. With these, the parameters of *sound frequency, melodic pattern (including polarity)*, *number of sounds*, *total sound length*, *tempo of the sound*, and *repetitions within patterns* can be adjusted [105, 106]:

Nowadays, high quality auditory signals can also be generated using most mobile and wearable devices, including compressed formats like MP3 or MPEG-4 AAC. In this case, additional hardware is required, including an amplifier and a speaker system. Of course, the above variables can be manipulated. Moreover, timbre, which is a critical factor in mapping data to sounds [107] or musical instruments can represent particular functions or events on the device. More musical variables can also be controlled, such as chord, chord progression, key change, etc. in this high level configuration. Auditory signals can be generated as a pre-recorded sound file or in real-time through the software oscillator or MIDI (musical instrument digital interface). Currently, all these sound formats are supported by graphic programming environments (e.g., Max/MSP or PureData) or traditional programming languages via a sound specific library (e.g. JFugue [108] in Java).

Tactile signals Mobile phones and tablets have integrated vibrotactile motors for haptic feedback. Usually, there is only one small vibrator installed in the device that can be accessed by the applications. The parameters of *vibration length*, *durations of patterns* and *repetitions within patterns* can be set through high-level object-oriented abstractions. Many vibration motor configurations are by design not suited to the real-time modification of vibration amplitude or frequency. Therefore, while such solutions offer easy accessibility and programmability, they do so with important restrictions and only for a limited number of channels (usually one channel). As a result, very few assistive applications for blind users make use of vibration for purposes other than explorative functions (such as 'zooming in' or 'panning') or alerting users, and even less use it to convey augmentative information feedback.

From the signal processing point of view, smartphones and tablets run pre-emptive operating systems, meaning that more than a single application can run "simultaneously", and even processes that would be critical for the assistive application can be interrupted. Although accessing a built-in vibration motor is not a critical application in and of itself, a comprehensive application incorporating other features (such as audio and/or visual rendering) together with haptics can be problematic. For example, spatial signals can be easily designed in a congruent way between audio and haptic (e.g., left auditory/haptic feedback vs. right auditory/haptic feedback). When it comes to frequency, the frequency range does not match with each other in one-on-one mapping. This is why we need to empirically assess the combination of multimodal feedback in a single device.

Latency and memory usage

Latency is a relatively short delay, usually measured in milliseconds, between the time when an audio signal enters and when it outputs from a system. This includes hardware processing times and any additional signal processing tasks (filtering, calculations etc.). The most important contributors to latency are DSP, ADC/DAC, buffering, and in some cases, travelling time of sound in the air. For audio circuits and processing pipelines, a latency of 10 milliseconds or less is sufficient for real-time experience [109], based on the following guidelines [110]:

Less than 10 ms - allows real-time monitoring of incoming tracks including effects.

At 10 ms - latency can be detected but can still sound natural and is usable for monitoring.

11-20 ms - monitoring starts to become unusable, smearing of the actual sound source, and the monitored output is apparent.

20-30 ms - delayed sound starts to sound like an actual delay rather than a component of the original signal.

In a virtual acoustic environment, the total system latency (TSL) refers to the time elapsed from the transduction of an event or action, such as movement of the head, until the consequences of that action cause the equivalent change in the virtual sound source location [111]. Problems become more significant if signal processing includes directional sound encoding and rendering, synthesizing and dynamic controlling of reverberation, room and distance effects. Several software applications have been recently developed to address this problem for various platforms such as Spat [112], Sound Lab (SLAB) [113], DirAC [114], etc.

It has been noted that real-time audio is a challenging task in VM-based garbage collected programming languages such as Java [115]. It was demonstrated that a low latency of a few milliseconds can nevertheless be obtained, even if such performance is highly dependent on the hardware-software context (i.e., drivers and OS) and cannot always be guaranteed [115]. Since version 4.1 ("Jelly Bean"), Android has included support for audio devices with low-latency playback through a new software mixer and other API improvements. Thus, improved functionality in terms of latency targets below 10 ms, as well as others such as multichannel audio via HDMI is gaining prevalence. The use of specialized audio synthesis software such as Csound [116], Supercollider, Max/MSP, and PureData can also help achieve low latency.

Training challenges

An important challenge when deploying assistive applications lies in how to train prospective users in applying them to their own real-life settings. In such cases, serious gaming is one of the options used for training and maintaining the interest of users. The term refers to a challenging and motivating environment for training in which users adapt to the system in an entertaining, and as a result, almost effortless way. Examples might include giving users the task of finding collectibles, earning rankings, or challenging other competitors by improving in the task. Especially blind users welcome audio-only games as a source of entertainment [117, 118]. As far as mobile applications are concerned, it can be stated as a general rule that regardless of subject matter, serious games usually do not challenge memory and processing units further, as they use limited visual information (if at all), and also often rely on reduced resolution in audio rendering.

5 Summary

This paper provided an overview of the state-of-the-art that is relevant to the key design phases behind portable and wearable assistive technologies for the visually impaired. Specifically, we focused on aspects of stimulus synthesis, semantically informed mapping between data / information and auditory / tactile feedback, as well as signal processing techniques that are useful either for curbing computational demands, or for manipulating the information properties of the feedback signals. Through a broad survey of existing applications, the paper demonstrated that audio-tactile feedback is increasingly relevant to transforming the daily lives of users with visual impairments.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643636 "Sound of Vision".

References

[1] F. Hosseini, S. M. Riener, A. Bose and M. Jeon, "Listen2dRoom": Helping Visually Impaired People Navigate Indoor Environments using an Ultrasonic Sensor-based Orientation Aid, in *Proc. of the 20th International Conference on Auditory Display (ICAD2014)*, NY, 2014, 6 pages

[2] M. Jeon, A. Nazneen, O. Akanser, A. Ayala-Acevedo and B. N. Walker, "Listen2dRoom": Helping Blind Individuals Understand Room Layouts. In *Extended Abstracts Proc. of the SIGCHI Conference on Human Factors in Computing Syst. (CHI'12),* Austin, TX pp. 1577-1582, 2012

[3] A. M. Ali and M. J. Nordin, "Indoor Navigation to Support the Blind Person using Weighted Topological Map," *in Proc. of the Int'l Conf. on Elect. Eng. and Inf.*, Malaysia, pp. 68-72, 2009

[4] B. N. Walker and J. Lindsay, "Development and Evaluation of a System for Wearable Audio Navigation," *in Proc. of Hum. Fact. and Erg. Society*

Annual Meeting Proceedings 09/2005, Vol. 49, No. 17, Orlando, USA, pp. 1607-1610, 2005

[5] G. V. Elli, S. Benetti and O. Collignon, "Is There a Future for Sensory Substitution Outside Academic Laboratories?," *Multisensory Research*, Vol. 27, No. 5/6, pp. 271-291, 2014

[6] M. Sivak, "The Information that Drivers Use: Is It Indeed 90% Visual?," *Perception*, Vol. 25, pp. 1081-1089, 1996

[7] S. O. Nielsen, "Auditory Distance Perception in Different Rooms," *Auditory Engineering Society 92nd Convention*, Vienna, Austria, paper nr. 3307, 1992

[8] P. Zahorik and F. L. Wightman, "Loudness Constancy with Varying Sound Source Distance," *Nature Neuroscience*, Vol. 4, pp. 78-83, 2001

[9] R. Klatzky, S. Lederman and C. Reed C, "There's More to Touch than Meets the Eye: The Salience of Object Attributes for Haptics With and Without Vision," *J. of Exp. Psychology*, Vol. 116, No. 4, pp. 356-369, 1987

[10] R. Klatzky and S. Lederman, "Intelligent Exploration by the Human Hand," *Dextrous Robot Hands*, pp. 66-81, 1990

[11] S. Lederman and R. Klatzky, "Hand Movements: A Window into Haptic Object Recognition," *Cognitive Psychology*, Vol. 19, pp. 342-368, 1987

[12] K. A. Kaczmarek and S. J. Haase, "Pattern Identification and Perceived Stimulus Quality as a Function of Stimulation Current on a Fingertip-scanned Electrotactile Display," *IEEE Trans Neural Syst Rehabil Eng*, Vol. 11, pp. 9-16, 2003

[13] D. Dakopoulos and N. G. Bourbakis, "Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey," *IEEE Trans. on Syst. Man and Cybernetics Part C*, 40(1), pp. 25-35, 2010

[14] J. Blauert, *Spatial Hearing*, The MIT Press, MA, 1983

[15] C. I. Cheng and G. H. Wakefield, "Introduction to HRTFs: Representations of HRTFs in Time, Frequency, and Space," *J. Audio Eng. Soc.*, Vol. 49, pp. 231-249, 2001

[16] H. Möller, M. F. Sorensen, D. Hammershoi and C. B. Jensen, "Head-Related Transfer Functions of Human Subjects," *J. Audio Eng. Soc.*, Vol. 43, pp. 300-321, 1995

[17] D. Hammershoi and H. Möller, "Binaural Technique – Basic Methods for Recording, Synthesis, and Reproduction," in J. Blauert (Editor) *Comm. Acoustics*, pp. 223-254, Springer, 2005

[18] F. Wightman and D. Kistler, "Measurement and Validation of Human HRTFs for Use in Hearing Research," *Acta acustica united with Acustica*, Vol. 91, pp. 429-439, 2005

[19] M. Vorländer, *Auralization: Fundamentals of Acoustics, Modeling, Simulation, Algorithms, and Acoustic Virtual Reality*, Springer, Berlin, 2008

[20] A. Dobrucki, P. Plaskota, P. Pruchnicki, M. Pec, M. Bujacz and P. Strumiłło, "Measurement System for Personalized Head-Related Transfer Functions and Its Verification by Virtual Source Localization Trials with Visually Impaired and Sighted Individuals," *J. Audio Eng. Soc.*, Vol. 58, pp. 724-738, 2010

[21] Gy. Wersényi, "Virtual localization by blind persons," *Journal of the Audio Engineering Society*, Vol. 60, No. 7/8, pp. 579-586, 2012

[22] Gy. Wersényi, "Localization in a HRTF-based Virtual Audio Synthesis using additional High-pass and Low-pass Filtering of Sound Sources," *Journal of the Acoust. Science and Technology Japan*, Vol. 28, pp. 244-250, 2007

[23] P. Minnaar, S. K. Olesen, F. Christensen and H. Möller, "Localization with Binaural Recordings from Artificial and Human Heads," *J. Audio Eng. Soc.*, Vol. 49, pp. 323-336, 2001

[24] D. R. Begault, E. Wenzel and M. Anderson, "Direct Comparison of the Impact of Head Tracking Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source," *J. Audio Eng. Soc.*, Vol. 49, pp. 904-917, 2001

[25] H. Möller, "Fundamentals of Binaural Technology," *Appl. Acoustics*, Vol. 36, pp. 171-218, 1992

[26] P. A. Hill, P. A. Nelson and O. Kirkeby, "Resolution of Front-Back Confusion in Virtual Acoustic Imaging Systems," *J. Acoust. Soc. Am.*, vol. 108, pp. 2901-2910, 2000

[27] H. Möller, D. Hammershoi, C. B. Jensen and M. F. Sorensen, "Transfer Characteristics of Headphones Measured on Human Ears," *J. Audio Eng. Soc.*, Vol. 43, pp. 203-216, 1995

[28] R. Algazi and R. O. Duda, "Headphone-based Spatial Sound," *IEEE Signal Processing Magazine*, Vol. 28, No. 1, pp. 33-42, 2011

[29] R. M. Stanley, "Toward Adapting Spatial Audio Displays for Use with Bone Conduction". M.S. Thesis. Georgia Institute of Technology. Atlanta, GA. 2006

[30] R. M. Stanley, "Measurement and Validation of Bone-Conduction Adjustment Functions in Virtual 3D Audio Displays," Ph.D. Dissertation. Georgia Institute of Technology. Atlanta, GA. 2009

[31] O. Balan, A. Moldoveanu, F. Moldoveanu, "A Systematic Review of the Methods and Experiments Aimed to Reduce Front-Back Confusions in the

Free-Field and Virtual Auditory Environments," submitted to *International Journal of Acoustics and Vibration*

[32] E. M. Wenzel, "Localization in Virtual Acoustic Displays," *Presence*, Vol. 1, pp. 80-107, 1992

[33] P. A. Hill, P. A. Nelson and O. Kirkeby, "Resolution of Front-Back Confusion in Virtual Acoustic Imaging Systems," *J. Acoust. Soc. Am.*, Vol. 108, pp. 2901-2910, 2000

[34] M. M. Boone, E. N. G. Verheijen and P. F. van Tol, "Spatial Sound-Field Reproduction by Wave-Field Synthesis," *J. Audio Eng. Soc.*, Vol. 43, No. 12, pp. 1003-1012, 1995

[35] J. Daniel, J. B. Rault and J. D. Polack, "Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions," *in Proc. of 105th AES Convention*, paper 4795, 1998

[36] W. Gaver, "Auditory Icons: Using Sound in Computer Interfaces," *Human Comput Interact,* Vol. 2, No. 2, pp. 167-177, 1986

[37] W. Gaver, "Everyday Listening and Auditory Icons," Ph.D. dissertation, University of California, San Diego, 1998

[38] W. Gaver, "The SonicFinder: an Interface that Uses Auditory Icons," *Human Comput Interact,* Vol. 4, No. 1, pp. 67-94, 1989

[39] D. Smith, "Pygmalion: a Computer Program to Model and Stimulate Creative Thought," Ph.D. dissertation, Stanford University, Dept. of Computer Science, 1975

[40] M. Blattner, D.Sumikawa and R. Greenberg, "Earcons and Icons: Their Structure and Common Design Principles," *Human Comput Interact,* Vol. 4, No. 1, pp. 11-44, 1989

[41] G. Kramer (ed), *Auditory Display: Sonification, Audification, and Auditory Interfaces*. Santa Fe Institute Studies in the Sciences of Complexity. Proceedings volume XVIII. Addison-Wesley, 1994

[42] T. Hermann, A. Hunt and J. G. Neuhoff, *The Sonification Handbook,* Logos, Berlin, 2011

[43] Á. Csapó and Gy. Wersényi, "Overview of Auditory Representations in Human-Machine Interfaces," *Journal ACM Computing Surveys (CSUR),* Vol. 46, No. 2, Art.nr. 19., 19 pages, 2013

[44] B. N. Walker, "Magnitude Estimation of Conceptual Data Dimensions for Use in Sonification," *Journal of Experimental Psychology*: *Applied,* Vol. 8, pp. 211-221, 2002

[45] See 72

[46] P. Lennox, J. M. Vaughan and T. Myatt, "3D Audio as an Information-Environment: Manipulating Perceptual Significance for Differentiation and Pre-Selection," *in Proc. of the 7th International Conference on Auditory Display (ICAD2001),* Espoo, Finland, 2001

[47] P. Lennox, and T. Myatt, "Perceptual Cartoonification in Multi-Spatial Sound Systems," *in Proc. of the 17th Int. Conference on Auditory Display (ICAD2011*), Budapest, Hungary, 2011

[48] See 71

[49] See 70

[50] T. Hermann, Sonification for Exploratory Data Analysis. Ph.D. dissertation. Bielefeld University. Bielefeld, Germany. 2002

[51] T. Hermann, "Taxonomy and Definitions for Sonification and Auditory Display," *Proc of the 14th International Conference on Auditory Display,* Paris, France, 2008

[52] G. Kramer, B. N. Walker, T. Bonebright, P. Cook, J. H. Flowers, N. Miner and J. Neuhoff, "Sonification Report: Status of the Field and Research Agenda," Faculty Publications, Department of Psychology. Paper 444. http://digitalcommons.unl.edu/psychfacpub/444, 2010

[53] M. Jeon, S. Gupta and B. N. Walker, "Advanced Auditory Menus II: Speech Application for Auditory Interface," Georgia Inst. of Technology Sonification Lab Technical Report. February, 2009

[54] M. Jeon, "Two or Three Things You Need to Know about AUI Design or Designers," *in Proc. of the 16th International Conference on Auditory Display (ICAD2010),* Washington, D.C., USA, 2010

[55] M. Jeon, "Exploration of Semiotics of New Auditory Displays: A Comparative Analysis with Visual Displays," *in Proc. of the 21st Int. Conference on Auditory Display (ICAD2015*), Graz, 2015

[56] Gy. Wersényi, "Auditory Representations of a Graphical User Interface for a Better Human-Computer Interaction," in S. Ystad et al. (Eds.): Auditory Display. CMMR/ICAD 2009 post proceedings edition, LNCS 5954, Springer, Berlin, pp. 80-102, 2010

[57] B. Gygi and V. Shafiro, "From Signal to Substance and Back: Insights from Environmental Sound Research to Auditory Display Design," *in Proc. of the International Conference on Auditory Display ICAD 09,* pp. 240-251, 2009

[58] B. N. Walker and A. Kogan, "Spearcon Performance and Preference for Auditory Menus on a Mobile Phone," *Universal Access in Human-Computer Interaction - Intelligent and Ubiquitous Interaction Environments,* Lecture Notes in Computer Science Volume 5615, pp. 445-454, 2009

[59] See 64

[60] M. Jeon and B. N. Walker, "Spindex (Speech Index) Improves Auditory Menu Acceptance and Navigation Performance," *Journal ACM Transactions on Accessible Computing (TACCESS)*, Vol. 3, No. 3, article no. 10, 2011

[61] M. Jeon, and Y. Sun, "Design and Evaluation of Lyricons (Lyrics + Earcons) for Semantic and Aesthetic Improvements of Auditory Cues," *in Proc. of the 20th International Conference on Auditory Display (ICAD2014),* 2014

[62] D. J. Hejna, "Real-Time Time-Scale Modification of Speech via the Synchronized Overlap-Add Algorithm," M.S. Thesis, Dept. of Electrical Engineering and Computer Science, MIT, 1990

[63] S. Roucos, and A. M. Wilgus, "High Quality Time-Scale Modification for Speech," *in Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Proc.*, pp. 493-496, New York, NY, USA. 1985

[64] G. Németh, G. Olaszy and T. G. Csapó, "Spemoticons: Text to Speech Based Emotional Auditory Cues," *in Proc. of ICAD2011*, Budapest, Hungary, 2011

[65] P. Yalla, and B. N. Walker, "Advanced Auditory Menus: Design and Evaluation of Auditory Scrollbars," *in Proc. of ACM Conference on Assistive Technologies (ASSETS'08*), pp. 105-112, 2008

[66] M. McGee-Lennon, M. K. Wolters, R. McLachlan, S. Brewster and C. Hall, "Name that Tune: Musicons as Reminders in the Home," *in Proc. of the SIGCHI Conference on Human Factors in Computing System*, Vancouver, BC, Canada, pp. 2803-2806, 2011

[67] B. N. Walker, J. Lindsay, A. Nance, Y. Nakano, D. K. Palladino, T. Dingler and M. Jeon, "Spearcons (Speech-based earcons) improve navigation performance in advanced auditory menus," *Human Factors*, Vol. 55, No. 1, pp. 157-182. 2013

[68] M. Jeon, T. M. Gable, B. K. Davison, M. Nees, J. Wilson and B. N. Walker, "Menu Navigation with In-Vehicle Technologies: Auditory Menu Cues Improve Dual Task Performance, Preference, and Workload," *Int. J. of Human-Computer Interaction*, Vol. 31, No. 1, pp. 1-16. 2015

[69] M. Jeon and J.-H. Lee, "The Ecological AUI (Auditory User Interface) Design and Evaluation of User Acceptance for Various Tasks on Smartphones," in M. Kurosu (Ed.), Human-Computer Interaction: Interaction Modalities and Techniques (Part IV), HCII2013, LNCS, Vol. 8007, pp. 49-58, Heidelberg, Springer. 2013

[70] Á. Csapó and P. Baranyi, "CogInfoCom Channels and Related Definitions Revisited," *in Proc. Intelligent Systems and Informatics (SISY), 2012 IEEE 10th Jubilee International Symposium*, Subotica, Serbia, pp. 73-78, 2012

[71] Á. Csapó, J. H. Israel and O. Belaifa, "Oversketching and Associated Audio-based Feedback Channels for a Virtual Sketching Application," *in Proc. 4th IEEE International Conference on Cognitive Infocommunications*, Budapest, Hungary, pp. 509-513, 2013

[72] S. Yantis, *Sensation and Perception*, NY: Worth Publishers, 2014

[73] R. van Bowen, K. O. Johnson, "The Limit of Tactile Spatial Resolution in Humans," *Neurology*, Vol. 44, No. 12, pp. 2361-2366, 1994

[74] S. Allin, Y. Matsuoka and R. Klatzky, "Measuring Just Noticeable Differences for Haptic Force Feedback: Implications for Rehabilitation," in *Proc. 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems,* Orlando, USA, pp. 299-302, 2002

[75] M. A. Srinivasan and J. S. Chen, "Human Performance in Controlling Normal Forces of Contact with Rigid Objects," *Adv. in Robotics, Mechat. and Haptic Interfaces*, pp. 119-125, 1993

[76] K. Maclean and M. Enriquez, "Perceptual Design of Haptic Icons," in *Proc. of Eurohaptics*, pp. 351-363, 2003

[77] M. Enriquez and K. MacLean, "The Hapticon Editor: a Tool in Support of Haptic Communication Research," *in Proc. of the 11th symposium on haptic interfaces for virtual environment and teleoperator systems (HAPTICS'03),* Los Angeles, USA, pp. 356-362, 2003

[78] S. Brewster and L. Brown, "Tactons: Structured Tactile Messages for Non-Visual Information Display," *in Proc. of the 5th Conf. on Australasian User Int. (AUIC'04*), Vol. 28, pp. 15-23, 2004

[79] L. M. Brown and T. Kaaresoja, "Feel Who's Talking: Using Tactons for Mobile Phone Alerts," pp. 1-6, CHI 2006

[80] Á. Csapó, Gy. Wersényi, H. Nagy and T. Stockman, "A Survey of Assistive Technologies and Applications for Blind Users on Mobile Platforms: a Review and Foundation for Research," *J Multimodal User Interfaces*, Vol. 9, No. 3, 11 pages, 2015

[81] Á. Csapó, Gy. Wersényi and H. Nagy, "Evaluation of Reaction Times to Sound Stimuli on Mobile Devices," in *Proc. of ICAD15*, Graz, Austria, pp. 268-272, 2015

[82] S. A. Panëels, D. Varenne, J. R. Blum and J. R. Cooperstock, "The Walking Straight Mobile Application: Helping the Visually Impaired avid Veering," i*n Proc. of ICAD13*, Lódz, Poland, pp. 25-32, 2013

[83] Gy. Wersényi, "Evaluation of a Navigational Application Using Auditory Feedback to Avoid Veering for Blind Users on Android Platform," *Acoust. Soc. Am.*, Vol. 137, pp. 2206, 2015

[84] L. Dunai, G. P. Fajarnes, V. S. Praderas, B. D. Garcia and I. L. Lengua, "Real-Time Assistance Prototype—A New Navigation Aid for Blind People," *in Proc. IECON 2010-36th Annual Conference on IEEE Industrial Electronics Society*, pp. 1173-1178, 2010

[85] https://cycling74.com/

[86] https://puredata.info/

[87] C. Capelle, C. Trullemans, P. Arno and C. Veraart, "A Real-Time Experimental Prototype for Enhancement of Vision Rehabilitation Using Auditory Substitution," *IEEE Transactions on Biomedical Engineering*, Vol. 45, No. 10, pp. 1279-1293, 1998

[88] L. Kay, "A Sonar Aid to Enhance Spatial Perception of the Blind: Engineering Design and Evaluation," *IEEE Radio and Electronic Engineer*, Vol. 44, No. 11, pp. 605-627, 1974

[89] E. F. Murphy, "The VA – Bionic Laser Can for the Blind. In The National Research Council,", Evaluation of Sensory Aids for the Visually Handicapped, NAS, pp. 73-82, 1971

[90] D. Bissitt and A. D. Heyes, "An Application of Bio-Feedback in the Rehabilitation of the Blind," *Applied Ergonomics*, Vol. 11, No. 1, pp. 31-33 1980

[91] L. Dunai, G. P. Fajarnes, V. S. Praderas, B. D. Garcia and I. L. Lengua, "Real-Time Assistance Prototype—A New Navigation Aid for Blind People," *in Proc. IECON 2010-36th Annual Conference on IEEE Industrial Electronics Society*, pp. 1173-1178, 2010

[92] P. Meijer, "An Experimental System for Auditory Image Representations," *IEEE Transactions on Biomedical Engineering*, Vol. 39, No. 2, pp. 112-121, 1992

[93] B. N. Walker and J. Lindsay, "Navigation Performance with a Virtual Auditory Display: Effects of Beacon Sound, Capture Radius, and Practice," *Hum Factors*, Vol. 48, No. 2, pp. 265-278, 2006

[94] J. Wilson, B. N. Walker, J. Lindsay, C. Cambias and F. Dellaert, "SWAN: System for Wearable Audio Navigation," *in Proc. of the 11th International Symposium on Wearable Computers (ISWC 2007)*, USA, 8 pages, 2007

[95] V. Levesque, J. Pasquero and V. Hayward, "Braille Display by Lateral Skin Deformation with the STReSS Tactile Transducer," *in Proc. of the 2nd joint eurohaptics conference and symposium on haptic interfaces for virtual environment and teleoperator systems*, Tsukuba, pp. 115-120, 2007

[96] P. Bach-y Rita, M. E. Tyler and K. A. Kaczmarek, “Seeing with the Brain,” *Int J Hum Comput Interact*, Vol. 15, No. 2, pp. 285-295, 2003

[97] J. C. Bliss, M. H. Katcher, C. H. Rogers, R. P. Shepard,” Optical-to-Tactile Image Conversion for the Blind,” *IEEE Trans Man Mach Syst* Vol. 11, pp. 58-65, 1970

[98] J. C. Craig, “Vibrotactile Pattern Perception: Extraordinary Observers,” *Science*, Vol. 196, No. 4288, pp. 450-452, 1977

[99] D. Hislop, B. L. Zuber and J. L. Trimble, “Characteristics of Reading Rate and Manual Scanning Patterns of Blind Optacon Readers,” *Hum Factors* Vol. 25, No. 3, pp. 379-389, 1983

[100] http://www.4thtdev.com. Accessed Mar 2015

[101] http://unreasonableatsea.com/artificial-vision-for-the-blind/. Accessed Mar 2015

[102] http://www.dgcs.unam.mx/ProyectoUNAM/imagenes/080214.pdf. Accessed Mar 2015

[103] D. Dewhurst, “Accessing Audiotactile Images with HFVE Silooet,” *in Proc. Fourth Int. Workshop on Haptic and Audio Interaction Design*, Springer-Verlag, pp. 61-70, 2009

[104] D. Dewhurst, “Creating and Accessing Audio-Tactile Images with “HFVE” Vision Substitution Software,” *in Proc. of 3rd Interactive Sonification Worksh*, Stockholm, pp. 101-104, 2010

[105] M. Jeon, “Two or Three Things You Need to Know about AUI Design or Designers,” *in Proc. of the 16th International Conference on Auditory Display (ICAD2010*), Washington, D.C., USA, 2010

[106] M. Jeon, “Auditory User Interface Design: Practical Evaluation Methods and Design Process Case Studies,” *The International Journal of Design in Society*, Vol. 8, No. 2, pp. 1-16, 2015

[107] B. N. Walker and G. Kramer, “Ecological Psychoacoustics and Auditory Displays: Hearing, Grouping, and Meaning Making,” Ed. by J. Neuhoff, Ecol. Psychoacoustics. NY, Academic Press. 2004

[108] http://www.jfugue.org/

[109] S. M. Kuo, B. H. Lee and W. Tian, *Real-Time Digital Signal Processing,* Wiley, 2013

[110] Adobe Audition User’s Manual

[111] E. M. Wenzel, “Effect of Increasing System Latency on Localization of Virtual Sounds,” *in Proc. of 16th International Conference: Spatial Sound Reproduction*, 1999

[112] J.-M. Jot, "Real-Time Spatial Processing of Sounds for Music, Multimedia and Interactive Human-Computer Interfaces," *Multimedia Systems*, Vol. 7, No. 1, pp. 55-69, 1999

[113] E. M. Wenzel, J. D. Miller and J. S. Abel, "Sound Lab: A Real-Time, Software-Based System for the Study of Spatial Hearing," *in Proc. of the 108 AES Convention*, Paris, France, 2000

[114] V. Pulkki, "Spatial Sound Reproduction with Directional Audio Coding," *Journal Audio Eng. Soc*., Vol. 55, No. 6, pp. 503-516, 2007

[115] N. Juillerat, S. Muller Arisona, S. Schubiger-Banz, "Real-Time, Low Latency Audio Processing in JAVA," *in Proc. of Int. Comp. Music Conference ICMC2007*, Vol. 2, pp. 99-102, 2007

[116] http://www.csounds.com/manual/html/UsingRealTime.html

[117] http://www.audiogames.net/

[118] O. Balan, A. Moldoveanu, F. Moldoveanu, M-I. Dascalu," Audio Games- a Novel Approach towards Effective Learning in the Case of Visually-impaired People," *In Proc. 7th Int. Conference of Education, Research and Innovation*, Seville, Spain, 7 pages, 2014