

Review Papers

Sonification: Review of Auditory Display Solutions in Electronic Travel Aids for the Blind

Michał BUJACZ, Paweł STRUMIŁŁO

*Institute of Electronics
Lodz University of Technology*

Wolczanska 211/215, 90-924 Łódź, Poland; e-mail: {bujaczm, pawel.strumillo}@p.lodz.pl

(received May 29, 2015; accepted May 31, 2016)

Sonification is defined as presentation of information by means of non-speech audio. In assistive technologies for the blind, sonification is most often used in electronic travel aids (ETAs) – devices which aid in independent mobility through obstacle detection or help in orientation and navigation. The presented review contains an authored classification of various sonification schemes implemented in the most widely known ETAs. The review covers both those commercially available and those in various stages of research, according to the input used, level of signal processing algorithm used and sonification methods. Additionally, a sonification approach developed in the Naviton project is presented. The prototype utilizes stereovision scene reconstruction, obstacle and surface segmentation and spatial HRTF filtered audio with discrete musical sounds and was successfully tested in a pilot study with blind volunteers in a controlled environment, allowing to localize and navigate around obstacles.

Keywords: sonification; auditory display; electronic travel aid; visual impairment; blindness; assistive technologies.

Notations

ETA – electronic travel aid,
HRTF – head related transfer function.

1. Introduction

The term sonification was coined by researchers dealing with auditory display of data (HERMANN *et al.*, 2011) and is an important aspect of many assistive technologies (CSAPO, WERSENYI, 2013). Sonification is one of the key techniques applied in the sensory substitution systems for the visually impaired (MAIDENBAUM *et al.*, 2014). The idea is to generate an auditory representation of the environment and create so called “auditory images” (BREGMAN, 1990) in the blind listener’s mind. Sensory substitution can be defined as a technique augmenting the blind user’s perception capabilities by using other than visual sensory modalities (hearing and touch) to generate information about the surrounding environment (ELLI *et al.*, 2014). The majority of current efforts for building electronic sensory substitution systems focus on devices using auditory interfaces. This is mainly due to a richer infor-

mation channel offered by the human hearing systems in comparison to touch. Also, sound rendering devices are more compact, light weight and cheaper than haptic actuators available nowadays (VISELL, 2009).

The authors’ research encompassed the design of sonification algorithms for a prototype electronic travel aid (ETA) for the blind called Naviton, the work which is being continued in the Sound of Vision H2020 European project (BUJACZ *et al.*, 2012). Part of the design process was an extensive review of existing output solutions in devices which aid the blind in independent travel, by providing information on their immediate surroundings.

One of the main observations from the review was that there existed a type of polarization in the ETA auditory output complexity. The devices either provided very limited information (usually obstacle detection with a single range-finding sensor) through simple though intuitive sounds, or they provided an overabundance of auditory data generated basing on more complex sensors (such as video or stereo cameras). The latter required the user to learn how to select and process information useful for travel, usually over months of training (DAKOPOULOS, BOURBAKIS, 2010). That

is why the presented review will be divided into the two main sections: obstacle detectors and environmental imaging devices.

The input of the obstacle detectors is usually a distance reading from an ultrasonic or laser range sensor, while the environmental imagers can use arrays of ultrasonic transducers, 2D images or 3D imaging systems such as stereovision or Time-of-Flight (ToF) camera technology. The output of an ETA is either auditory (monaural or binaural) or tactile (pressure, vibration or electric) (CSAPO, WERSENYI, 2015).

A worth noting recently engineered ETA solution that employs tactile presentation is the I-Cane Mobilo (2015). It uses a patented tactile arrow embedded into the white cane's handle to alert about the near obstacles and indicate the navigated path direction. A comprehensive review of systems using the tactile modality in various human-computer interfaces can be found in (VISELL, 2009).

The presented review focuses on auditory non-visual presentation techniques of spatial information to blind individuals in electronic travel aids. The auditory outputs used in such systems range from simple binary alerts indicating the presence of an obstacle in the range of a sensor, to complex sound patterns with varied spectra carrying almost as much information as a graphical image.

In general, there are two basic techniques that use the auditory modality for conveying information about the environment to a visually impaired person:

- verbal description (termed also visual description), i.e. a technique using speech to narrate the video content (also using speech synthesizers),
- sonification, in which non-speech synthesized audio signals are used for presenting information about the environment.

Audio description is used by guides of the blind persons, it is also applied for describing any visual content, e.g. in education and the verbal description of works of art (movies, paintings, sculptures).

Sonification is a technique widely used in human-system interaction devices in ETAs for the blind people in particular. Comprehensive reports on these techniques can be found in (KRAMER *et al.*, 1994) and (EDWARDS, 2011). Sonification is also used in the virtual audio displays as part of virtual reality systems (LOKKI *et al.*, 2003). By the terminology used in (MCGOOKIN, BREWSTER, 2004), the sound messages corresponding to real-life events that produce non-verbal sounds (e.g. a sound of a crushed sheet of paper informing about placing it in a trash folder) are referred to as auditory icons. Earcons, on the other hand, are virtual sounds having no equivalent in natural sounds (e.g. various beeping alerts). MCGOOKIN and BREWSTER (2004) proposed to interpret sets of such sounds in terms of a “musical language” with ap-

propriate semantic and syntactic rules allowing to combine them into longer meaningful messages in a similar way as used in natural languages.

The review of auditory display devices used in the ETA for the blind is followed by a short description of a sonification scheme proposed by the authors in a system named Naviton. An important aspect of the proposed ETA is spatial audio utilizing personalized head related transfer functions (HRTFs) (DOBRUCKI *et al.*, 2010).

It should be stressed that so far, no ETA has even remotely reached the scale of use as the white cane, used by millions of blind persons worldwide, with top obstacle detectors numbering their users in tens of thousands (e.g. SmartCane – smartcane.saksham.org).

2. Obstacle detectors

The white cane is the most common, primary obstacle detector used by the visually impaired. It is a simple device that extends the sense of touch of the blind user by ca. 1.5 m. The vibrations of the white cane handle are clearly perceived by the user as originating from the tip of the cane. This effect, known as attribution of the distal stimulus was coined by LOOMIS (1992) (see also a more recent discussion on this physiological phenomenon in (SIEGE, WARREN, 2010)). This observation is important since the objective of the ETA systems is to generate an effect of “mental images”, which the visually impaired build for perceiving the environment surrounding them.

The most basic, but also most common ETAs are called obstacle detectors, as they consist of a sensor that either simply detects or measures the distance to objects within a specified range (HERSH, 2008). The most common detection angular range is within a 30° wide ultrasonic cone, with distance readings ranging from several cm to 5 m. The sonification in such devices is limited to mono sounds and commonly consists either of simple auditory alerts or some form of a pitch transform, where the distance to an obstacle is either directly or inversely proportional to the frequency of a cyclically repeated sound.

The Kay Sonic Torch was the first widely used modern ETA, developed in 1959 (KAY, 1964). Its designer – Dr. Leslie Kay is one of the most recognizable names in assistive technology. The Sonic Torch was a flashlight-like device that utilized a narrow-beam ultrasonic sensor and a mono earphone output. The torch transmitted a wide-bandwidth ultrasonic signal (40–80 kHz) four times per second, which was multiplied by the received reflected signal, creating a complex sound in the auditory range (approximately 400–8000 Hz). The resulting signal was quite rich in information. The loudness and pitch corresponded to the distance (louder lower sounds meant a nearby obstacle, high pitched quieter sounds more distant reflections). The timbre

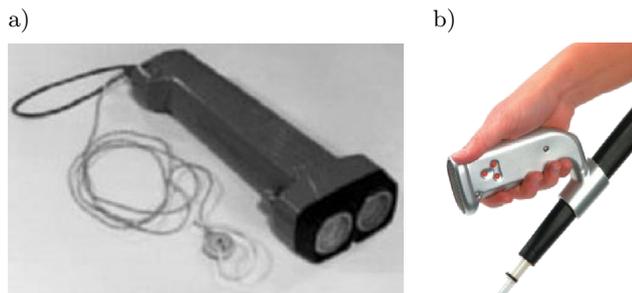


Fig. 1. The Kay Sonic Torch (KAY, 1964) and its modern version – Kay Sonic Cane (KAY, 1974).

of the sounds corresponded to how well the target surface reflected the sonar, allowing an experienced user to distinguish between different surfaces, e.g. a sidewalk and grass. The Kay Sonic Torch had many later versions – it was made attachable to a traditional long cane and greatly decreased in size, and its most current version is known as the BAT K-Sonar Cane. The same sonification scheme was later utilized by the binaural Sonicguide, which will be discussed in the environmental imagers section.

A much simpler early obstacle detector was the Russell Pathsounder (FARMER, 1978). It was a small box worn suspended on the neck and it had an ultrasonic beam sensor (30°). The output was a simple sound alert when there was an obstacle present in front of the sensor within 2 m, with an additional vibration alert when an obstacle was closer than 1 m.

The Laser Cane, first introduced in 1973, is one of the most popular and successful ETAs (FARMER, 1978; MALVERN, NAZIR, 1973). It consists of three laser range-finders placed on an adjustable long cane, at three distinct angles. The top sensor aimed upwards, after calibration to the user's height, warns of obstacles close enough to be at head-level, the middle sensor is directed horizontally and set to alert of obstacles 2–4 m away. The bottom sensor is aimed downward along the cane and warns of oncoming curbs or drop-offs by differentiating the range signal and detecting sudden changes. The output of the Laser Cane consists of simple alerts, either auditory or tactile, with a distinct sound or vibration frequency assigned to each one of the three sensors.

Mims' Seeing Aid constructed in 1966 is worth noting as it has utilized an uncommon type of sensor – several IR diodes and a photosensitive cell that measured the reflected light (FARMER, 1978). The sonification was a simple proximity to loudness transform, with the distance to an obstacle determined from the intensity of reflected IR light pulses. The aid could be hand-held or worn on eyeglasses and the audio output was delivered through a small tube placed in the ear.

The Nottingham Obstacle Detector (FARMER, 1978) was the first travel aid to utilize musical tones in the sonification of range. The input was a typical

$\sim 30^\circ$ wide ultrasonic sensor. The output consisted of 8 major scale notes, corresponding to 8 ranges which were multiples of about 0.3 m (from 0 to 2.1 m). The higher the note, the closer was the nearest obstacle. The device had a loudspeaker or earphone output.

A very good example of an advanced modern obstacle detector that utilizes sonification is the Teletact (FARCY, BELLIK, 2002). A triangulating laser telemeter is used as the input and the user can choose between auditory or tactile output. The laser is often regarded as a much better range-finding solution, as its narrow beam allows for accurate detection of doors or accurate estimation of the width of obstacles. For sonification, the distance information is converted into one of 28 notes of inversely proportional pitch, i.e. the smaller the distance to an obstacle, the higher pitched the sound. The Teletact also offered a tactile output, with a lower resolution (8 different vibration patterns). A similar solution was explored in a computer simulation by the first author in his M.Sc. thesis.

An interesting laboratory study of sonification of range was performed by MILIOS *et al.* (2003). In their prototype ETA readings from a laser rangefinder were transformed onto discrete pitches, more precisely MIDI piano sounds ranging from 4200 Hz at 0.3 m to 106 Hz at 15 m at the default rate of 8 notes per second. An alternative mode was also tested in which not the distance, but its derivative was transformed to the pitch. This allowed to emphasize changes in consecutive range measurements. The prototype was never developed into a portable ETA device, however the extensive study demonstrated that using a simple sonified rangefinder allowed trial participants to quickly learn to perceive scene elements such as wall corners.

The EyeCane (MAIDENBAUM *et al.*, 2014) is an original solution that combines both auditory output and tactile vibrations for obstacle warnings. Infra-red emitters and sensors are embedded into the white cane handle. The device emits two narrow beams, one directly ahead and one towards the ground at a 45° angle. The pitch of the sound increases while the user approaches an obstacle (no concrete frequency values given in the reference though), simultaneously the strength of vibrations also increases.

A summary of the sonification methods used by the listed obstacle detectors can be found in Table 1. Most devices either use simple alerts or a pitch transform that convert a range reading to the frequency of sound. Interestingly, the older ultrasound devices that transformed the actual received signal into the auditory domain used a proportional pitch transform – a large distance reading meant higher pitch (though also decreased loudness). All other devices had an inversely proportional pitch transform – nearby obstacles corresponded to a higher pitched sound, which seems more natural from the psychoacoustic point of view. Also most devices that used a pitch transform utilized

Table 1. Summary of ETAs and main characteristics.

ETA name (creator, date)	Device input and description	Sonification summary	Description (reference)
1	2	3	4
Kay Sonic Torch (KAY, 1964) a.k.a. BAT 'K' Sonar Cane	Ultrasound torch (emitter + receiver) hand held or cane-mounted	Distance to pitch transform.	The closer an obstacle, the lower the pitch. Due to direct transformation of reflected ultrasound into audible range, the texture of an obstacle's surface is evident in the sound's timbre.
Mims' Seeing Aid (MIMS, 1966)	Infrared beam and sensor, could be hand held or attached to glasses	Distance to loudness transform.	The IR pulses reflected from the environment were directly transformed into an audible noise-like signal, that varied in loudness depending on the proximity (and reflectivity) of the nearest obstacle caught in the IR beam.
Sonic Guide Other names: KASPA (KAY, 1974)	Ultrasounds	Distance to pitch transform, binaural amplitude difference.	A binaural version of the Kay Sonic Torch. The ultrasounds received by two separated ultrasound receivers are transformed into the audible range and independently played in each stereo channel, creating a natural binaural effect.
Pulsed Ultrasonic Binaural Aid (ORLOWSKI, 1976)	Handheld ultrasound emitter and stereo receivers on glasses with headphones	Binaural clicks. Distance transformed to discrete frequencies of repetition. 1 s cycles.	Directionality of obstacles coded via binaural time difference. Distance to obstacles determines the frequency of clicks (from 16 Hz at 0–60 cm to 1 Hz at 5–10 m)
Nottingham Obstacle Detector (FARMER, 1978)	Ultrasound torch (emitter + receiver) hand held	Distance to pitch transform, discrete musical tones.	The closer an obstacle, the higher the pitch. 8 major scale notes, corresponding to 8 ranges which were multiples of about 0.3 m (from 0 to 2.1 m).
Sonic Pathfinder (HEYES, 1984)	Ultrasound sensors (2 emitters and 3 receivers)	Distance to pitch transform. Discrete musical sounds. Three discrete stereo directions (left, right, center).	Only the nearest obstacle is sonified. The closer an obstacle, the lower the pitch. Speed of repetition of the sound alerts depends on the walking speed and probability of collision. Alternative shoreline following mode – pitch changes signify if the user is straying towards (lower) or away from (higher) a wall or a sidewalk edge that is parallel to the travel direction.
The vOICe (MEIJER, 1992)	Grayscale webcam image or 2.5D depth image from stereovision	Mono or stereo inverse spectrogram transform. 1 s cycles.	A sweep transform that maps columns of pixels onto sinusoidal frequency components (synthesizing a 1s sound the spectrogram of which looks like the image). The sounds can be mono, or the left-right image sweep can correspond to a stereo pan.

Table 1 [Cont.]

1	2	3	4
Navbelt (SHOVAL, 1998)	Ultrasound sensor array worn on a belt	Distance to pitch and loudness transform. Stereo panning.	The closer an obstacle, the higher the pitch and loudness. Distance mapped to loudness. Sounds virtually placed in the horizontal axis. Frequency of repetition adapts to the travel speed and scene complexity.
Laser ETA (MILIOS, 1999)	Laser range finder	Distance to pitch and loudness transform. Discrete MIDI piano sounds. 8 notes/s.	The closer an obstacle, the higher the pitch and amplitude. Discrete MIDI notes are used, ranging from 4200 Hz at 0.3 m to 106 Hz at 15 m. In the derivative mode, change in consecutive range measurements is signified by higher or lower pitch.
EAV (GONZÁLEZ-MORA, 1999)	2.5D images	Spatial audio (HRTF). Distance to loudness transform.	Virtual sound sources spatialized through HRTFs are projected onto the scene. They click in unison creating an illusion of obstacles producing the sounds.
Teletact (FARCY, 2002)	Laser telemeter	Distance to pitch transform. Discrete notes.	The smaller the distance to an obstacle, the higher pitched the sound. 28 major notes. Alternative tactile output.
Cross-modal ETA (FONTANA <i>et al.</i> , 2002)	2.5D images + laser pointer	Spatialized (HRTF). Reverb.	The user hears a virtual sound source originating in the spot illuminated by a laser pointer. Additional reverb is introduced into the sound proportionally to the distance.
NAVI (SAINARAYANAN, 2007)	2D images	Inverse spectrogram transform, binaural amplitude difference.	Processing to separate objects from background. 32 × 32 pixels and four brightness levels. Columns of pixels read left to right, sweeping in stereo, row of pixel corresponds to a frequency component. The sounds can be limited to a musical tones.
SVETA (BALAKRISHNAN, 2006)	2.5D images from stereo cameras	Simultaneous stereo sweep transforms for two halves of image.	Stereovision version of the NAVI. Rows of pixels correspond to frequency components. Simultaneous scanning from left and right towards centre of the image.
CASBliP (FAJARNES <i>et al.</i> , 2010)	2.5D images from stereo cameras + GPS	Spatial audio (HRTF). Distance to loudness transform.	Spatialized virtual sound source moves along a horizontal line projected onto the scene.
Sidewalk Detector (JIE, 2010)	Handheld PDA's built in camera (2D HSV image)	Speech with varied loudness.	Sidewalk edges are detected. Spoken commands direct the user to walk between them. The more a person veers off course, the louder the commands.

Table 1 [Cont.]

1	2	3	4
AudioGuider (ZHIGANG, 2010)	2D images + GPS and GIS	Auditory icons for obstacles. Stereo directionality. Synthesized speech.	Auditory icons represent obstacles and their directionality and distance are coded through stereo and loudness curves. Synthesized speech relays information such as street names from the Geographic Information System(22).
HiFiVE (DEWHURST <i>et al.</i> , 2010)	Video camera	Speech-like syllables reflecting colour and location of objects; motion of objects is represented through binaural panning.	The device is using both audio and haptic output to create the so called audiotactile objects. The multisyllable audio output is augmented by tactile Braille-like matrix reflecting object shape.
Naviton (BUJACZ <i>et al.</i> , 2012)	2.5D + RGB images from stereo cameras	Sonar sweep with spatialized (HRTF) discrete (MIDI) musical sounds. Pitch, loudness and temporal delay map distance.	Reconstructed 3D scenes are segmented into parametrically described surfaces and objects. The distance is mapped onto loudness, frequency and temporal delay (a scanning surface moves away from the observer, releasing sounds in scene elements as it intersects them). Due to HRTF filtering, all sounds appear to originate from the scene elements.
EyeCane (MAIDENBAUM <i>et al.</i> , 2014)	A pair of IR range sensors	Distance to pitch.	The device simultaneously emits two beams: one to detect obstacles directly ahead (up to 5 m range) and the second towards the ground at a 45° angle (range up to 1.5 m). Two outputs are simultaneously generated: auditory and tactile. The closer the detected object the higher the sound frequency and the stronger the vibrations.
EyeMusic (LEVY-TZEDEK, 2014)	2D color images	Inverse spectrogram transform. Discrete musical notes. Colours coded by instruments.	A sweep transform that maps columns of pixels onto musical notes. Row corresponds to pitch, brightness to loudness, colour to instrument.
SeeColor (GOMEZ VALENCIA, 2014)	Stereovision camera	Auditory timbres represent colours and rhythmic patterns represent distance.	There are specialised modules applied for local perception (e.g. colour), global perception (distance) and alerting (for objects threatening the user). The user can interact with the modules by using a tactile tablet to indicate regions of the image for non-visual presentation.

musical notes in an attempt to make the sounds more pleasant to the ear. Many simple obstacle detectors, such as the Miniguide (MINIGUIDE, 2015), were not listed in this review, as they all work in a very similar fashion – playing a sound alert when an obstacle is within a sensor’s detection range (usually up to 4 m) which is sometimes adjustable.

3. Environmental imagers

The concept behind environmental imaging ETAs is to go a step beyond obstacle detection and provide some degree of information on the respective layout of obstacles in the near proximity of a blind person without the need to manually scan the environment using a narrow beam sensor (MILIOS *et al.*, 2003). The development of microprocessor and computer technology allowed for creation of more complex travel aids, though most of the most recent listed devices are only research prototypes.

All environmental imagers either utilize multiple ultrasonic sensors, mono or stereovision cameras. In the device descriptions the terms 1.5D and 2.5D will frequently be used. The terms 2.5D refers to an image in which the intensity of each pixel represents the distance to the nearest scene element along a line projected from a hypothetical pinhole camera through the image plane, while 1.5D would be a single row of pixels from such an image.

The Sonicguide (KAY, 1974), considered by most to be the first environmental imager, was a stereophonic expansion of a well known obstacle detector – the Kay Sonic Torch (KAY, 1964). While the torch used a single narrow beam ultrasonic sensor, the Sonicguide used a wide beam emitter and two angled receivers. Each received signal was transformed independently into the auditory range using the same principle as the sonic torch, and passed to a separate stereo channel, creating a natural binaural effect. There were several versions of the Sonicguide with its most modern version renamed KASPA (Kay’s advanced sensory perception aid). In an effort to minimize the blocking of environmental sounds, the headphones for the Sonicguide ended with

narrow tubes which did not cover the ear channel. The stereo output of the Sonicguide provided no information on the vertical placement of obstacles, allowing only to approximate the directions towards the nearest objects in the wide area covered by the ultrasonic sensor.

The Pulsed Ultrasonic Binaural Aid constructed by ORLOWSKI (1976) utilized a click-based sonification similar to a Geiger counter. It had an uncommon construction as it consisted of a handheld ultrasonic emitter and stereo receivers worn on glasses with headphones. The distance to the nearest obstacle was sonified by varying the frequency of pulses (from 16 Hz at 0–60 cm to 1 Hz at 5–10 m). Additionally, the direction towards the obstacle (difference between the signals observed by the stereo receivers) was transformed into a binaural time difference between the clicks.

Another early ETA of note is the Sonic Pathfinder (HEYES, 1984). It was the first ETA that used a microprocessor to attempt to intelligently limit the presented information to the most useful and necessary minimum. It had two ultrasonic emitters and three receivers, the input of which was analyzed and a stereophonic output informed the user of just the nearest obstacle (the nearest peak in the sonar ranging input), with one of eight musical notes corresponding to the distance and with a binaural amplitude difference indicating the direction. Objects near the centre of the travel direction or moving towards the observer were given priority. Additionally, the system varied the output depending on the estimated speed of travel, as well as the number, position and movement of the detected obstacles. The more cluttered the scene or the faster the travel speed, the more frequent were the sounds coming from the device. The Pathfinder has an alternate mode that can be used for shoreline following (e.g. moving along a wall). In this mode pitch changes signal whether the wearer is straying towards or away from a shoreline that should remain parallel to the direction of travel.

The Navbelt is frequently mentioned in ETA reviews, even though it has never gone past a cumbersome prototype stage and mainly an attempt to verify if navigation sensors designed for an autonomous robot could be used for the benefit of the blind user (SHOVAL *et al.*, 1998). An array of ultrasonic sensors creates a 1.5D map of the nearest environment (up to approximately 5 m). The Navbelt can operate in two very different modes. In the Imaging Mode, the user is presented with an acoustic panoramic image of the environment in front by using a stereophonic sound pattern: a signal appears to sweep through the user’s head from the right ear to the left, the direction to an obstacle being indicated by the binaural amplitude difference, while the distance is represented by the signal’s volume and pitch – nearer obstacles produce louder and higher pitched sounds. In an alterna-

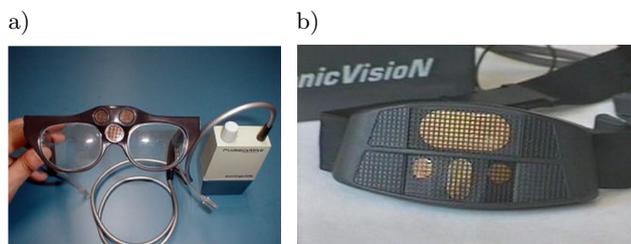


Fig. 2. Sonicguide (KAY, 1974) – one of the first binaural ETAs and one of its modern version KASPA. The sonification method is identical to the Sonic Torch (KAY, 1964), but the angled ultrasonic receivers allow producing a natural binaural effect.

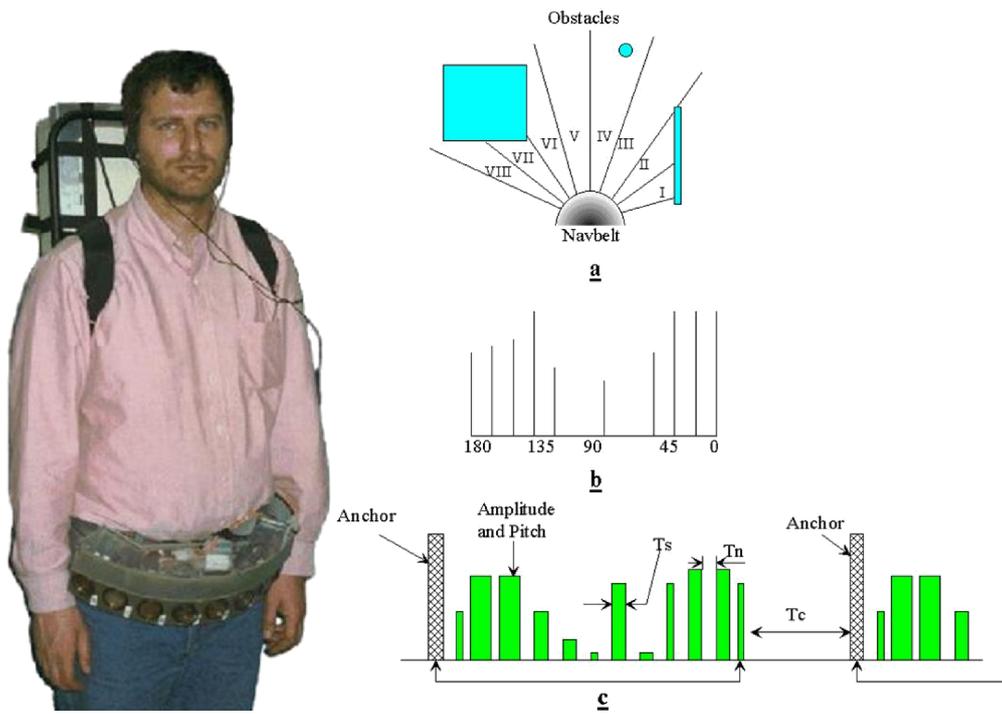


Fig. 3. Navbelt (SHOVAL *et al.*, 1998) with a visualization of its sonification algorithm. Distances to the nearest object detected by an array of ultrasonic sensors are transformed into the pitch and loudness of a short sequence of sounds.

tive Guidance Mode, only a single stereophonic signal is produced, pointing to a recommended direction of travel, which is calculated so that the current direction of travel is not changed greatly, but veers just enough to avoid any obstacles in the direct path. If the number of possible collisions with obstacles near the travel path is low, then the audio signal is low in frequency and quiet. To warn the user of a more cluttered environment, a higher pitched sound is emitted. Interestingly, in both modes the frequency with which the information is provided is determined by the user's travel speed and the probability of collisions with obstacles in the travel path.

The vOICe (MEIJER, 1992) is definitely the most known algorithm for sonification of 2D grayscale images. It has been utilized in several PC-based or even mobile-phone based prototypes (vOICe, 2015). The sonification can best be described as a reversed spectrogram transform, i.e. the algorithm synthesizes a 1 s sound, the spectrogram of which would be identical to the input grayscale image. The 174×64 pixel image is read column by column from left to right. The vertical coordinate of every pixel corresponds to a specific frequency component from roughly 500 Hz at the bottom of the image to 5 kHz in the top row. The brightness of each pixel is translated as the amplitude of its assigned sinusoid, which changes as the image is scanned. The sound lasts for a second and also pans stereophonically as the image is read from left to right. Although recognition of useful information in the com-

plex sound stream is very difficult to learn, it has been demonstrated that after prolonged training (3 months) a blind user's brain can significantly adapt itself to the

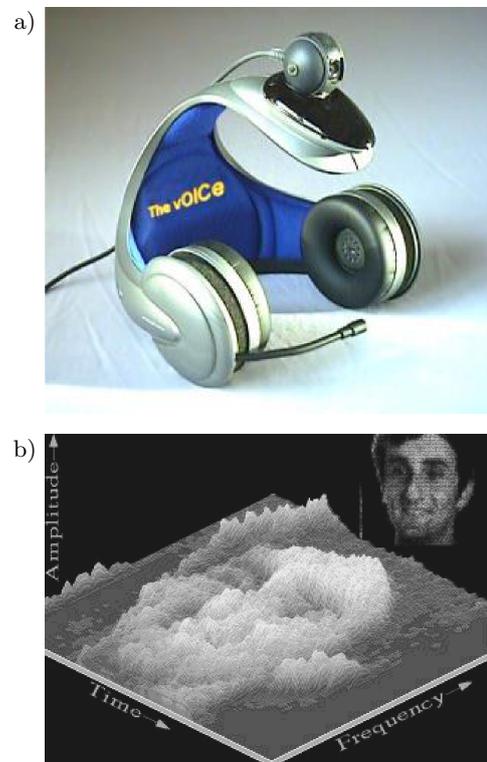


Fig. 4. The vOICe: a) headset, b) audio-plot representation of a 2D image (MEIJER, 1992).

sensory substitution (MERABET *et al.*, 2009). The 2D input devoid of depth information greatly limits the vOICE's use as a mobility aid, but the algorithm can be easily applied to 2.5D images and the potential use of stereovision technology has been mentioned by the system's designers (vOICE, 2015).

A similar solution to the vOICE was implemented in the SVETA system by BALKARISHNAN *et al.* (2006). The input for the sonification is the 2.5D depth map from a stereovision system. As in the vOICE, each column of pixels is treated as a momentary spectrum of the synthesized sound, except in SVETA two simultaneous sound streams are created for two halves of the image. The left half of the image is scanned left to right, while the right half in the opposite direction. The two synthesized sounds are independently output to their respective stereo channels. The training to understand this complex sound stream is as difficult or even more difficult than for the vOICE; however, the sonification seems more suited for aiding in independent travel.

The HiFiVE is an example of an ETA that combines sound with touch for presenting visual information to the blind (DEWHURST, 2010). Visual data is mapped onto speech-like (but non-verbal) auditory sound messages. The messages comprise three syllables which correspond to different image regions: one syllable for colour and two for layout, e.g. "way-lair-roar" might correspond to "white-grey" and "left-to-right". Changes in texture are mapped onto fluctuations of volume, while motion is represented through binaural panning.

Another ETA solution that combines auditory display and tactile interaction is the See Color system (GOMEZ VALENCIA, 2014). Coloured pixels represented in a Hue, Saturation, Luminosity (HSL) system are transformed into spatialized classical instrument sounds lasting approx. 3000 ms. Hue is sonified by the timbre, saturation by musical notes, and luminosity is represented by different musical instruments, e.g. by bass for dark regions and singing voice for bright image regions. The user can indicate regions of the image for sonification. Finally, distance is represented by rhythmic sound patterns.

A very different sonification approach focusing on natural psychoacoustic cues can be found in the Spanish prototype ETA called EAV – Espacio Acustico Virtual (Virtual Acoustic Space) (GONZALES-MORA *et al.*, 1999). The EAV uses stereoscopic cameras to create a low resolution ($16 \times 16 \times 16$) depth map of the observed environment in front of the user. All occupied voxels become virtual sound sources that pulse (Dirac delta function) in unison with each other. Due to spatial filtering with the user's individual HRTFs and calculated time delays, the sources create an illusion of "sound clouds" emanating from all scene elements. It is worthy to note that the EAV project, just like the

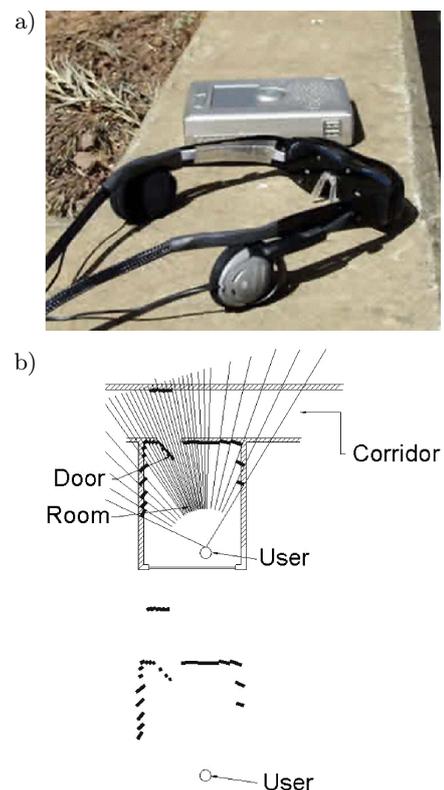


Fig. 5. EAV – Espacio Acustico Virtual (GONZALES-MORA *et al.*, 1999) and a visualization of its sonification method. A cloud of spatialized virtual sound sources is projected onto the scene (black rectangles) and click in unison.

vOICE, has demonstrated the adaptability of its users' brains to the sensory substitution, with fMRI showing activation of vision-related brain regions when listening to the audio output of the device.

The Sidewalk Detector is a recent ETA that is difficult to classify, as it is neither an obstacle detector nor an environmental imager (JIE, 2010). It is a handheld PDA running an application that utilizes the device's built in camera to observe the path in front of a blind user. The video stream is processed to detect the edges of the sidewalk. The user is provided with instructions in the form of spoken commands that prevent the user from straying off of the sidewalk. The commands are played with varying loudness dependant on how much the user strays.

The AudioGuider is a recent prototype that combines an ETA with a navigation aid (ZHIGANG, TING, 2010). The proposed system uses image recognition to assign auditory icons to obstacles with their position and distance being coded through stereo and loudness curves, although this aspect has only been tested virtually. Additionally, the system incorporates travel directions from global positioning system (GPS) and geographic information system (GIS), such as street names, using synthesized speech.

The Cognitive Aid System for Blind People (CAS-BliP) is an ETA project by a large Spanish and EU consortium (FERNÁNDEZ TOMÁS *et al.*, 2007). Although it utilizes stereovision input, the 2.5D image sequence undergoes significant simplification and segmentation to detect obstacles and create a 1.5D distance map. The distance map is sonified using HRTFs to generate a virtual source moving along the horizontal shoreline created by the observed obstacles. This essentially transforms distance to loudness and direction to a binaural amplitude and time difference. A later version of the system was the previously described SVETA (BALAKRISHNAN *et al.*, 2006). A new prototype stemming from the same project is called EYE21.

EyeMusic is a prototype that was developed more than a decade ago (LEVY-TZEDEK *et al.*, 2004). It is a musical variation of the inverse spectrogram approach popularized by the vOICE. The input of the system consists of 2D colour images recorded by a small digital camera worn on eyeglasses. The images are sonified left to right, with pixel vertical positions corresponding to musical notes spanning five octaves, and their brightness to the loudness. What differentiates the algorithm from similar solutions is the use of multiple musical instruments to represent different colours present in the image: blue – trumpet, red – reggae organ, green – reed, yellow – violin, white – vocals and black – silence.

The overall observation is that several environmental imagers (KASPA, vOICE, EAV) provide an overabundance of data and require substantially more focus and training than simple obstacle detectors. This is why environmental imagers are often referred to as sensory substitution devices (SSDs), as the sense of hearing is almost completely occupied by replacing the lost sight.

A common way to sonify a 2D or 2.5D image is using the inverse spectrogram concept – the row of a pixel corresponds to a frequency or a musical tone, it's brightness or closeness represents amplitude, and the image is sonified column by column in short sweeps, often coupled with a stereo pan.

Information in 1.5D form is usually sonified as a single changing sound or a short sequence of sounds with distance transformed into loudness and/or pitch (in almost all cases higher pitched sounds mean closer obstacles).

Another approach is to use natural psychoacoustic cues through spatial audio filtered through HRTFs. This is also the only way to provide information on vertical position of obstacles using stereophonic sound, though very few systems utilize this property. HRTFs are quite effective in generating an illusion that a virtual sound source is located in space in front of a listener, however, the spatial filtering is computationally complex and some argument arises as to

whether individually measured HRTFs are needed or can generic/modelled ones substitute them.

Several environmental imagers (Sonic Pathfinder, Navbelt) attempt to select and present only the most important information and/or vary the presented information depending on the travel speed and number of possible obstacles detected in front of an ETA user.

4. The naviton ETA prototype

The prototype developed by the authors was intended as an environmental imaging ETA that combined the strengths of obstacle detectors and environmental imagers. The goal was to make the sonification algorithm easy and intuitive to learn by simplifying the information used as its input. A scene segmentation algorithm was used to produce an approximated model of scene surfaces and obstacles as illustrated in Fig. 6. The sonification could then assign sounds to the segmented scene elements.

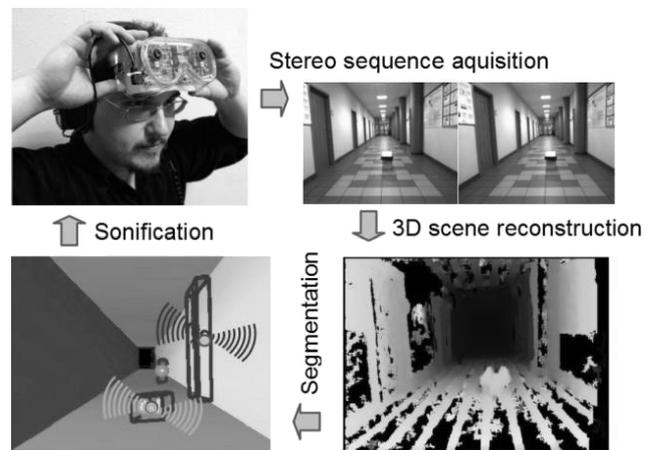


Fig. 6. The Naviton electronic travel aid concept (BUJACZ *et al.*, 2012).

4.1. Scene reconstruction and segmentation

The ETA solution proposed by the authors utilizes stereovision input with custom made glasses (OSTROWSKI *et al.*, 2011), though a Bumblebee Point Grey camera set was used for the published trials. In a parallel project a successful attempt was made to utilize GPU processing to significantly speed up this process (STRUMILLO *et al.*, 2009). DirectShow was used to interface the different prototype software modules (SZCZYPINSKI *et al.*, 2010).

The segmentation is divided into two main tasks: detection of large planes (floor, walls, etc.) and after their removal from the image approximation of remaining obstacles by cuboids is carried out (SKULIMOWSKI *et al.*, 2009). This process allowed to use a relatively simple stream of parameters (coordinates and dimensions of the cuboids and the plane equations) as the

input for the sonification algorithm. The default model limits the input to four surfaces and four closest obstacles and the algorithm updates this data 5–10 times per second.

4.2. Naviton sonification algorithm

The sonification algorithm proposed by the authors was shaped by several earlier simulation trials and surveys with the blind (BUJACZ, STRUMILLO, 2008). The main concept was to translate the parameters of the segmented scene elements (surfaces and obstacles) into parameters of virtual sound sources localized in space using HRTFs.

The chosen sound code represents the distance to an object with both pitch and amplitude. The duration was made proportional to the size (width) of an object. Using individualized HRTF based spatial audio (DOBRUCKI *et al.*, 2010), the virtual sound sources are localized to provide the illusion of originating from the scene elements they represent. The surfaces and obstacles are assigned different instrument types and/or pitch ranges and the sonification algorithm includes classes for elements not yet implemented in the segmentation algorithm (e.g. humans or doorways classified using image recognition). The sounds are played back in order of proximity to the observer, creating a sonar-like effect. See Fig. 7 for explanation of the adopted sonification scheme of the environment.

To enable utilization of both HRTF filtering and MIDI instruments, banks of wave sound files were generated using the Microsoft General MIDI synthesizer and modulated with 5% noise (14 dB SNR). This is because previous studies had showed that such virtual sources had been better localized than clean instrument sounds (PEC *et al.*, 2008). The banks contained 5 s long tones from the diatonic scale (octaves 2 through 4), although the default pitches used in the sonification of obstacles spanned from tone G2 (98 Hz) to B4 (493 Hz) and for walls from G3 (196 Hz) to G4 (392 Hz). The default instrument for obstacles was a grand piano (General MIDI program 1), while a calliope (General MIDI program 83) was used for walls.

The concepts of sound stream segregation and integration as described by BREGMAN (1990) heavily influenced the sonification algorithm's design. The instruments were chosen to differ spectrally in both pitch and tone, and a minimum delay between the onsets of two sounds was set to 0.2 s (artificial delay was introduced if two obstacles were at the same distance from the observer). The selection of sounds was also influenced by surveys with ten blind testers conducted during development, with most participants preferring musical full tones of recognizable instruments (BUJACZ, STRUMILLO, 2012).

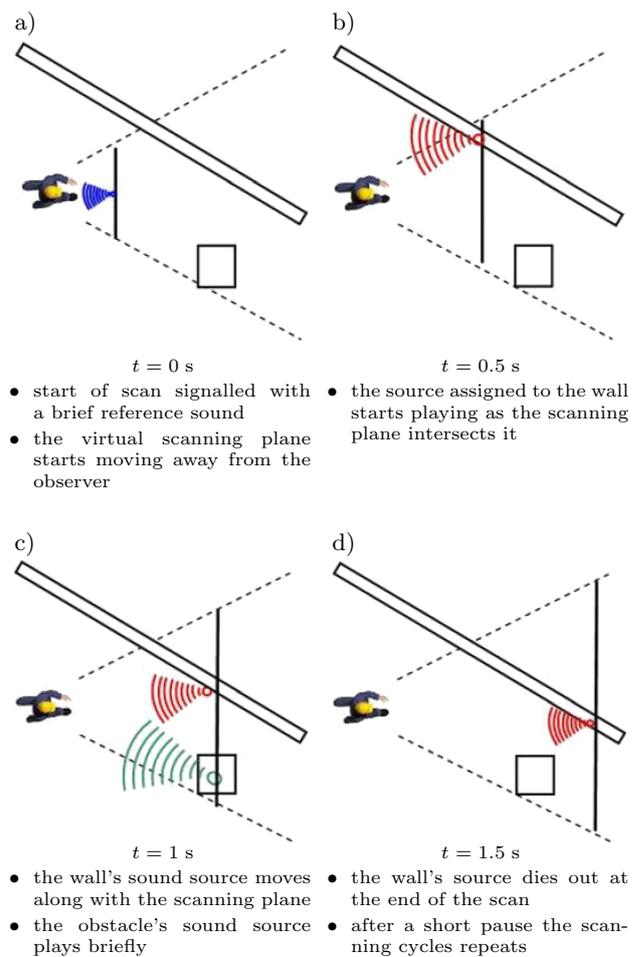


Fig. 7. The Naviton's sonification algorithm can be visualized using a virtual scanning surface that cyclically travels through the scene and releases spatialized sound sources assigned to segmented scene elements (BUJACZ *et al.*, 2012).

4.3. Prototype tests

The sonification algorithm was tested in a number of ways suggested for ETA assessments (HERSH, JOHNSON, 2008): surveys, simulations, emphatic trials and tests in real environments. The virtual reality simulation test was performed by ten sighted subjects (BUJACZ *et al.*, 2009), and later in controlled real world conditions in a pilot study with 5 blind and 5 blindfolded volunteers (BUJACZ *et al.*, 2012).

The trials used a Point Grey Bumblebee2 Fire-Wire stereovision module, although since then custom made stereovision glasses were developed; however since then several custom stereovision modules were built.

The trials tested basic obstacle avoidance and orientation in scenes arranged from coloured cardboard boxes (the colours and texturing significantly improved the accuracy of scene reconstruction and segmentation). The tests completed by the blind and blindfolded volunteers showed that the sonification algorithm was straightforward to understand and use for obstacle de-

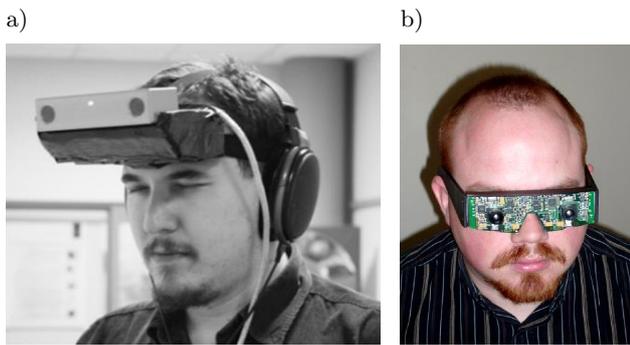


Fig. 8. The early prototype with a Point Grey camera (a) and the custom made stereovision module (b).

tection and orientation in a scene. For example the blind volunteers could easily navigate between several boxes to reach a small radio serving as a destination marker (Fig. 9). A detailed description of the trials and a summary of the results can be found in (BUJACZ *et al.*, 2012).

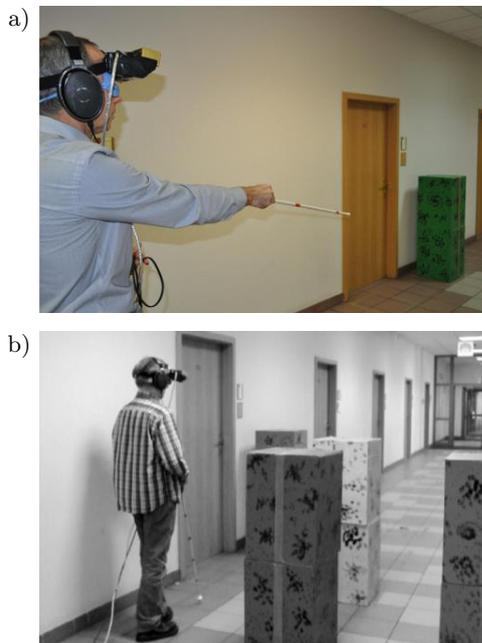


Fig. 9. Naviton prototype trials: blindfolded participant locates an obstacle (a), a visually impaired volunteer navigates between obstacles (b) (BUJACZ *et al.*, 2012).

Although, the number of participants in the pilot study was small, several useful qualitative observations were made that strongly influenced the authors' future focus on designing the electronic travel aids. First of all the use of HRTFs was a very subjective experience – despite the use of individually measured HRTFs for all participants (DOBRUCKI *et al.*, 2010) only half of the testers reported they experienced immediate clear and intuitive externalization of the sounds. The remainder perceived the sounds in simple stereo; however, most remarked that after the initial training they started to

intuitively attribute the sounds with an external location. Secondly, all the blind participants remarked on the type of headphones used in the test – the high quality reference headphones were suitable for laboratory conditions, but any ETA used on the street would need to be designed with headphones that do not block environmental sounds. Two types of patterns were observed in the participant movements – as the sounds were presented in 2 s cycles, the participants either moved in short “bursts” – listening for 2–3 cycles than taking several fast steps, or they moved continuously, but much slower than their usual walking speed.

Another important qualitative assessment was that of the cognitive load and tiredness associated with the prolonged perception of the sonification. On one hand the conscious effort to attribute sounds to scene obstacles quickly decreased with training, on the other the prolonged exposure to repetitive sounds was reported as tiresome. A common suggestion was to develop different sonification modes that could be chosen by the ETA user – e.g. a mostly silent mode for travelling that only alerted of dangerous obstacles or a detailed imaging mode to perceive the environmental layout and all nearby objects, but released manually and not looped.

The conclusions from the Naviton prototype tests are being used as a base for the EU Horizon 2020 project called Sound of Vision (www.soundofvision.net).

5. Conclusions and future work

The review of the sonification methods in electronic travel aids summarized the most common ways of representing obstacles through sound. Nearly all reviewed devices used loudness to represent the distance to scene elements. Many used pitch to represent distance, and most used an inverted form of the transform (higher pitch meant shorter distance), although a few used a proportional one (those utilizing direct transforms of ultrasonic signals). When sonifying 2D or 2.5D images a popular sensory substitution approach was to perform an inverse spectrogram transform – have rows of pixels correspond to frequency components (in all cases lowest row = lowest frequency) and use pixel brightness for the amplitudes of those components. Nearly all environmental imagers utilize a binaural amplitude difference for sonifying directionality to obstacles, while several recent devices use spatial audio (through generic or personalized HRTFs) to attach virtual sound sources to scene elements. A number of devices attempt various levels of processing to decrease the amount of sonified data, e.g. detection of the nearest obstacle, segmentation of obstacles from background, detection of shorelines and/or varying the sonification pattern depending on travel speed and/or the amount of obstacles.

The author's proposed algorithm is novel in that the sonification is based on a simplified parametric

model of the 3D scene. At the proposed sonification scheme the scene elements are divided into two classes: surfaces and obstacles, however both the segmentation and the sonification algorithms are developed with intent of expansion. The parametric model is sonified in a novel sonar-inspired way, with the virtual sound sources representing scene elements released in order of proximity and spatialized using individually measured HRTFs. Feedback gathered during the pilot study with blind participants confirmed the authors' observations about ETAs in general – the testers remarked that they would envision very different modes: one for use while walking, producing as few sounds as possible with a simple sonification scheme, and a different one for imaging of the environment, which could use a more complex sonification approach that would likely require a greater cognitive effort while standing still.

Acknowledgments

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. BALAKRISHNAN G., SAINARAYANAN G., NAGARAJAN R., SAZALI Y. (2006), *A stereo image processing system for visually impaired*, International Journal of Signal Processing, **2**, 3, 136–145.
2. BREGMAN A. (1990), *Auditory scene analysis: the perceptual organization of sound*, The MIT Press.
3. BUJACZ M., STRUMILLO P. (2008), *Synthesizing a 3D auditory scene for use in an electronic travel aid for the blind*, Signal Processing Symposium, Proceedings of the SPIE, Vol. 6937, pp. 693737–693737–8, Jachranka, Poland.
4. BUJACZ M., SKULIMOWSKI P., WROBLEWSKI G., WOJCIECHOWSKI A., STRUMILLO P. (2009), *A Proposed Method for Sonification of 3D Environments Using Scene Segmentation and Personalized Spatial Audio*, Conference and Workshop on Assistive Technology for People with Vision and Hearing Impairments CVHI2009, pp. 1–6, Wrocław, Poland.
5. BUJACZ M., SKULIMOWSKI P., STRUMILLO P. (2012), *Naviton – a prototype mobility aid for auditory presentation of 3D scenes*, Journal of Audio Engineering Society, **60**, 9, 696–708.
6. CAPP M., PICTON P. (2000), *The optophone: an electronic blind aid*, Engineering Science and Education Journal, **9**, 3, 37–143.
7. CSAPO A., WERSENYI G. (2013) *Overview of auditory representations in human-machine interfaces*, ACM Computing Surveys, **46**, 2, 19:1–19:23.
8. CSAPO A., WERSENYI G., NAGY H., STOCKMAN T. (2015), *A survey of assistive technologies and applications for blind users on mobile platforms – a review and foundation for research*, Journal of Multimodal User Interfaces, **9**, 3, 11 pages.
9. DAKOPOULOS D., BOURBAKIS N.G. (2010), *Wearable obstacle avoidance electronic travel aids for blind: a survey*, IEEE Transactions on Systems Man and Cybernetics – Part C: Applications and Reviews, **40**, 1, 25–35.
10. DEWHURST D. (2010), *Creating and Accessing Audio-Tactile Images with "HFVE" Vision Substitution Software*, [in:] Proc. of the 3rd Interactive Sonification Workshop, pp. 101–104, KTH, Stockholm.
11. DOBRUCKI A., PLASKOTA P., PRUCHNICKI P., PEC M., BUJACZ M., STRUMILLO P. (2010), *Measurement system for personalized head-related transfer functions and its verification by virtual source localization trials with visually impaired and sighted individuals*, Journal of Audio Engineering Society, **58**, 9, 724–738.
12. EDWARDS A.D.N. (2011), *Auditory Display in Assistive Technology*, [in:] The Sonification Handbook, Hermann T., Hunt A., Neuhoff J.G. [Eds.], Logos Publishing House, Berlin, pp. 431–453.
13. ELLI G.V., BENETTI S., COLLINGTON O. (2014), *Is There a Future for Sensory Substitution Outside Academic Laboratories?*, Multisensory Research, **27**, 271–291.
14. FAJARNES G.P., DUNAI L., PRADERAS V.S., DUNAI I. (2010), *CASBLiP-a new cognitive object detection and orientation system for impaired people*, 4th International Conference on Cognitive Systems, Zurich, Switzerland.
15. FARCY R., BELLIK Y. (2002), *Locomotion assistance for the blind*, [in:] *Universal Access and Assistive Technology*, Keates S., Langdom P., Clarkson P., Robinson P. [Eds.], pp. 277–284, Springer.
16. FARMER L. (1978), *Mobility Devices*, [in:] *Foundation of Orientation and Mobility*, American Foundation for the Blind Inc. NY.
17. FERNÁNDEZ TOMÁS M., PERIS-FAJARNÉS G., DUNAI L., REDONDO J. (2007), *Convolution application in environment sonification for blind people*, VIII Jornadas de Matemática Aplicada, UPV.
18. FONTANA F., FUSIELLO A., GOBBI M., MURINO V., ROCCHESO D., SARTOR L., PANUCCIO A. (2002) *A Cross-Modal Electronic Travel Aid Device*, Human Computer Interaction with Mobile Devices, Lecture Notes in Computer Science Volume 2411, pp 393–397.
19. GOMEZ VALENCIA J.D. (2014), *A computer-vision based sensory substitution device for the visually impaired (See ColOr)*, PhD thesis, University of Geneva.
20. GONZÁLEZ-MORA J., RODRÍGUEZ-HERNÁNDEZ A., RODRÍGUEZ-RAMOS L., DÍAZ-SACO L., SOSA N. (1999), *Development of a new space perception system for blind people, based on the creation of a virtual acoustic space*, Engineering Applications of Bio-Inspired Artificial Neural Networks, pp. 321–330, Springer Berlin/Heidelberg.
21. HERMANN T., HUNT A., NEUHOFF J.G. [Eds.], (2011), *The Sonification Handbook*, Logos Publishing House, Berlin.

22. HERSH M., JOHNSON M. [Eds.], (2008), *Assistive technology for visually impaired and blind people*, Springer, London.
23. HEYES D. (1984), *The Sonic Pathfinder: a new electronic travel aid*, Journal of Visual Impairment and Blindness, **77**, 200–202.
24. I-Cane Mobilo, (2015), www.i-cane.org (accessed May 2015).
25. JIE X., XIAOCHI W., ZHIGANG F. (2010), *Research and Implementation of Blind Sidewalk Detection in Portable ETA System*, International Forum on Information Technology and Applications, pp. 431–434.
26. KAY L. (1964), *An ultrasonic sensing probe as a mobility aid for the blind*, Ultrasonics, **2**, 2, 53–56.
27. KAY L. (1974), *A sonar aid to enhance spatial perception of the blind: engineering design and evaluation*, Radio and Electronic Engineering, **44**, 605–627.
28. KRAMER G. [Ed.], (1994), *Auditory Display: Sonification, Audification, and Auditory interfaces*, Santa Fe Institute Studies in the Sciences of Complexity, vol. XVIII, Reading, MA: Addison-Wesley.
29. LEVY-TZEDEK S., HANASSY S., ABBOUD S., MAIDENBAUM S., AMEDI A. (2012), *Fast, accurate reaching movements with a visual-to-auditory sensory substitution device*, Restorative Neurology and Neuroscience, **30**, 313–323.
30. LOKKI T., SAVIOJA L., VAANANEN R., HUOPANIEMI J., TAKALA T. (2003), *Creating interactive virtual auditory environments*, IEEE Computer Graphics and Applications, **22**, 4, 49–57.
31. LOOMIS J.M. (1992), *Distal attribution and presence*, Forum Spotlight on: The Concept of Telepresence, **1**, 113–119.
32. MAIDENBAUM S., ABBOUD S., AMEDI A. (2014), *Sensory substitution: Closing the gap between basic research and widespread practical visual rehabilitation*, Neuroscience and Biobehavioral Reviews, **14**, 3–15.
33. MAIDENBAUM S., HANASSY S., ABBOUD S., BUCHS G., CHEBAT D.R., LEVY-TZEDEK S., AMEDI A. (2014), *The “EyeCane”, a new electronic travel aid for the blind: technology, behavior & swift learning*, Restorative Neurology and Neuroscience, **32**, 813–824.
34. MALVERN B., NAZIR A. (1973), *An improved laser cane for the blind*, [in:] Developments in Laser Technology II, SPIE Proceedings 0041.
35. MCGOOKIN D.K., BREWSTER S.A. (2004), *Understanding concurrent earcons: applying auditory scene analysis principles to concurrent earcon recognition*, ACM Transactions on Applied Perception, **1**, 2, 130–155.
36. MERABET L., BATTELLI L., OBRETEANOVA S., MAGUIRE S., MEIJER P., PASCUAL-LEONE A. (2009), *Functional recruitment of visual cortex for sound encoded object identification in the blind*, Neuroreport, **20**, 2, 132–138.
37. MEIJER P. (1992), *An experimental system for auditory image representations*, IEEE Transactions on Biomedical Engineering, **39**, 112–121.
38. MILIOS E., KAPRALOS B., KOPINSKA A., STERGIPOULOS S. (2003), *Sonification of range information for 3-D space perception*, IEEE Transactions on Neural Systems and Rehabilitation Engineering, **11**, 4, 416–421.
39. Miniguide mobility aid (2015), www.gdp-research.com.au (accessed May 2015).
40. MOLDOVEANU A., BALAN O., MOLDOVEANU F. (2014), *Training system for improving sound localization*, 10th International Conference “eLearning and Software for Education”, 24–25 April, pp. 32, Bucharest, Romania.
41. ORLOWSKI R. (1976), *Ultrasonic echo reinforcement for the blind*, Ph.D. thesis University of Nottingham.
42. OSTROWSKI B., STRUMILLO P., PELCZYNSKI P., DANYCH R. (2011), *A wearable stereovision unit in an electronic travel-aid system for the visually impaired*, Image Processing and Communications Challenges 2, Advances in Intelligent and Soft Computing, **102**, 191–198.
43. PEC M., BUJACZ M., STRUMILLO P., MATERKA A. (2008), *Individual HRTF measurements for accurate obstacle sonification in an electronic travel aid for the blind*, Proceedings International Conference on Signals and Electronic Systems, pp. 235–238, Cracow, Poland.
44. SAINARAYANAN G., NAGARAJAN R., YAACOB S. (2007), *Fuzzy image processing scheme for autonomous navigation of human blind*, Applied Soft Computing, **7**, 1, 257–264.
45. SHOVAL S., BORENSTEIN J., KOREN Y. (1998), *Auditory guidance with the navbelt – a computerized travel aid for the blind*, IEEE Transactions on Systems, Man, and Cybernetics, **28**, 3, 459–467.
46. SKULIMOWSKI P., BUJACZ M., STRUMILLO P. (2009), *Detection and parameter estimation of objects in a 3D scene image*, Processing & Communications Challenges, **1**, pp. 308–316, Academy Publ. House EXIT, Warsaw.
47. STRUMILLO P., SZAJERMAN D., PELCZYNSKI P., MATERKA A. (2009), *Implementation of stereo matching algorithms on graphics processing units*, Image Processing and Communications Challenges, pp. 286–293, Bydgoszcz, Poland.
48. SZCZYPINSKI P., PELCZYNSKI P., SZAJERMAN D., STRUMILLO P. (2010) *Implementation of computer vision algorithms in DirectShow technology*, Image Processing and Communications Challenges 2, Advances in Intelligent and Soft Computing, **84**, 31–38.
49. vOICE (2015), www.seeingwithsound.com (last accessed: May 2015).
50. VISELL Y. (2009), *Tactile sensory substitution: models for enactment in HCI*, Interacting with Computers, **21**, 1–2, 38–53.
51. ZHIGANG F., TING L. (2010), *Audification-based electronic travel aid system*, IEEE International Conference on Computer Design and Applications (ICDDA 2010), pp. 137–141, Qinhuangdao, China.