Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 3918284, 17 pages
https://doi.org/10.1155/2018/3918284

Review Article

Current Use and Future Perspectives of Spatial Audio Technologies in Electronic Travel Aids

Simone Spagnol,[1,2] **György Wersényi**,[3] **Michał Bujacz**,[4] **Oana Bălan**,[5] **Marcelo Herrera Martínez**,[2,6] **Alin Moldoveanu**,[5] **and Runar Unnthorsson**[2]

[1]*Department of Information Engineering, University of Padova, Via Gradenigo 6B, 35131 Padova, Italy*
[2]*Department of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland, Reykjavík, Iceland*
[3]*Department of Telecommunications, Széchenyi István University, Győr, Hungary*
[4]*Institute of Electronics, Technical University of Lodz, Lodz, Poland*
[5]*Faculty of Automatic Control and Computers, Politehnica University of Bucharest, Bucharest, Romania*
[6]*Faculty of Engineering, University of San Buenaventura, Carrera 8H #172-20, Bogotá, Colombia*

Correspondence should be addressed to Simone Spagnol; spagnols@hi.is

Received 23 August 2017; Revised 23 January 2018; Accepted 12 February 2018; Published 21 March 2018

Academic Editor: Tao Gu

Electronic travel aids (ETAs) have been in focus since technology allowed designing relatively small, light, and mobile devices for assisting the visually impaired. Since visually impaired persons rely on spatial audio cues as their primary sense of orientation, providing an accurate virtual auditory representation of the environment is essential. This paper gives an overview of the current state of spatial audio technologies that can be incorporated in ETAs, with a focus on user requirements. Most currently available ETAs either fail to address user requirements or underestimate the potential of spatial sound itself, which may explain, among other reasons, why no single ETA has gained a widespread acceptance in the blind community. We believe there is ample space for applying the technologies presented in this paper, with the aim of progressively bridging the gap between accessibility and accuracy of spatial audio in ETAs.

1. Introduction

Spatial audio rendering techniques have various application areas ranging from personal entertainment, through teleconferencing systems, to real-time aviation environments [1]. They are also used in health care, for instance, in motor rehabilitation systems [2], electronic travel aids (ETAs, i.e., devices which aid in independent mobility through obstacle detection or help in orientation and navigation) [3], and other assistive technologies for visually impaired persons [4].

In the case of ETAs, the hardware has to be portable, lightweight, and user-friendly, allow for real-time operation, and be able to support long-term operation. All these issues put designers and developers to a challenge where state-of-the-art technology literally comes at hand in the form of high-tech mobile devices, smartphones, and so on. Furthermore, if ETAs are designed for the visually impaired (The term Electronic Travel Aid was born and is almost exclusively used to describe systems developed to help visually impaired persons with navigating their surroundings safely and efficiently. Nevertheless, visually impaired persons are not strictly the only group who might benefit from ETAs: for instance, non-visual interaction focused towards navigation is of interest to firefighters operating in smoke-filled buildings [5].), even more aspects have to be considered. Beyond the aforementioned, the devices should have a special user interface as well as alternative input and output solutions, where feedback in the form of sound can enhance the functionality of the device. Most of the developments of ETAs for the visually impaired aim at safety during navigation, such as avoiding obstacles, recognizing objects, and extending the auditory information by spatial cues [6, 7]. Since visually impaired persons rely on spatial audio cues as their primary sense of orientation [8], providing them with an accurate virtual auditory representation of the environment is essential.

ETAs evolved considerably over the past years, and a variety of virtual auditory displays [9] were proposed, using different spatial sound techniques and sonification approaches, as well as basic auditory icons, earcons, and speech [10]. Available ETAs for the visually impaired provide various information that ranges from simple obstacle detection with a single range-finding sensor, to more advanced feedback employing data generated from visual representations of the scenes, acquired through camera technologies. The auditory outputs of such systems range from simple binary alerts indicating the presence of an obstacle in the range of a sensor, to complex spatial sound patterns aiming at sensory substitution and carrying almost as much information as a graphical image [7, 11].

A division can also be made between *local mobility aids* (environmental imagers or obstacle detectors, with visual or ranging sensors) that present only the nearest surroundings to the blind traveler and *navigation aids* (usually GPS- or beacon-based) that provide information on path waypoints [12] or geographical points of interest [13]. While the latter group focuses on directions towards the next waypoint, meaning that a limited spatial sound rendering could be used (e.g., just presenting sounds in the horizontal plane) [14], the former group primarily provides information on obstacles (or the lack of them) and near scene layouts (e.g., walls and shorelines), supporting an accurate spatial representation of the scene [6].

Nevertheless, most of these systems are still in their infancy and at a prototype stage. Moreover, no single electronic assistive device has gained a widespread acceptance in the blind community, for different reasons: limited functionalities, ergonomics, small scientific/technological value, limited end-user involvement, high cost, and potential lack of commercial/corporate interest in pushing high-quality electronic travel aids [3].

While many excellent recent reviews on ETA solutions are available (see, e.g., [3, 4, 6, 7]), to our knowledge none of these works critically discusses or analyzes in depth the important aspect of spatial audio delivery. This paper gives an overview about existing solutions for delivering spatial sound, focusing on wearable technologies suitable for use in electronic travel aids for the visually impaired. The analysis reported in this paper indicates a significant potential to achieve accurate spatial sound rendering through state-of-the-art audio playback devices suitable for visually impaired persons and advances in customization of virtual auditory displays. This review was carried out within the European Horizon 2020 project named Sound of Vision (http://www.soundofvision.net). Sound of Vision focuses on creating an ETA for the blind that translates 3D environment models, acquired in real-time, into their corresponding real-time auditory and haptic representations [15].

The remainder of the paper is organized as follows. Section 2 reviews the basics of 3D sound localization, with a final focus on blind localization. Section 3 introduces the available state-of-the-art software solutions for customized binaural sound rendering, while Section 4 presents the available state-of-the-art hardware solutions suitable for the visually impaired. Finally, in Section 5 we discuss current uses and future perspectives of spatial audio in ETAs.

2. Basics of 3D Sound Localization

Localizing a sound source means determining the location of the sound's point of origin in the three-dimensional sound space [16]. Location is defined according to a head-related coordinate system, for instance, the interaural polar system. In the interaural polar coordinate system the origin coincides with the interaural midpoint and the elevation angle ϕ goes from -180° to 180° with negative values below the horizontal plane and positive values above, while the azimuth angle θ ranges from -90° at the left ear to 90° at the right ear. The third dimension, distance r, is the Euclidean distance between the sound source and the origin. In the following we will refer to the three planes that divide the head into halves as the *horizontal* plane (upper/lower halves), the *median* plane (left/right halves), and the *frontal* plane (front/back halves).

Spatial cues for sound localization can be categorized according to polar coordinates. As a matter of fact, each coordinate is thought to have one or more dominant cues in a certain frequency range associated with a specific body component, in particular the following:

(i) Azimuth and distance cues at all frequencies are associated with the head.

(ii) Elevation cues at high frequencies are associated with the pinnae.

(iii) Elevation cues at low frequencies are associated with torso and shoulders.

Based on well-known concepts and results, the most relevant cues for sound localization are now discussed [17].

2.1. Azimuth Cues. At the beginning of the twentieth century, Lord Rayleigh studied the means through which a listener is able to discriminate at a first level the horizontal direction of an incoming sound wave. Following his Duplex Theory of Localization [18], azimuth cues can be reduced to two basic quantities thanks to the active role of the head in the differentiation of incoming sound waves, that is, the following:

(i) *Interaural Time Difference* (ITD), defined as the temporal delay between sound waves at the two ears

(ii) *Interaural Level Difference* (ILD), defined as the ratio between the instantaneous amplitudes of the same two sounds.

ITD is known to be frequency-independent below 500 Hz and above 3 kHz, with an approximate ratio of low-frequency ITD by high-frequency ITD of 3/2, and slightly variable at middle range frequencies [19]. Conversely, frequency-dependent shadowing and diffraction effects introduced by the human head cause ILD to greatly depend on frequency.

Consider a low-frequency sinusoidal signal (up to 1 kHz approximately). Since its wavelength is greater than the head dimensions, ITD is no more than a phase lag $\Delta\phi < 2\pi$ between the signals arriving at the ears and therefore a reliable cue for horizontal perception in the low-frequency range [16]. Conversely, the considerable shielding effect of the human

head on high-frequency waves (above 1 kHz) makes ILD the most relevant cue in such spectral range.

Still, the information provided by ITD and ILD can be ambiguous. If one assumes a spherical geometry of the human head, a sound source located in front of the listener at azimuth θ and a second one located at the rear, at azimuth $180 - \theta$, provide in theory identical ITD and ILD values. In practice, ITD and ILD will not be identical at these two azimuth angles because the human head is clearly not spherical, and all subjects exhibit slight asymmetries with respect to the median plane. Nonetheless their values will be very similar, and *front-back confusion* is in fact often observed experimentally [20]: listeners erroneously locate sources at the rear instead of the front (or less frequently, vice versa).

2.2. Elevation Cues. Directional hearing in the median vertical plane is known to have lower resolution compared with that in the horizontal plane [21]. For the record, the smallest change of position of a sound source producing a just-noticeable change of position of the auditory event (known as "localization blur") along the median plane was found to be never less than 4°, reaching a much larger threshold (≈17°) for unfamiliar speech sounds, as opposed to a localization blur of approximately 1°-2° in the frontal part of the horizontal plane for a vast class of sounds [16]. Such a poor resolution is due to

(i) the need of high-frequency content (above 4-5 kHz) for accurate vertical localization [22, 23];

(ii) mild interaural differences between the signals arriving at the left and right ear for sources in the median plane.

If a source is located outside the horizontal plane, ITD- and ILD-based localization becomes problematic. As a matter of fact, sound sources located at all possible points of a conic surface pointing towards the ear of a spherical head produce the same ITD and ILD values. These surfaces, which generalize the aforementioned concept of front-back confusion for elevation angles, are known as *cones of confusion* and represent a potential difficulty for accurate perception of sound direction.

Nonetheless, it is undisputed that vertical localization ability is brought by the presence of the pinnae [24]. Even though localization in any plane involves pinna cavities of both ears [25], determination of the perceived vertical angle of a sound source in the median plane is essentially a monaural process [26]. The external ear plays an important role by introducing peaks and notches in the high-frequency spectrum of the incoming sound, whose center frequency, amplitude, and bandwidth greatly depend on the elevation angle of the sound source [27, 28], to a remarkably minor extent on azimuth [29], and are almost independent of distance between source and listener beyond a few centimeters from the ear [30, 31]. Such spectral effects are physically due to reflections on pinna edges as well as resonances and diffraction inside pinna cavities [26, 29, 32].

In general, both pinna peaks and notches are thought to play an important function in vertical localization of a sound source [33, 34]. Contrary to notches, peaks alone are not sufficient vertical localization cues [35]; however, the addition of spectral peaks supports the improvement of localization performance at upper directions with respect to notches alone [36]. It is also generally considered that a sound source has to contain substantial energy in the high-frequency range for accurate judgement of elevation, because wavelengths significantly longer than the size of the pinna are not affected. Since wavelength λ and frequency f are related as $\lambda = c/f$ (Here c is the speed of sound, typically $c = 343.2$ m/s in dry air at 20°C.), we could roughly state that pinnae have relatively little effect below $f = 3$ kHz, corresponding to an acoustic wavelength of $\lambda \approx 11$ cm.

While the role of the pinna in vertical localization has been extensively studied, the role of torso and shoulders is less understood. Their effects are relatively weak if compared to those due to the head and pinnae, and experiments to establish the perceptual importance of the relative cues have produced mixed results in general [23, 37, 38]. Shoulders disturb incident sound waves at frequencies lower than those affected by the pinna by providing a major additional reflection, whose delay is proportional to the distance from the ear to the shoulder when the sound source is directly above the listener. Complementarily, the torso introduces a shadowing effect for sound waves coming from below. Torso and shoulders are also commonly seen to perturb low-frequency ITD, even though it is questionable whether they may help in resolving localization ambiguities on a cone of confusion [39].

However, as Algazi et al. remarked [38], when a signal is low-passed below 3 kHz, elevation judgement is very poor in the median plane if compared to a broadband source but proportionally improves as the source is progressively moved away from the median plane, where performance is more accurate in the back than in the front. This result suggests the existence of low-frequency cues for elevation that although being overall weak is significant away from the median plane.

2.3. Distance and Dynamic Cues. Distance estimation of a sound source (see [40] for a comprehensive review on the topic) is even more troublesome than elevation perception. At a first level, when no other cue is available, sound intensity is the first variable that is taken into account: the weaker the intensity is, the farther the source should be perceived. Under anechoic conditions, sound intensity reduction with increasing distance can be predicted through the inverse square law: intensity of an omnidirectional sound source will decay by approximately 6 dB for each doubling distance [41]. Still, a distant blast and a whisper at few centimeters from the ear could produce the same sound pressure level at the eardrum. Having a certain familiarity with the involved sound is thus a second fundamental requirement [42].

However, the apparent distance of a sound source is systematically underestimated in an anechoic environment [43]. On the other hand, if the environment is reverberant, additional information can be given by the direct to reflected energy ratio, or DRR, which functions as a stronger cue for distance than intensity: a sensation of changing distance occurs if the overall intensity is constant but the DRR is altered [41]. Furthermore, distance-dependent spectral effects also

have a role in everyday environments: higher frequencies are increasingly attenuated with distance due to air absorption effects.

Literature on source direction perception is generally based on a fundamental assumption; that is, the sound source is sufficiently far from the listener. In particular, previously discussed azimuth and elevation cues are distance-independent when the source is in the so-called *far-field* (approximately more than 1.5 m from the center of the head) where sound waves reaching the listener can be assumed to be planar. On the other hand, when the source is in the *near field* some of the previously discussed cues exhibit a clear dependence on distance. By gradually approaching the sound source to the listener's head in the near field, it was observed that low-frequency gain is emphasized; ITD slightly increases; and ILD dramatically increases across the whole spectrum for lateral sources [20, 30, 44]. The following conclusions were drawn:

(i) Elevation-dependent features are not correlated to distance-dependent features.

(ii) ITD is roughly independent of distance even when the source is close.

(iii) Low-frequency ILDs are the dominant auditory distance cues in the near field.

It should be then clear that ILD-related information needs to be considered in the near field, where dependence on distance cannot be approximated by a simple inverse square law.

Finally, it has to be remarked that, switching from a static to a dynamic environment where the source and/or the listener move with respect to each other, both source direction and distance perception improve. The tendency to point towards the sound source in order to minimize interaural differences, even without visual aid, is commonly seen and aids in disambiguating front/back confusion [45]. Active motion helps especially in azimuth estimation and to a lesser extent in elevation estimation [46]. Furthermore, thanks to the *motion parallax* effect, slight translations of the listener's head on the horizontal plane can help discriminate source distance [47, 48]: if the source is near, its angular direction will drastically change after the translation (reflecting itself onto interaural differences), while for a distant source this will not happen.

2.4. Sound Source Externalization. Real sound sources are typically *externalized*, that is, perceived to be located outside our own head. However, when virtual 3D sound sources are presented through headphones (see next section), in-the-head localization may typically occur and have a major impact on localization ability. Alternatively, listeners may perceive the direction of the sound source and be able to make accurate localization judgements yet accompanied with perception of the source being way closer to the head than otherwise intended (e.g., on the surface of the skull [49]). However, when relevant constraints are taken into account, such as the use of individually measured head-related transfer functions as explained in Section 3, virtual sound sources can be externalized almost as efficiently as real sound sources [50, 51]. Externalization is, along with other attributes such as coloration, immersion, and realism, one of the key perceptual attributes that go beyond the basic issue of localization recently proposed for the evaluation of virtually rendered sound sources [52].

In-the-head localization is mainly introduced by the loss of accuracy in interaural level differences and spectral profiles in virtually rendered sound sources [49]. Another extremely important factor is given by the interaural and spectral changes triggered by natural head movements in real-life situations: correctly tracked head movements can indeed substantially enhance externalization in virtual sonic environments, especially for sources close to the median plane (hardest to externalize statically in anechoic conditions, due to minimal interaural differences [53]), and even relatively small movements of a few degrees can efficiently reduce in-the-head localization [54]. Furthermore, it has been recently showed that externalization can persist once coherent head movement with the virtual auditory space is stopped [55].

Finally, factors related to sound reverberation contribute to a strong sense of externalization, as opposed to dry anechoic sound. The introduction of artificial reverberation [56] through image-source model-based early reflections, wall and air absorption, and late reverberation can significantly contribute to sound image externalization in headphone-based 3D audio systems [57], as well as congruence between the real listening room and the virtually recreated reverberating environment [58].

2.5. Auditory Localization by the Visually Impaired. A number of previous studies showed that sound source localization by visually impaired persons can be different from that of sighted persons. It has to be first highlighted that previous investigations on visually impaired subjects indicated neither better auditory sensitivity [59–61] nor lower auditory hearing thresholds [62] compared to normally sighted subjects. On the other hand, visually impaired subjects acquire the ability to use auditory information more efficiently thanks to the plasticity of the central nervous system, as, for instance, in speech discrimination [63], temporal resolution [64], or spatial tuning [65].

Experiments with real sound sources suggest that visually impaired (especially early blind) subjects map the auditory environment with equal or better accuracy than sighted subjects on the horizontal plane [62, 66–68] but are less accurate in detecting elevation [67] and show an overly compressed auditory distance perception beyond the near field [69]. However, unlike sighted subjects, visually impaired subjects can correctly localize sounds monaurally [66, 70], which suggests a trade-off in the localization proficiency between the horizontal and median planes taking place [71]. By comparing behavioral and electrophysiological indices of spatial tuning within the central and peripheral auditory space in congenitally blind and normally sighted but blindfolded adults, it was found that blind participants displayed localization abilities that were superior to those of sighted controls, but only when attending to sounds in peripheral auditory space [72]. Still, it has to be taken into account that early blind subjects have no possibility of learning the mapping between auditory events and visual stimuli [73].

While localizing, adapting to the coloration of the signals is a relevant component for both sighted and blind subjects. Improved obstacle sense of the blind is also mainly due to enhanced sensitivity to echo cues [74], which allows so-called echolocation [75, 76]. Thanks to this obstacle sensing ability, which can be improved by training, distance perception in blind subjects may be enhanced [68, 76–78]. In addition, some blind subjects are able to determine size, shape, or even texture of obstacles based on auditory cues [70, 77, 79, 80].

Switching to virtual auditory displays, that is, the focus of this paper, a detailed comparative evaluation of blind and sighted subjects [81] confirmed some of the previously discussed results in the literature on localization with real sound sources. Better performance in localizing static frontal sources was obtained in the blind group due to a decreased number of front-back reversals. In the case of moving sources, blind subjects were more accurate in determining movements around the head in the horizontal plane. Sighted participants, however, performed better during listening to ascending movements in the median plane and in identifying sound sources in the back. In-the-head localization rates and the ability to detect descending movements were almost identical for the two groups. In a further experiment [82] error rates of about 6 to 14 degrees horizontally and 9 to 24 degrees vertically were measured for a pool of blind subjects. Improvements in localization by blind persons were observed mainly in the horizontal plane and in case of a broadband stimulus.

Finally, although visual information corresponding to auditory information significantly aids localization and creation of correct spatial mental mappings, it has to be remarked that visually impaired subjects can benefit from off-site representations in order to gain spatial knowledge of a real environment. For instance, results of recent studies showed that interactive exploration of virtual acoustic spaces [83–85] and audio-tactile maps [86] can provide relevant information for the construction of coherent spatial mental maps of a real environment in blind subjects and that such mental representations preserve topological and metric properties, with performances comparable or even superior to an actual navigation experience.

3. Binaural Technique

The most basic method for simulating sound source direction over loudspeakers is to use panning. This usually refers to amplitude panning using two channels (stereo panning). In this case, only level information is used as a balance between the channels, and the virtual source is shifted towards the louder channel. However, ILD and spectral cues are determined by the actual speaker locations. In traditional stereo setups, where loudspeakers and listener form a triangle, sources can be correctly simulated on the line ideally connecting the two speakers. However, although traditional headphones also use two channels, correct directional information is not maintained due to a different arrangement of the speakers with respect to the listener and by the loss of crosstalk between the channels.

Spatial features of virtual sound sources can be more realistically rendered through headphones by processing an input sound with a pair of filters, each simulating all the linear transformations undergone by the acoustic signal during its path from the sound source to the corresponding listener's eardrum. These filters are known in the literature as head-related transfer functions (HRTFs) [87], formally defined as the frequency-dependent ratio between the sound pressure level (SPL) $\Phi(\theta, \phi, \omega)$ at the eardrum and the free field SPL at the center of the head $\Phi_f(\omega)$ as if the listeners were absent:

$$H(\theta, \phi, \omega) = \frac{\Phi(\theta, \phi, \omega)}{\Phi_f(\omega)}, \quad (1)$$

where (θ, ϕ) indicates the angular position of the source relative to the listener and ω is angular frequency. The HRTF contains all of the information relative to sound transformations caused by the human body, in particular by the head, external ears, torso, and shoulders.

HRTF measurements are typically conducted in large anechoic rooms. Usually, a set of loudspeakers is arranged around the subject, pointing towards him/her and spanning an imaginary spherical surface. The listener is positioned so that the center of the interaural axis coincides with the center of the sphere defined by the loudspeakers and their rotation (or, equivalently, the subject's rotation). A probe microphone is inserted into each ear, either at the entrance or inside the ear canal. The measurement technique consists in recording and storing the signal arriving at the microphones. Consequently, these signals are processed in order to remove the effects of the room and the recording equipment (especially speakers and microphones), leaving only the HRTF [87, 88].

By processing a desired monophonic sound signal with a pair of individual HRTFs, one per channel, and by adequately accounting for headphone-induced spectral coloration (see next Section), authentic 3D sound experiences can take place. Virtual sound sources created with individual HRTFs can be localized almost as accurately as real sources and efficiently externalized [50], provided that head movements can be made and that the sound is sufficiently long [89]. As a matter of fact, localization of short broadband sounds without head movements is less accurate for virtual sources than for real sources, especially in regard to vertical localization accuracy [90], and front/back reversal rates are higher for virtual sources [89].

Unfortunately, the individual HRTF measurement technique requires the use of dedicated research facilities. Furthermore, the process can take up to several hours, depending on the used measurement system and on the desired spatial grid density, being uncomfortable and tedious for subjects. As a consequence, most practical applications use nonindividual (or generic) HRTFs, for instance, measured on *dummy heads*, that is, mannequins constructed from average anthropometric measurements. Several generic HRTF sets are available online. The most popular are based on measurements using the KEMAR mannequin [91] or the Neumann KU-100 dummy head (see the Club Fritz study [92]). Alternatively, an HRTF set can be taken from one of many public databases of individual measurements (see, e.g., [93]); many of these databases were recently unified in a common HRTF format

known as Spatially Oriented Format for Acoustics (SOFA) (https://www.sofaconventions.org/).

On the other hand, while nonindividual HRTFs represent the cheapest means of providing 3D perception in headphone reproduction, especially in the horizontal plane [94, 95], listening to nonindividual spatial sounds is more likely to result in evident sound localization errors such as incorrect perception of source elevation, front-back reversals, and lack of externalization [96] that cannot be fully counterbalanced by additional spectral cues, especially in static conditions [46]. In particular, individual elevation cues cannot be characterized through generic spectral features.

For the above reasons, different alternative approaches towards HRTF-based synthesis were proposed throughout the last decades [37, 97]. These are now reviewed and presented sorted by increasing level of customization.

3.1. HRTF Selection Techniques. HRTF selection techniques typically use specific criteria in order to choose the best HRTF set for a particular user from a database. Seeber and Fastl [98] proposed a procedure according to which one HRTF set is selected based on multiple criteria such as spatial perception, directional impression, and externalization. Zotkin et al. [99] selected the HRTF set that best matched an anthropometric data vector of the pinna. Geronazzo et al. [100] and Iida et al. [101] selected the HRTF set whose extracted pinna notch frequencies were closest to the hypothesized frequencies of the user according to a reflection model and an anthropometric regression model, respectively.

Similarly, selection can be targeted at detecting a subset of HRTFs in a database that fit the majority of a pool of listeners. Such an approach was pursued, for example, by So et al. [102] through cluster analysis and by Katz and Parseihian [103] through subjective ratings. The choice of the personal best HRTF among this reduced set is left to the user. Even different selection approaches were undertaken by Hwang et al. [104] and Shin and Park [105]. They modeled HRIRs on the median plane as linear combinations of basis functions whose weights were then interactively self-tuned by the listeners themselves.

Results of localization tests included in the majority of these works show a general decrease of the average localization error as well as of the front/back reversal and inside-the-head localization rates using selected HRTFs rather than generic HRTFs.

3.2. Analytical Solutions. These methods try to find a mathematical solution for the HRTF, taking into account the size and shape of the head and torso in particular. The most recurring head model in the literature is that of a rigid sphere, where the response related to a fixed observation point on the sphere's surface can be described by means of an analytical transfer function [106]. Brown and Duda [37] proposed a first-order approximation of this transfer function for sources in the far-field as a minimum-phase analog filter. Near-field distance dependence can be accounted for through an additional filter structure [107].

Although the spherical head model provides a satisfactory approximation to the low-frequency magnitude of a measured HRTF [108], it is far less accurate in predicting ITD, which is actually variable around a cone of confusion by as much as 18% of the maximum interaural delay [109]. ITD estimation accuracy can be improved by considering an ellipsoidal head model that can account for the ITD variation and be adapted to individual listeners [110]. It has to be highlighted, however, that ITD estimation from HRTFs is a nontrivial operation, given the large variability of objective and perceptual ITD results produced by different common calculation methods for the same HRTF dataset [111, 112].

A spherical model can also approximate the contribution of the torso to the HRTF. Coaxial superposition of two spheres of different radii, separated by a distance accounting for the neck, results in the snowman model [113]. The far-field behavior of the snowman model was studied in the frontal plane both by direct measurements on two rigid spheres and by computation through multipole reexpansion [114]. A filter model was also derived from the snowman model [113]; its structure distinguishes the two cases where the torso acts as a reflector or as a shadower, switching between the two filter substructures as soon as the source enters or leaves the torso shadow zone, respectively. Additionally, an ellipsoidal model for the torso was studied in combination with the usual spherical head [38]. Such model is able to account for different torso reflection patterns; listening tests confirmed that this approximation and the corresponding measured HRTF gave similar results, showing larger correlations away from the median plane.

A drawback of these techniques is that since they do not consider the contribution of the pinna, the generated HRTFs match measured HRTFs at low frequencies only, lacking spectral features at higher frequencies [115].

3.3. Structural HRTF Models. According to the structural modeling approach, the contributions to the HRTF of the user's head, pinnae, torso, and shoulders, each accounting for some well-defined physical phenomena, are treated separately and modeled with a corresponding filtering element [37]. The global HRTF model is then constructed by combining all the considered effects [116]. Structural modeling opens to an interesting form of content adaptation to the user's anthropometry, since parameters of the rendering blocks can be estimated from physical data, fitted, and finally related to anthropometric measurements.

Structural models typically assume a spherical or ellipsoidal geometry for both the head and torso, as discussed in the previous subsection. Effective customizations of the spherical head radius given the head dimensions were proposed [117, 118], resulting in a close agreement with experimental ITDs and ILDs, respectively. Alternatively, ITD can be synthesized separately using individual morphological data [119]. An ellipsoidal torso can also be easily customized for a specific subject by directly defining control points for its three axes on the subject's torso [114]. Furthermore, a great variety of pinna models is available in the literature, ranging from simple reflection models [120] and geometric models [121] to more complex physical models that treat the pinna either as a configuration of cavities [122] or as a reflecting surface [29]. Structural models of the pinna, simulating its resonant and reflective behaviors in two separate filter blocks, were also proposed [123–125].

Algazi et al. [93] suggested using a number of one-dimensional anthropometric measurements for HRTF fitting through regression methods or other machine learning techniques. This approach was recently pursued in a number of studies [126–129] investigating the correspondence between anthropometric parameters and HRTF shape. When suitable processing is performed on HRTFs, clear relations with anthropometry emerge. For instance, Middlebrooks [130] reported a correlation between pinna size and center frequencies of HRTF peaks and notches and argued that similarly shaped ears that differ in size just by a scale factor produce similarly shaped HRTFs that are scaled in frequency. Further evidence of the correspondence between pinna shape and HRTF peaks [123, 131, 132] and notches [125, 133, 134] is provided in a number of following works. The use of such knowledge leads to the effective parametrization of structural pinna models based on anthropometric parameters, which suggests an improvement in median plane localization with respect to generic HRTFs [135, 136].

3.4. Numerical HRTF Simulations. Numerical methods typically require as input a 3D mesh of the subject, in particular the head and torso, and include approaches such as finite-difference time domain (FDTD) methods [108], the finite element method (FEM) [137], and the boundary element method (BEM) [138].

Recent literature has focused on the BEM. It is known that high-resolution meshes are needed in order to effectively simulate HRTFs with the BEM, especially for the pinna area. Low mesh resolution results indeed in simulated HRTFs that greatly differ from acoustically measured HRTFs at high frequencies, thus destroying elevation cues [139]. However, as the number of mesh elements grows, memory requirements and computational load grow even faster [140]. Recent works introduced the fast multipole method (FMM) and the reciprocity principle (i.e., interchanging sources and receivers) in order to face BEM efficiency issues [140, 141]. Ultimately, localization performances of simulated HRTFs through the BEM were found to be similar to those observed with acoustically measured HRTFs [142], and databases of simulated HRTFs [143] as well as open-source tools for calculating HRTFs through the BEM given a head mesh as input [144] are available online.

On the other hand, image-based 3D modeling, based on the reconstruction of 3D geometry from a set of user pictures, is a fast and cost-effective alternative to obtaining mesh models [145]. Furthermore, the advent of consumer level depth cameras and the availability of huge computational power on consumer computers open new perspectives towards very cheap and yet very accurate calculation of individualized HRTFs.

4. Headphone Technologies

One of the crucial variables for generating HRTF-based binaural audio is the headphone itself. Headphones are of different types (e.g., circumaural, supra-aural, extra-aural, and in-ear) and can have transfer functions that are far from linear. The main issue with classic headphones is that the transfer function between headphone and eardrum heavily varies from person to person and with small displacements of the headphone itself [146, 147]. Such variation is particularly marked in the high-frequency range where important elevation cues generally lie. As a consequence, headphone playback introduces significant localization errors, such as in-the-head localization, front-back confusion, and elevation shift [148].

In order to preserve the relevant localization cues provided by HRTF filtering during headphone listening, various headphone equalization techniques, usually based on a prefiltering with the inverse of the average headphone transfer function, are used [149]. However, previous research suggests that these techniques are little to no effective when nonindividual (even selected) HRTFs are used [149, 150]. On the other hand, several authors support the use of individual headphone compensation in order to preserve localization cues in the high-frequency range [146, 147].

In the case of travel aids for the visually impaired, additional factors need to be considered in the design and choice of the headphone type. Most importantly, ears are essential to provide information about the environment, and visually impaired persons refuse to use headphones during navigation if these either partially or fully cover the ears, therefore blocking environmental noises. The results of a survey of the preferences of visually impaired subjects for a possible personal navigation device [151] showed indeed that the majority of participants rated headphones worn over the ears as the least acceptable output device, compared to other technologies such as bone-conduction and small tube-like headphones, or even a single headphone worn over one ear. Furthermore, those fully blind had much stronger negative feelings about headphones that blocked ambient sounds than those who were partially sighted.

This important consideration shifts our focus to alternative state-of-the-art solutions for spatial audio delivery such as unconventional headphone configurations, bone-conduction headsets, or active transparent headsets.

4.1. Unconventional Headphone Configurations. The problem of ear occlusion can be tackled by decentralizing the point of sound delivery from the entrance of the ear canal to positions around the ear, with one or more transducers per ear. In this case, issues arise regarding the proper direction and distance of each transducer with respect to the ear canal, as well as their types and dimensions. Furthermore, there is a challenge in the spatial rendering technique in that no research results support the application of traditional loudspeaker-based spatial audio techniques (such as Vector Base Amplitude Panning [152] or Ambisonics [153]) to multispeaker headsets and that traditional HRTF measurements do not match with decentralized speaker positions.

The first attempts in delivering spatial audio through multispeaker headphones were performed by König. A decentralized 4-channel arrangement placed on a pair of circumaural earcups for frontal surround sound reproduction was implemented [154] (an alternative small supra-aural configuration was also proposed [155]). Results showed that this speaker arrangement induces individual direction-dependent pinna

cues as they appear in real frontal sound irradiation in the free field for frequencies above 1 kHz [156]. Psychoacoustic effects introduced by the headphone revealed that frontal auditory events are achieved, as well as effective distance perception [154].

The availability of individual pinna cues at the eardrum is imperative for accurate frontal localization [157]. Accordingly, Sunder et al. [158] later proposed the use of a 2-channel frontal projection headphone which customizes nonindividual HRTFs by introducing idiosyncratic pinna cues. Perceptual experiments validated the effectiveness of frontal headphone playback over conventional headphones with reduced front-back confusions and improved frontal localization. It was also observed that the individual spectral cues created by the frontal projection are self-sufficient for front-back discrimination even with the high-frequency pinna cues removed from the nonindividual HRTF. However, additional transducers are needed if virtual sounds behind the head have to be delivered, and timbre differences with respect to the frontal transducers need to be solved.

Greff and Katz [159] extended the above solutions to a multiple transducer array placed around each ear (8 speakers per ear) recreating the pinna-related component of the HRTF. Simulations and subjective evaluations showed that it is possible to excite the correct localization cues provided by the diffraction of the reconstructed wave front on the listener's own pinnae, using transducer driving filters related to a simple spherical head model. Furthermore, different speaker configurations were investigated in a preliminary localization test, the one with transducers placed at grazing incidence all around the pinna showing the best results in terms of vertical localization accuracy and front/back confusion rate.

Recently, Bujacz et al. [160] proposed a custom headphone solution for a prospective ETA with four proximaural speakers positioned above and below the ears, all slightly to the front. Amplitude panning was then used as spatial audio technique to shift the power of the output sound between pairs of speakers, both horizontally and vertically. Results of a preliminary localization test showed a localization accuracy comparable to HRTF-based rendering through high-quality circumaural headphones, both in azimuth and in elevation.

4.2. Bone-Conduction Headsets. The use of a binaural bone-conduction headset (also known as *bonephones*) is an extremely attractive solution for devices intended for the blind as the technology does not significantly interfere with sounds received through the ear canal, allowing for natural perception of environmental sounds. The typical solution is to place vibrational actuators, also referred to as bone-conduction transducers, on each mastoid (the raised portion of the temporal bone located directly behind the ear) or alternatively on the cheek bones just in front of the ears [161]. Pressure waves are sent through the bones in the skull to the cochlea, with some amount of natural sound leakage through air into the ear canals still occurring.

There are some difficulties in using bone conduction for delivering spatial audio. The first is the risk of crosstalk impeding an effective binaural separation: because of the high propagation speed and low attenuation of sound in the human skull, both the ITD and ILD cues are significantly softened. Walker et al. [162] still observed some degree of spatial separation with interaural cues provided through bone conduction and ear canals either free or occluded, especially relative to ILD. Perceived lateralization is even comparable between air conduction and bone conduction with unoccluded ear canals [163]. However, the degradation relative to standard headphones suggests the difficulty to produce large enough interaural differences to simulate sound sources at extreme lateral locations [162].

The second problem is the need to introduce additional transfer functions for correct equalization of HRTF-based spatial audio: the frequency response of the transducer [164] and the transfer function to the bones themselves, referred to as bone-conduction adjustment function (BAF) [165], which takes into account high-frequency attenuation by the skin [166] and differs between individuals, similar to HRTFs. Walker et al. [167, 168] proposed the use of appropriate bone-related transfer functions (BRTFs) in replacement of HRTFs. Stanley [165] derived individual BAFs from equal-loudness judgements on pure tones, showing that individual BAF adjustments to HRTF-based spatial sound delivery were effective in restoring the spectral cues altered by the bone-conduction pathway. This allowed for effective localization in the median plane by reducing up/down reversals with respect to the BAF-uncompensated stimuli. However, there is no way to measure BAFs empirically, and it is unclear whether the use of a generic, average BAF could lead to the same conclusions.

MacDonald et al. [164] reported similar localization results in the horizontal plane between bone conduction and air conduction, using individual HRTFs as the virtual auditory display and headphone frequency response compensation. Lindeman et al. [169, 170] compared localization accuracy between bone conduction with unoccluded ear canals and an array of speakers located around the listener. The results showed that although the best accuracy was achieved with the speaker array in the case of stationary sounds, there was no difference in accuracy between the speaker array and the bone-conduction device for sounds that were moving, and that both devices outperformed standard headphones for moving sounds.

Finally, Barde et al. [171] recently investigated the minimum discernable angle difference in the horizontal plane with nonindividual HRTFs over a bone-conduction headset, resulting in an average value of 10°. Interestingly, almost all participants reported actual sound externalization.

4.3. Active Transparent Headsets. An active headset is able to detect and process environmental sounds through analog circuits or digital signal processing. One of the most important fields of application of active headsets is noise reduction, where the headset uses active noise control [172, 173] to reduce unwanted sound by the addition of an antiphase signal to the output sound. In the case of ETAs, the environmental signal should not be canceled but provided back to the listener (*hear-through* signal) mixed with the virtual auditory display signal in order for the subject to be aware of the surroundings. Binaural hear-through headsets (in-ear headphones with integrated microphones) are typically used

in augmented reality audio (ARA) applications [174], where a combination of real and virtual auditory objects in a real environment is needed [175].

The hear-through signal is a processed version of the environmental sound and should produce similar auditory perception to natural perception with unoccluded ears. Thus, equalization is needed to make the headset acoustically transparent, since it affects the acoustic properties of the outer ear [176]. The most important problem here is poor fit on the head causing leaks and attenuation problems. The fit of the headphone affects isolation and frequency response as well. Using internal microphones inside the headset in addition to the external ones, a controlled adaptive equalization can be realized [177].

The second basic requirement for a hear-through system is that processing of the recorded sound should have minimal latency [175]. As a matter of fact, when the real signal (leaked to the eardrum) is summed up with the hear-through signal, the delayed version can cause audible comb-filtering effects, especially at lower frequencies where leakage is higher. The audibility of comb-filtering effects depends on both the time and amplitude difference between the hear-through signal and the leaked signal [178]. Using digital realizations, which are preferable over analog circuits in the case of an ETA in terms of both cost and size, suitable latencies of less than 1.4 ms, for which the comb-filtering effect was found to be inaudible when the attenuation of the headset is 20 dB or more, can be achieved with a DSP board [179].

Finally, the hear-through signal should preserve localization cues at the ear canal entrance. Since sound transmission from the microphone to the eardrum is independent of direction whether the microphone is inside or at most 6 mm outside the ear canal [180], having binaural microphones just outside the ear canal entrance is sufficient for obtaining the correct listener-dependent spatial information.

5. Spatial Audio in ETAs

From the multitude of ETAs, two main trends in selecting sound cues can be observed, one to provide very limited yet easily interpretable data, typically from a range sensor, and the other to provide an overabundance of auditory data and let the user learn to extract useful information from it (e.g., the vOICe [181]). A third approach, taken for instance by the authors in the Sound of Vision project [15], is to limit the data from a full-scene representation to just the most useful information, for example, by segmenting the environment and identifying the nearest obstacles or detecting special dangerous scene elements such as stairs. Surveys show that individual preferences among the blind can vary greatly, and all three approaches have users that prefer them [182].

In a recent literature review, Bujacz and Strumiłło [6] classified the auditory display solutions implemented in the most widely known ETAs, either commercially available or in various stages of research and development. Of the 22 considered ETAs, 12 use a spatial representation of the environment. However, breaking the list of ETAs down to obstacle detectors (mostly hand-held) and environmental imagers (mostly head-mounted), ETAs that use a spatial representation almost all belong to the second category. Some of them, such as the vOICe [181], Navbelt [183], SVETA [184], and AudioGuider [185], use stereo panning to represent directions, whereas elevation information is either ignored or coded into sound pitch. ETAs (including works not included in the above cited review) that use HRTFs as the spatial rendering method are now summarized. All of the systems presented in the following are laboratory prototypes.

5.1. Available ETAs Using HRTFs. The *EAV (Espacio Acustico Virtual)* system [186] uses stereoscopic cameras to create a low resolution ($16 \times 16 \times 16$) 3D stereopixel map of the environment in front of the user. Each occupied stereopixel becomes a virtual sound source filtered with the user's individual HRTFs, measured in a reverberating environment. The sonification technique employs spatial audio cues (synthesized with HRTFs) and a distance-to-loudness encoding. Sounds were presented through a pair of individually equalized Sennheiser HD-580 circumaural headphones. Classic localization tests with the above virtual auditory display and tests with multiple sources were performed on 6 blind and 6 normally sighted subjects. Subjects were accurate in identifying the objects' position and recognizing shapes and dimensions within the limits imposed by the system's resolution.

The *cross-modal ETA* device [187] is a wearable prototype that consists of low-cost hardware: earphones (no further information provided), sunglasses fitted with two CMOS micro cameras, and a palm-top computer. The system is able to detect the light spot produced by a laser pointer, compute its angular position and depth, and generate a corresponding sound to the position and distance of the pointed surface. The sonification encoding uses directional auditory cues provided through Brown and Duda's structural HRTF model [37], and distance cues through loudness control and reverberation effects. The subjective effectiveness of the sonification technique was evaluated by several volunteers who were asked to use the system and report their opinions. The overall result was satisfactory, with some problems related to the lack of elevation perception. Targets very high and very low were perceived correctly, whereas those laying in the middle were associated with wrong elevations.

The *Personal Guidance System* [12] receives information from a GPS receiver and was evaluated in five different types of configurations involving different types of auditory displays, spatial sound delivery methods (either via classic headphones or through a speaker worn on the shoulder), and tracker locations. No details about the binaural spatialization engine or the headphones used were provided. Fifteen visually impaired subjects traveled a 50 m long pathway with each of the 5 configurations. Results showed that the configuration using binaurally spatialized virtual speech led to the shortest travel times and highest subjective ratings. However, there were many negative comments about the headphones blocking environmental sounds.

The SWAN system [8, 188] aids navigation and guidance through a set of navigation beacons (earcon-like sounds), object-related sounds (provided through spatial auditory icons), location information, and brief prerecorded speech

samples. Sounds are updated in real-time by tracking the subject's orientation and accordingly spatialized through nonindividual HRTFs. Sounds were played either through a pair of Sony MDR-7506 closed-ear headphones or an equalized bone-conduction headset (see [165]). In an experimental procedure, 108 sighted subjects were required to navigate three different maps. Results showed good navigation skills for almost all the participants in both time and path efficiency.

The main idea of the *Virtual Reality Simulator for the visually impaired people* [189] consists in calculating the distance between the user and nearby objects (depth map) and converting it into sound. The depth map is transformed into a spatial auditory map by using 3D sound cues synthesized with individually measured HRTFs from 1003 positions in the frontal field. Sounds were provided through a standard pair of stereophonic headphones (no further information provided). The Virtual Reality Simulator proved to be helpful for visually impaired people in different research experiments performed indoors and outdoors, in virtual and real-life situations. Among the main limitations of the simulator are tracking accuracy and the lack of a real-time HRTF convolver.

The *Real-Time Assistance Prototype* [190], an evolution of the CASBliP prototype [191], encodes objects' position in space based on their distance (inversely proportional to sound frequency), direction (3D binaural sounds synthesized with nonindividual HRTFs), and speed (proportional to pitch variation). Nonindividual HRTFs of a KEMAR mannequin were measured for different spatial points in a 64° azimuth range, a 30° elevation range, and a 15 m distance range. Sounds were provided through a pair of SONY MDR-EX75SL in-ear headphones. Two experiments were performed with four totally blind subjects, one requiring subjects to identify the sound direction and the other one to detect the position of a moving source and to follow. Despite providing encouraging results in static conditions for objects moving in the detected area, its main limitations reside in the inability to detect objects at ground level and in the reduced 64° field of view.

The NAVITON system [192, 193] processes stereo images to segment out key elements for auditory presentation. For each segmented element, the sonification approach uses discrete pitched sounds, whose pitch, loudness, and temporal delay (depth scanning) depend on object distance, and whose duration is proportional to the depth of the object. Sounds are spatialized with individual HRTFs, custom measured in the full azimuth range and in the vertical plane from −54° to 90°, in 5° steps. Sounds were provided through high-quality open-air reference headphones without headphone compensation. Ten blindfolded participants reported their auditory perception about the sonified virtual 3D scenes in a virtual reality trial, proving to be capable of grasping the general spatial structure of the environment and accurately estimate scene layouts. A real-world navigation scenario was also tested with 5 blind and 5 blindfolded volunteers, who could accurately estimate the spatial position of single obstacles or pairs of obstacles and walk through simple obstacle courses.

The *NAVIG (Navigation Assisted by Artificial VIsion and GNSS)* system [194, 195] aims to enhance mobility and orientation, navigation, object localization, and grasping, both indoors and outdoors. It uses a Global Navigation Satellite System (GNSS) and a rapid visual recognition algorithm. Navigation is ensured by real-time nonindividual HRTF-based rendering, text-to-speech, and semantic sonification metaphors that provide information about the trajectory, position, and the important landmarks in the environment. The 3D audio scenes are conveyed through a bone-conduction headset whose complex frequency response is equalized in order to properly render all the spectral cues of the HRTF. Preliminary experiments have shown that it is possible to design a wearable device that can provide fully analyzed information to the user. However, thorough evaluations of the NAVIG prototype have not been published yet.

5.2. Discussion and Conclusions. The use of HRTFs to code directional information in the above summarized ETAs suggests the importance of a high-fidelity spatial auditory representation of the environment for blind users. However, most of the above works fail to address the hardware- and/or software-related aspects we discussed in Sections 3 and 4, presenting results of performance and usability tests that are based on binaural audio rendering setups that either are ideal yet unrealistic (e.g., [186]) or underestimate the potential of spatial sound itself (e.g., [190]).

As a matter of fact, the preferred choice for the virtual auditory display within the 8 listed ETAs is either individually measured HRTFs or nonindividual, generic HRTFs. Only the cross-modal ETA [187] proposes the use of structural HRTF modeling as a trade-off between localization accuracy and measurement cost. As a result, the evaluation of these systems (often performed through proper localization performance tests) is based either on the best scoring yet unfeasible solution (individually measured HRTFs) or on a costless yet inaccurate one (generic HRTFs), overlooking important aspects in the fidelity of the virtual auditory display such as elevation accuracy and front/back confusion avoidance. Furthermore, the aforementioned monaural localization ability by visually impaired persons (especially early blind) suggests the use of individual pinna cues for azimuth perception, which would make a visually impaired person more vulnerable to degraded localization from nonindividual HRTFs than a sighted person.

Even more unfortunately, the headphones chosen for these tests were in the majority of cases classic circumaural or in-ear headphones that block environmental sounds and thus, as discussed before, are not acceptable for the visually impaired community. The use of a bone-conduction headset is reported only for the SWAN and NAVIG systems [188, 194], where the importance of headphone equalization, although forced to be nonindividual, is also stressed. None of the remaining works, except one [186], even mentions headphone equalization. Effective externalization of the virtual sounds provided to the users is therefore questionable.

It is difficult to rank the importance of the various factors influencing a satisfactory virtual acoustic experience (e.g., externalization, localization accuracy, and front-back confusion rate). Most studies check for only one or two factors and can confirm their influence on one or more spatial sound

perception parameters. Besides the choice of the HRTF set, headphone type, and equalization, and type of sound source (frequency content, familiar/unfamiliar sound, and temporal aspects) [16, 44, 196], other important factors have to be considered. For instance, as explained in Section 2.4, rendering environmental reflections increases externalization, as well as the use of a proper head-tracking method, which also helps in resolving front/back confusion [95]. This may be why most of the above cited studies chose to use high-quality headphones with generic or individual HRTFs, without applying headphone equalization as long as head-tracking or real-time obstacle tracking is implemented. It is also relevant to notice that those systems that use head-mounted cameras to render sounds at locations relative to current head orientation do not even strictly require head-tracking to work dynamically [197].

We believe there is ample space for applying the technologies presented in this review paper to the case of ETAs for the blind. Basic research in HRTF customization techniques is currently in a prolific stage, thanks to advances in computational power and the widespread availability of technologies such as 3D scanning and printing allowing researchers to investigate in detail the relation between individual anthropometry and HRTFs. Although a full and thorough understanding of the mechanisms involved in spatial sound perception still has to be reached, techniques such as HRTF selection, structural HRTF modeling, or HRTF simulations are expected to progressively bridge the gap between accessibility and accuracy of individual binaural audio.

Still it has to be noted that many experiments proved that subjective training to nonindividual HRTFs, especially through cross-modal and game-based training methods, can significantly reduce localization errors in both free field and virtual listening conditions [198]. Feedback can be provided through visual stimuli [199, 200], proprioceptive cues [201, 202], or haptic information [203]. Reductions in front-back confusion rates as large as 40% were reported, as well as improvements in sound localization accuracy in the horizontal and vertical planes regardless of head movement.

On the other hand, the headphone technologies discussed in Section 4 are expected to reach widespread popularity in the blind community. Bone-conduction and active headsets are growing in the consumer market thanks to their affordable price. External multispeaker headsets are still at a prototype stage but from a research point of view open the attractive possibility of introducing individualized binaural playback without the need of fully individual HRTFs. Efforts in the design of such headphones have been produced within the Sound of Vision project [160].

A final comment regards the cosmetic acceptability of the playback device. While bone-conduction and binaural headsets are relatively discreet and portable, external multispeaker headsets may require a bulky and unconventional design. There is considerable variation within the blind community when assessing the cosmetic acceptability of a wearable electronic device, even if it works well. Nevertheless, the visually impaired participants to the survey by Golledge et al. [151] showed overwhelming support for the idea of traveling more often with such a device, independently of its appearance.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement no. 643636.

References

[1] B. D. Simpson, D. S. Brungart, R. C. Dallman, J. Joffrion, M. D. Presnar, and R. H. Gilkey, "Spatial audio as a navigation aid and attitude indicator," in *Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society, HFES '05*, vol. 49, pp. 1602–1606, September 2005.

[2] F. Avanzini, S. Spagnol, A. Rodá, and A. De Götzen, "Designing interactive sound for motor rehabilitation tasks," in *Sonic Interaction Design*, K. Franinovic and S. Serafin, Eds., chapter 12, pp. 273–283, MIT Press, Massachusetts, Mass, USA, 2013.

[3] D. Dakopoulos and N. G. Bourbakis, "Wearable obstacle avoidance electronic travel aids for blind: a survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 40, no. 1, pp. 25–35, 2010.

[4] A. Bhowmick and S. M. Hazarika, "An insight into assistive technology for the visually impaired and blind people: State-of-the-art and future trends," *Journal on Multimodal User Interfaces*, vol. 2017, no. 2, pp. 1–24, 2017.

[5] D. I. Ahlmark, *Haptic navigation aids for the visually impaired [Ph.D. thesis]*, Lulea University of Technology, Lulea, Sweden, 2016.

[6] M. Bujacz and P. Strumiłło, "Sonification: review of auditory display solutions in electronic travel aids for the blind," *Archives of Acoustics*, vol. 41, no. 3, pp. 401–414, 2016.

[7] Á. Csapó, G. 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," *Journal on Multimodal User Interfaces*, vol. 9, no. 4, pp. 275–286, 2015.

[8] B. N. Walker and J. Lindsay, "Navigation performance with a virtual auditory display: effects of beacon sound, capture radius, and practice," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 48, no. 2, pp. 265–278, 2006.

[9] D. S. Brungart, "Near-field virtual audio displays," *Presence: Teleoperators and Virtual Environments*, vol. 11, no. 1, pp. 93–106, 2002.

[10] Á. Csapó and G. Wersényi, "Overview of auditory representations in human-machine interfaces," *ACM Computing Surveys*, vol. 46, no. 2, pp. 1–23, 2013.

[11] Á. Kristjánsson, A. Moldoveanu, Ó. I. Jóhannesson et al., "Designing sensory-substitution devices: principles, pitfalls and potential," *Restorative Neurology and Neuroscience*, vol. 34, no. 5, pp. 769–787, 2016.

[12] J. M. Loomis, J. R. Marston, R. G. Golledge, and R. L. Klatzky, "Personal guidance system for people with visual impairment: a comparison of spatial displays for route guidance," *Journal of Visual Impairment & Blindness*, vol. 99, no. 4, pp. 219–232, 2005.

[13] P. Skulimowski, P. Korbel, and P. Wawrzyniak, "POI explorer - A sonified mobile application aiding the visually impaired in

urban navigation," in *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems, FedCSIS '14*, pp. 969–976, September 2014.

[14] A. Garcia, V. Finomore, G. Burnett, A. Calvo, C. Baldwin, and C. Brill, "Evaluation of multimodal displays for waypoint navigation," in *Proceedings of the 2012 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support, CogSIMA '12*, pp. 134–137, IEEE, Louisiana, La, USA, March 2012.

[15] P. Strumiłło, M. Bujacz, P. Baranski et al., "Different approaches to aiding blind persons in mobility and navigation in the Naviton and Sound of Vision projects," in *Mobility of Visually Impaired People*, E. Pissaloux and R. Velazquez, Eds., pp. 435–468, Springer International Publishing, Cham, Switzerland, 2018.

[16] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, Massachusetts, Mass, USA, 2nd edition, 1996.

[17] S. Spagnol, *Techniques for customized binaural audio rendering with applications to virtual rehabilitation [Ph.D. thesis]*, University of Padova, Padova, Italy, 2012.

[18] J. W. Strutt, "On our perception of sound direction," *Philosophical Magazine*, vol. 13, no. 1907, pp. 214–232, 1907.

[19] G. F. Kuhn, "Model for the interaural time differences in the azimuthal plane," *The Journal of the Acoustical Society of America*, vol. 62, no. 1, pp. 157–167, 1977.

[20] D. S. Brungart, N. I. Durlach, and W. M. Rabinowitz, "Auditory localization of nearby sources. II. Localization of a broadband source," *The Journal of the Acoustical Society of America*, vol. 106, no. 4, pp. 1956–1968, 1999.

[21] A. Wilska, *Studies on Directional Hearing, English translation, Aalto University School of Science and Technology, Department of Signal Processing and Acoustics [Ph.D. thesis]*, University of Helsinki, Helsinki, Finland, 2010.

[22] J. Hebrank and D. Wright, "Spectral cues used in the localization of sound sources on the median plane," *The Journal of the Acoustical Society of America*, vol. 56, no. 6, pp. 1829–1834, 1974.

[23] F. Asano, Y. Suzuki, and T. Sone, "Role of spectral cues in median plane localization," *The Journal of the Acoustical Society of America*, vol. 88, no. 1, pp. 159–168, 1990.

[24] M. B. Gardner and R. S. Gardner, "Problem of localization in the median plane: effect of pinnae cavity occlusion," *The Journal of the Acoustical Society of America*, vol. 53, no. 2, pp. 400–408, 1973.

[25] M. Morimoto, "The contribution of two ears to the perception of vertical angle in sagittal planes," *The Journal of the Acoustical Society of America*, vol. 109, no. 4, pp. 1596–1603, 2001.

[26] J. Hebrank and D. Wright, "Are two ears necessary for localization of sound sources on the median plane?" *The Journal of the Acoustical Society of America*, vol. 56, no. 3, pp. 935–938, 1974.

[27] E. A. G. Shaw and R. Teranishi, "Sound pressure generated in an external-ear replica and real human ears by a nearby point source," *The Journal of the Acoustical Society of America*, vol. 44, no. 1, pp. 240–249, 1968.

[28] S. Spagnol, M. Hiipakka, and V. Pulkki, "A single-azimuth Pinna-Related Transfer Function database," in *Proceedings of the 14th International Conference on Digital Audio Effects, DAFx '11*, pp. 209–212, September 2011.

[29] E. A. Lopez-Poveda and R. Meddis, "A physical model of sound diffraction and reflections in the human concha," *The Journal of the Acoustical Society of America*, vol. 100, no. 5, pp. 3248–3259, 1996.

[30] D. S. Brungart and W. M. Rabinowitz, "Auditory localization of nearby sources. Head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1465–1479, 1999.

[31] S. Spagnol, "On distance dependence of pinna spectral patterns in head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 137, no. 1, pp. EL58–EL64, 2015.

[32] E. A. G. Shaw, "Acoustical features of human ear," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R. H. Gilkey and T. R. Anderson, Eds., pp. 25–47, Lawrence Erlbaum Associates, New Jersey, NJ, USA, 1997.

[33] B. C. J. Moore, S. R. Oldfield, and G. J. Dooley, "Detection and discrimination of spectral peaks and notches at 1 and 8 khz," *The Journal of the Acoustical Society of America*, vol. 85, no. 2, pp. 820–836, 1989.

[34] K. Iida, M. Itoh, A. Itagaki, and M. Morimoto, "Median plane localization using a parametric model of the head-related transfer function based on spectral cues," *Applied Acoustics*, vol. 68, no. 8, pp. 835–850, 2007.

[35] R. Greff and B. F. G. Katz, "Perceptual evaluation of HRTF notches versus peaks for vertical localisation," in *Proceedings of the 19th International Congress on Acoustics*, 2007.

[36] K. Iida and Y. Ishii, "Roles of spectral peaks and notches in the head-related transfer functions in the upper median plane for vertical localization," *The Journal of the Acoustical Society of America*, vol. 140, no. 4, pp. 2957-2957, 2016.

[37] C. P. Brown and R. O. Duda, "A structural model for binaural sound synthesis," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 6, no. 5, pp. 476–488, 1998.

[38] V. R. Algazi, C. Avendano, and R. O. Duda, "Elevation localization and head-related transfer function analysis at low frequencies," *The Journal of the Acoustical Society of America*, vol. 109, no. 3, pp. 1110–1122, 2001.

[39] O. Kirkeby, E. T. Seppala, A. Karkkainen, L. Karkkainen, and T. Huttunen, "Some effects of the torso on head-related transfer functions," in *Proceedings of the 122nd Audio Engineering Society Convention*, pp. 1045–1052, May 2007.

[40] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst, "Auditory distance perception in humans: a summary of past and present research," *Acta Acustica United with Acustica*, vol. 91, no. 3, pp. 409–420, 2005.

[41] D. R. Begault, *3-D Sound for Virtual Reality and Multimedia*, Academic Press Professional, Inc., Massachusetts, Mass, USA, 1994.

[42] M. B. Gardner, "Distance estimation of 0◦ or apparent 0◦-oriented speech signals in anechoic space," *The Journal of the Acoustical Society of America*, vol. 45, no. 1, pp. 47–53, 1969.

[43] D. H. Mershon and J. N. Bowers, "Absolute and relative cues for the auditory perception of egocentric distance," *Perception*, vol. 8, no. 3, pp. 311–322, 1979.

[44] D. S. Brungart, "Auditory localization of nearby sources. III. Stimulus effects," *The Journal of the Acoustical Society of America*, vol. 106, no. 6, pp. 3589–3602, 1999.

[45] F. L. Wightman and D. J. Kistler, "Resolution of front-back ambiguity in spatial hearing by listener and source movement," *The Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2841–2853, 1999.

[46] W. R. Thurlow and P. S. Runge, "Effect of induced head movements on localization of direction of sounds," *The Journal of the Acoustical Society of America*, vol. 42, no. 2, pp. 480–488, 1967.

[47] J. M. Speigle and J. M. Loomis, "Auditory distance perception by translating observers," in *Proceedings of the IEEE Research Properties in Virtual Reality Symposium*, pp. 92–99, IEEE, California, Calif, USA, 1993.

[48] M. Rebillat, X. Boutillon, E. T. Corteel, and B. F. G. Katz, "Audio, visual, and audio-visual egocentric distance perception by moving subjects in virtual environments," *ACM Transactions on Applied Perception*, vol. 9, no. 4, article no. 19, 2012.

[49] W. M. Hartmann and A. Wittenberg, "On the externalization of sound images," *The Journal of the Acoustical Society of America*, vol. 99, no. 6, pp. 3678–3688, 1996.

[50] G. Plenge, "On the differences between localization and lateralization," *The Journal of the Acoustical Society of America*, vol. 56, no. 3, pp. 944–951, 1974.

[51] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. II: psychophysical validation," *The Journal of the Acoustical Society of America*, vol. 85, no. 2, pp. 868–878, 1989.

[52] L. S. R. Simon, N. Zacharov, and B. F. G. Katz, "Perceptual attributes for the comparison of head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 140, no. 5, pp. 3623–3632, 2016.

[53] D. R. Begault and E. M. Wenzel, "Headphone localization of speech," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 35, no. 2, pp. 361–376, 1993.

[54] G. Wersényi, "Effect of emulated head-tracking for reducing localization errors in virtual audio simulation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 17, no. 2, pp. 247–252, 2009.

[55] E. Hendrickx, P. Stitt, J. Messonnier, J. Lyzwa, B. F. Katz, and C. de Boishéraud, "Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis," *The Journal of the Acoustical Society of America*, vol. 141, no. 3, pp. 2011–2023, 2017.

[56] V. Välimäki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "Fifty years of artificial reverberation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, no. 5, pp. 1421–1448, 2012.

[57] Y. Yuan, L. Xie, Z.-H. Fu, M. Xu, and Q. Cong, "Sound image externalization for headphone based real-time 3D audio," *Frontiers of Computer Science*, vol. 11, no. 3, pp. 419–428, 2017.

[58] S. Werner, G. Götz, and F. Klein, "Influence of head tracking on the externalization of auditory events at divergence between synthesized and listening room using a binaural headphone system," in *Proceedings of the 142nd Audio Engineering Society International Convention*, 2017.

[59] L. H. Benedetti and M. Loeb, "A comparison of auditory monitoring performance in blind subjects with that of sighted subjects in light and dark," *Perception & Psychophysics*, vol. 11, no. 1, pp. 10–16, 1972.

[60] I. Starlinger and W. Niemeyer, "Do the blind hear better? investigations on auditory processing in congenital or early acquired blindness I. Peripheral functions," *International Journal of Audiology*, vol. 20, no. 6, pp. 503–509, 1981.

[61] M. Bross and M. Borenstein, "Temporal auditory acuity in blind and sighted subjects: a signal detection analysis," *Perceptual and Motor Skills*, vol. 55, no. 3, pp. 963–966, 1982.

[62] H.-H. Lai and Y.-C. Chen, "A study on the blind's sensory ability," *International Journal of Industrial Ergonomics*, vol. 36, no. 6, pp. 565–570, 2006.

[63] W. Niemeyer and I. Starlinger, "Do the blind hear better? investigations on auditory processing in congenital or early acquired blindness II. Central functions," *International Journal of Audiology*, vol. 20, no. 6, pp. 510–515, 1981.

[64] C. Muchnik, M. Efrati, E. Nemeth, M. Malin, and M. Hildesheimer, "Central auditory skills in blind and sighted subjects," *Scandinavian Audiology*, vol. 20, no. 1, pp. 19–23, 1991.

[65] J. P. Rauschecker, "Compensatory plasticity and sensory substitution in the cerebral cortex," *Trends in Neurosciences*, vol. 18, no. 1, pp. 36–43, 1995.

[66] N. Lessard, M. Paré, F. Lepore, and M. Lassonde, "Early-blind human subjects localize sound sources better than sighted subjects," *Nature*, vol. 395, no. 6699, pp. 278–280, 1998.

[67] M. P. Zwiers, A. J. Van Opstal, and J. R. Cruysberg, "A spatial hearing deficit in early-blind humans," *The Journal of Neuroscience*, vol. 21, no. 9, pp. 1–5, 2001.

[68] M. Ohuchi, Y. Iwaya, Y. Suzuki, and T. Munekata, "A comparative study of sound localization acuity of congenital blind and sighted people," *Acoustical Science and Technology*, vol. 27, no. 5, pp. 290–293, 2006.

[69] A. J. Kolarik, S. Pardhan, S. Cirstea, and B. C. J. Moore, "Auditory spatial representations of the world are compressed in blind humans," *Experimental Brain Research*, vol. 235, no. 2, pp. 597–606, 2017.

[70] C. E. Rice, "Human echo perception," *Science*, vol. 155, no. 3763, pp. 656–664, 1967.

[71] P. Voss, V. Tabry, and R. J. Zatorre, "Trade-off in the sound localization abilities of early blind individuals between the horizontal and vertical planes," *The Journal of Neuroscience*, vol. 35, no. 15, pp. 6051–6056, 2015.

[72] B. Röder, W. Teder-Salejarvi, A. Sterr, F. Rosler, S. A. Hillyard, and H. J. Neville, "Improved auditory spatial tuning in blind humans," *Nature*, vol. 400, no. 6740, pp. 162–166, 1999.

[73] B. J. Cratty, *Movement and Spatial Awareness in Blind Children and Youth*, C. C. Thomas, Ed., Springfield, Illinois, Ill, USA, 1971.

[74] A. Dufour, O. Després, and V. Candas, "Enhanced sensitivity to echo cues in blind subjects," *Experimental Brain Research*, vol. 165, no. 4, pp. 515–519, 2005.

[75] I. G. Bassett and E. J. Eastmond, "Echolocation: measurement of pitch versus distance for sounds reflected from a flat surface," *The Journal of the Acoustical Society of America*, vol. 36, no. 5, pp. 911–916, 1964.

[76] L. D. Rosenblum, M. S. Gordon, and L. Jarquin, "Echolocating distance by moving and stationary listeners," *Ecological Psychology Journal*, vol. 12, no. 3, pp. 181–206, 2000.

[77] W. N. Kellogg, "Sonar System of the Blind," *Science*, vol. 137, no. 3528, pp. 399–404, 1962.

[78] T. Miura, T. Muraoka, and T. Ifukube, "Comparison of obstacle sense ability between the blind and the sighted: A basic psychophysical study for designs of acoustic assistive devices," *Acoustical Science and Technology*, vol. 31, no. 2, pp. 137–147, 2010.

[79] C. E. Rice and S. H. Feinstein, "Sonar system of the blind: size discrimination," *Science*, vol. 38, no. 148, pp. 1107-1108, 1965.

[80] T. A. Stoffregen and J. B. Pittenger, "Human echolocation as a basic form of perception and action," *Ecological Psychology Journal*, vol. 7, no. 3, pp. 181–216, 1995.

[81] G. Wersényi, "Virtual localization by blind persons," *Journal of the Audio Engineering Society*, vol. 60, no. 7-8, pp. 568–579, 2012.

[82] 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," *Journal of the Audio Engineering Society*, vol. 58, no. 9, pp. 724–738, 2010.

[83] A. Afonso, A. Blum, B. F. G. Katz, P. E. Tarroux, G. Borst, and M. Denis, "Structural properties of spatial representations in blind people: scanning images constructed from haptic exploration or from locomotion in a 3-D audio virtual environment," *Memory & Cognition*, vol. 38, no. 5, pp. 591–604, 2010.

[84] L. Picinali, A. Afonso, M. Denis, and B. F. G. Katz, "Exploration of architectural spaces by blind people using auditory virtual reality for the construction of spatial knowledge," *International Journal of Human-Computer Studies*, vol. 72, no. 4, pp. 393–407, 2014.

[85] A. Cobo, N. E. Guerrón, C. Martín, F. del Pozo, and J. J. Serrano, "Differences between blind people's cognitive maps after proximity and distant exploration of virtual environments," *Computers in Human Behavior*, vol. 77, no. 12, pp. 294–308, 2017.

[86] K. Papadopoulos, E. Koustriava, and M. Barouti, "Cognitive maps of individuals with blindness for familiar and unfamiliar spaces: construction through audio-tactile maps and walked experience," *Computers in Human Behavior*, vol. 75, no. 10, pp. 376–384, 2017.

[87] C. I. Cheng and G. H. Wakefield, "Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space," *Journal of the Audio Engineering Society*, vol. 49, no. 4, pp. 231–249, 2001.

[88] J. Blauert, *The Technology of Binaural Listening*, Springer, New York, NY, USA, 2013.

[89] A. W. Bronkhorst, "Localization of real and virtual sound sources," *The Journal of the Acoustical Society of America*, vol. 98, no. 5, pp. 2542–2553, 1995.

[90] G. Wersényi, "Localization in a head-related transfer function-based virtual audio synthesis using additional high-pass and low-pass filtering of sound sources," *Acoustical Science and Technology*, vol. 28, no. 4, pp. 244–250, 2007.

[91] M. D. Burkhard and R. M. Sachs, "Anthropometric manikin for acoustic research," *The Journal of the Acoustical Society of America*, vol. 58, no. 1, pp. 214–222, 1975.

[92] A. Andreopoulou, D. R. Begault, and B. F. G. Katz, "Inter-laboratory round robin HRTF measurement comparison," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 895–906, 2015.

[93] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proceedings of the IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pp. 99–102, IEEE, New York, NY, USA, 2001.

[94] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 94, no. 1, pp. 111–123, 1993.

[95] D. R. Begault, E. M. Wenzel, and M. R. 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," *Journal of the Audio Engineering Society*, vol. 49, no. 10, pp. 904–916, 2001.

[96] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, "Binaural technique: do we need individual recordings?" *Journal of the Audio Engineering Society*, vol. 44, no. 6, pp. 451–464, 1996.

[97] M. Geronazzo, S. Spagnol, and F. Avanzini, "A modular framework for the analysis and synthesis of head-related transfer functions," in *Proceedings of the 134th AES Convention - Audio Engineering Society*, 2013.

[98] B. U. Seeber and H. Fastl, "Subjective selection of non-individual head-related transfer functions," in *Proceedings of the International Conference on Auditory Display (ICAD '03)*, pp. 259–262, 2003.

[99] D. N. Zotkin, R. Duraiswami, and L. S. Davis, "Rendering localized spatial audio in a virtual auditory space," *IEEE Transactions on Multimedia*, vol. 6, no. 4, pp. 553–564, 2004.

[100] M. Geronazzo, S. Spagnol, A. Bedin, and F. Avanzini, "Enhancing vertical localization with image-guided selection of non-individual head-related transfer functions," in *Proceedings of the 2014 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '14*, pp. 4496–4500, May 2014.

[101] K. Iida, Y. Ishii, and S. Nishioka, "Personalization of head-related transfer functions in the median plane based on the anthropometry of the listener's pinnae," *The Journal of the Acoustical Society of America*, vol. 136, no. 1, pp. 317–333, 2014.

[102] R. H. Y. So, B. Ngan, A. Horner, J. Braasch, J. Blauert, and K. L. Leung, "Toward orthogonal non-individualised head-related transfer functions for forward and backward directional sound: Cluster analysis and an experimental study," *Ergonomics*, vol. 53, no. 6, pp. 767–781, 2010.

[103] B. F. G. Katz and G. Parseihian, "Perceptually based head-related transfer function database optimization," *The Journal of the Acoustical Society of America*, vol. 131, no. 2, pp. EL99–EL105, 2012.

[104] S. Hwang, Y. Park, and Y.-S. Park, "Modeling and customization of head-related impulse responses based on general basis functions in time domain," *Acta Acustica united with Acustica*, vol. 94, no. 6, pp. 965–980, 2008.

[105] K. H. Shin and Y. Park, "Enhanced vertical perception through head-related impulse response customization based on pinna response tuning in the median plane," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E91-A, no. 1, pp. 345–356, 2008.

[106] W. M. Rabinowitz, J. Maxwell, Y. Shao, and M. Wei, "Sound localization cues for a magnified head: implications from sound diffraction about a rigid sphere," *Presence: Teleoperators and Virtual Environments*, vol. 2, no. 2, pp. 125–129, 1993.

[107] S. Spagnol, E. Tavazzi, and F. Avanzini, "Distance rendering and perception of nearby virtual sound sources with a near-field filter model," *Applied Acoustics*, vol. 115, pp. 61–73, 2017.

[108] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Acoustic simulation of KEMAR's HRTFs: verification with measurements and the effects of modifying head shape and pinna concavity," in *Proceeding of the Interenational Workshop Principles Apply Spatial Hearing (IWPASH '09)*, 2009.

[109] R. O. Duda, C. Avendano, and V. R. Algazi, "An adaptable ellipsoidal head model for the interaural time difference," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 965–968, March 1999.

[110] R. Bomhardt, M. Lins, and J. Fels, "Analytical ellipsoidal model of interaural time differences for the individualization of head-related impulse responses," *Journal of the Audio Engineering Society*, vol. 64, no. 11, pp. 882–894, 2016.

[111] B. F. G. Katz and M. Noisternig, "A comparative study of interaural time delay estimation methods," *The Journal of the Acoustical Society of America*, vol. 135, no. 6, pp. 3530–3540, 2014.

[112] A. Andreopoulou and B. F. Katz, "Identification of perceptually relevant methods of inter-aural time difference estimation," *The*

Journal of the Acoustical Society of America, vol. 142, no. 2, pp. 588–598, 2017.

[113] V. R. Algazi, R. O. Duda, and D. M. Thompson, "The use of head-and-torso models for improved spatial sound synthesis in," in *Proceedings of the 113th Convention of the Audio Engineering Society*, pp. 1–18, 2002.

[114] V. R. Algazi, R. O. Duda, R. Duraiswami, N. A. Gumerov, and Z. Tang, "Approximating the head-related transfer function using simple geometric models of the head and torso," *The Journal of the Acoustical Society of America*, vol. 112, no. 5, pp. 2053–2064, 2002.

[115] A. Meshram, R. Mehra, H. Yang, E. Dunn, J.-M. Franm, and D. Manocha, "P-HRTF: Efficient personalized HRTF computation for high-fidelity spatial sound," in *Proceedings of the 13th IEEE International Symposium on Mixed and Augmented Reality, ISMAR '14*, pp. 53–61, IEEE, Munich, Germany, September 2014.

[116] V. Algazi, R. Duda, R. Morrison, and D. Thompson, "Structural composition and decomposition of HRTFs," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 103–106, IEEE, New York, NY, USA, 2001.

[117] V. R. Algazi, C. Avendano, and R. O. Duda, "Estimation of a spherical-head model from anthropometry," *Journal of the Audio Engineering Society*, vol. 49, no. 6, pp. 472–479, 2001.

[118] S. Spagnol and F. Avanzini, "Anthropometric tuning of a spherical head model for binaural virtual acoustics based on interaural level differences," in *Proceedings of the 21st International Conference on Auditory Display ICAD '15*, pp. 204–209, 2015.

[119] M. Aussal, F. Alouges, and B. F. Katz, "HRTF interpolation and ITD personalization for binaural synthesis using spherical harmonics," in *Proceedings of the 25th UK Conference Audio Engineering Society*, 2012.

[120] A. J. Watkins, "Psychoacoustical aspects of synthesized vertical locale cues," *The Journal of the Acoustical Society of America*, vol. 63, no. 4, pp. 1152–1165, 1978.

[121] R. Teranishi and E. A. G. Shaw, "External-ear acoustic models with simple geometry," *The Journal of the Acoustical Society of America*, vol. 44, no. 1, pp. 257–263, 1968.

[122] E. A. G. Shaw, "The acoustics of the external ear," in *Acoustical Factors Affecting Hearing Aid Performance*, G. A. Studebaker and I. Hochberg, Eds., pp. 109–125, University Park Press, Maryland, Md, USA, 1980.

[123] P. Satarzadeh, R. V. Algazi, and R. O. Duda, "Physical and filter pinna models based on anthropometry," in *Proceedings of the 122nd Audio Engineering Society Convention '07*, pp. 718–737, May 2007.

[124] K. J. Faller, A. Barreto, and M. Adjouadi, "Augmented Hankel total least-squares decomposition of head-related transfer functions," *Journal of the Audio Engineering Society*, vol. 58, no. 1-2, pp. 3–21, 2010.

[125] S. Spagnol, M. Geronazzo, and F. Avanzini, "On the relation between pinna reflection patterns and head-related transfer function features," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 21, no. 3, pp. 508–520, 2013.

[126] L. Li and Q. Huang, "HRTF personalization modeling based on RBF neural network," in *Proceedings of the 2013 38th IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '13*, pp. 3707–3710, IEEE, Vancouver, BC, Canada, May 2013.

[127] Q. Huang and L. Li, "Modeling individual HRTF tensor using high-order partial least squares," *Journal on Advances in Signal Processing*, vol. 58, pp. 1–14, 2014.

[128] F. Grijalva, L. Martini, S. Goldenstein, and D. Florencio, "Anthropometric-based customization of head-related transfer functions using Isomap in the horizontal plane," in *Proceedings of the 39th IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '14*, pp. 4506–4510, IEEE, Firenze, Italy, May 2014.

[129] P. Bilinski, J. Ahrens, M. R. P. Thomas, I. J. Tashev, and J. C. Platt, "HRTF magnitude synthesis via sparse representation of anthropometric features," in *Proceedings of the 39th IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '14*, pp. 4501–4505, IEEE, Firenze, Italy, May 2014.

[130] J. C. Middlebrooks, "Individual differences in external-ear transfer functions reduced by scaling in frequency," *The Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1480–1492, 1999.

[131] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Frequency and amplitude estimation of the first peak of head-related transfer functions from individual pinna anthropometry," *The Journal of the Acoustical Society of America*, vol. 137, no. 2, pp. 690–701, 2015.

[132] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Vertical normal modes of human ears: Individual variation and frequency estimation from pinna anthropometry," *The Journal of the Acoustical Society of America*, vol. 140, no. 2, pp. 814–831, 2016.

[133] V. C. Raykar, R. Duraiswami, and B. Yegnanarayana, "Extracting the frequencies of the pinna spectral notches in measured head related impulse responses," *The Journal of the Acoustical Society of America*, vol. 118, no. 1, pp. 364–374, 2005.

[134] S. Spagnol and F. Avanzini, "Frequency estimation of the first pinna notch in head-related transfer functions with a linear anthropometric model," in *Proceedings of the 18th International Conference on Digital Audio Effects*, pp. 231–236, 2015.

[135] S. Spagnol, M. Geronazzo, D. Rocchesso, and F. Avanzini, "Synthetic individual binaural audio delivery by pinna image processing," *International Journal of Pervasive Computing and Communications*, vol. 10, no. 3, pp. 239–254, 2014.

[136] S. Spagnol, S. Scaiella, M. Geronazzo, and F. Avanzini, "Subjective evaluation of a low-order parametric filter model of the pinna for binaural sound rendering," in *Proceedings of the 22nd International Congress on Sound and Vibration, ICSV '15*, July 2015.

[137] T. Huttunen, E. T. Seppälä, O. Kirkeby, A. Kärkkäinen, and L. Kärkkäinen, "Simulation of the transfer function for a head-and-torso model over the entire audible frequency range," *Journal of Computational Acoustics*, vol. 15, no. 4, pp. 429–448, 2007.

[138] B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation," *The Journal of the Acoustical Society of America*, vol. 110, no. 5, pp. 2440–2448, 2001.

[139] Y. Kahana and P. A. Nelson, "Boundary element simulations of the transfer function of human heads and baffled pinnae using accurate geometric models," *Journal of Sound and Vibration*, vol. 300, no. 3-5, pp. 552–579, 2007.

[140] W. Kreuzer, P. Majdak, and Z. Chen, "Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range," *The Journal of the Acoustical Society of America*, vol. 126, no. 3, pp. 1280–1290, 2009.

[141] N. A. Gumerov, A. E. O'Donovan, R. Duraiswami, and D. N. Zotkin, "Computation of the head-related transfer function via the fast multipole accelerated boundary element method

and its spherical harmonic representation," *The Journal of the Acoustical Society of America*, vol. 127, no. 1, pp. 370–386, 2010.

[142] H. Ziegelwanger, P. Majdak, and W. Kreuzer, "Numerical calculation of listener-specific head-related transfer functions and sound localization: microphone model and mesh discretization," *The Journal of the Acoustical Society of America*, vol. 138, no. 1, pp. 208–222, 2015.

[143] C. T. Jin, P. Guillon, N. Epain et al., "Creating the Sydney York morphological and acoustic recordings of ears database," *IEEE Transactions on Multimedia*, vol. 16, no. 1, pp. 37–46, 2014.

[144] H. Ziegelwanger, W. Kreuzer, and P. Majdak, "MESH2HRTF: an open-source software package for the numerical calculation of head-related transfer functions," in *Proceedings of the 22nd International Congress on Sound and Vibration 2015 (ICSV '22)*, IEEE, Firenze, Italy, 2015.

[145] M. Dellepiane, N. Pietroni, N. Tsingos, M. Asselot, and R. Scopigno, "Reconstructing head models from photographs for individualized 3D-audio processing," *Computer Graphics Forum*, vol. 27, no. 7, pp. 1719–1727, 2008.

[146] H. Møller, D. Hammershøi, C. B. Jensen, and M. F. Sørensen, "Transfer characteristics of headphones measured on human ears," *Journal of the Audio Engineering Society*, vol. 43, no. 4, pp. 203–217, 1995.

[147] D. Pralong and S. Carlile, "The role of individualized headphone calibration for the generation of high fidelity virtual auditory space," *The Journal of the Acoustical Society of America*, vol. 100, no. 6, pp. 3785–3793, 1996.

[148] B. Masiero and J. Fels, "Perceptually robust headphone equalization for binaural reproduction," in *Proceedings of the 130th Audio Engineering Society Convention*, 2011.

[149] Z. Schärer and A. Lindau, "Evaluation of equalization methods for binaural signals," in *Proceedings of the 126th Audio Engineering Society Convention*, 2009.

[150] D. Schonstein, L. Ferré, and B. F. Katz, "Comparison of headphones and equalization for virtual auditory source localization," *The Journal of the Acoustical Society of America*, vol. 123, no. 5, pp. 3724-3724, 2008.

[151] R. G. Golledge, J. R. Marston, J. M. Loomis, and R. L. Klatzky, "Stated Preferences for Components of a Personal Guidance System for Nonvisual Navigation," *Journal of Visual Impairment & Blindness*, vol. 98, no. 3, pp. 135–147, 2004.

[152] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *Journal of the Audio Engineering Society*, vol. 45, no. 6, pp. 456–466, 1997.

[153] M. A. Gerzon, "Ambisonics in multichannel broadcasting and video," *Journal of the Audio Engineering Society*, vol. 33, no. 11, pp. 859–871, 1985.

[154] F. M. König, "4-canal headphone for in-front localization and HDTV- or Dolby-surround use," in *Proceedings of the 96th Convention Audio Engineering Society*, 1994.

[155] F. M. König, "A new supra-aural dynamic headphone system for in-front localization and surround reproduction of sound," in *Proceedings of the 102nd Convention Audio Engineering Society*, 1997.

[156] F. M. König, "New measurements and psychoacoustic investigations on a headphone for TAX/HDTV/Dolby-surround reproduction of sound," in *Proceedings of the 98th Convention Audio Engineering Society*, 1995.

[157] S. G. Weinrich, "Improved externalization and frontal perception of headphone signals," in *Proceedings of the 92nd Convention Audio Engineering Society*, 1992.

[158] K. Sunder, E.-L. Tan, and W.-S. Gan, "Individualization of binaural synthesis using frontal projection headphones," *Journal of the Audio Engineering Society*, vol. 61, no. 12, pp. 989–1000, 2013.

[159] R. Greff and B. F. Katz, "Circumaural transducer arrays for binaural synthesis," *The Journal of the Acoustical Society of America*, vol. 123, no. 5, pp. 3562-3562, 2008.

[160] M. Bujacz, K. Kropidlowski, G. Ivanica et al., "Sound of Vision - Spatial audio output and sonification approaches," in *Proceedings of the 15th International Conference of Computers Helping People with Special Needs (ICCHP '16)*, K. Miesenberger, C. Bühler, and P. Penaz, Eds., vol. 9759, pp. 202–209, Springer, Linz, Austria, 2016.

[161] B. N. Walker and R. M. Stanley, "Thresholds of audibility for bone-conduction headsets," in *Proceedings of the International Conference on Auditory Display*, pp. 218–222, IEEE, Limerick, Ireland, 2005.

[162] B. N. Walker, R. M. Stanley, N. Iyer, B. D. Simpson, and D. S. Brungart, "Evaluation of bone-conduction headsets for use in multitalker communication environments," in *Proceedings of the 49th Annual Meeting of Human Factors and Ergonomics Society*, pp. 1615–1619, 2005.

[163] R. M. Stanley and B. N. Walker, "Lateralization of sound using bone-conduction headsets," in *Proceedings of the 50th Annual Meeting of Human Factors and Ergonomics Society*, pp. 1571–1575, 2006.

[164] J. A. MacDonald, P. P. Henry, and T. R. Letowski, "Spatial audio through a bone conduction interface," *International Journal of Audiology*, vol. 45, no. 10, pp. 595–599, 2006.

[165] R. M. Stanley, *Measurement and validation of bone-conduction adjustment functions in virtual 3D audio displays [Ph.D. thesis]*, Georgia Institute of Technology, Georgia, Ga, USA, 2009.

[166] S. Reinfeldt, B. Håkansson, H. Taghavi, and M. Eeg-Olofsson, "New developments in bone-conduction hearing implants: a review," *Medical Devices: Evidence and Research*, vol. 8, pp. 79–93, 2015.

[167] B. N. Walker and J. L. Lindsay, "Navigation performance in a virtual environment with bonephones," in *Proceedings of the 11th International Conf.erence Auditory Display (ICAD '05)*, pp. 260–263, IEEE, Limerick, Ireland, 2005.

[168] B. N. Walker, R. M. Stanley, and A. Przekwas, "High fidelity modeling and experimental evaluation of binaural bone conduction communication devices," in *Proceedings of the 19th International Congress on Acoustics*, 2007.

[169] R. W. Lindeman, H. Noma, and P. G. De Barros, "Hear-through and mic-through augmented reality: Using bone conduction to display spatialized audio," in *Proceedings of the 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR '07*, November 2007.

[170] R. W. Lindeman, H. Noma, and P. G. De Barros, "An empirical study of Hear-Through augmented reality: Using bone conduction to deliver spatialized audio," in *Proceedings of the IEEE Virtual Reality*, pp. 35–42, March 2008.

[171] A. Barde, W. S. Helton, G. Lee, and M. Billinghurst, "Binaural spatialization over a bone conduction headset: Minimum discernable angular difference," in *Proceedings of the 140th Convention of the Audio Engineering Society*, 2016.

[172] B. Rafaely, "Active noise reducing headset - An overview," in *Proceedings of the International Congress and Exhibition on Noise Control Engineering*, 2001.

[173] A. V. Oppenheim, E. Weinstein, K. C. Zangi, M. Feder, and D. Gauger, "Single-sensor active noise cancellation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 2, no. 2, pp. 285–290, 1994.

[174] M. Tikander, "Usability issues in listening to natural sounds with an augmented reality audio headset," *Journal of the Audio Engineering Society*, vol. 57, no. 6, pp. 430–441, 2009.

[175] A. Härmä, J. Jakka, M. Tikander et al., "Augmented reality audio for mobile and wearable appliances," *Journal of the Audio Engineering Society*, vol. 52, no. 6, pp. 618–639, 2004.

[176] M. Tikander, M. Karjalainen, and V. Riikonen, "An augmented reality audio headset," in *Proceedings of the 11th International Conference on Digital Audio Effects, DAFx '08*, 2008.

[177] J. Liski, *Adaptive hear-through headset [M.S. thesis]*, Aalto University School of Electrical Engineering, Espoo, Finland, 2016.

[178] S. Brunner, H. J. Maempel, and S. Weinzierl, "On the audibility of comb filter distortions," in *Proceedings of the 122nd Convention Audio Engineering Society*, 2007.

[179] J. Rämö and V. Välimäki, "Digital augmented reality audio headset," *Journal of Electrical and Computer Engineering*, vol. 2012, Article ID 457374, 13 pages, 2012.

[180] D. Hammershøi and H. Møller, "Sound transmission to and within the human ear canal," *The Journal of the Acoustical Society of America*, vol. 100, no. 1, pp. 408–427, 1996.

[181] P. B. L. Meijer, "An experimental system for auditory image representations," *IEEE Transactions on Biomedical Engineering*, vol. 39, no. 2, pp. 112–121, 1992.

[182] M. A. Hersh and M. A. Johnson, *Assistive Technology for Visually Impaired and Blind People*, Springer, London, UK, 1st edition, 2008.

[183] S. Shoval, J. Borenstein, and Y. Koren, "Auditory guidance with the Navbelt-a computerized travel aid for the blind," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 28, no. 3, pp. 459–467, 1998.

[184] G. Balakrishnan, G. Sainarayanan, R. Nagarajan, and S. Yaacob, "A stereo image processing system for visually impaired," *International Journal of Signal Processing*, vol. 2, no. 3, pp. 136–145, 2008.

[185] F. Zhigang and L. Ting, "Audification-based electronic travel aid system," in *Proceedings of the International Conference on Computer Design and Applications, ICCDA '10*, vol. 5, pp. 137–141, IEEE, Qinhuangdao, China, 2010.

[186] J. L. González-Mora, A. Rodríguez-Hernández, L. F. Rodríguez-Ramos, L. Díaz-Saco, and N. Sosa, "Development of a new space perception system for blind people, based on the creation of a virtual acoustic space," in *Engineering Applications of Bio-Inspired Artificial Neural Networks*, vol. 1607 of *Lecture Notes in Computer Science*, pp. 321–330, Springer, Berlin, Germany, 1999.

[187] F. Fontana, A. Fusiello, M. Gobbi et al., "A cross-modal electronic travel aid device," in *Human Computer Interaction with Mobile Devices*, vol. 2411 of *Lecture Notes in Computer Science*, pp. 393–397, Springer, Berlin, Germany, 2002.

[188] J. Wilson, B. N. Walker, J. Lindsay, C. Cambias, and F. Dellaert, "SWAN: system for wearable audio navigation," in *Proceedings of the 11th IEEE International Symposium on Wearable Computers, ISWC '07*, pp. 91–98, IEEE, Massachusetts, Mass, USA, October 2007.

[189] M. A. Torres-Gil, O. Casanova-Gonzalez, and J. L. Gonzalez-Mora, "Applications of virtual reality for visually impaired people," *WSEAS Transactions on Computers*, vol. 9, no. 2, pp. 184–193, 2010.

[190] 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 *Proceedings of the 36th Annual Conference of the IEEE Industrial Electronics Society, IECON '10*, pp. 1173–1178, IEEE, Arizona, Ariz, USA, November 2010.

[191] G. P. Fajarnes, L. Dunai, and V. S. Praderas, "CASBLiP - A new cognitive object detection and orientation system for impaired people," in *Proceedings of the 4th International Conference on Cognitive Systems*, IEEE, Zurich, Switzerland, 2010.

[192] M. Bujacz, P. Skulimowski, and P. Strumiłło, "Sonification of 3D scenes using personalized spatial audio to aid visually impaired persons," in *Proceedings of the 17th International Conference on Auditory Display*, 2011.

[193] M. Bujacz, P. Skulimowski, and P. Strumillo, "Naviton-a prototype mobility aid for auditory presentation of three-dimensional scenes to the visually impaired," *Journal of the Audio Engineering Society*, vol. 60, no. 9, pp. 696–708, 2012.

[194] B. F. G. Katz, S. Kammoun, G. Parseihian et al., "NAVIG: augmented reality guidance system for the visually impaired," *Virtual Reality*, vol. 16, no. 4, pp. 253–269, 2012.

[195] B. F. G. Katz, F. Dramas, G. Parseihian et al., "NAVIG: Guidance system for the visually impaired using virtual augmented reality," *Technology and Disability*, vol. 24, no. 2, pp. 163–178, 2012.

[196] J. Vliegen and A. J. Van Opstal, "The influence of duration and level on human sound localization," *The Journal of the Acoustical Society of America*, vol. 115, no. 4, pp. 1705–1713, 2004.

[197] S. Spagnol, R. Hoffmann, M. Herrera Martínez, and R. Unnthorsson, "Blind wayfinding with physically-based liquid sounds," *International Journal of Human-Computer Studies*, vol. 115, pp. 9–19, 2018.

[198] C. Mendonça, "A review on auditory space adaptations to altered head-related cues," *Frontiers in Neuroscience*, vol. 8, no. 219, Article ID Article 219, pp. 1–14, 2014.

[199] P. Zahorik, P. Bangayan, V. Sundareswaran, K. Wang, and C. Tam, "Perceptual recalibration in human sound localization: Learning to remediate front-back reversals," *The Journal of the Acoustical Society of America*, vol. 120, no. 1, pp. 343–359, 2006.

[200] S. Carlile, "The plastic ear and perceptual relearning in auditory spatial perception," *Frontiers in Neuroscience*, vol. 8, no. 237, pp. 1–13, 2014.

[201] G. Parseihian and B. F. G. Katz, "Rapid head-related transfer function adaptation using a virtual auditory environment.," *The Journal of the Acoustical Society of America*, vol. 131, no. 4, pp. 2948–2957, 2012.

[202] A. Honda, H. Shibata, S. Hidaka, J. Gyoba, Y. Iwaya, and Y. Suzuki, "Effects of head movement and proprioceptive feedback in training of sound localization," *i-Perception*, vol. 4, no. 4, pp. 253–264, 2013.

[203] O. Balan, A. Moldoveanu, H. Nagy et al., "Haptic-auditory perceptual feedback based training for improving the spatial acoustic resolution of the visually impaired people," in *Proceedings of the 21st International Conference on Auditory Display (ICAD '15)*, pp. 21–28, 2015.