Chapter 2
Human Factors Research in Audio Augmented Reality

Nicholas Mariette

1 Introduction

To discuss research into human factors in audio augmented reality, it is helpful to specify the context and motivation for such investigations. Elucidating this scope requires some discretionary definitions and limitations due to the broad and finely-variable nature of the medium. We will start with the fundamentals.

AR itself is a medium for which it can be problematic to make a comprehensive definition. One widely-cited publication by Azuma [1] defines generic AR as having the following three characteristics:

1. Combines real and virtual.
2. Interactive in real time.
3. Registered in 3D.

A looser definition by Mallem [33] defines AR as enabling "spatial and temporal virtual and real worlds [to] co-exist, which aims to enhance user perception in his real environment." While the former definition might be considered the canonical one, the latter might better describe many recent instances of technology that have come to be known as AR.

A significant current trend in technology and for AR applications is the rapid development and adoption of mobile computing devices such as current generation smart phones that are capable of bringing AR "apps" to mass-markets. Aspects of mobile AR including design, cognition and user experience are discussed further in this book in Chaps. 5–7 and 9. It is these kinds of accessible AR instances, such as the common format of visual augmentations to live video from smart-phone cameras that are simultaneously bringing AR to larger audiences, as well as stretching the bounds of Azuma's three characteristics of AR. In particular, the "real" is often substantially

---

N. Mariette (✉)
University of New South Wales, Sydney, NSW, Australia
e-mail: n.mariette@unswalumni.com

W. Huang et al. (eds.), *Human Factors in Augmented Reality Environments*,
DOI 10.1007/978-1-4614-4205-9_2,

mediated (e.g. by the use of live video), while 3D registration might also be of low quality, or somewhat closer to 2D registration. As such, these new applications often conform better to Mallem's broader definition of AR. The same trend is occurring to the characteristics of *audio AR* technology, with many interesting and exciting new developments happening in the smart-phone application market.

As a curious aside, the broadest definition of AR is manifest in the literal meaning of the words, augmented reality. In other words, AR is any augmentation or enhancement to plain old, unadorned reality. Interestingly, a broader definition such as Mallem's that satisfactorily encompasses the present evolution of AR also embraces *older* technology not previously identified as AR, yet comfortably described by the literal phrase. For instance, the visual use of static or moving images projected onto unconventional surfaces, or the use of mobile transistor radios or early portable tape players with headphones could both be understood as presenting an augmented reality. Equally, these examples of realities augmented by technologically-delivered media both in some way enhance the perception of the person apprehending them.

With such broad possible meanings of the term AR, it is clearly difficult to commence a focussed discussion of human factors of these media. Nevertheless, it is clear that human factors are important to discuss. If any given augmentation of reality is to be understood as enhancing a user's perception within their real environment, this logically invites questions as to *how well* their perception is enhanced, in *what manner* it occurs, and *why* and *how* is it useful.

This chapter explores the study of human factors in audio augmented reality by first defining the media that exist under this name. This scope is limited somewhat to those kinds of audio AR that have significant human factors and consequential performance effects. We briefly look at human factors research in AR in general, then we survey how human factors have been investigated in research on audio AR, and what these studies might be able to contribute to the further development of the medium.

2 What Defines *Audio* Augmented Reality?

As is the case for general AR, a definition of the medium of *audio* augmented reality is both problematic, and necessary for a survey of human factors research.

2.1 A Broad Definition

In the most general sense, audio AR is simply the introduction of artificial sound material into the real world. In discussing telepresence applications, Cohen et al. [12] noted:

> One common example of augmented audio reality is sound reinforcement, as in a public address system.

Novo [43, p. 291] made a similar observation:

> The idea of mixing real and virtual environments, “joint reality” or “augmented reality,” was probably first realized with assisted reverberation systems.

This observation was also made by Härmä [16, pp. 795–823], who considered voice-only teleconference systems dating back to the 1950s as AR systems.

Note that this general concept, nevertheless has some specificity to it, given that the application of assisted reverberation (and others) considers the augmentation of the *aural* reality (the real sound sources and room acoustics). This suggests that augmented reality in an aural context means augmentation of the *aural* reality—i.e. that the sensory mode of the augmentation matches that of the augmented, however this may not always be the case. Next, we consider the range of characteristics that define the different types of audio AR, of which the above is only one example.

2.1.1 Predominant Sensory Modality

The modality of the reality that is being augmented with virtual audio—aural, visual, haptic, olfactory, gustatory, or any combination of these. This accounts for the possibility of cross-modal augmentation, the most common of which would be virtual audio augmenting the visual reality. Audio augmentation of the haptic reality has also been investigated. In cross-modal situations it is necessary for the link between stimuli to be clearly referenced to the same object.

2.1.2 Spatial Characteristic

Virtual audio may be processed as a mono, stereo, two-dimensional (2D) or three-dimensional (3D) audio signal. In fact, this is a somewhat artificial distinction, given that there is a complex connection between the creation of a signal—a *sound event*—and a listener’s auditory perception of the signal—the resulting *auditory event* [7, p. 2]. A sound event with one spatial characteristic might be perceived as an auditory event with a different spatial characteristic, depending on the circumstances. For instance, a mono signal presented to only one ear could be perceived as a near-field, 3D positional sound source.

For the purposes of this discussion, we shall consider only the situation where the spatial characteristic of the perceived auditory event matches the intention of the processing that generated the sound event. The nature of the auditory event is an important human factor in spatial audio reproduction. In general, mono and stereo virtual audio are perceived on the left–right axis through the centre of the head, while 2D and 3D virtual audio are perceived to be in a given direction and distance no closer than the perimeter of the head.

2.1.3 Presentation Means

This describes whether the virtual audio is rendered specifically for each individual user, or once for a group audience. Audio AR has been conceived both for individual users (generally presented over headphones), or for a collective group audience (generally presented via a loudspeaker array e.g. [57, 58]). This category of classification partly aligns with the classification by Lindeman and Noma [26] based on the “location of mixing,” which can be in the environment (via speakers, or using passive, acoustic mixing through headphones), in the sensory subsystem (e.g. using bonephones or implant technology), or in a computer system (e.g. microphone feedthrough).

2.1.4 User Tracking Extent or Interactivity Mode

This describes whether or not the user’s head-orientation is tracked and used in addition to position, to control rendering. It relates to Azuma’s AR characteristic of real-time interaction and 3D registration of the stimuli. Many audio AR applications on smart phones only provide virtual audio in relation to position. This does not preclude 3D audio from being presented—e.g. a spatial sound-field at a given location, without a fixed orientation. However it does preclude *positional* 3D audio—which requires the 3D audio scene orientation to remain fixed relative to the physical world, even while the user turns their head. The availability of the user’s head-orientation thus has a great effect on the affordances that may be realised by the medium. This in turn has a great effect on the tasks, usability, and higher-level applications that are possible. For this reason, the interactivity mode is important in the discussion of human factors research in audio AR media.

2.1.5 User Mobility

This describes the amount of freedom of the user to move around. It depends on whether the tracking system requires built infrastructure within a local area, or whether it enables free-roaming, as for global positioning system (GPS) receivers. This is an important, yet basic factor, however we have not used this to define the types of audio AR for this discussion. Some literature specifies whether or not a system is mobile, because much research to date has used tethered or limited-area tracking systems (e.g. [59]).

2.2 Taxonomy

Now we use the defining characteristics described above to develop a taxonomy of audio AR media terms from the literature, and to aid the discussion of human factors research. In practice, particular combinations of characteristics are more common than others. This chapter will focus on the most common audio AR media.

2.2.1 Augmented Reality Audio

Augmented Reality Audio (ARA) specifies augmentation of the auditory sensory modality, generally with virtual audio that has 2D or 3D spatial characteristics, individual presentation, and full position and orientation tracking extent. For instance, the medium known as Augmented Reality Audio (ARA; [17]) is distinctly focused on audio augmentations to the *audible* reality. The augmented sensory modality is important because new human factors arise during multi-modal or cross-modal perception.

2.2.2 Audio AR

This term has been used without precise identification of the augmented sensory modality, although augmentations of visual or auditory modes are the most common. Tools going by the term Audio AR have used all spatial characteristics from mono to full 3D, both individual or group presentation, and both extents of user tracking. Media known as Audio AR are usually more general, and often augment the *visible* reality, or have a less critical focus on audio-feedthrough [32, 42].

2.2.3 Augmented Audio Reality

This less common term was first used to describe the augmentation of *audible* reality [11, 12], a meaning that is supported by its literal interpretation, however it has also been used interchangeably with audio AR [24]. For this discussion, we will not include the term Augmented Audio Reality, which has some overlap with the other terms, yet is probably best aligned with the term ARA.

2.2.4 Spatial Audio AR

An important sub-group of Audio AR only uses 2D or 3D spatial characteristics, with full position and orientation tracking, denoted herein as Spatial Audio AR (SAAR).

2.2.5 Personal, Location-Aware Spatial Audio

A further sub-group of SAAR is the medium presented only to individual users, which I have previously termed Personal Location-Aware Spatial Audio (PLASA; [37]). PLASA by definition requires 2D or 3D spatial audio, with position and orientation tracking, and rendering for individual audition to headphones, or other personal transducers such as bonephones [51].

2.2.6 Locative Audio

This term groups a broader category of media that use only positional user tracking, thereby not facilitating 2D or 3D audio with absolute directionality in the world reference frame. This medium does not specify *how* the sound relates to location or how it is processed. As noted earlier, this does not preclude the use of spatial audio, however it will remain *oriented* within the *user's* reference frame, while being *positioned* relative to the absolute, *world* reference frame.

The term derives from "locative media" [14], which refers to electronic media that relate to the user's locational context. Locative media are closely related to the media of pervasive/ubiquitous computing, whereby services are made constantly available, and sensitive to user *contexts* (locational, social, technological and others). Locative audio can exist using relatively unsophisticated technology, since by definition it only requires mono audio and rough user position information.

During the last decade, many projects with artistic or utilitarian applications could be classed as locative audio, or non-spatial audio AR. Fewer projects could be classified as PLASA or SAAR.

3 Early Audio AR Media

This section presents a brief review of applications and evaluations of mobile audio AR.

Audio augmented reality was first proposed no later than 1993 by Cohen et al., who demonstrated a system for a stationary user that employed binaural spatial audio to augment the sight of a real telephone, with the virtual auditory image of a ringing telephone presented over headphones. The system used head-tracking, so the virtual sound was registered to the world reference frame.

This static AR concept has since been extended as mobile AR, by introducing indoors or outdoors body-motion-tracking. Also, even before the completion in early 1994 of the 24 satellite constellation for the Global Positioning System (GPS), Loomis et al. [29] proposed using GPS position tracking in their personal guidance system (PGS) for the visually impaired.

The PGS concept was to present a virtual acoustic display of externalised sound beacons within the traveler's auditory space, with an aim to "allow blind users to travel without assistance over unfamiliar territory." They also hoped to "permit the user to develop better cognitive representations of the environments" [29]. Although the PGS was not conceived as an audio AR system, it is a good example both of PLASA and SAAR.

Cohen [11] also identified GPS as an appropriate positioning technology for audio AR applications, with a design for a GPS Personal Guidance System. He made the visionary statement that "further synergetic effects are obtained by eventual leverage off emerging ubiquity of telecommunication networks (more bandwidth [like high-fidelity audio]) and GIS (geographic information system) databases, accessing terrain or street data.” Even present day audio AR systems are yet to fully realise the potential of these synergies.

Over the following decade, several outdoor, GPS-based audio AR projects were implemented as fairly bulky packages, for example, backpack-based systems [18, 19], or roll-around cases [46]. In 2001–2004, the indoors LISTEN project [53] used high resolution, sub-decimetre tracking, and further reduced the worn system to passive tracking signal emitters and headphones, with remote tracking and spatial audio rendering. Mariette [37] reviews several other audio AR systems with sound art and pure research applications. The form factor of a system often has important repercussions on human factors and usage simply due to the affordances associated with a device’s size and weight.

With cheap digital compasses, powerful portable computers, lightweight consumer GPS receivers (and soon, Galileo receivers,) affordable, portable outdoors audio AR systems may now be implemented. However, despite this great potential, there has been relatively little investigation of the usability and perceptual performance of many of these systems. Evaluation of audio AR systems has often been limited to basic verification of functionality.

4 Introduction to Human Factors in Audio AR

This section discusses the relationship between affordances of audio AR media, human characteristics and perceptual abilities, technical characteristics of devices and their combined meaning in regards to the human factors of the resulting media ecosystems. This provides a conceptual introduction to the specific survey of research that follows.

4.1 Affordances and Human Factors

An affordance of an object, device or system is something that it is possible to do with that thing. Gibson [15] defined an affordance as “a specific combination of the properties of [an object’s] substance and its surfaces taken with reference to an animal.” According to Jones [22], Gibson further suggested that an affordance concerns personal and environmental properties taken in reference to each other. Clearly, affordances involve inherent human capability and inherent physical possibilities of an object and its environment.

For example, a *sounding object* (i.e. an object that is radiating acoustic energy), along with the human ability for spatial hearing and the physical acoustics of air and the surrounding environment affords the act of localising the sound and navigating to its origin if so desired.

A human factor is a characteristic of one person or many people in relation to their use of an object/device or system, such as audio AR. Typically, a human factor describes any effect on task performance that is unique to the *interaction* between a human and a given implementation of a device that affords the task in question. Components of the task may then be broken down into fundamental tasks that are afforded by natural human physiology and physics, in a situation *without* a technical device or intervention.

4.2 Background: Human Factors in Visual AR

In a study of human factors in visual AR, Livingston [27] performed a *domain analysis* (from software engineering) to identify tasks performed using an AR system at various levels of human functionality such as perceptual, cognitive and higher level tasks. Chapter 4 in this book provides an example of this approach towards the pursuit of "X-ray vision" using visual AR. Chapter 5 also discusses the human factors of mobile AR that are related to embodied cognition, such as physical interaction.

Livingston claims that the two most important means of studying AR applications are: to compare them with traditional methods for achieving a given cognitive task (such as navigation by way-finding); and to understand the perceptual bases of the AR interface in order to improve its suitability for given tasks from a more fundamental level. This approach led to a strategy of testing the perception and task-based performance using only the well-designed features of an AR interface. By identifying "canonical" perceptual and cognitive sub-tasks of common higher-level cognitive tasks, techniques can be transferred "between applications and across perceptual mechanisms."

Given examples of canonical tasks are recognition, resolution or description of an object by itself or amongst others, prediction of future positions of an object, navigation to or in-relation-to an object, distance and direction identification, identification of alert signals, and manipulation of the position or orientation of objects. Most or all of these tasks can be performed to some degree via different senses or perceptual mechanisms such as visual, auditory or tactile modes.

After identifying the canonical perceptual tasks, a set of AR system and user performance requirements can be determined for those tasks, and thus the successful use of the system for higher-level tasks. In order to understand performance of perceptual tasks in AR, it is often necessary to compare with performance of an equivalent real-world task. For instance, resolution and identification of objects can be tested in using visual acuity tests such as the standard Snellen eye chart. In audition, the localization test is an analogous example.

If a canonical task is isolated from extraneous factors so as to function as a control condition in an experiment, then it can be analysed in terms of the user and system parameters that effect performance. While the performance of an isolated perceptual task using real stimuli relies only on an individual's ability, in AR, the equivalent perceptual task is mediated by the technology that implements the AR system. Thus, task performance in AR involves both the individual's inherent ability, and some effect of the technological and system parameters of the specific AR implementation.

In order to discuss the research of human factors of audio AR, we need knowledge of the human characteristics that are involved in interactions with an audio AR system. We also need to understand the technical characteristics of a given audio AR system implementation. Human factors then arise during interaction of natural human ability and system characteristics, for all tasks afforded by the human-device-environment "ecosystem."

4.3 Characteristics of Natural Spatial Hearing

Amongst audio AR media, the most sophisticated are those that use 2D or 3D spatial audio for the augmentation of reality in any sensory modality. For sonic stimuli that contain no spatial information (mono sound sources), humans have the fundamental ability to detect individual sound sources within the sonic gestalt. Higher-level abilities then include the ability to comprehend dynamic pitch, timbre, rhythm, loudness and higher-level semantic information from the detected sound sources.

When some spatial information is available in the sonic stimuli (as for stereo, 2D and 3D stimuli), in addition to source detection, humans have the ability to *localize* (determine the direction and distance of) the sound source in space to a varying degree of accuracy, depending on the particular circumstances. The fundamentals of human spatial hearing have been examined in detail in the research literature, and thoroughly reviewed by Blauert [7].

The innate human ability to hear sounds in space has been characterized beginning with the fundamental situation of a single, stationary sound source in *free-field, anechoic* conditions (an unenclosed, un-obstructed environment). Further studies have examined more complex situations involving multiple sound sources, source and/or listener motion, and reverberant environments. This has established the human ability and limitations involved in achieving particular tasks in relation to *real* sound sources, in ideal, and certain non-ideal situations.

Perception of dynamic, real sources or motion-interactive, virtual sources is affected by a further range of factors that require new ways of measuring perceptual performance in comparison to static spatial audio.

4.4 Human Factors of Synthetic Spatial Audio Presentation

The characterization of human spatial hearing abilities becomes more complicated when they are mediated by technology. Perceptual experiments occur in laboratories with seated, stationary listeners, and spatial stimuli that are either static, or interactive with head-turns only.

Free-field stimuli have been simulated on headphones using detailed measurements of the acoustic response between sound sources and listeners' ear canals. Such filters are known as *head-related transfer functions* (HRTFs) when represented in the frequency domain. These techniques enabled subsequent researchers to implement systems that produced real-time, head-turn-interactive synthetic binaural audio [3, p. 57] that is used in PLASA systems.

The range of factors is wide indeed, encompassing any environmental variable that might alter spatial hearing perception for real or synthetic situations. For example, recently studied factors include: multiple simultaneous source masking [5], stimulus temporal features [8] or spectral features [21], source motion [10], the reverberant acoustic environment [4], multi-modal stimuli (usually visual and aural; [60]), and the numerous limitations of spatial sound synthesis techniques [25, 39].

These factors become highly influential in the human factors of audio AR. New factors often require a new performance measure to quantify the perceptible effects of the stimulus factor in question. The range of performance measures could be organised along an axis from fundamental (such as source azimuth, elevation or distance) to high-level (such as the "presence," "realism" or "stability" of a stimulus). Some measures of performance are described below.

Lateralization describes the kind of localization that occurs when auditory events are heard within the head—e.g. for headphone-delivered synthetic sources generated using only interaural time and intensity differences [20].

Absolute localization relates points in the sound source space to those in the auditory space—i.e. from the physical to the perceptual [6, p. 38]. This ability can be further characterised by localization and localization blur in the horizontal plane, median plane and in distance. Factors noted to influence localization include the source position itself, signal type, familiarity, spectral content, amplitude envelope and duration, previous sonic events, listener head movements, critical listening experience and individual differences. Even the response method used to measure localization is itself a potentially influential factor [44].

Ambiguous localization or the "cone of confusion" is a phenomenon in human spatial hearing whereby the left–right symmetry of the human head and resulting ambiguity in the fundamental cues to localization means that errors can occur as *front–back* or *up–downconfusion*. The main cue to resolve these confusions is the variation of the binaural signals with head-movements. Wightman and Kistler [55] also showed that front–back disambiguation was achievable if intended sound source movements are known by the listener—for instance if they control the source position. This shows that front–back resolution is to some degree a cognitive process.

Externalization is the perceived quality of an auditory event by which the associated sound source seems located at some distance away from the listener's head. The converse perceptual quality is called "inside-the-head locatedness" (IHL; [6], 116–160). Blauert notes that while this occurs often for headphone-presented signals, it also occurs for particular natural situations, such as the experience of humming with one's mouth closed and ears blocked. Factors noted to affect IHL and distance perception include signal type, familiarity, prior sounds, and listener individual differences, experience and expectations. A classic example of the influence of signal type, familiarity and experience is that whispered speech is systematically perceived to be at closer distances than shouted speech, regardless of other distance cues, or the amount of experience with the system [45].

5 Human Factors of Motion-Interactive Synthetic Spatial Audio

This category of spatial audio synthesis includes mobile sources with or without head-tracking, or static sources using head-tracking. The human factors of head-tracked spatial audio all apply to SAAR and PLASA as well.

Minimum audible movement angle (MAMA) measures the smallest detectable source motion below which a source appears to be stationary. MAMA depends on source velocity, source azimuth and duration, with typical values between 5 and 30°. It also has implications for the minimum detectable latency of spatial audio in response to head-turns [54].

Localization response time (RT) is the time to localize a sound source. This depends on source azimuth, the listener's head's rotational velocity, stimulus duration, the definition of *when* localization has been achieved, and the total system latency to head-turns. RTs can also be used to assess performance of tasks other than localization, such as various pointing tasks. Fitts' Law [13] predicts that RTs for general pointing tasks increase with decreasing target size and increasing target distance. The pointing task is an analogous task to that of using a PLASA system to navigate to the location of a virtual sound source by walking, for which mean velocity (related to RT) has been shown to depend on the "capture radius" of the sound source [38].

Perceptible effects of total system latency to head-turns (TSL) are an important factor in head-tracked binaural rendering systems. TSL to head-turns is the time between when the user turns their head and the corresponding change of spatial audio rendering. This interactive rendering is relative to the world reference frame, rather than the listener's reference frame. The TSL is a technical system parameter that translates into a human factor in terms of its *detection threshold* (DT, or just-noticeable effect) and *difference limen* (DL, or just-noticeable difference). Various studies have found measurable perceptual effects of TSL to head-turns between 32 and 250.4 ms, for different tasks, covering DTs and DLs, subjective latency ratings, localization errors, front–back error rates and RTs. For instance, Brungart et al. [9] found that latencies up to 243 ms for continuous stimuli barely affected localization error, while RT increased significantly for latencies of 73 ms or greater, with stimulus durations over 500 ms. For continuous stimuli, listeners can achieve accurate localization by adapting their behaviour and reducing head-turn speed, which increases RTs, while for stimuli briefer than the TSL, latency has little effect on localization RTs, but increases localization error.

5.1 Perception in Applied Contexts

When spatial audio is used as an interface or display of information, as in SAAR, task-based performance becomes an important measure of quality. If task performance were not to be affected by variation of a fundamental performance measure (e.g. localization blur) up to some threshold value, there might be no reason to ensure system performance to that level. Functional assessment of this kind can vary substantially depending on the user tasks or perceptual aspects being studied, yet these evaluations are often revealing and valuable.

5.1.1 Research by Loomis et al.

Since 1985 [30], Loomis and colleagues have evaluated auditory localization performance afforded by the Personal Guidance System (introduced in Sect. 2.3), which was designed for outdoors, audio-based navigation assistance for visually impaired people. The use of "virtual acoustic display" is considered a potentially powerful interface thanks to direct spatial mapping of positional information, and research on spatial hearing abilities in specific mobile task contexts is directed towards achieving this potential.

Task-based performance investigation has included comparisons between spatial audio presentation and other modes such as speech-only interfaces. Further research sought to evaluate the human ability to use spatial audio navigation cues during motion. An active localization experiment [31] compared navigation to real and virtual sound sources and validated the concept of navigation using simple binaural spatial audio synthesis. It was found that azimuth bearing of synthetic spatial sounds is perceived relatively accurately, especially when utilising head-turn interaction. However, distance is often poorly estimated [49], therefore many later experiments examined distance perception of real and virtual sources.

Work by Loomis and colleagues is important to the research of mobile audio AR because it provides patterns for the design of novel experiments on motion-interactive, multimodal spatial audio perception.

6 Human Factors in Audio AR

Many projects have been developed and written up in audio AR since the early days of Cohen et al. [12] and Bederson [2]. Still there are relatively few that have strongly evaluated the relationship between various system design choices and the resulting user experience (which is discussed in Chap. 9). Most projects have been content with simple validation of the functionality of the system, discussion of potential usage, or a qualitative evaluation that doesn't link system parameters to performance outcomes. This section presents a broad, selective survey of works that have provided more substantial investigation into human factors of audio AR media.

6.1 Locative Audio

Of the audio AR media, locative audio is probably the most prevalent, as well as presenting the most potential for popularization, given that without the requirement for head-tracking, it is the easiest to implement on mass-market devices such as smart-phones. At the same time, this medium is represented by relatively little literature containing any substantial evaluation, possibly due to its relative accessibility, ease of implementation and perhaps an impression of self-evidential functionality. Many cite the basic human factor that audio interfaces (in comparison to visual interfaces) provide a sensory information channel that may be received in a passive, background manner, as described by Sawhney and Schmandt [48] for instance.

Mantell et al. [34] introduced an interesting application for smart-phones called Navinko, designed as "a social network and navigation system enabled with audio augmented reality for cyclists in Tokyo." In this unique application, it was necessary to create an interface that was usable while riding a bicycle, and could inform users about nearby landmarks (Points Of Interest) and other users in terms of their momentary relative distance, direction and speed. They found that prior art systems were unable to provide comprehensible information about all these characteristics of multiple POIs simultaneously. One solution they found was to use a simulated Doppler effect on the sounds assigned to POIs, employing a familiar physical sonic effect as a metaphor. This provided a way of mapping extra physical interaction parameters (relative speed and distance) to sounds without adding extraneous simultaneous sounds. However, this innovation had the disadvantage of requiring continuous sounds, which prohibits some sounds that might be more semantically appropriate for some objects.

Williams et al. [56] evaluated a system that enabled school children to design and create "soundscapes" in the outdoor environment. The potential future impact of the technology on children's spatial practice is discussed and the concept of children "tagging" environmental hazards is raised.

This study provides a great deal of qualitative feedback on human factors of the individual and social user experience. For example, with a visual, screen-based interface on the user device, as well as the auditory interface, the visual feedback became a significant part of the experience:

> The children watched the screen often as they moved around, and enjoyed trying to steer the little dot that represented their position into the circles that represented sounds, rather than just using the headphones and coming across sounds.

Whilst the use of the screen as an extra user interface element in audio AR detracts from the purity of the medium, it is often desired for convenience or other reasons. When including this element, it is important to realise that it raises human factors of its own in terms of usability, attention and interaction quality.

6.2 ARA

As discussed previously in Sect. 2.2.2, an important question is which sensory modality of real-world stimuli is augmented. Different problems exist for augmentation of visible or audible objects. Both situations require perceptually accurate synthetic spatial audio, with precise and timely registration to real-world objects. However, augmentation of *acoustic* real objects requires perceptually matching the surrounding acoustical environment as well.

An example is the Wearable Augmented Reality Audio (WARA) research by Härmä et al. [17], in which synthetic audio augments the real acoustic environment around the listener. To achieve this *augmented audio reality*, Härmä et al. use headphones with embedded microphones facing outwards for active, real-time transmission of the sonic environment, mixed with the virtual audio. Perceptual evaluation in Härmä et al. [17] was limited but showed that in some cases, listeners found it very difficult to determine which sound sources were real or virtual.

Tikander [50] presented a study of the usability of an ARA headset, for long periods of use (20–40 h) in everyday life conditions, focused on issues with the active, equalised "hear-through" functionality of the device. While this device can be used to present virtual audio augmentations to the natural acoustic world, in this study no synthetic audio was added—this was suggested for future work. At the time, only the naturalness of the feed-through "pseudoacoustics" was investigated. Nevertheless, this is a valuable study in the human factors of a vital component of a complete ARA presentation system.

Many details of the usability of the pseudoacoustics were gathered. The most common annoyances related to sounds of the user being boosted due to an "occlusion effect" of internal sounds resonating in the ear canal—for example, while eating crunchy food. There were also handling inconveniences, social issues (due to the appearance of hearing being blocked by headphones), some discomfort with the in-ear headphones over time, and amplification of mechanical noises of the device, such as movement of the headset wires. Some loud sounds also caused undesirable distortion in the hear-through pseudoacoustic sound. Tests showed that mean opinion scores for annoyance of the experience improved after the field-trial period, in comparison to initial impressions. Acoustically, the headset was almost neutral and worked well for most occasions. Some difference to natural, open-ear listening was detectable in various circumstances, but was not clearly definable.

Martin et al. [40] validated two ARA headset systems with different earpieces, using localization errors, in comparison with standard virtual auditory space (VAS) technique and natural spatial hearing of real, free-field sound sources. They found that both headsets afforded the participants ability to localize sound sources in VAS as well as normal headsets. In regards to free-field localization, one earpiece afforded performance similar to normal hearing, although with poor low-frequency response, while the other earpiece interfered with normal localization, but provided a good low-frequency response.

Moustakas et al. [41] introduced an audio-only, two-player (or multiplayer) game prototype that was experienced as an ARA environment. The game consisted of hunting creatures that could only be identified by their sound, with further ambient sound objects in the real and virtual spaces, as well as the sonic presence of the other player(s).

They used questionnaires to assess the human factors of perceived user immersion; the perceptual significance of the real and virtual audio in terms of "game-play realization"; the effectiveness of interaction between players; and the overall expected and achieved user experiences.

The study showed that the game design achieved high levels of perceived user immersion, and great novelty to the players. Source localization ability was widely varied, and sound source identification was rated lower than other user experience factors that were examined. Self-reported (questionnaire results) and objective performance levels (ability to win games) were generally better for players with a background studying sound or music. This shows that listening skills may be important for faster adaptation and/or greater ability to use ARA systems.

In terms of presence, 80 % of players considered the presence of the remote player was significant within their immediate augmented audio environment. A slighter majority of 65 % felt that the real and virtual audio components equally supported the game scenario, while the rest felt the virtual component was dominant. This is evidence of the potential for high quality, immersive game-play built on ARA interaction environments.

Overall, the results of this evaluation were considered positive and encouraged the researchers to plan the development of more sophisticated game scenarios for ARA.

6.3 Spatial Audio AR and PLASA

In comparison with other forms of audio AR, relatively few SAAR or PLASA systems have been reported, particularly those with unrestricted mobility and outdoor functionality. There have been fewer published reports of human factors evaluation, particularly of a quantitative nature. However, several technical and perceptual factors are unique to this medium, so new experiments are required to understand the human factors of the new kinds of interactions that it affords.

6.3.1 Disambiguation of Front–Back Confusions by Body Motion inSAAR

In spatial hearing, front–back confusions are a well-known phenomenon whereby the listener incorrectly localizes a sound source to its mirrored position through the frontal plane (the vertical plane on the axis through the ears). This type of localization error can occur for real and synthetically spatialised sound sources. Experiments have shown that the listener can resolve front–back ambiguities by rotating their head; also that sound source movement can be used to resolve confusions if the listener is aware of the intended direction of source movement.

Mariette [36, 38] presented an experiment that shows mitigation of front–back confusions for synthetic binaural spatial audio interactive with body movement but not head-turns—i.e. SAAR without head-orientation tracking. This partly disabled mobile augmented reality system renders sound source *positions* relative to the *world* reference frame, (so the listener may walk past a stationary spatialised sound), but it renders instantaneous source *bearing* relative to the *listener's* reference frame.

As expected, front–back localization improves after the listener interacts with the spatialised sound by walking forward on a straight line past the source, so that it either looms towards, passes or recedes away from the listener. Dynamic localization cues of increasing source azimuth and changing source range enable the listener to constantly revise their judgment of a sound's location in front or behind them. An angular change in the sound source bearing while walking of between 12 and 16 □ give a significant improvement rate. This variation was observed to be similar to the expected static localization error rate for the rendering method that was used.

This suggests that higher resolution spatial sound synthesis will allow listeners to use smaller source azimuth and range changes to disambiguate front–back confusions.

Since the geometry of dynamic localization cues scales linearly with distance, the minimum source range for an acceptable localization correctness rate will then depend on the position tracking resolution. Expressed in another way, interactions exist between the audio AR system performance limitations of position tracking accuracy and computation power available for more accurate rendering. The weaker of these two specifications will determine the minimum source distance that allows acceptable front/back localization performance.

6.3.2 Validation of Navigation Using Motion-Interactive Spatial Audio

The first relevant study of navigation affordance was before the earliest use of the term audio AR, and involved motion-interactive, basic spatial audio used to represent world-stationary sound sources. Loomis et al. [31] created a "simple virtual sound display" with analogue hardware controlled by a 12MHz 80286 computer, video position tracking using a light source on the user's head, and head-orientation tracking using a fluxgate compass. Basic binaural spatial audio was rendered by synthesis of interaural time and intensity differences, with distance simulated using a filter for atmospheric attenuation and a variable ratio of direct to reverberated audio.

This experiment investigated human factors with regards to a navigation task, where participants had to "home-into" real or virtual sound sources around them. The real/virtual sound source condition had no significant effect on any of the performance measures: time to localise; distance error at perceived location; and absolute angular change during navigation. The path angular change was much larger for virtual stimuli (33°) than real stimuli (14 °), but this difference was not found to be statistically significant. Loomis et al. concluded that a simple virtual display could be "effective in creating the impression of external sounds to which subjects can readily locomote," but that more sophisticated displays might improve space perception and navigation performance.

That, and much other research has focused on basic verification of navigation ability. As Loomis et al. stated: "homing can be accomplished merely by keeping the sound in the median plane until maximum sound intensity is achieved."

A study by Jones et al. [23] investigated outdoors navigation of three routes on a university campus, using directional guidance provided by applying stereo-panning to music tracks according to the user's position, which was tracked using a GPS receiver. This study measured completion rate and mean time to complete each route, as well as ratings for the NASA *task load index* questionnaire. They found that simple panning was adequate for navigation of complicated routes, and users had a positive experience on several scales. They also found that the navigation completion rate suffered for routes on an open field in comparison to routes limited by pathways and buildings.

Some other studies such as those by Lokki et al. [28], Rutherford and Withington [47] and Walker and Lindsay [52] have used virtual auditory environments (VAEs) to evaluate navigation tasks using spatialised auditory beacons. However, these studies occurred indoors, with seated participants, so while they use related technology, they are not able to inform on human factors of audio AR.

6.3.3 Human Factors with Respect to Rendering Method and Head-TurnLatency in Navigation with SAAR

More recent studies premise that navigation *is* afforded by SAAR systems, then consider the human factors relating to technical specifications of the system, such as latency and rendering technique.

Mariette [35, 36] assessed human factors of an outdoor navigation task using a mobile SAAR system. One pilot study and a main experiment attempted to measure variations in the navigation performance afforded by the system. Two common technological limitations of such systems are the binaural rendering resolution, and latency between head-turns and corresponding rendered audio changes. The experiments investigated how several quantitative navigation performance measures and one subjective measure might be affected by the binaural rendering technique and head-turn latency. These are human factors unique to the SAAR environment, and technical parameters that are often limited by other aspects of the system design, such as available computing power, battery life, hardware financial budget, and the mobile device operating system.

In order to study the human factors of this situation, the task was to navigate from a central base position to the location of multiple virtual sound sources that were stationary in the world reference frame. This task was designed as a generalization of any navigation task using spatial audio beacons, since any continuous series of way-points (A–B–C–D–…) can be generalised as a series of point-to-point navigations (A–B, A–C, …). Thus the same experiment protocol could be used in the future to examine the effect of many other technical factors on human performance, and the supported affordance for any navigation activity.

For each stimulus, system parameters were varied, providing the technical factors under examination. Simultaneously, body position/orientation and head-orientation sensor data were recorded for later analysis. Novel objective performance measures were then used to assess the degradation of participants' navigation performance from the ideal, due to tested system parameter values.

The pilot experiment used only one participant, and found that a source *capture radius* of two meters significantly affected the user's navigation distance efficiency compared to other radii. The capture radius is specified as the distance from the exact source location within which the source is considered to be localised correctly. Decreasing capture radius significantly reduced distance efficiency ($p < 0.05$), showing that straighter navigation paths were supported by larger capture circles. Therefore, for better navigation performance in SAAR applications using similar technical specifications, source capture radius should be three meters or more.

The main experiment, using eight participants, found that render method significantly affected all performance measures except subjective stability rating, while head-turn latency only affected mean track curvature and subjective stability.

Overall results showed that regardless of sometimes severe system performance degradation, all eight participants successfully navigated to most source positions within a reasonable time limit of 60 s.

This experiment shows the unique human factors of mobile, SAAR in comparison to static spatial audio displays. When navigation performance is compared to static sound source localization, it is clear that improved localization performance is enabled by the perceptual feedback provided by system interactivity to user position movements and head-turns. While front–back confusions are common and azimuth errors are large in static experiments, in the SAAR, participants usually began their navigation in approximately the correct direction, and were almost always successful in walking to the source location.

Higher-precision binaural rendering significantly improved most navigation performance measures, while excessive head-turn latency only showed significant effects on stability rating and mean track curvature, and only then for the greatest latency level of 976 ms. No significant performance effects were found for latencies of 376 ms or worse.

The most interesting result was a statistically significant interaction effect between rendering method and latency, in which subjective stability degradation occurred for increased head-turn latency only when the higher-precision rendering was used. Apparently the lower resolution rendering mitigated the detrimental effect of high head-turn latency on perceived stability.

The primary conclusion for human factors of navigation using SAAR was that even systems with high head-turn latency or relatively low resolution rendering afforded successful navigation to positional sound sources. However, degradation of both specifications does have a significant detrimental effect on objective and subjective performance. Improved navigation performance is best supported by improving both system specifications at the same time, but within the parameters of this study, greater performance benefits were achievable by increasing rendering resolution than by reducing system latency. Mid-range latencies up to 376 ms TSL can be tolerated for any of the rendering methods used in the study.

6.4 Audio AR Implications for Human Factors of General Sound inSpace

Finally, in an almost circular extension of human factors research, audio AR has been used to understand the more general experience of sound in space. Kinayoglu [24] used audio AR to investigate how sound influences sense of place in terms of emotive, synaesthetic effects; attention and gaze behaviour; spatial orientation and sense of scale; audio–visual congruence in sense of place; and perception of personal and social space.

The experimental protocol compared the experience of space during free exploration of several zones within a university campus, separately using natural hearing and mobile, head-tracked SAAR. Participants rated their acoustic and multi-modal experience of each zone in terms of pleasantness, vibrancy/impressiveness, noisiness, relaxation, orientation, intimacy and familiarity of the environment. In some places, the sound design deliberately played with congruence with the surrounding visible environment. Results showed that soundscapes influence the experience of place, with effects on climatic, emotive, attentive, social and behavioural impressions of the spaces. In particular, sonic aesthetic qualities most strongly influenced emotional response to the environment, and semantic compatibility most strongly influenced the audio–visual congruence ratings, with incongruence resulting in anxiety and disconnectedness, even when using aesthetically pleasing sounds. This research shows that the semantic and aesthetic qualities of virtual audio content contribute to important human factors in the AR experience.

7 Conclusions

Human factors of audio AR media is a vital front of investigation in the field. Many experimental audio AR systems of various classifications have been presented to date, yet greater availability of consumer and commercial applications has only recently begun. At this stage in the development of the medium, it is important to understand how the component technologies, and variations of system design and specifications affect the user experience. This chapter has provided an introduction through classification of the media involved, a review of selected prior art, and a survey of human factors investigations. With continuing development along these lines, the utility and effectiveness of audio AR will further improve, and the field is likely to produce some thrilling future applications.

References

1. Azuma, R.: A survey of augmented reality. Presence: Teleoperators and Virtual Environment 6(4), 355–385 (1997)
2. Bederson, B.B.: Audio augmented reality: A prototype automated tour guide. In: CHI. pp. 210–211 (May 7–11 1995)
3. Begault, D.R.: 3-D Sound for Virtual Reality and Multimedia. Academic Press Professional, Cambridge, MA (1994)
4. Begault, D.R., Wenzel, E.M., Anderson, M.R.: Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. J of the Audio Eng Soc 49(10), 904–916 (2001)
5. Best, V.: Spatial Hearing with Simultaneous Sound Sources: A Psychophysical Investigation. Ph.D. thesis, Faculty of Medicine, University of Sydney (April 2004)
6. Blauert, J.: Spatial Hearing: The Psychophysics of Human Sound Localization. MIT Press, Cambridge, MA (1983)
7. Blauert, J.: Spatial hearing: The psychophysics of human sound localization (rev. ed.). MIT Press, Cambridge, MA (1997)
8. Brungart, D.S., Simpson, B.D.: Effects of temporal fine structure on the localization of broadband sounds: potential implications for the design of spatial audio displays. In: Int Conf on Aud Disp. Paris, France (June 24–27 2008)
9. Brungart, D.S., Simpson, B.D., McKinley, R.L., Kordik, A.J., Dallman, R.C., Ovenshire, D.A.: The interaction between head-tracker latency, source duration, and response time in the localization of virtual sound sources. In: Int Conf on Aud Disp. Sydney, Australia (July 6–9 2004)
10. Carlile, S., Best, V.: Discrimination of sound source velocity in human listeners. J of the Acoustical Soc of America 111(2), 1026 (February 2002)
11. Cohen, M.: Augmented Audio Reality: Design for a Spatial Sound GPS PGS. In: Center on Disabilities Virtual Reality Conf. California State University, Northridge (1994)
12. Cohen, M., Aoki, S., Koizumi, N.: Augmented audio reality: Telepresence/ar hybrid acoustic environments. In: IEEE Int Workshop on Robot and Human Communication. pp. 361–364 (1993)
13. Fitts, P.: The information capacity of the human motor system in controlling the amplitude of movement. J of Experimental Psychology 47(6), 381–91 (1954)
14. Galloway, A., Ward, M.: Locative Media as Socialising and Spatialising Practice: Learning from Archaeology. Leonardo Electronic Almanac 14(3) (2006)
15. Gibson, J.J.: In Perceiving, Acting, and Knowing: Toward an Ecological Psychology, chap. The theory of affordances. Lawrence Erlbaum Associates, Hillsdale, NJ (1977)
16. Härmä, A.: Ambient human-to-human communication. Handbook of Ambient Intelligence and Smart Environments pp. 795–823 (2010)
17. Härmä, A., Jakka, J., Tikander, M., Karjalainen, M., Lokki, T., Nironen, H., Vesa, S.: Techniques and applications of wearable augmented reality audio. In: 114th Conv Audio Eng Soc. Amsterdam, The Netherlands (March 22–25 2003)
18. Helyer, N.: Sonic landscapes (22/8/2006 1999–2001), http://www.sonicobjects.com/index.php/sonicobjects/more/sonic_landscapes/
19. Holland, S., Morse, D.R., Gedenryd, H.: AudioGPS: spatial audio in a minimal attention interface. Personal and Ubiquitous Computing 6(4), 253–259 (2002)
20. Jeffress, L.A., Taylor, R.W.: Lateralization vs Localization. J of the Acoustical Soc of America 32, 936 (1960)
21. Jin, C., Corderoy, A., Carlile, S., van Schaik, A.: Contrasting monaural and interaural spectral cues for human sound localization. J of the Acoustical Soc of America 115, 3124 (2004)
22. Jones, K.S.: What is an affordance? Ecological Psychology 15(2), 107–114 (2003)
23. Jones, M., Bradley, G., Jones, S., Holmes, G.: Navigation-by-Music for Pedestrians: an Initial Prototype and Evaluation. Proc of the Int Symp on Intelligent Environments: Improving the quality of life in a changing world (April 5–7 2006)

24. Kinayoglu, G.: Using audio-augmented reality to assess the role of soundscape in environmental perception. In: eCAADe27. Istanbul (2009)
25. Kyriakakis, C.: Fundamental and technological limitations of immersive audio systems. Proc of the IEEE 86(5), 941–951 (1998)
26. Lindeman, R.W., Noma, H.: A classification scheme for multi-sensory augmented reality. In: Proc 2007 ACM symposium on Virtual reality software and technology. pp. 175–178. ACM (2007)
27. Livingston, M.A.: Evaluating human factors in augmented reality systems. Computer Graphics and Applications, IEEE 25(6), 6–9 (November/December 2005)
28. Lokki, T., Grohn, M., Savioja, L., Takala, T.: A case study of auditory navigation in virtual acoustic environments. In: Int Conf on Aud Disp (2000)
29. Loomis, J.M., Golledge, R.G., Klatzky, R.L.: Personal Guidance System for the Visually Impaired using GPS, GIS, and VR Technologies. In: VR Conf. California State University, Northridge (1993)
30. Loomis, J.: Digital map and navigation system for the visually impaired (July 25 1985), http://www.geog.ucsb.edu/pgs/papers/loomis_1985.pdf
31. Loomis, J., Hebert, C., Cicinelli, J.: Active localization of virtual sounds. J of the Acoustical Soc of America 88(4), 1757–1764 (October 1990)
32. Lyons, K., Gandy, M., Starner, T.: Guided by voices: An audio augmented reality system (2000)
33. Mallem, M.: Augmented reality: Issues, trends and challenges. In: 2nd Int Conf on Image Processing Theory Tools and Applications (IPTA). p. 8 (July 2010)
34. Mantell, J., Rod, J., Kage, Y., Delmotte, F., Leu, J.: Navinko: Audio augmented reality-enabled social navigation for city cyclists. In: Pervasive (2010)
35. Mariette, N.: CMMR/ICAD 2009, chap. Navigation Performance Effects of Render Method and Head-Turn Latency in Mobile Audio Augmented Reality, pp. 239–265. Lecture Notes in Computer Science 5954, Springer, Berlin/Heidelberg (2010)
36. Mariette, N.: Perceptual Evaluation of Personal, Location-Aware Spatial Audio. Ph.D. thesis, University of New South Wales, Sydney, Australia (March 2011)
37. Mariette, N.: From backpack to handheld: The recent trajectory of personal location aware spatial audio. In: 7th Int Digital Arts and Culture Conf. Perth, Australia (15th–18th September 2007)
38. Mariette, N.: Mitigation of binaural front–back confusions by body motion in audio augmented reality. In: 13th Int Conf on Aud Disp. Montréal, Canada (June 26–29 2007)
39. Martens, W.L.: Perceptual evaluation of filters controlling source direction: Customized and generalized HRTFs for binaural synthesis. Acoustical Science and Technology 24(5), 220–232 (2003)
40. Martin, A., Jin, C., Schaik, A.: Psychoacoustic evaluation of systems for delivering spatialized augmented-reality audio. J of the Audio Eng Soc 57(12), 1016–1027 (December 2009)
41. Moustakas, N., Floros, A., Grigoriou, N.: Interactive audio realities: An augmented / mixed reality audio game prototype. In: 130th Conv Audio Eng Soc (May 13–16 2011)
42. Mynatt, E.D., Back, M., Want, R., Baer, M., Ellis, J.B.: Designing audio aura. In: Proc SIGCHI Conf on Human factors in Comp Sys. pp. 566–573. ACM Press/Addison-Wesley Publishing Co. (1998)
43. Novo, P.: Communication Acoustics, chap. Auditory Virtual Environments, pp. 277–297. Springer, Berlin–Heidelberg–New York (2005)
44. Pernaux, J.M., Emerit, M., Nicol, R.: Perceptual evaluation of binaural sound synthesis: the problem of reporting localization judgments. In: 114th Conv Audio Eng Soc (March 22–25 2003)
45. Philbeck, J.W., Mershon, D.H.: Knowledge about typical source output influences perceived auditory distance. J of the Acoustical Soc of America 111, 1980–1983 (2002)
46. Rozier, J., Karahalios, K., Donath, J.: Hear&There: An Augmented Reality System of Linked Audio. In: Int Conf on Aud Disp. Atlanta, Georgia (April 2000)

47. Rutherford, P., Withington, D.: The application of virtual acoustic techniques for the development of an auditory navigation beacon used in building emergency egress. In: Proc 2001 Int Conf on Aud Disp. Espoo, Finland (July 29–August 1 2001)
48. Sawhney, N., Schmandt, C.: Design of Spatialized Audio in Nomadic Environments. In: Proc Int Conf on Aud Disp (1997)
49. Speigle, J.M., Loomis, J.M.: Auditory distance perception by translating observers. In: Proc, IEEE Symp on Research Frontiers in Virtual Reality. pp. 92–99 (1993)
50. Tikander, M.: Usability issues in listening to natural sounds with an augmented reality audio headset. J Audio Eng Soc 57(6), 430–441 (2009)
51. Walker, B.N., Lindsay, J.: Navigation performance in a virtual environment with bonephones. In: Proc 11th Int Conf on Aud Disp. pp. 260–263. Limerick, Ireland (2005)
52. Walker, B., Lindsay, J.: Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. Human Factors 48(2), 265 (2006)
53. Warusfel, O., Eckel, G.: LISTEN – augmenting everyday environments through interactive soundscapes. In: IEEE Virtual Reality (2004)
54. Wenzel, E.M.: Analysis of the role of update rate and system latency in interactive virtual acoustic environments. In: 103rd Conv Audio Eng Soc. New York, NY, USA (September 26–29 1997)
55. Wightman, F.L., Kistler, D.J.: Resolution of front-back ambiguity in spatial hearing by listener and source movement. J of the Acoustical Soc of America 105(5), 13 (1999)
56. Williams, M., Jones, O., Fleuriot, C., Wood, L.: Children and emerging wireless technologies: investigating the potential for spatial practice. In: Proc SIGCHI Conf on Human factors in Comp Sys, April. pp. 02–07 (2005)
57. Woo, D., Mariette, N., Helyer, N., Rizos, C.: Syren – a ship based location-aware audio experience. J of global positioning systems, Int Assoc of Chinese Professionals in GPS 4(1–2), 41–47 (2005), alberta, Canada
58. Woo, D., Mariette, N., Salter, J., Rizos, C., Helyer, N.: Audio Nomad. In: ION GNSS. Fort Worth, Texas (September 26–29 2006)
59. Zhou, Z., Cheok, A.D., Qiu, Y., Yang, X.: The role of 3-D sound in human reaction and performance in augmented reality environments. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on 37(2), 262–272 (March 2007)
60. Zwiers, M., Van Opstal, A., Paige, G.: Plasticity in human sound localization induced by compressed spatial vision. Nature Neuroscience 6(2), 175–181 (2003)