J. Yang, A. Barde and M. Billinghurst, “Audio Augmented Reality:
A Systematic Review of Technologies, Applications, and Future Research Directions”
J. Audio Eng. Soc., vol. 70, no. 10, pp. 788–809, (2022 October).
DOI: https://doi.org/10.17743/jaes.2022.0048

Audio Augmented Reality: A Systematic Review of Technologies, Applications, and Future Research Directions

JING YANG,[1] AMIT BARDE,[2] AND MARK BILLINGHURST[2]
(jing.yang@inf.ethz.ch) (amit.barde@auckland.ac.nz)(mark.billinghurst@auckland.ac.nz)

[1] *Department of Computer Science, ETH Zurich, Switzerland*
[2] *The Empathic Computing Laboratory, Auckland Bioengineering Institute, The University of Auckland, New Zealand*

Audio Augmented Reality (AAR) aims to augment people’s auditory perception of the real world by synthesizing virtual spatialized sounds. AAR has begun to attract more research interest in recent years, especially because Augmented Reality (AR) applications are becoming more commonly available on mobile and wearable devices. However, because audio augmentation is relatively under-studied in the wider AR community, AAR needs to be further investigated in order to be widely used in different applications. This paper systematically reports on the technologies used in past studies to realize AAR and provides an overview of AAR applications. A total of 563 publications indexed on Scopus and Google Scholar were reviewed, and from these, 117 of the most impactful papers were identified and summarized in more detail. As one of the first systematic reviews of AAR, this paper presents an overall landscape of AAR, discusses the development trends in techniques and applications, and indicates challenges and opportunities for future research. For researchers and practitioners in related fields, this review aims to provide inspirations and guidance for conducting AAR research in the future.

0 INTRODUCTION

Augmented Reality (AR) technology aims to seamlessly blend computer-generated virtual content with the physical environment so that the virtual content appears fused with the real world [1]. AR can enhance people’s perception of and interaction with their surroundings and can also help them more easily perform real-world tasks [2]. Over the past few decades, technological advancements have significantly increased the adoption of AR technology in a wide range of application domains, including industrial maintenance [3, 1, 4, 5], education [6–8], gaming [9, 10], and collaborative work [11–15].

An important affordance of AR technology is its ability to augment human senses [16] so that people can interact with virtual objects and scenes as easily as they do with the physical world. However, the overwhelming majority of AR research has been focusing on visual augmentation [17–19]. Audio Augmented Reality (AAR) remains relatively under-explored. In AAR, virtual auditory content is blended into the physical world to augment the user’s real acoustic environment. To mimic real-world auditory perception, virtual sounds are usually binaurally spatialized (along with reverberation if required) to create a realistic sense of direction and distance.

AAR technology has a remarkable potential for creating effective and immersive AR experiences, for the following reasons:

1. Diversity of sound content: Different types of audio content can provide the user with rich information. For example, speech conveys information to address questions and issue commands, whereas non-speech beacons and alerts inform operation status of applications and notify users of new messages [20]. Sounds created for AAR applications are thus capable of conveying a range of information based on the context in which they occur.
2. Localization and immersive experience: Given human binaural hearing in three dimensional space [21], AAR can provide users with an enhanced sense of immersion by spatializing virtual sounds that recreate distance, direction, and spectral cues, which can play a critical role in several situations.

For example, if AAR applications provide alarms, users may first hear a threat to their safety and move away from the danger promptly.

3. Ubiquitous hardware: Hardware capable of delivering AAR experiences is readily available. For example, mobile devices with powerful computing capabilities (like smartphones) can be used to compute virtual sounds and deliver high-quality immersive audio experiences. Users can access these experiences using readily available off-the-shelf headphones [20]. These devices can be conveniently used to deliver AAR experiences [20].

Given these reasons, it is unsurprising that AAR technology has begun to attract a greater amount of research interest. The availability of devices that are capable of delivering real-time AAR experiences has further spurred interest in this field. For example, Apple AirPods Pro,[1] Samsung Galaxy Buds Pro,[2] and JBL Quantum ONE[3] can enable accurate spatialization of virtual sounds because of their integrated modules for head tracking.

Overall, AAR is a promising field yet still relatively under-studied in the wider AR community. Moreover, technologies required to realize AAR make it more difficult to implement audio augmentation in AR scenarios than in Virtual Reality (VR) scenarios. More specifically, in VR, using pre-designed virtual scenes can simplify the rendering of audio content, whereas creating virtual sounds in the physical world and adapting them to the user in real time is more complicated in AR. For example, the user's pose with respect to the environment should be tracked to spatialize sounds properly, and the environmental acoustics should be updated according to the user's movements in the space. AAR technologies and AAR usability still need to be investigated and considerably improved in order to be widely applied and accepted by end users.

This paper, as one of the first surveys of its type, aims to provide a systematic overview of AAR. It aims to motivate the wider AR community to actively consider audio augmentation in the delivery of informative and immersive experiences. This review focuses on spatialized rather than monophonic virtual sounds. To better reflect the auralization process in real-world situations, the integration of simulating the real environmental acoustics is also considered in this review. In summary, this paper makes the following contributions:

1. Providing one of the first comprehensive summaries to facilitate a systematic understanding of AAR technology and its development over the past few decades.
2. With a focus on spatial sound-related AAR, this paper identifies five functional components of AAR systems and discusses techniques to implement these functions.
3. Based on published studies, this paper identifies seven application domains where AAR has shown to be practical or has the potential to make a significant difference.
4. Discussing future research challenges and opportunities for making AAR more beneficial and acceptable.

The rest of this paper first explains the methods employed for paper selection and the review process. Following this, technologies used for developing AAR systems are reviewed. The application domains of AAR technology are then reviewed, and finally, future research directions to advance AAR development are discussed.

1 METHODOLOGY FOR PAPER SELECTION AND REVIEWING

This section outlines the process for selecting and reviewing papers. The potential limitations of this process are also discussed.

1.1 Paper Selection and Review

This survey paper aims to provide a comprehensive review of the existing AAR landscape. The Scopus bibliographic database was first searched, and then Google Scholar was searched to include more related work, both of which have been commonly used for previous AR reviews [17, 22, 23].

Papers published in conferences and journals up until November 2021 were considered. A start date for the search was not specified in order to cover early works as well. Table 1 lists the search terms used for paper collection. Note that the search terms cover two distinct aspects of AAR:

1. Technologies: This part covers the different technologies that have been used to realize AAR. To binaurally spatialize virtual sounds, AAR systems should include three functional components: *user-object pose tracking*, *room acoustics modeling*, and *spatial sound synthesis*. Two other important technologies for creating AAR systems are also reviewed: *interaction technology* and *display technology*. Interaction technology refers to how the user provides input (e.g., touch screen input) to enable or adjust AAR applications. Display technology refers to how the virtual sounds are output to end users (e.g., only audio via earphones, through handheld displays together with visual content).
2. Application domains: This part covers AAR applications over a given time period in a number of real-world use cases. The use of generic search terms such as "user study" and "experiments" allowed for gathering a larger set of papers and examine the various types of AAR applications proposed and/or implemented by researchers.

---

[1] https://www.apple.com/airpods-pro/.

[2] https://www.samsung.com/us/mobile/audio/galaxy-buds-pro/.

[3] https://www.jbl.com.sg/gaming/QUANTUMONE.html.

Table 1. Search terms used for collecting publications.

Technologies	"Audio Augmented Reality" "Augmented Reality" AND "Audio Augmentation" "Augmented Reality" AND "Head Pose Tracking" "Augmented Reality" AND "User Pose Tracking" "Augmented Reality" AND "Pose Tracking" "Augmented Reality" AND "Acoustics Modeling" "Augmented Reality" AND "Room Acoustics" "Augmented Reality" AND "Acoustic Effect(s)" "Augmented Reality" AND "3D/Spatial Sound Synthesis" "Augmented Reality" AND "3D Audio/Sound" "Augmented Reality" AND "Spatial Audio/Sound" "Augmented Reality" AND "HRTF" "Augmented Reality" AND "Audio/Auditory Interaction" "Audio Augmented Reality" AND "Interaction" "Augmented Reality" AND "Audio/Auditory Display" "Audio Augmented Reality" AND "Display"
Application domains	"Audio Augmented Reality" AND "Study/-ies" "Audio Augmented Reality" AND "User Study/-ies" "Audio Augmented Reality" AND "Pilot Study/-ies" "Audio Augmented Reality" AND "Experiment(s)"

HRTF = head-related transfer function.

The terms were searched in the title, abstract, and keywords fields to identify relevant literature. The full text of each paper was read to identify its suitability for this survey, and papers were excluded for two reasons: 1) The primary research theme/objective of the paper had little to do with the reviewed topics. 2) A few works added monophonic sounds, whereas this survey focuses on spatial virtual sounds. For example, [24] provided monophonic audio description in a tourist guide application that users heard through earphones.

After filtering out papers according to their research theme and content, the impact of each remaining paper was considered to ensure that representative and influential work was being reviewed. To this end, every paper's average citation count (ACC) [17] was calculated. For papers published before 2020, 96 papers with ACC ≥ 2.0 were included. The remaining papers published in 2020 and 2021 were chosen to be included regardless of their ACC because most of them were still too new to accrue a significant citation count.

Overall, 117 papers were reviewed, of which 62 presented a complete AAR system through which the user can perceive virtual spatialized sounds in the given scenes. From these 62 papers, the technologies used to implement AAR and the domains in which they were applied are summarized. The remaining 55 papers did not demonstrate the development and/or use of complete AAR systems. Instead, they proposed algorithms, methods, or techniques for implementing one specific AAR component. The proposed techniques were not adequately covered in the existing complete AAR systems, therefore, these papers are also reviewed in the related sections.

1.2 Limitations

Two limitations that might influence the thoroughness of this review are identified. First, this review focuses on research studies rather than commercial practices, so some works that can also been involved (e.g., some white papers and patents) might not be covered by the Scopus database and Google Scholar. Second, the authors strove to collect all related papers by using the search terms in Table 1. However, some papers might only use other keywords such as "Mixed Reality" to describe AR-related research. Nevertheless, the search terms should have covered a large proportion of the work that is relevant to AAR technology and its applications.

2 MAJOR TECHNOLOGIES FOR CREATING AAR

This section discusses the technologies used to develop the AAR systems mentioned in the reviewed literature. Previous research [25] has indicated that *tracking*, *interaction*, and *displays* form the main components of typical AR systems. In addition to these, two technology components are specifically needed for AAR: *room acoustics modeling* and *spatial sound synthesis*. For papers that presented a complete AAR system, the technologies used for each component are summarized in Table 2. Note that some works did not specify or include some technology components, so these were represented by "···" in the table.

Fig. 1 shows percentages of the methods that were used to realize each technology component of the AAR systems listed in Table 2. In the remainder of this section, a detailed review for each of these five technologies is provided.

2.1 User-Object Pose Tracking

In the context of AAR, to guarantee that virtual sounds are correctly spatialized from real locations, *tracking* specifically refers to tracking the user's location, orientation, and relative pose to the desired audio source. Table 2 shows that 43% of the works implemented tracking using a *single type of sensor*, among which visual tracking is the most commonly used method (58%). For AAR systems using visual tracking, a typical approach is to detect and track visual features from input video frames to calculate the current pose. Although some works employ natural images captured from unmodified environments (e.g., [26, 27, 28]), others exploit specifically designed fiducial markers (i.e., image markers that serve as references) that are pre-allocated in the space (e.g., [29, 30, 31]).

Table 2. Summary of the technologies used in the 62 complete AAR systems. Categories in each column are clarified in more detail in the corresponding sections.

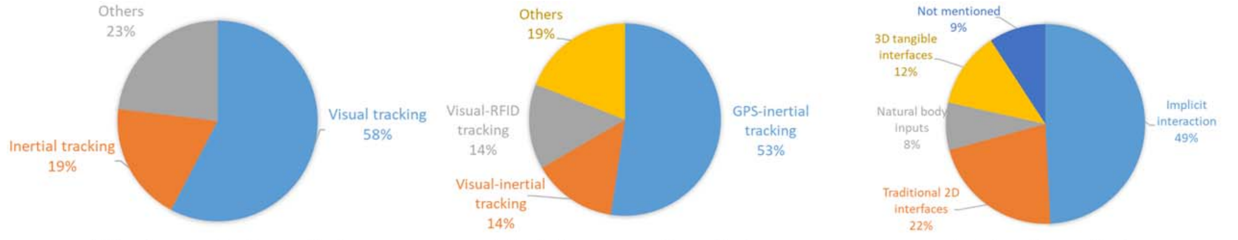
Paper	Tracking method	Interaction method	Display method	Acoustics modeling	Sound spatialization
Bederson [55]	Infrared tracking	Implicit	Audio only	···	···
Mynatt et al. [56]	Infrared tracking	Implicit	Audio only	···	···
Behringer et al. [88]	GPS-inertial tracking	Voice input/game pad	HMD	···	AudioTechnica ATW-R100 receivers
Sawhney and Schmandt [89]	Static head position	Voice input/button	Audio only	···	···
Walker et al. [83]	···	Mouse scroll	Audio only	···	Microsoft DirectX with generic HRTFs
Härmä et al. [102, 103]	···	Implicit	Audio only	Artificial reverberation	Measured BRIRs
Sundareswaran et al. [98]	GPS-inertial tracking	Implicit	HMD	···	Off-the-shelf engines with generic HRTFs
Tachi et al. [99]	Visual tracking	Implicit	HMD	···	···
Hatala et al. [68]	RFID-visual tracking	3D tangible interface	Audio only	···	···
Terrenghi and Zimmermann [57]	RFID tracking	Implicit	Audio only	Artifical reverberation	Generic HRTFs
Zhou et al. [29]	Visual/acoustic-inertial tracking	Implicit	HMD	···	OpenAL with generic HRTFs
Zhou et al. [90]	Visual tracking	Foldable AR book	HMD	···	···
Zotkin et al. [105]	Visual tracking	···	PC display	Computed IRs	Selected HRTFs
Hatala and Wakkary [69]	RFID-visual tracking	3D tangible interface	Audio only	···	···
Walker and Lindsay [81]	···	Keyboard input	Audio only	Artificial reverberation	Generic HRTFs
Fröhlich et al. [145]	GPS-inertial tracking	Implicit	Audio only	Outdoor	···
Sodnik et al. [30]	Visual tracking	Implicit	HMD	···	OpenAL using generic HRTFs
Tonnis and Klinker [79]	···	Implicit	Head-up display	···	Surrounding speakers
Walker and Lindsay [61]	GPS-inertial tracking	Implicit	Audio only	···	Generic HRTFs
Grasset et al. [38]	Visual tracking	3D tangible interface/gaze input	Handheld display	···	···
Liarokapis [82]	Visual tracking	Keyboard/mouse/touch screen	HMD	···	OpenAL using generic HRTFs
Stahl [62]	GPS-inertial tracking	Slider on GUI	Mobile device	Outdoor	···
Wakkary and Hatala [70]	RFID-visual tracking	3D tangible interface	Audio only	···	···
Wilson et al. [146]	GPS-inertial tracking	2D scrolling interface	Audio only	Outdoor	···
Zimmermann and Lorenz [58]	RFID tracking	Implicit	Audio only	Artifical reverberation	···
Heller et al. [72]	UWB-inertial tracking	Implicit	Audio only	···	OpenAL using generic HRTFs
Kern et al. [80]	···	Implicit	PC display	···	···
Blum et al. [91]	GPS-inertial tracking	3D tangible interface	Audio only	Outdoor	OpenAL using generic HRTFs
Katz et al. [65]	Visual-inertial tracking	Implicit	Audio only	Outdoor	···

(*continued on next page*)

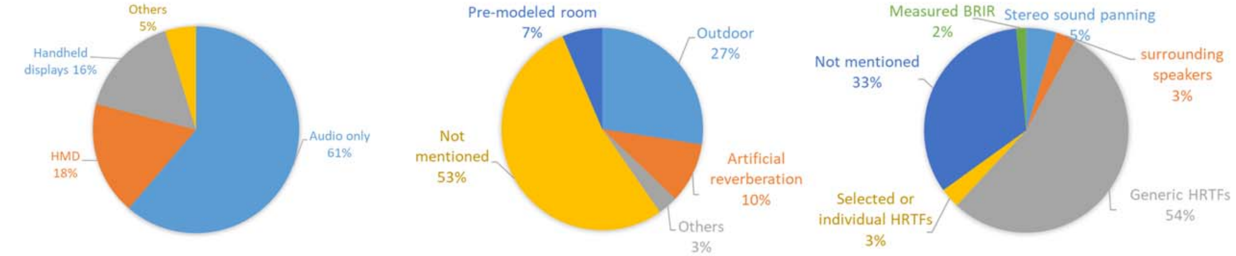
Table 2. *(Continued)*

Paper	Tracking method	Interaction method	Display method	Acoustics modeling	Sound spatialization
McGookin et al. [84]	GPS-inertial tracking	Touch screen	Mobile device	Outdoor	···
Ribeiro et al. [66]	Visual-inertial tracking	Implicit	Audio only	Pre-modeled room	Generic HRTFs
Vazquez-Alvarez et al. [92]	GPS-inertial tracking	3D tangible interface	Audio only	Outdoor	JAVA JSR-234 using generic HRTFs
Blum et al. [51]	Inertial tracking	3D tangible interface	Audio only	···	PureData using generic HRTFs
Langlotz et al. [71]	GPS-visual tracking	Touch screen	Mobile device	Outdoor	Stereo sound panning
de Borba Campos et al. [123]	···	···	Audio only	···	Stereo sound panning
Heller et al. [78]	Retroreflective tracking	Implicit	Audio only	Artificial reverberation	OpenAL using generic HRTFs
Blessenohl et al. [26]	Visual tracking	Implicit	Audio only	···	Generic HRTFs
Ruminski [31]	Visual tracking	···	Mobile device	···	···
Chatzidimitris et al. [59]	GPS tracking	Touch screen	Mobile device	Outdoor	OpenAL using generic HRTFs
Heller et al. [52]	Inertial tracking	Implicit	Audio only	···	KLANG using generic HRTFs
Russell et al. [73]	UWB-inertial tracking	Implicit	Audio only	Outdoor	3DCeption using generic HRTFs
Heller and Schöning [63]	GPS-inertial tracking	Implicit	Audio only	···	KLANG using generic HRTFs
Kim et al. [86]	···	Touch screen	HMD	···	···
Lim et al. [85]	···	Touch screen	Mobile device	Outdoor	···
Schoop et al. [27]	Visual tracking	Implicit	Audio only	Outdoor	Stereo sound panning
Sikora et al. [64]	GPS-inertial tracking	Touch screen	Audio only	Outdoor	Generic HRTFs
Huang et al. [100]	Visual tracking	···	HMD	Outdoor	Generic HRTFs
Rovithis et al. [60]	GPS tracking	Gesture control	Mobile device	Outdoor	SceneKit using generic HRTFs
Yang et al. [37]	Retroreflective tracking	Implicit	Audio only	Pre-modeled room	Generic HRTFs
Bandukda and Holloway [148]	···	Implicit	Audio only	···	···
Cliffe et al. [106]	Visual tracking	Implicit	Audio only	Pre-recorded soundscape	Generic HRTFs
Joshi et al. [149]	···	···	Audio only	···	···
Kaghat et al. [53]	Inertial tracking	Gesture control	Audio only	···	Generic HRTFs
Lawton et al. [122]	···	···	Audio only	Outdoor	Surrounding speakers
Mattheiss et al. [104]	···	Implicit	Audio only	Artificial reverberation	Individual and generic HRTFs
May et al. [147]	···	Implicit	Audio only	···	Generic HRTFs
Sagayam et al. [87]	Visual tracking	Touch screen	Mobile device	Pre-modeled room	Generic HRTFs
Yang et al. [101]	Visual tracking	Implicit	HMD	···	Generic HRTFs
Chong and Alimardanov [169]	···	Implicit	Audio only	Outdoor	Generic HRTFs
Comunita et al. [67]	Visual-inertial tracking	Implicit	Mobile device	Pre-modeled room	Generic HRTFs
Guarese et al. [54]	Inertial tracking	Implicit	HMD	···	Generic HRTFs
Kaul et al. [28]	Visual tracking	Implicit	Audio only	···	Generic HRTFs

*AAR = audio augmented reality; HMD = head-mounted display; HRTF = head-related transfer function; IR = impulse response; RFID = radio frequency identification; UWB = ultra wideband.



(a) Single-sensor tracking methods (b) Hybrid-sensor tracking methods (c) Interaction methods



(d) Display methods (e) Room acoustics modeling (f) Spatial sound synthesis

Fig. 1. Overview of the technologies used in the reviewed audio augmented reality (AAR) systems. BRIR = binaural room impulse response; HMD = head-mounted display; HRTF = head-related transfer function; RFID = radio frequency identification.

The fact that visual tracking is most popular could be largely due to the development of computer vision techniques. More specifically, advancements in vision sensors (e.g., stereoscopic depth cameras [32] and RGB-D camera sensors [33–36]) have enabled increasingly accurate pose tracking. Moreover, computer vision tracking algorithms have been studied for decades and are relatively mature. Furthermore, visual tracking can be implemented in many forms (e.g., inside-out tracking [26], outside-in tracking [37], tracking for small-scale desktop scenes [38] and large-scale outdoor places [27], etc.). Therefore, visual tracking can be used in many different application scenarios. In the AR community, some visual tracking methods have been used in AR systems for vision augmentation [39, 32, 40–43, 33, 34, 44, 45, 35, 46–50, 36], and they are also feasible for AAR systems.

Inertial sensors appear to be the second-most–used sensors in AAR systems that used a single sensor type for pose tracking (19%) [29, 51, 52, 53, 54]. Inertial sensors usually involve accelerometers and gyroscopes to track the position and orientation of an object relative to a known starting point, orientation, and velocity. Note that because the drift issue may cause significant errors especially for position tracking, these AAR systems used inertial sensors for the three-degrees-of-freedom orientation tracking in a small area where the user barely changed location.

In addition to visual and inertial sensors, some other sensors have also been employed for AAR tracking. Two indoor AAR systems used infrared transmitters and receivers [55, 56] and another two indoor AAR systems used radio frequency identification (RFID) technology [57, 58]. Furthermore, GPS has been used in two outdoor AAR systems to determine the user's location [59, 60].

Although some AAR systems have successfully used a single sensor type to track the user's pose, each sensor type has limitations that might restrict its deployment in some scenarios. For example, visual tracking is not feasible in low-light environments, in spaces with occlusions, or when power consumption is critical [20]. To achieve robust tracking in various environments, researchers have also explored hybrid techniques that fuse several kinds of sensors.

Table 2 shows that 34% of the AAR systems employed hybrid pose tracking approaches, with the most popular combination being GPS and inertial sensors (53%; e.g., [61–64]). These implementations were typically designed for large spaces or outdoor environments, where GPS sensors can be used for localizing the user's position and the inertial sensors can be used to determine the user's head orientation. The reviewed AAR systems demonstrate another five hybrid tracking approaches. A total of 14% of the AAR systems fused visual and inertial sensors [65–67], and 14% fused visual and RFID sensors [68–70]. Additionally, one AAR system used GPS-visual tracking [71], whereas another demonstrated acoustic-inertial tracking [29]. Two AAR systems used an implementation of ultra-wideband–inertial tracking [72, 73].

It can be seen that hybrid tracking improves tracking accuracy by combining the strengths of different sensor types. For example, inertial sensors are commonly used for measuring orientation in the wider tracking community, and 81% of the hybrid methods exploited inertial sensors. GPS and visual sensors are more commonly used for determining the user's position. Although visual sensors are more flexibly used for different scales of scenes, GPS fits better in large environments or outdoor spaces.

Overall, the AAR systems show that mainstream user-object pose tracking approaches are based on visual, inertial, and GPS sensors. This trend is, in general, aligned with the popular tracking methods in the wider AR community. However, the authors find it surprising that acoustic tracking and acoustic-inertial tracking, which have been used in a few AR systems [74–76], were rarely exploited by AAR systems. When using acoustic sensors for pose tracking, acoustic signals can be emitted from one or several sound sources and received by microphones that are attached on the object or user to be tracked. The authors see the potential to employ acoustic sensors for pose tracking in AAR systems. As some earbuds are now equipped with acoustic and motion sensors [77], using acoustic or acoustic-inertial tracking might be a lightweight and convenient solution for implementing AAR systems on wearable devices. More will be discussed about acoustic sensor-based tracking as a future research direction in SEC. 4.

2.2 Interaction Technology

Interaction refers to how the user initiates or responds to an AAR system and how the system dynamically reacts to the user's actions. To this end, AAR systems must incorporate input methods that allow users to choose and activate the audio augmentation content or adjust the presentation of the audio augmentation (e.g., adjust the volume).

Among the reviewed AAR systems, 49% of them implemented "implicit" interaction. Implicit interactions with AAR systems refer to those that are not actively initiated by the user. Instead, the system reacts to the user's surroundings and their actions in the environment and provides the desired virtual sounds. Users' actions serve as the only inputs to the AAR system. Such implicit interaction is typically used in localization and navigation services (e.g., [30, 72, 78, 27]). Because the user does not need to control devices or objects to provide explicit commands, implicit interaction is convenient and helpful for visually impaired individuals (e.g., [26, 54]) or when the user is engaged in an attention-critical task (e.g., driving [79, 80]). However, implicit interaction does not allow users to flexibly control the audio augmentation at will.

More proactive interaction modes are shown in the rest of the AAR systems. Traditional 2D user interfaces were implemented in 22% of the AAR systems, such as a GUI [62], keyboard [81, 82], mouse [83, 82], and touch-screen input [82, 84]. Some AAR systems (8%) designed a mobile application or game for the user's interaction on the touch screen [59, 85, 86, 64, 87]. By actions such as clicking buttons and selecting items on menus, the user can activate the AAR service, change virtual audio content, and adjust the audio presentation as they prefer.

Some AAR systems (8%) employed more natural user interfaces. For example, the user can directly provide voice commands to select or adjust the audio content [88, 89]. Hand gestures [60], head gestures [53], and eye gaze [38] were other common types of natural body input for AAR systems. For example, the user can control the audio volume by swinging their head to the left or right [53].

Finally, the authors noticed that 12% of the AAR systems employed novel, application-associated 3D tangible user interfaces [68, 90, 69, 38, 70, 91, 92, 51]. For example, in an exhibition, users could choose the virtual audio content at a specific position by rotating a physical cube to that direction [68, 69]. Some other AAR systems used mobile devices (e.g., smartphones) as the manipulating interface [91, 92, 51]. For example, users can tip the phone to switch between "stop" and "listen" mode [51]. In an AR book application, the book was designed to be foldable for interaction [90]. More specifically, appropriate audio content will be played accompanying visual animations in a pre-defined sequence when the user unfolds the book into a specific state [90].

Overall, a variety of interaction methods have been demonstrated in the reviewed AAR systems. These interaction methods can be used to choose and activate the desired audio augmentation content or adjust the presentation of the audio augmentation. However, most AAR systems integrated pre-designed virtual audio content that was not editable during run-time by using the interaction techniques.

2.3 Display Technology

Display technology, in the context of audio reproduction, refers to the hardware used to present sounds to a user. This may be in the form of loudspeakers, headphones, and earphones, among other methods used to deliver sound. Among the reviewed AAR systems, 61% of them included only audio augmentation ("audio only" in Table 2). In these cases, virtual sounds were rendered on computing devices (e.g., smartphones, tablets, and PCs) and then delivered to the user via wired/wireless headphones or earbuds or bone-conducted headsets. The virtual sounds can help users finish some tasks or provide users with a better experience. For example, spatialized sounds can indicate the direction and distance in a navigation application [92].

For AAR displays, *acoustic hear-through* is an important functionality for some applications in which there already exist real environmental sounds apart from those added virtually. If the real sounds are wanted, acoustically-transparent devices can be used so that real sounds can pass through unaltered for natural fusion with the virtual sounds [93–95]. In some other cases, real sounds might be unwanted and thus supposed to be reduced or removed. For example, for users to clearly hear spatialized navigation cues through earphones when riding bicycles outdoors, environmental wind noise was attenuated [96, 97].

Some AAR systems also integrated visual augmentation in addition to audio augmentation, which require other display methods for the user to comprehensively experience the augmented environment. Some systems (18%) employed head-mounted displays (HMDs) [88, 98, 99, 29, 90, 30, 82, 86, 100, 101, 54], whereas others (16%) implemented handheld displays (such as smartphones, tablets, and some other handheld devices) [38, 62, 84, 71, 31, 59, 85, 60, 87, 67] to enable users to perceive virtual visual and auditory content together. For applications like driving simulation, head-up display [79] and PC display [80] were equipped in the environment to simulate an inside-car situation. When visual augmentation was also included, users could perceive virtual sounds using the device-integrated speakers (e.g., Magic Leap One [101]). Alternatively, additional headsets or earbuds could be connected to the display devices for delivering virtual sounds like in the audio-only cases.

Overall, AAR systems mainly included audio-only displays or audio-visual displays. Off-the-shelf headsets or earbuds have been most commonly used to deliver virtual sounds. Newer display devices such as miniature loudspeaker arrays placed close to ears, as seen in the HoloLens [54] and Magic Leap [101], also started to gain popularity. The choice(s) of display technology is heavily dependent on the nature of the application and whether the augmentation of other sense(s) is needed.

2.4 Room Acoustics Modeling

For the purpose of this review, room acoustics modeling technology specifically pertains to AAR systems used in indoor environments. The need for this technology component stems from peoples' natural auditory perception of the real world. The perception of the same sound in different environments can vary drastically, even if the relative pose stays the same. This is because the acoustic properties of an environment (e.g., room geometry, surface materials, etc.) influence sound propagation and affect how the user perceives the sound source width, externalization, spectral characteristics, etc. These acoustic properties are unique to each environment. For users to perceive virtual sounds like they physically "belong" in the environment, an AAR system should model the room acoustics when rendering virtual sounds.

A typical approach to room acoustics modeling is to acquire the impulse response (IR) of the environment. An IR is a function that describes how the environment would influence the sound propagation from the source to the listener. Convolving the IR and a "dry" version of the sound results in a virtual sound source that is colored by the acoustic properties of the environment it occupies.

Among the reviewed AAR systems, 27% focused on outdoor environments, which did not include room acoustics modeling. In fact, to embed virtual sounds seamlessly in a real environment, appropriate reflection modeling is also important for outdoor applications. However, these works did not involve such reflection modeling in their systems.

Among the remaining 45 AAR systems, 73% of them did not include or specify the room acoustics modeling component, which might be because of the following three reasons: 1) Some systems used virtual spatial sounds for vivid presentations (e.g., to play messages at different spatial locations around the user's head based on their time of arrival [89]). In these applications, a strong association with the environment in which the user was located was not required. Therefore, integrating room acoustics did not add much value to the user experience. 2) Some AAR systems aimed to provide localization or navigation services (e.g., [72, 26, 52]). Researchers focused on rendering 3D locations or directions, and room acoustics modeling did not significantly influence the user's perception of the source location, especially when the application was designed for small-scale scenes (e.g., a desktop [30]). 3) Some AAR systems that integrated acoustics effects did not specify how or what was used to simulate acoustic environments.

The remaining 12 systems involved room acoustics modeling, but most of them did not provide their implementation details. Some works mentioned that they modeled some artificial reverberation [102, 103, 57, 81, 58, 78, 104]. Artificial reverberation simulates sound wave propagation phenomena in an environment such as reflection and diffusion, which can create the feeling of being in an indoor environment. However, because they did not mention the implementation details of creating such artificial reverberation, it is difficult to determine how well the added reverberation matched the real environment.

Four works implemented room acoustics modeling by first creating a virtual 3D room model that corresponded to the real environment and then simulating room acoustics based on the room model [66, 37, 87, 67]. One work mentioned that they computed IRs in a rectangular room by using an extended image source method for several source-receiver pairs [105]. Another work [106] integrated pre-recorded audio clips with their room acoustics (e.g., concert hall and drama scenes). Switching between these audio clips could then create the feeling of being immersed in a different indoor environment.

In general, room acoustics has been ignored or probably not modeled well enough in many of the systems. Although the reviewed AAR systems did not present much about room acoustics modeling, research on acoustics modeling has been conducted for years and some methods have the potential to be further explored and integrated into future AAR systems.

As mentioned above, one can first create a desired 3D room model that includes geometry modeling and surface material identification and then model the environmental IRs by simulating the sound wave propagation in the space. To this end, visual inputs from cameras can be used to reconstruct 3D environment models [107–111] and recognize materials [109, 111]. Apart from vision-based approaches, acoustics-based methods can also be used for geometry modeling and material classification. For example, smartphones can be used to receive ultrasonic chirps to reconstruct the environment geometry and estimate the sound absorption coefficients of indoor surface materials [112]. Based on the modeled geometry and the classification of materials, some computational techniques can be applied to generate the desired IRs [113]. In addition to the geometry and material-based sound propagation simulation approaches, it is also possible to exploit parametric methods for statistically coding desired IRs [114]. It is possible to statistically code a sound field because much of the perceptual quality of virtually rendered sounds can be quantified by a few critical acoustic parameters (e.g., reverberation time) [114]. Compared to a complete sound propagation simulation based on the geometry and surface material properties, the parametric coding method may run faster and have lower computational requirements.

In summary, although research on room acoustics modeling has been active for a long time, little of it has been implemented in AAR systems. One reason could be that some room acoustics modeling methods are computationally expensive [109–111] or require additional measurements in the application environments [111, 114], which makes them difficult to be used for interactive AAR applications in arbitrary environments. From another perspective, it has been analyzed that room acoustics modeling played a less important role in some AAR systems, and thus this component was little considered when designing the system. The authors argue that appropriate room acoustics modeling can significantly improve the user's immersive experience in some applications, such as AR-facilitated remote collaboration and exhibition tour. Thus, exploring computationally-efficient and conveniently-implementable solutions remains an important future research topic. More discussions about future work will be presented in SEC. 4.

2.5 Spatial Sound Synthesis

Spatial sound synthesis technology aims to synthesize virtual sounds that can be externalized like they are present in the user's real environment versus "inside-the-head" [115]. A key aspect of this is being able to process sounds in a way that the spectral and binaural characteristics of sounds delivered through headphones or earphones mimic those of sounds that are incident on the ears in the real world.

A classic technique is ambisonics recording and reproduction [116–118]. The sound of interest is first recorded using microphone arrays that typically consist of a large number of homogeneously distributed microphones [119–121]. Thereafter, depending on the user's real-time location in the environment, the recorded sound field can be reproduced from the desired source location. Such reproduction can be implemented through spatially distributed real loudspeakers such as [79] and [122] did in their AAR systems.

In order to conveniently use AAR applications in environments that are not pre-equipped with real loudspeakers, virtual sounds are better delivered through off-the-shelf or specifically designed headsets or earphones. Three works implemented a stereo sound panning [71, 123, 27] technique. This technique assigns a piece of monophonic sound to the left and right audio channels with time and level differences, which creates the illusion of width and space for the user. However, stereo panning sounds are usually perceived as localized inside the head rather than outside in the space, causing an unnatural fusion of virtual sounds with the real environment. In these works, because of their specific application settings [71] or the use of bone-conduction headset in streets [27], the stereo panning technique was able to provide a reasonable performance. In more general cases, more precise localization and externalization of virtual sounds should be achieved through binaural spatialization.

Binaural spatialization reproduces audio in a manner that mimics auditory perception with two ears in the real world. Binaural spatialization can be achieved through various processes such as equalization, delay filters, or convolution with head-related transfer functions (HRTFs) [124]. An HRTF is typically formulated as a function of the sound source position and its spectral distribution [125]. More specifically, an HRTF describes how a sound emitted from a location in the space will reach the eardrum after the sound waves interact with the listener's anatomical structure such as head and torso [125]. There is a pair of HRTFs, one for each ear. Because people are anthropometrically different, the HRTFs associated with each individual tend to be unique. Acquiring precise HRTFs for each person usually requires laborious measurements in strictly controlled environments, along with a lot of equipment such as multiple speakers, etc. Fortunately, because many people's anatomical structures are largely similar, previous works have shown that using a pair of generic HRTFs of an average head or those of a "good localizer" [126, 127] can produce perceptually adequate virtual sounds for a large group of users [128–130].

Of the reviewed AAR systems, 54% implemented binaural spatialization with generic HRTFs. Some of these systems used open-source or freely available spatial audio engines, such as OpenAL [29, 30, 82, 72, 91, 78, 59] and KLANG [52, 63]. However, the auralization details of the spatial audio engines are not available, and some of these systems did not specify the auralization principles or the audio engines they used for binaural spatialization.

Instead of using generic HRTFs, users may select the most suitable HRTFs for themselves from a dataset of premeasured HRTFs (e.g., MIT HRTF database [131], CIPIC HRTF database [132], SADIE II HRTF database [133]) [134]. Among the AAR systems that have been reviewed, Zotkin et al. [105] selected HRTFs for each user by measuring the user's anthropometric parameters and then finding the closest match in their HRTF database.

Although generic or similar HRTFs can work well for many users, some studies have shown that personalized HRTFs demonstrate significantly better results, especially if users demonstrate a high sensitivity of auditory localization or if their anatomical structures are far from average [135–137]. Among the AAR systems that have been reviewed, one work [104] measured HRTFs for some users and used these individual HRTFs to render binaural audio for these users.

So far, it has been summarized how most of the reviewed AAR systems used binaural spatialization (with generic, similar, or individual HRTFs) to synthesize virtual auditory sources. For users to experience an immersive auditory experience in the environment, or to have a better localization performance [138], one should integrate room acoustics modeling into spatial sound synthesis. One approach to this is by combining environmental IRs and HRTFs [139] when rendering virtual sounds. Alternatively, researchers can directly measure a user's binaural room impulse responses (BRIRs) in the environment of interest, which integrate room acoustics and the user's personal auditory perception [140, 141] in one measurement. To avoid the complexity of in-situ measurements, it is also possible to simulate perceptually plausible BRIRs [142]. Among the reviewed AAR systems, only one of them chose to measure each user's BRIRs for binaural audio rendering [102, 103].

Overall, it can be seen that most AAR systems have used open-source engines and generic HRTFs when creating binaural sounds that users can perceive through normal headsets or earphones. In comparison, personalized HRTFs or BRIRs have not been widely used in the existing AAR systems, which could be partly because of the difficulty of acquiring personalized HRTFs or BRIRs in reality. Because room acoustics modeling was ignored in most of the AAR systems, spatialized virtual sounds might be perceived without appropriate engagement in the real environment. To seamlessly blend virtual sounds with the physical world, the technologies of room acoustics modeling and spatial sound synthesis should go hand in hand in AAR implementations.

2.6 Summary

In this section, the five major technologies for creating AAR systems have been reviewed. From the above discussions, some general trends of AAR technologies over the past decades are summarized.

User-object pose tracking: Around 43% of the AAR systems used a single type of sensor for pose tracking. However, to guarantee a more robust pose tracking in different environments, hybrid tracking methods that exploit the strengths of different sensor types were more favored. For hybrid tracking, visual sensors, inertial sensor, and GPS have been used most.

Interaction technology: Implicit interaction was most commonly used in the reviewed AAR systems, followed by traditional 2D interfaces, 3D tangible interfaces and natural body inputs. Interaction technologies have been mainly used to activate or adjust a pre-designed virtual sound clip rather than editing the sound signal during run-time.

Display technology: Most AAR systems only augmented the user's auditory sense, for which the virtual sounds were delivered to the user via headsets or earbuds. Around 40% of the AAR systems combined audio and visual augmentation, for which other display methods (e.g., HMDs and handheld displays) were used.

Room acoustics modeling: Around half of the reviewed AAR systems did not include the room acoustics modeling component. Moreover, those that included room acoustics modeling tended to only create some approximate artificial reverberation effects. Additionally, 27% of the AAR systems aimed for outdoor applications, and none of them modeled outdoor reflections.

Spatial sound synthesis: Around half of the reviewed AAR systems created spatial sounds using generic HRTFs. Only a few works employed the user's BRIRs or the user's individual HRTFs.

SEC. 4 will identify and discuss important future research directions to promote AAR technologies and advance AAR applications.

3 APPLICATION DOMAINS OF AAR TECHNOLOGY

In SEC. 2, technologies used to create AAR systems were reviewed. In this section, a range of different real-world applications for AAR technology is reviewed.

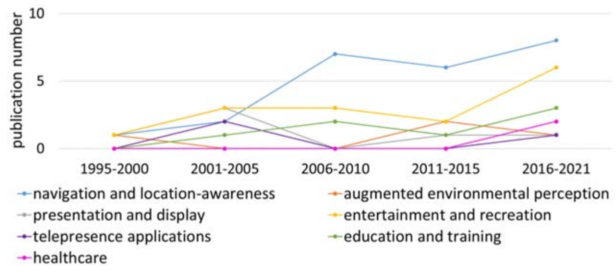


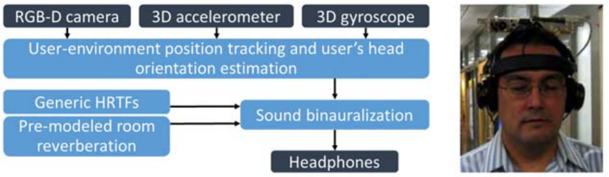
Fig. 2. Number of publications from 1995 to 2021 for each application domain.

Based on the literature reviewed, applications for AAR can be broadly divided into seven categories. Most of the reviewed AAR systems have been lab work, and a handful of these are available to be used as systems or services in the real world. These application domains include navigation and location-awareness assistance, augmented environmental perception, presentation and display, entertainment and recreation, telepresence applications, education and training, and healthcare. There can be some overlap between these categories. For example, in some *telepresence applications* (e.g., Mixed Reality remote collaboration), the user needs to localize objects or the other user [101], which involves *navigation*. In such cases, the work is categorized mainly according to its goal and targeted application scenarios of the original study.

To provide an overview of AAR applications, Fig. 2 shows the number of publications from 1995 to 2021 for each application domain, highlighting an overall increase in the number of AAR studies. This is especially true for applications in navigation and location awareness, followed by entertainment and recreation. This reflects, in many ways, the gradual commercial and research interest in these domains and how it has grown over time with the advent of technology capable of supporting AAR. It was also noticed that several application domains have begun to attract more research interest in recent years, such as education and training and healthcare. In the following subsections, each application domain will be reviewed in more detail.

3.1 Navigation and Location-Awareness Assistance

Given the nature of binaural hearing, navigation and location-awareness assistance appears to be the most popular application of AAR technology. Human beings have a limited field-of-view (FoV) of approximately 120° either side of the median plane and about 60° above and below the plane that passes through the eyes. However, this range encompasses the entire FoV, which also includes peripheral vision [143, 144]. Generally, human vision is the best within an approximately 60° horizontal and vertical arc in the front. Most of the understanding of the environment outside this "window" tends to come from surrounding sounds [143, 144]. A large part of AAR applications attempts to



(a) Integration of AAR technologies in the AAR system (b) Prototype system

Fig. 3. Audio augmented reality (AAR) system presented in [66]. This system used visual-inertial tracking, implicit interaction, audio-only display, pre-modeled room reverberation for acoustics modeling, and generic head-related transfer functions (HRTFs) for binauralization. (a) is adapted from Fig. 2 and (b) is Fig. 3 in [66].

exploit this natural form of auditory perception to provide information to users.

Spatialization of virtual sound sources can assist with prompt localization of targets, in particular when the targets are outside the user's FoV. Some researchers have used spatialized auditory beacons to provide coarse directional guidance [26, 63, 80], whereas more work aimed to direct the user to specific objects or landmarks [88, 98, 105, 145, 30, 79, 61, 62, 146, 65, 92, 51, 78, 31, 73, 27, 37, 104, 147, 54]. Some experiments have demonstrated adequate localization accuracy when directed to an object or position by spatialized virtual sound sources [30, 78, 26, 52, 73, 63, 37, 104]. These experiments demonstrated a promising and intuitive use of AAR technology for navigation and also established the foundation for some other applications, such as augmented environmental perception and telepresence applications.

In this application domain, several projects specifically focused on warning systems. They used spatial sounds to indicate the locations of safety threats [98] or the directions of imminent dangers while driving [79]. Furthermore, the function of navigation and location-awareness assistance could especially help visually impaired people [26, 54].

3.2 Augmented Environmental Perception

AAR technology could assist visually impaired people in understanding the ongoing events or notice surrounding objects in their environment [56, 146, 91, 66, 147, 28]. Note that although environmental perception is largely based on the awareness of direction or object locations, it is more about providing users with an overall understanding of their environment. For example, [66] developed an AAR system (see Fig. 3) that can describe surrounding objects (e.g., stuff on a table and furniture in a hallway) to users, and the audio messages are rendered from the object locations.

3.3 Presentation and Display

Some research used spatial sounds to vividly convey messages [89, 71] or create a certain ambience [100]. For example, calendar items could be played to the user from different directions to remind them of events at corresponding hours [83]. Similarly, messages could be played at different spatial locations around the user's head based on their time of arrival [89]. In these applications, spatial sounds are probably not associated with specific objects or the space where the user is, but such a presentation or display could enrich the user's perception of their surrounding activities or events so that they could conveniently select items, receive messages, or monitor progress [89, 83].

Some other researchers have created spatial soundscapes to immerse users in specific scenes [100]. For example, the soundscape of a city scene could be rendered to augment a 360° visual panorama [100]. Overall, AAR technology can enrich the presentation of information and enhance interaction by engaging the user more effectively.

3.4 Entertainment and Recreation

Entertainment and recreation has also been a popular application for AAR technology. Two major scenarios in this category are summarized.

The first popular scenario is exhibition scenes, including museums [55, 68, 57, 69, 58, 123, 106, 53], cultural heritage displays [72, 67], and archaeological sites [84, 64]. In such scenarios, AAR technology could be used to spatialize the preface or background knowledge and vividly introduce exhibits or re-direct visitors' interest [57, 68, 69, 70, 84, 106, 67]. Alternatively, some implementations use AAR technology to augment the exhibits or scenes themselves by adding content-related virtual sounds [72, 64, 53]. For example, the operating sound of printers was virtually added and spatialized to accompany an old printer exhibit in an exhibition at the MAM Museum [53].

Moreover, AAR technology could also be used in mobile games [59, 60]. In these applications, spatial sounds were used to either provide navigation cues [59] or render content-related soundscapes [60].

3.5 Telepresence Applications

AAR technology has also been implemented in several telepresence applications [99, 29, 101]. In such applications, spatial audio was commonly used to augment the virtual avatars/objects and enable people to easily distinguish who is speaking in a multi-party setting [99, 29, 101]. This assists with remote collaboration tasks and enhances the feeling of human presence.

3.6 Education and Training

AAR technology has been applied in several educational scenarios to impart knowledge and convey information more vividly and effectively [90, 82, 84, 85, 122, 87]. The most popular application is storytelling using AR/MR books [90, 38]. The content and characters in the book could be augmented by story-related spatial sounds, thus improving reader experience and retention of the book. Such a vivid storytelling application was also deployed at cultural heritage sites to enhance visitor experience [85].

Another educational application is to augment the teaching material to help students acquire a better understanding of underlying concepts [82, 87]. For example, the revolution pattern of the solar system was presented with 3D audio effects to help students understand the concepts with impressive illustrations [87]. Moreover, AAR technology has also been used to augment natural soundscapes to enhance public understanding of the natural world [122].

Another novel application of AAR technology is telecoaching for fast-paced tasks such as training users to play tennis [86]. This coaching application is based on the function of navigation and localization using spatial sounds. More specifically, the user's coach could initiate a spatial audio instruction that guided the user to hit the ball toward a specific direction or a spot. In the future, it might inspire more applications in sports training, given the advantage of directional and timing guidance by spatial sounds.

3.7 Healthcare

Healthcare is a relatively new application domain of AAR technology. For example, a spatial soundscape that demonstrates natural elements in open spaces could be used to enhance people's connection with the nature, which could benefit their mental and physical well-being [148]. Such an implementation creates the illusion of staying in outdoor environments, which can be especially useful when venturing outside is not possible. Furthermore, this might help visually impaired people who find it difficult navigating outdoor environments to remain indoors and experience some aspects of an outdoor natural surrounding [148]. In another example, researchers proposed to bring the restorative benefits of outdoor environments to indoor spaces by creating virtual natural soundscapes, which can help to deal with geriatric depression [149].

4 FUTURE RESEARCH DIRECTIONS

The previous two sections reviewed technologies used to build AAR systems and the application domains in which AAR has been studied. Specifically, five technology components are needed for implementing AAR, namely, user-object pose tracking, interaction technology, display technology, room acoustics modeling, and spatial sound synthesis. The increasing research interest in AAR technology has prompted the exploration of its application across seven domains. Among these, navigation and location-awareness assistance has been the most popular application type. In recent years, a significant number of novel applications have also been developed for entertainment and recreation, education and training, healthcare, etc.

From the papers that have been reviewed, a number of important future research directions, which will be discussed in this section, are identified.

4.1 Future AAR Technologies

4.1.1 Tracking

Based on the papers reviewed, it is seen that a number of different sensor types have been used for pose tracking in AAR systems, including visual sensors, inertial sensors, and GPS. However, some pose tracking approaches require an environment to be equipped in advance to enable tracking (e.g., retroreflective tracking using a Vicon system [78, 37]), which limits the use of AAR applications in arbitrary environments. Some pose tracking approaches ask the user to put on some form of obtrusive tracking apparatus (e.g., visual tracking using HMDs [29, 30, 101]). This is necessary in some application scenarios such as when visual and auditory augmentation is needed together. However, using HMDs might modify the user's HRTFs, thus impacting their AAR experience [150, 151] Furthermore, using HMDs might also be physically uncomfortable and not socially acceptable [152], which limits the use of AAR.

Therefore, the authors suggest that future research on pose tracking could investigate approaches using lightweight but powerful wearable devices that have already been adopted by consumers. For example, earables, which refer to wearable devices around the ear and head such as hearing aids, earbuds, and electronics-embedded glasses [153], are becoming increasing popular. Example devices include Nokia eSense [77, 154] and Bose Frames.[4] These devices are typically equipped with acoustic and motion sensors that can be exploited for tracking the user's position and orientation. Moreover, such devices are typically in the form of earbuds or glasses, which can be conveniently used for audio delivery in everyday work and life. Therefore, it is possible to integrate pose tracking and spatial audio delivery into a single device and perform the required computation on the device too. However, there might be trade-offs in terms of power consumption and latency, which requires further research and development in the future.

4.1.2 Interaction and Display Technologies

Approximately half of the AAR systems that have been reviewed implemented implicit interaction. This form of interaction is well suited to certain scenarios, such as when performing an attention-critical task. Moreover, visually impaired people might find implicit interaction helpful, but an implementation of voice input or a well-designed tangible interface is recommended to allow for more flexible and personalized control of the system. Future research can also explore real-time audio editing using interaction technologies, which might open up more opportunities to enrich AAR systems and their applications.

Although interaction and display are two distinct aspects of AAR systems, there exists a close functional relationship between the two. For example, [85] deployed their AAR system on mobile device while enabling interaction via a mobile application. Most of the current AAR systems use audio-only displays. Future research direction for display technology could focus on the integration of several senses into one AAR system. For example, if using electronics-embedded glasses in AAR applications, visual augmentation and vibrations might be added to enrich the user's experience.

4.1.3 Room Acoustics Modeling

As covered in Sec. 2, most of the AAR systems that have been reviewed did not include a room acoustics modeling

---

[4] https://www.bose.com/en_us/products/frames.html.

component. Those that included this component tended to model rough artificial reverberation to only create the feeling of being in a room. However, as has also been discussed, some research has studied room acoustics modeling methods, but these methods have not been employed in the AAR systems yet. When modeling environmental acoustics for AAR applications, one must take into account the computational costs and quality of the modeled acoustic environment. Computationally efficient approaches are favored because the desired room acoustics could vary in real time and need to be adapted to the user's movement in the environment. In other words, online methods for room acoustics modeling can better fit most AAR applications. To obtain satisfactory room acoustics modeling in an efficient manner, it could be worth exploring parametric approaches to code sound fields and improve computational techniques on wearable or mobile devices.

The time efficiency in computation might sacrifice the quality of the modeled room acoustics to some extent. However, it is also not necessary to achieve perfect modeling for two reasons. First, since human auditory perception is only sensitive to a certain level of difference between sounds, users might not recognize the differences between the modeled sounds and those present in the real world as long as the differences fall within certain perceptual limits. These perceptual limits are the "just noticeable difference" (JND) [155]. JND thresholds are typically measured for different parameters (e.g., reverberation time, early decay time, center time, etc.) [156, 157]. Moreover, when measuring JNDs under different conditions (e.g., different rooms, participants, audio frequencies, etc.), the resultant JNDs may also be different. Therefore, there exist different JND standards. From one perspective, the authors suggest more investigations into JNDs to provide more detailed standards for different parameters under different conditions. From another perspective, researchers can follow relatively strict standards from literature when designing and evaluating their AAR systems.

The second reason that it is not necessary to achieve perfect room acoustics modeling is because the visual sense tends to work in concert with the auditory sense to perceive the environment as a whole. More specifically, research [158] shows that a perceptually adequate acoustic environment is likely to suffice in AR applications in which the user can also perceive their surroundings by seeing the real space. In the future, more studies are needed to investigate the required precision of acoustics modeling in different application scenarios. Additionally, more efforts should be made to develop new acoustics modeling algorithms, especially in situations in which the movement of the user and the surrounding objects is arbitrary.

4.1.4 Spatial Sound Synthesis

From the reviewed AAR systems, it can be seen that most of them used generic HRTFs to synthesize binaural sounds. As mentioned earlier, individualized HRTFs can produce better localization and immersive experience for users. One important future research direction is to explore methods that can enable the convenient capture and implementation of individual HRTFs. To this end, there have been some attempts in recent years. Because a person's anthropometric features (e.g., head width, shoulder width, and pinna height) and the corresponding HRTFs are closely related [159], some researchers have explored techniques that first acquire a user's anthropometric parameters [160, 161] and then approximate the matching HRTFs using a numeric sound propagation solver [160], numerical acoustic simulations [162], or neural network-based regression algorithms [163, 164]. In the commercial space, Sony has implemented personalization using an application that visually scans the ears to enable tailor-made immersive experiences.[5] Future research could explore methods that can make this process faster, more accurate, and easy to implement.

SEC. 2.5 reviewed the use of BRIRs for spatial sound synthesis. Although directly measuring BRIRs provides an alternative to acquiring room acoustic effects and individual HRTFs separately, it has some limitations. For example, the measurement needs to be conducted in the desired environment for a specific user. To address this restriction, future research can explore techniques that adapt BRIRs measured in one room to a different room, different listener, and arbitrary sound source. Previous work has presented an adaptive algorithm that applied different equalizations to different reverberation stages [165], and more research is needed to advance the generalization of BRIRs.

This review paper has discussed room acoustics modeling and spatial sound synthesis separately. This is because many of the reviewed AAR systems did not include room acoustics modeling. However, combining room acoustics modeling and spatial sound synthesis is necessary for providing an appropriate sense of space and engagement in an environment. The acquisition or simulation of BRIRs could be a way of combining these two approaches, and more research is needed to promote individualized binaural spatialization with environmental acoustics.

4.2 Future AAR Applications

In previous sections, it was seen that the most fundamental affordance of AAR technology is navigation and localization. Given the human nature of binaural hearing, and as the foundation in some other applications (e.g., sports training and telepresence applications), navigation and location-awareness assistance is anticipated to remain one of the most intuitive and popular AAR applications in the future. Along with the development of AAR technologies, such as more comfortable tracking apparatus with a long battery life, navigation-related and localization-related applications might be widely accepted in everyday life.

Another popular AAR application in the future could lie in the field of AR-mediated remote work, education, gaming, and social activities. Nowadays, using video-audio conferencing technologies has become a new normal for

---

[5] https://www.sony.co.nz/electronics/360-reality-audio.

conducting business and social activities remotely. The COVID-19 pandemic has resulted in the widespread adoption of conferencing platforms in everyday life. To create the feeling of belonging and connectedness that people experience in face-to-face interactions, AR technologies, involving both visual and audio augmentation, are being explored to enable "natural interactions" that transcend distances.

AR technology has been widely adopted in industrial applications, but audio augmentation is usually ignored. Spatialized sounds could be used for reporting device status in routine maintenance and error diagnostics systems [166, 167]. Compared to vision augmentation, audio augmentation still remains under-employed in industrial scenarios. Future studies can be conducted to investigate the potential of using AAR technology to enhance industrial activities.

In this review, a few novel trials of using AAR for healthcare have been discussed. In the future, the authors see great potential to extend explorations in this direction. For example, mobile music therapy could be a field worth exploring. Clinical research has shown that spatial configuration of physical instruments could help to attract users' focus and guide their movements in some therapeutic exercises [168]. These insights indicate remarkable potential for exploring AAR technology to create virtual spatial soundscapes for music therapeutic applications.

Overall, existing work has shown a broad landscape of AAR applications, and more extensive use of AAR technology is expected to be seen in the coming years. With the ubiquity of mobile and/or wearable devices on the rise, AAR has the potential to significantly help everyday work and life in several ways.

5 CONCLUSION

In this paper, the development of AAR technology was summarized by reviewing a range of research papers published over the last few decades. Five techniques for implementing AAR were first reviewed. Overall, the quality of audio augmentation appears to have steadily improved over time. This is a result of a greater amount of AAR research that has contributed to a number of modeling methods to replicate human auditory perception, model room acoustics, etc. The development of allied sensing technology, availability of low-cost high-performance hardware, and exponential increases in computing power have also played a significant role in the advance of AAR technology. Technical advancements have enabled AAR systems to be integrated into mobile devices, such as smartphones and hearables.[6] These advances have contributed to, and continue to contribute to, the development of AAR across a range of applications.

This review also demonstrated that there appear to be seven domains within which the application of AAR has been studied. The most fundamental and popular application of AAR is navigation and location-awareness assistance, which also provides the basis for extended applications in some other fields. More recently, AAR appears to be gaining a foothold in the healthcare industry. There is also a huge untapped potential for using AAR in remote collaborative environments for work, study, and social activities.

Overall, this survey provided a systematic review of the research that has been conducted in the domain of AAR. After reviewing existing AAR systems, the relevant technological methods, areas that may benefit from the application-based research, and future research directions for advancing each AAR technology component and applications were also identified. The authors hope researchers and practitioners can derive inspiration from this review when they plan for related work in the future. AAR has the potential to benefit numerous aspects of peoples' lives. The authors hope that it becomes more widely used in the future to enable working better, staying more connected, and living healthier.

6 REFERENCES

[1] R. T. Azuma, "A Survey of Augmented Reality," *Presence (Camb)*, vol. 6, no. 4, pp. 355–385 (1997 Aug.). https://doi.org/10.1162/pres.1997.6.4.355.

[2] F. P.Jr. Brooks,, "The Computer Scientist as Toolsmith II," *Commun. ACM*, vol. 39, no. 3, pp. 61–68 (1996 Mar.). https://doi.org/10.1145/227234.227243.

[3] S. Feiner, B. Macintyre, and D. Seligmann, "Knowledge-Based Augmented Reality," *Commun. ACM*, vol. 36, no. 7, pp. 53–62 (1993 Jul.). https://doi.org/10.1145/159544.159587.

[4] M. Funk, T. Kosch, R. Kettner, O. Korn, and A. Schmidt, "motionEAP: An Overview of 4 Years of Combining Industrial Assembly With Augmented Reality for Industry 4.0," in *Proceedings of the International Conference on Knowledge Technologies and Data-Driven Business*, paper 4 (Graz, Austria) (2016 Oct.).

[5] M. Funk, A. Bächler, L. Bächler, et al., "Working With Augmented Reality? A Long-Term Analysis of In-Situ Instructions at the Assembly Workplace," in *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments*, pp. 222–229 (Rhodes, Greece) (2017 Jun.). https://doi.org/10.1145/3056540.3056548.

[6] S. Claudino Daffara, A. Brewer, B. Thoravi Kumaravel, and B. Hartmann, "Living Paper: Authoring AR Narratives Across Digital and Tangible Media," in *Extended Abstracts of the Conference on Human Factors in Computing Systems*, pp. 1–10 (Honolulu, HI) (2020 Apr.). https://doi.org/10.1145/3334480.3383091.

[7] D. Pérez-López and M. Contero, "Delivering Educational Multimedia Contents Through an Augmented Reality Application: A Case Study on Its Impact on Knowledge Acquisition and Retention," *Turkish Online J. Educ. Technol.*, vol. 12, no. 4, pp. 19–28 (2013 Oct.).

[8] M. Rusiñol, J. Chazalon, and K. Diaz-Chito, "Augmented Songbook: An Augmented Reality Educational Application for Raising Music Awareness," *Multimed. Tools*

---

[6] https://www.bragi.com/.

Appl., vol. 77, no. 11, pp. 13773–13798 (2018 Jul.). https://doi.org/10.1007/s11042-017-4991-4.

[9] W. Piekarski and B. Thomas, “ARQuake: The Outdoor Augmented Reality Gaming System,” *Commun. ACM*, vol. 45, no. 1, pp. 36–38 (2002 Jan.). https://doi.org/10.1145/502269.502291.

[10] B. H. Thomas, “A Survey of Visual, Mixed, and Augmented Reality Gaming,” *Comput. Entertain.*, vol. 10, no. 1, paper 3 (2012 Oct.). https://doi.org/10.1145/2381876.2381879.

[11] S. R. Fussell, R. E. Kraut, and J. Siegel, “Coordination of Communication: Effects of Shared Visual Context on Collaborative Work,” in *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, pp. 21–30 (Philadelphia, PA) (2000 Dec.). https://doi.org/10.1145/358916.358947.

[12] G. A. Lee, T. Teo, S. Kim, and M. Billinghurst, “A User Study on MR Remote Collaboration Using Live 360 Video,” in *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, pp. 153–164 (Munich, Germany) (2018 Oct.). https://doi.org/10.1109/ISMAR.2018.00051.

[13] T. Piumsomboon, G. A. Lee, J. D. Hart, et al., “Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration,” in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, paper 46 (Montreal, Canada) (2018 Apr.). https://doi.org/10.1145/3173574.3173620.

[14] T. Teo, L. Lawrence, G. A. Lee, M. Billinghurst, and M. Adcock, “Mixed Reality Remote Collaboration Combining 360 Video and 3D Reconstruction,” in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, paper 201 (Glasgow, UK) (2019 May). https://doi.org/10.1145/3290605.3300431.

[15] T. Teo, G. A. Lee, M. Billinghurst, and M. Adcock, “Hand Gestures and Visual Annotation in Live 360 Panorama-Based Mixed Reality Remote Collaboration,” in *Proceedings of the 30th Australian Conference on Computer-Human Interaction*, pp. 406–410 (Melbourne, Australia) (2018 Dec.). https://doi.org/10.1145/3292147.3292200.

[16] R. Azuma, Y. Baillot, R. Behringer, et al., “Recent Advances in Augmented Reality,” *IEEE Comput. Graph. Appl.*, vol. 21, no. 6, pp. 34–47 (2001 Dec.). https://doi.org/10.1109/38.963459.

[17] A. Dey, M. Billinghurst, R. W. Lindeman, and II J. E. Swan, “A Systematic Review of 10 Years of Augmented Reality Usability Studies: 2005 to 2014,” *Front. Robot. AI*, vol. 5, paper 37 (2018 Apr.). https://doi.org/10.3389/frobt.2018.00037.

[18] F. Zhou, H. B.-L. Duh, and M. Billinghurst, “Trends in Augmented Reality Tracking, Interaction and Display: A Review of Ten Years of ISMAR,” in *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 193–202 (Washington, D.C.) (2008 Sep.). https://doi.org/10.1109/ISMAR.2008.4637362.

[19] K. Kim, M. Billinghurst, G. Bruder, H. B.-L. Duh, and G. F. Welch, “Revisiting Trends in Augmented Reality Research: A Review of the 2nd Decade of ISMAR (2008–2017),” *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 11, pp. 2947–2962 (2018 Nov.). https://doi.org/10.1109/TVCG.2018.2868591.

[20] J. Yang, *Audio-Facilitated Human Interaction with the Environment: Advancements in Audio Augmented Reality and Auditory Notification Delivery*, Ph.D. thesis, ETH Zurich, Zurich, Switzerland (2021 Nov.).

[21] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA, 1997). https://doi.org/10.7551/mitpress/6391.001.0001.

[22] P. Vávra, J. Roman, P. Zonča, et al., “Recent Development of Augmented Reality in Surgery: A Review,” *J. Healthc. Eng.*, vol. 2017, paper 4574172 (2017 Aug.). https://doi.org/10.1155/2017/4574172.

[23] X. Li, W. Yi, H.-L. Chi, X. Wang, and A. P. C. Chan, “A Critical Review of Virtual and Augmented Reality (VR/AR) Applications in Construction Safety,” *Autom. Constr.*, vol. 86, pp. 150–162 (2018 Feb.). https://doi.org/10.1016/j.autcon.2017.11.003.

[24] D. Szymczak, K. Rassmus-Gröhn, C. Magnusson, and P.-O. Hedvall, “A Real-World Study of an Audio-Tactile Tourist Guide,” in *Proceedings of the 14th International Conference on Human-Computer Interaction With Mobile Devices and Services*, pp. 335–344 (San Francisco, CA) (2012 Sep.). https://doi.org/10.1145/2371574.2371627.

[25] M. Billinghurst, A. Clark, and G. Lee, “A Survey of Augmented Reality,” *Found. Trends Hum.-Comput. Interact.*, vol. 8, no. 2-3, pp. 73–272 (2015 Mar.). http://doi.org/10.1561/1100000049.

[26] S. Blessenohl, C. Morrison, A. Criminisi, and J. Shotton, “Improving Indoor Mobility of the Visually Impaired With Depth-Based Spatial Sound,” in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 418–426 (Santiago, Chile) (2015 Dec.). https://doi.org/10.1109/ICCVW.2015.62.

[27] E. Schoop, J. Smith, and B. Hartmann, “HindSight: Enhancing Spatial Awareness by Sonifying Detected Objects in Real-Time 360-Degree Video,” in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, paper 143 (Montreal, Canada) (2018 Apr.). https://doi.org/10.1145/3173574.3173717.

[28] O. B. Kaul, K. Behrens, and M. Rohs, “Mobile Recognition and Tracking of Objects in the Environment Through Augmented Reality and 3D Audio Cues for People With Visual Impairments,” in *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, paper 394 (Yokohama, Japan) (2021 May). https://doi.org/10.1145/3411763.3451611.

[29] Z. Zhou, A. D. Cheok, X. Yang, and Y. Qiu, “An Experimental Study on the Role of 3D Sound in Augmented Reality Environment,” *Interact. Comput.*, vol. 16, no. 5, pp. 989–1016 (2004 Oct.). https://doi.org/10.1016/j.intcom.2004.06.014.

[30] J. Sodnik, S. Tomazic, R. Grasset, A. Duenser, and M. Billinghurst, “Spatial Sound Localization in an Augmented Reality Environment,” in *Proceedings of the 18th Australia Conference on Computer-Human Interaction: Design: Activities, Artefacts and Environ-*

ments, pp. 111–118 (Sydney, Australia) (2006 Nov.). https://doi.org/10.1145/1228175.1228197.

[31] D. Rumiński, "An Experimental Study of Spatial Sound Usefulness in Searching and Navigating Through AR Environments," *Virtual Real.*, vol. 19, no. 3, pp. 223–233 (2015 Nov.). https://doi.org/10.1007/s10055-015-0274-4.

[32] G. Gordon, M. Billinghurst, M. Bell, et al., "The Use of Dense Stereo Range Data in Augmented Reality," in *Proceedings of the International Symposium on Mixed and Augmented Reality*, pp. 14–23 (Darmstadt, Germany) (2002 Oct.). http://doi.org/10.1109/ISMAR.2002.1115063.

[33] R. A. Newcombe, S. Izadi, O. Hilliges, et al., "KinectFusion: Real-Time Dense Surface Mapping and Tracking," in *Proceedings of the 10th IEEE International Symposium on Mixed and Augmented Reality*, pp. 127–136 (Basel, Switzerland) (2011 Oct.). https://doi.org/10.1109/ISMAR.2011.6092378.

[34] B. W. Babu, S. Kim, Z. Yan, and L. Ren, "σ-DVO: Sensor Noise Model Meets Dense Visual Odometry," in *Proceedings of the International Symposium on Mixed and Augmented Reality*, pp. 18–26 (Merida, Mexico) (2016 Oct.). http://doi.org/10.1109/ISMAR.2016.11.

[35] H. Jiang, D. Weng, Z. Zhang, Y. Bao, Y. Jia, and M. Nie, "HiKeyb: High-Efficiency Mixed Reality System for Text Entry," in *Proceedings of International Symposium on Mixed and Augmented Reality Adjunct*, pp. 132–137 (Munich, Germany) (2018 Oct.). http://doi.org/10.1109/ISMAR-Adjunct.2018.00051.

[36] Z. Yuan, K. Cheng, J. Tang, and X. Yang, "RGB-D DSO: Direct Sparse Odometry With RGB-D Cameras for Indoor Scenes," *IEEE Trans. Multimed*, vol. 24, pp. 4092–4101 (2021 Sep.). http://doi.org/10.1109/TMM.2021.3114546.

[37] J. Yang, Y. Frank, and G. Sörös, "Hearing Is Believing: Synthesizing Spatial Audio From Everyday Objects to Users," in *Proceedings of the 10th Augmented Human International Conference*, paper 28 (Reims France) (2019 Mar.). https://doi.org/10.1145/3311823.3311872.

[38] R. Grasset, A. Duenser, H. Seichter, and M. Billinghurst, "The Mixed Reality Book: A New Multimedia Reading Experience," in *Extended Abstracts on CHI Human Factors in Computing Systems*, pp. 1953–1958 (San Jose, CA) (2007 Apr.). https://doi.org/10.1145/1240866.1240931.

[39] U. Neumann and S. You, "Natural Feature Tracking for Augmented Reality," *IEEE Trans. Multimed.*, vol. 1, no. 1, pp. 53–64 (1999 Mar.). http://doi.org/10.1109/6046.748171.

[40] Y. K. Yu, K. H. Wong, M. M. Y. Chang, and S.-H. Or, "Recursive Camera-Motion Estimation With the Trifocal Tensor," *IEEE Trans. Syst. Man Cybern.*, vol. 36, no. 5, pp. 1081–1090 (2006 Oct.). http://doi.org/10.1109/TSMCB.2006.874133.

[41] G. Klein and D. Murray, "Parallel Tracking and Mapping for Small AR Workspaces," in *Proceedings of the 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 225–234 (Nara, Japan) (2007 Oct.). https://doi.org/10.1109/ISMAR.2007.4538852.

[42] K. Xu, K. W. Chia, and A. D. Cheok, "Real-Time Camera Tracking for Marker-Less and Unprepared Augmented Reality Environments," *Image Vis. Comput.*, vol. 26, no. 5, pp. 673–689 (2008 May). https://doi.org/10.1016/j.imavis.2007.08.015.

[43] F. Ababsa and M. Mallem, "A Robust Circular Fiducial Detection Technique and Real-Time 3D Camera Tracking," *J. Multimed.*, vol. 3, no. 4, pp. 34–41 (2008 Oct.).

[44] A. Loquercio, M. Dymczyk, B. Zeisl, et al., "Efficient Descriptor Learning for Large Scale Localization," in *Proceedings of International Conference on Robotics and Automation*, pp. 3170–3177 (Singapore) (2017 May). https://doi.org/10.1109/ICRA.2017.7989359.

[45] G. Younes, D. Asmar, I. Elhajj, and H. Al-Harithy, "Pose Tracking for Augmented Reality Applications in Outdoor Archaeological Sites," *J. Electron. Imag.*, vol. 26, no. 1, paper 011004 (2016 Oct.). https://doi.org/10.1117/1.JEI.26.1.011004.

[46] W. Ma, H. Xiong, X. Dai, X. Zheng, and Y. Zhou, "An Indoor Scene Recognition-Based 3D Registration Mechanism for Real-Time AR-GIS Visualization in Mobile Applications," *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 3, paper 112 (2018 Mar.). https://doi.org/10.3390/ijgi7030112.

[47] J. Rambach, C. Deng, A. Pagani, and D. Stricker, "Learning 6DoF Object Poses From Synthetic Single Channel Images," in *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality Adjunct*, pp. 164–169 (Munich, Germany) (2018 Oct.). https://doi.org/10.1109/ISMAR-Adjunct.2018.00058.

[48] C.-Y. Tsai, K.-J. Hsu, and H. Nisar, "Efficient Model-Based Object Pose Estimation Based on Multi-Template Tracking and PnP Algorithms," *Algorithms*, vol. 11, no. 8, paper 122 (2018 Aug.). https://doi.org/10.3390/a11080122.

[49] H. Huang, F. Zhong, Y. Sun, and X. Qin, "An Occlusion-Aware Edge-Based Method for Monocular 3D Object Tracking Using Edge Confidence," in *Comput. Graph. Forum*, vol. 39, no. 7, pp. 399–409 (2020 Nov.). https://doi.org/10.1111/cgf.14154.

[50] M. Ortega, E. Ivorra, A. Juan, et al., "MANTRA: An Effective System Based on Augmented Reality and Infrared Thermography for Industrial Maintenance," *Appl. Sci.*, vol. 11, no. 1, paper 385 (2021 Jan.). https://doi.org/10.3390/app11010385.

[51] J. R. Blum, M. Bouchard, and J. R. Cooperstock, "Spatialized Audio Environmental Awareness for Blind Users With a Smartphone," *Mobile Netw. Appl.*, vol. 18, no. 3, pp. 295–309 (2013 Jun.). https://doi.org/10.1007/s11036-012-0425-8.

[52] F. Heller, J. Jevanesan, P. Dietrich, and J. Borchers, "Where Are We?: Evaluating the Current Rendering Fidelity of Mobile Audio Augmented Reality Systems," in *Proceedings of the 18th International Conference on Human-Computer Interaction With Mobile Devices and Services*, pp. 278–282 (Florence, Italy) (2016 Sep.). http://doi.org/10.1145/2935334.2935365.

[53] F. Z. Kaghat, A. Azough, M. Fakhour, and M. Meknassi, "A New Audio Augmented Reality Interaction and Adaptation Model for Museum Visits," *Com-*

put. Electr. Eng., vol. 84, paper 106606 (2020 Jun.). https://doi.org/10.1016/j.compeleceng.2020.106606.

[54] R. Guarese, F. Bastidas, J. Becker, et al., “Cooking in the Dark: Exploring Spatial Audio as MR Assistive Technology for the Visually Impaired*Human-Computer Interaction – INTERACT 2021*, Lecture Notes in Computer Science, vol. 12936, pp. 318–322 (Springer, Cham, Switzerland, 2021). http://doi.org/10.1007/978-3-030-85607-6_29.

[55] B. B. Bederson, “Audio Augmented Reality: A Prototype Automated Tour Guide,” in *Conference Companion on Human Factors in Computing Systems*, pp. 210–211 (Denver, CO) (1995 May). https://doi.org/10.1145/223355.223526.

[56] E. D. Mynatt, M. Back, R. Want, and R. Frederick, “Audio Aura: Light-Weight Audio Augmented Reality,” in *Proceedings of the 10th Annual ACM Symposium on User Interface Software and Technology*, pp. 211–212 (Banff, Canada) (1997 Oct.). https://doi.org/10.1145/263407.264218.

[57] L. Terrenghi and A. Zimmermann, “Tailored Audio Augmented Environments for Museums,” in *Proceedings of the 9th International Conference on Intelligent User Interfaces*, pp. 334–336 (Funchal, Portugal) (2004 Jan.). https://doi.org/10.1145/964442.964523.

[58] A. Zimmermann and A. Lorenz, “LISTEN: A User-Adaptive Audio-Augmented Museum Guide,” *User Model. User-Adap. Interact.*, vol. 18, no. 5, pp. 389–416 (2008 Nov.). https://doi.org/10.1007/s11257-008-9049-x.

[59] T. Chatzidimitris, D. Gavalas, and D. Michael, “SoundPacman: Audio Augmented Reality in Location-Based Games,” in *Proceedings of the 18th Mediterranean Electrotechnical Conference*, pp. 1–6 (Limassol, Cyprus) (2016 Apr.). http://doi.org/10.1109/MELCON.2016.7495414.

[60] E. Rovithis, N. Moustakas, A. Floros, and K. Vogklis, “Audio Legends: Investigating Sonic Interaction in an Augmented Reality Audio Game,” *Multimodal Technol. Interact.*, vol. 3, no. 4, paper 73 (2019 Nov.). http://doi.org/10.3390/mti3040073.

[61] B. N. Walker and J. Lindsay, “Navigation Performance With a Virtual Auditory Display: Effects of Beacon Sound, Capture Radius, and Practice,” *Hum. Factors*, vol. 48, no. 2, pp. 265–278 (2006 Jun.). https://doi.org/10.1518/001872006777724507.

[62] C. Stahl, “The Roaring Navigator: A Group Guide for the Zoo With Shared Auditory Landmark Display,” in *Proceedings of the 9th International Conference on Human Computer Interaction With Mobile Devices and Services*, pp. 383–386 (Singapore) (2007 Sep.). https://doi.org/10.1145/1377999.1378042.

[63] F. Heller and J. Schöning, “NavigaTone: Seamlessly Embedding Navigation Cues in Mobile Music Listening,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, paper 637 (Montreal, Canada) (2018 Apr.). https://doi.org/10.1145/3173574.3174211.

[64] M. Sikora, M. Russo, J. Đerek, and A. Jurčević, “Soundscape of an Archaeological Site Recreated With Audio Augmented Reality,” *ACM Trans. Multimed. Comput. Commun. Appl.*, vol. 14, no. 3, paper 74 (2018 Aug.). https://doi.org/10.1145/3230652.

[65] B. F. Katz, S. Kammoun, G. Parseihian, et al., “NAVIG: Augmented Reality Guidance System for the Visually Impaired: Combining Object Localization, GNSS, and Spatial Audio,” *Virtual Real.*, vol. 16, no. 4, pp. 253–269 (2012 Nov.). https://doi.org/10.1007/s10055-012-0213-6.

[66] F. Ribeiro, D. Florêncio, P. A. Chou, and Z. Zhang, “Auditory Augmented Reality: Object Sonification for the Visually Impaired,” in *Proceedings of the IEEE 14th International Workshop on Multimedia Signal Processing*, pp. 319–324 (Banff, Canada) (2012 Sep.). http://doi.org/10.1109/MMSP.2012.6343462.

[67] M. Comunità, A. Gerino, V. Lim, and L. Picinali, “Design and Evaluation of a Web- and Mobile-Based Binaural Audio Platform for Cultural Heritage,” *Appl. Sci.*, vol. 11, no. 4, paper 1540 (2021 Feb.). http://doi.org/10.3390/app11041540.

[68] M. Hatala, L. Kalantari, R. Wakkary, and K. Newby, “Ontology and Rule Based Retrieval of Sound Objects in Augmented Audio Reality System for Museum Visitors,” in *Proceedings of the ACM Symposium on Applied Computing*, pp. 1045–1050 (Nicosia, Cyprus) (2004 Mar.). https://doi.org/10.1145/967900.968114.

[69] M. Hatala and R. Wakkary, “Ontology-Based User Modeling in an Augmented Audio Reality System for Museums,” *User Model. User-Adap. Interact.*, vol. 15, no. 3, pp. 339–380 (2005 Aug.). https://doi.org/10.1007/s11257-005-2304-5.

[70] R. Wakkary and M. Hatala, “Situated Play in a Tangible Interface and Adaptive Audio Museum Guide,” *Pers. Ubiquit. Comput.*, vol. 11, no. 3, pp. 171–191 (2007 Mar.). https://doi.org/10.1007/s00779-006-0101-8.

[71] T. Langlotz, H. Regenbrecht, S. Zollmann, and D. Schmalstieg, “Audio Stickies: Visually-Guided Spatial Audio Annotations on a Mobile Augmented Reality Platform,” in *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration*, pp. 545–554 (Adelaide, Australia) (2013 Nov.). https://doi.org/10.1145/2541016.2541022.

[72] F. Heller, T. Knott, M. Weiss, and J. Borchers, “Multi-User Interaction in Virtual Audio Spaces,” in *Extended Abstracts on CHI Human Factors in Computing Systems*, pp. 4489–4494 (Boston, MA) (2009 Apr.). https://doi.org/10.1145/1520340.1520688.

[73] S. Russell, G. Dublon, and J. A. Paradiso, “HearThere: Networked Sensory Prosthetics Through Auditory Augmented Reality,” in *Proceedings of the 7th Augmented Human International Conference*, paper 20 (Geneva, Switzerland) (2016 Feb.). https://doi.org/10.1145/2875194.2875247.

[74] R. Kapoor, S. Ramasamy, A. Gardi, and R. Sabatini, “A Bio-Inspired Acoustic Sensor System for UAS Navigation and Tracking,” in *Proceedings of the 36th Digital Avionics Systems Conference*, pp. 1–7 (St. Petersburg, FL) (2017 Sep.). https://doi.org/10.1109/DASC.2017.8102080.

[75] C. Evers and P. A. Naylor, "Acoustic SLAM," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 26, no. 9, pp. 1484–1498 (2018 Sep.). https://doi.org/10.1109/TASLP.2018.2828321.

[76] A. Terán Espinoza, *Acoustic-Inertial Forward-Scan Sonar Simultaneous Localization and Mapping*, Master's thesis, KTH Royal Institute of Technology, Stockholm, Sweden (2020 Sep.).

[77] F. Kawsar, C. Min, A. Mathur, and A. Montanari, "Earables for Personal-Scale Behavior Analytics," *IEEE Pervasive Comput.*, vol. 17, no. 3, pp. 83–89 (2018 Oct.). https://doi.org/10.1109/MPRV.2018.03367740.

[78] F. Heller, A. Krämer, and J. Borchers, "Simplifying Orientation Measurement for Mobile Audio Augmented Reality Applications," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 615–624 (Toronto, Canada) (2014 Apr.). https://doi.org/10.1145/2556288.2557021.

[79] M. Tonnis and G. Klinker, "Effective Control of a Car Driver's Attention for Visual and Acoustic Guidance Towards the Direction of Imminent Dangers," in *Proceedings of the IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 13–22 (Santa Barbara, CA) (2006 Oct.). https://doi.org/10.1109/ISMAR.2006.297789.

[80] D. Kern, P. Marshall, E. Hornecker, Y. Rogers, and A. Schmidt, "Enhancing Navigation Information With Tactile Output Embedded Into the Steering Wheel," in H. Tokuda, M. Beigl, A. Friday, A. J. B. Brush, Y. Tobe (Eds.), *Pervasive Computing: Pervasive 2009*, Lecture Notes in Computer Science, vol. 5538, pp. 42–58 (Springer, Berlin, Germany, 2009). https://doi.org/10.1007/978-3-642-01516-8_5.

[81] B. N. Walker and J. Lindsay, "Navigation Performance in a Virtual Environment With Bonephones," in *Proceedings of the 11th International Conference on Auditory Display*, pp. 260–263 (Limerick, Ireland) (2005 Jul.).

[82] F. Liarokapis, "An Augmented Reality Interface for Visualizing and Interacting With Virtual Content," *Virtual Real.*, vol. 11, no. 1, pp. 23–43 (2007 Mar.). https://doi.org/10.1007/s10055-006-0055-1.

[83] A. Walker, S. Brewster, D. McGookin, and A. Ng, "Diary in the Sky: A Spatial Audio Display for a Mobile Calendar," in A. Blandford, J. Vanderdonckt, P. Gray (Eds.), *People and Computers XV—Interaction Without Frontiers*, pp. 531–539 (Springer, London, UK, 2001). https://doi.org/10.1007/978-1-4471-0353-0_33.

[84] D. McGookin, Y. Vazquez-Alvarez, S. Brewster, and J. Bergstrom-Lehtovirta, "Shaking the Dead: Multimodal Location Based Experiences for Un-Stewarded Archaeological Sites," in *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, pp. 199–208 (Copenhagen, Denmark) (2012 Oct.). https://doi.org/10.1145/2399016.2399048.

[85] V. Lim, N. Frangakis, L. M. Tanco, and L. Picinali, "PLUGGY: A Pluggable Social Platform for Cultural Heritage Awareness and Participation," in M. Ioannides, J. Martins, R. Žarnić, and V. Lim (Eds.), *Advances in Digital Cultural Heritage*, Lecture Notes in Computer Science, vol. 10754, pp. 117–129 (Springer, Cham, Switzerland, 2018). https://doi.org/10.1007/978-3-319-75789-6_9.

[86] Y. Kim, S. Hong, and G. J. Kim, "Augmented Reality-Based Remote Coaching for Fast-Paced Physical Task," *Virtual Real.*, vol. 22, no. 1, pp. 25–36 (2018 Mar.). https://doi.org/10.1007/s10055-017-0315-2.

[87] K. M. Sagayam, A. J. Timothy, C. C. Ho, L. E. Henesey, and R. Bestak, "Augmented Reality-Based Solar System for E-Magazine With 3-D Audio Effect," *Int. J. Simul. Process. Model.*, vol. 15, no. 6, pp. 524–534 (2021 Jan.). http://doi.org/10.1504/IJSPM.2020.112460.

[88] R. Behringer, C. Tam, J. McGee, S. Sundareswaran, and M. Vassiliou, "A Wearable Augmented Reality Testbed for Navigation and Control, Built Solely With Commercial-off-the-Shelf (COTS) Hardware," in *Proceedings IEEE and ACM International Symposium on Augmented Reality*, pp. 12–19 (Munich, Germany) (2000 Oct.). https://doi.org/10.1109/ISAR.2000.880918.

[89] N. Sawhney and C. Schmandt, "Nomadic Radio: Speech and Audio Interaction for Contextual Messaging in Nomadic Environments," *ACM Trans. Comput. Hum. Interact.*, vol. 7, no. 3, pp. 353–383 (2000 Sep.). https://doi.org/10.1145/355324.355327.

[90] Z. Zhou, A. D. Cheok, J. Pan, and Y. Li, "Magic Story Cube: An Interactive Tangible Interface for Storytelling," in *Proceedings of the ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, pp. 364–365 (Singapore) (2004 Jun.). https://doi.org/10.1145/1067343.1067404.

[91] J. R. Blum, M. Bouchard, and J. R. Cooperstock, "What's Around Me? Spatialized Audio Augmented Reality for Blind Users With a Smartphone," in A. Puiatti and T. Gu (Eds.), *Mobile and Ubiquitous Systems: Computing, Networking, and Services*, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 104, pp. 49–62 (Springer, Berlin, Germany, 2011). http://doi.org/10.1007/978-3-642-30973-1_5.

[92] Y. Vazquez-Alvarez, I. Oakley, and S. A. Brewster, "Auditory Display Design for Exploration in Mobile Audio-Augmented Reality," *Pers. Ubiquit. Comput.*, vol. 16, no. 8, pp. 987–999 (2012 Dec.). https://doi.org/10.1007/s00779-011-0459-0.

[93] J. Rämö and V. Välimäki, "Digital Augmented Reality Audio Headset," *J. Electric. Comput. Eng.*, vol. 2012, paper 457374 (2012 Oct.). https://doi.org/10.1155/2012/457374.

[94] R. Gupta, R. Ranjan, J. He, and W. S. Gan, "Parametric Hear Through Equalization for Augmented Reality Audio," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1587–1591 (Brighton, UK) (2019 May). https://doi.org/10.1109/ICASSP.2019.8683657.

[95] R. Gupta, J. He, R. Ranjan, et al., "Augmented/Mixed Reality Audio for Hearables: Sensing, Control, and Rendering," *IEEE Signal Process. Mag.*, vol. 39, no. 3, pp. 63–89 (2022 May). https://doi.org/10.1109/MSP.2021.3110108.

[96] T. Kitagawa and K. Kondo, "On a Wind Noise Countermeasure for Bicycle Audio Augmented Reality Systems," in *Proceedings of the 6th Global Conference on Consumer Electronics*, pp. 1–2 (Las Vegas, NV) (2017 Oct.). https://doi.org/10.1109/GCCE.2017.8229227.

[97] T. Kitagawa and K. Kondo, "Detailed Evaluation of a Wind Noise Reduction Method Using DNN for 3D Audio Navigation System Audio Augmented Reality for Bicycles," in *Proceedings of the 8th Global Conference on Consumer Electronics*, pp. 863–864 (Osaka, Japan) (2019 Oct.). https://doi.org/10.1109/GCCE46687.2019.9015262.

[98] V. Sundareswaran, K. Wang, S. Chen, et al., "3D Audio Augmented Reality: Implementation and Experiments," in *Proceedings of the 2nd IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 296–297 (Tokyo, Japan) (2003 Oct.). https://doi.org/10.1109/ISMAR.2003.1240728.

[99] S. Tachi, K. Komoriya, K. Sawada, et al., "Telexistence Cockpit for Humanoid Robot Control," *Adv. Robot.*, vol. 17, no. 3, pp. 199–217 (2003 Apr.). https://doi.org/10.1163/156855303764018468.

[100] H. Huang, M. Solah, D. Li, and L.-F. Yu, "Audible Panorama: Automatic Spatial Audio Generation for Panorama Imagery," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, paper 621 (Glasgow, UK) (2019 May). https://doi.org/10.1145/3290605.3300851.

[101] J. Yang, P. Sasikumar, H. Bai, A. Barde, G. Sörös, and M. Billinghurst, "The Effects of Spatial Auditory and Visual Cues on Mixed Reality Remote Collaboration," *J. Multimodal User Interfaces*, vol. 14, no. 4, pp. 337–352 (2020 Dec.). https://doi.org/10.1007/s12193-020-00331-1.

[102] A. Härmä, J. Jakka, M. Tikander, et al., "Techniques and Applications of Wearable Augmented Reality Audio," presented at the *114th Convention of the Audio Engineering Society* (2003 Mar.), paper 5768.

[103] A. Härmä, J. Jakka, M. Tikander, et al., "Augmented Reality Audio for Mobile and Wearable Appliances," *J. Audio Eng. Soc.*, vol. 52, no. 6, pp. 618–639 (2004 Jun.).

[104] E. Mattheiss, G. Regal, C. Vogelauer, and H. Furtado, "3D Audio Navigation - Feasibility and Requirements for Older Adults," in K. Miesenberger, R. Manduchi, M. Covarrubias Rodriguez, and P. Peňáz (Eds.), *Computers Helping People With Special Needs*, Lecture Notes in Computer Science, vol. 12377, pp. 323–331 (Springer, Cham, Switzerland, 2020). https://doi.org/10.1007/978-3-030-58805-2_38.

[105] D. N. Zotkin, R. Duraiswami, and L. S. Davis, "Rendering Localized Spatial Audio in a Virtual Auditory Space," *IEEE Trans. Multimed.*, vol. 6, no. 4, pp. 553–564 (2004 Aug.). https://doi.org/10.1109/TMM.2004.827516.

[106] L. Cliffe, J. Mansell, C. Greenhalgh, and A. Hazzard, "Materialising Contexts: Virtual Soundscapes for Real-World Exploration," *Pers. Ubiquit. Comput.*, vol. 25, pp. 623–636 (2021 Aug.). https://doi.org/10.1007/s00779-020-01405-3.

[107] G. Arvanitis, K. Moustakas, and N. Fakotakis, "Real-Time Context Aware Audio Augmented Reality," in A. Ronzhin, R. Potapova, and N. Fakotakis (Eds.), *Speech and Computer (SPECOM)*, Lecture Notes in Computer Science, vol. 9319, pp. 333–340 (Springer, Cham, Switzerland, 2015). https://doi.org/10.1007/978-3-319-23132-7_41.

[108] H. Kim, R. J. Hughes, L. Remaggi, et al., "Acoustic Room Modelling Using a Spherical Camera for Reverberant Spatial Audio Objects," presented at the *142nd Convention of the Audio Engineering Society* (2017 May), paper 9705.

[109] H. Kim, L. Remaggi, P. J. B. Jackson, and A. Hilton, "Immersive Spatial Audio Reproduction for VR/AR Using Room Acoustic Modelling From 360° Images," in *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces*, pp. 120–126 (Osaka, Japan) (2019 Mar.). https://doi.org/10.1109/VR.2019.8798247.

[110] D. Li, T. R. Langlois, and C. Zheng, "Scene-Aware Audio for 360° Videos," *ACM Trans. Graph.*, vol. 37, no. 4, paper 111 (2018 Aug). https://doi.org/10.1145/3197517.3201391.

[111] C. Schissler, C. Loftin, and D. Manocha, "Acoustic Classification and Optimization for Multi-Modal Rendering of Real-World Scenes," *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 3, pp. 1246–1259 (2018 Mar.). https://doi.org/10.1109/TVCG.2017.2666150.

[112] O. Shih and A. Rowe, "Can a Phone Hear the Shape of a Room?" in *Proceedings of the 18th International Conference on Information Processing in Sensor Networks*, pp. 277–288 (Montreal, Canada) (2019 Apr.). https://doi.org/10.1145/3302506.3310407.

[113] V. Hulusic, C. Harvey, K. Debattista, et al., "Acoustic Rendering and Auditory–Visual Cross-Modal Perception and Interaction," in *Comput. Graph. Forum*, vol. 31, vol. 1, pp. 102–131 (2012 Feb.). https://doi.org/10.1111/j.1467-8659.2011.02086.x.

[114] J. Yang, F. Pfreundtner, A. Barde, K. Heutschi, and G. Sörös, "Fast Synthesis of Perceptually Adequate Room Impulse Responses from Ultrasonic Measurements," in *Proceedings of the 15th International Conference on Audio Mostly*, pp. 53–60 (Graz, Austria) (2020 Sep.). https://doi.org/10.1145/3411109.3412300.

[115] W. G. Gardner, *3-D Audio Using Loudspeakers*, The Springer International Series in Engineering and Computer Science, vol. 444 (Springer, New York. NY, 1998).

[116] M. A. Gerzon, "Periphony: With-Height Sound Reproduction," *J. Audio Eng. Soc.*, vol. 21, no. 1, pp. 2–10 (1973 Feb.).

[117] M. Gorzel, A. Allen, I. Kelly, et al., "Efficient Encoding and Decoding of Binaural Sound With Resonance Audio," in *Proceedings of the AES International Conference on Immersive and Interactive Audio* (2019 Mar.), paper 68.

[118] M. Kentgens and P. Jax, "Translation of a Higher-Order Ambisonics Sound Scene by Space Warping," in *Proceedings of AES International Conference on Audio for Virtual and Augmented Reality* (2020 Aug.), paper 2-3.

[119] S. Moreau, J. Daniel, and S. Bertet, "3D Sound Field Recording With Higher Order Ambisonics – Objective Measurements and Validation of a Spherical Microphone," presented at the *120th Convention of the Audio Engineering Society* (2006 May), paper 6857.

[120] S. Favrot, M. Marschall, J. Käsbach, J. Buchholz, and T. Weller, "Mixed-Order Ambisonics Recording and Playback for Improving Horizontal Directionality," presented at the *131st Convention of the Audio Engineering Society* (2011 Oct.), paper 8528.

[121] Y. Tanabe, G. Yamauchi, M. Atsushi, and T. Kamekawa, "Tesseral Array for Group Based Spatial Audio Capture and Synthesis," in *Proceedings of the AES International Conference on Audio for Virtual and Augmented Reality* (2020 Aug.), paper 2-7.

[122] M. Lawton, S. Cunningham, and I. Convery, "Nature Soundscapes: An Audio Augmented Reality Experience," in *Proceedings of the 15th International Conference on Audio Mostly*, pp. 85–92 (Graz, Austria) (2020 Sep.). https://doi.org/10.1145/3411109.3411142.

[123] M. de Borba Campos, J. Sánchez, A. Cardoso Martins, R. Schneider Santana, and M. Espinoza, "Mobile Navigation Through a Science Museum for Users Who Are Blind," in C. Stephanidis and M. Antona (Eds.), *Universal Access in Human-Computer Interaction. Aging and Assistive Environments*, Lecture Notes in Computer Science, vol. 8515, pp. 717–728 (Springer, Berlin, Germany, 2014). https://doi.org/10.1007/978-3-319-07446-7_68.

[124] D. A. Mauro, R. Mekuria, and M. Sanna, "Binaural Spatialization for 3D Immersive Audio Communication in a Virtual World," in *Proceedings of the 8th Audio Mostly Conference*, paper 8 (Piteå, Sweden) (2013 Sep.). https://doi.org/10.1145/2544114.2544115.

[125] B. Xie, *Head-Related Transfer Function and Virtual Auditory Display* (J. Ross Publishing, Plantation, FL, 2013), 2nd ed.

[126] E. Wenzel, F. Wightman, D. Kistler, and S. Foster, "Acoustic Origins of Individual Differences in Sound Localization Behavior," *J. Acoust. Soc. Am.*, vol. 84, no. S1, pp. S79–S79 (1988 Nov.). https://doi.org/10.1121/1.2026486.

[127] D. R. Begault, "Challenges to the Successful Implementation of 3-D Sound," presented at the *89th Convention of the Audio Engineering Society* (1990 Sep.), paper 2948.

[128] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization Using Nonindividualized Head-Related Transfer Functions," *J. Acoust. Soc. Am.*, vol. 94, no. 1, pp. 111–123 (1993 Sep.). https://doi.org/10.1121/1.407089.

[129] Z. Yang, Y.-L. Wei, S. Shen, and R. R. Choudhury, "Ear-AR: Indoor Acoustic Augmented Reality on Earphones," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, paper 56 (London, UK) (2020 Sep.). https://doi.org/10.1145/3372224.3419213.

[130] T. Suenaga, S. Kaneko, and H. Okumura, "Development of Shape-Based Average Head-Related Transfer Functions and Their Applications," *Acoust. Sci. Technol.*, vol. 41, no. 1, pp. 282–287 (2020 Jan.). https://doi.org/10.1250/ast.41.282.

[131] B. Gardner and K. Martin, "HRTF Measurements of a KEMAR Dummy-Head Microphone," Tech Rep. 280 (1994 May).

[132] 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 (New Paltz, NY) (2001 Oct.). https://doi.org/10.1109/ASPAA.2001.969552.

[133] C. Armstrong, L. Thresh, D. Murphy, and G. Kearney, "A Perceptual Evaluation of Individual and Non-Individual HRTFs: A Case Study of the SADIE II Database," *Appl. Sci.*, vol. 8, no. 11, paper 2029 (2018 Oct.). https://doi.org/10.3390/app8112029.

[134] S. Spagnol, "Auditory Model Based Subsetting of Head-Related Transfer Function Datasets," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, pp. 391–395 (Online) (2020 May). https://doi.org/10.1109/ICASSP40776.2020.9053360.

[135] L. S. Simon, N. Zacharov, and B. F. G. Katz, "Perceptual Attributes for the Comparison of Head-Related Transfer Functions," *J. Acoust. Soc. Am.*, vol. 140, no. 5, pp. 3623–3632 (2016 Nov.). https://doi.org/10.1121/1.4966115.

[136] D. Poirier-Quinot and B. F. G. Katz, "Impact of HRTF Individualization on Player Performance in a VR Shooter Game II," in *Proceedings of the AES International Conference on Audio for Virtual and Augmented Reality* (2018 Aug.), paper P4-1.

[137] Z. Ben-Hur, D. Alon, P. W. Robinson, and R. Mehra, "Localization of Virtual Sounds in Dynamic Listening Using Sparse HRTFs," in *Proceedings of the AES International Conference on Audio for Virtual and Augmented Reality* (2020 Aug.), paper 1-1.

[138] 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," *J. Audio Eng. Soc.*, vol. 49, no. 10, pp. 904–916 (2001 Oct.).

[139] C. Pörschmann, P. Stade, and J. M. Arend, "Binaural Auralization of Proposed Room Modifications Based on Measured Omnidirectional Room Impulse Responses," in *Proc. Mtgs. Acoust.*, vol. 30, no. 1, paper 015012 (2017 Jun.). https://doi.org/10.1121/2.0000622.

[140] B. G. Shinn-Cunningham, N. Kopco, and T. J. Martin, "Localizing Nearby Sound Sources in a Classroom: Binaural Room Impulse Responses," *J. Acoust. Soc. Am.*, vol. 117, no. 5, pp. 3100–3115 (2005 May). https://doi.org/10.1121/1.1872572.

[141] S. Werner, F. Klein, A. Neidhardt, et al., "Creation of Auditory Augmented Reality Using a Position-Dynamic Binaural Synthesis System—Technical Components, Psychoacoustic Needs, and Perceptual Evaluation," *Appl. Sci.*, vol. 11, no. 3, paper 1150 (2021 Jan.). https://doi.org/10.3390/app11031150.

[142] T. Wendt, S. van de Par, and S. D. Ewert, "A Computationally-Efficient and Perceptually-Plausible Algorithm for Binaural Room Impulse Response Simulation," *J. Audio Eng. Soc.*, vol. 62, no. 11, pp. 748–766 (2014 Nov.). https://doi.org/10.17743/jaes.2014.0042.

[143] D. R. Perrott, T. Sadralodabai, K. Saberi, and T. Z. Strybel, "Aurally Aided Visual Search in the Cen-

tral Visual Field: Effects of Visual Load and Visual Enhancement of the Target," *Hum. Factors*, vol. 33, no. 4, pp. 389–400 (1991 Aug.). https://doi.org/10.1177/001872089103300402.

[144] J. J. Gibson, *The Ecological Approach to Visual Perception: Classic Edition* (Psychology Press, New York. NY, 2014).

[145] P. Fröhlich, R. Simon, L. Baillie, and H. Anegg, "Comparing Conceptual Designs for Mobile Access to Geo-Spatial Information," in *Proceedings of the 8th International Conference on Human-Computer Interaction With Mobile Devices and Services*, pp. 109–112 (Helsinki, Finland) (2006 Sep.). https://doi.org/10.1145/1152215.1152238.

[146] 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*, pp. 91–98 (Boston, MA) (2007 Oct.). http://doi.org/10.1109/ISWC.2007.4373786.

[147] K. R. May, B. J. Tomlinson, X. Ma, P. Roberts, and B. N. Walker, "Spotlights and Soundscapes: On the Design of Mixed Reality Auditory Environments for Persons With Visual Impairment," *ACM Trans. Access. Comput.*, vol. 13, no. 2, pp. 8 (2020 Apr.). https://doi.org/10.1145/3378576.

[148] M. Bandukda and C. Holloway, "Audio AR to Support Nature Connectedness in People With Visual Disabilities," in *Adjunct Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the ACM International Symposium on Wearable Computers*, pp. 204–207 (Online) (2020 Sep.). https://doi.org/10.1145/3410530.3414332.

[149] S. Joshi, K. Stavrianakis, and S. Das, "Substituting Restorative Benefits of Being Outdoors Through Interactive Augmented Spatial Soundscapes," in *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, paper 80 (Online) (2020 Oct.). https://doi.org/10.1145/3373625.3418029.

[150] R. Gupta, R. Ranjan, J. He, and G. Woon-Seng, "Investigation of Effect of VR/AR Headgear on Head Related Transfer Functions for Natural Listening," in *Proceedings of the AES International Conference on Audio for Virtual and Augmented Reality* (2018 Aug.), paper P3-9.

[151] A. Genovese, G. Zalles, G. Reardon, and A. Roginska, "Acoustic Perturbations in HRTFs Measured on Mixed Reality Headsets," in *Proceedings of International Conference on Audio for Virtual and Augmented Reality* (2018 Aug.), paper P8-4.

[152] N. Kelly, "All the World's a Stage: What Makes a Wearable Socially Acceptable," *Interactions*, vol. 24, no. 6, pp. 56–60 (2017 Nov.). https://doi.org/10.1145/3137093.

[153] R. R. Choudhury, "Earable Computing: A New Area to Think About," in *Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications*, pp. 147–153 (Online) (2021 Feb.). https://doi.org/10.1145/3446382.3450216.

[154] F. Kawsar, C. Min, A. Mathur, et al., "eSense: Open Earable Platform for Human Sensing," in *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems*, pp. 371–372 (Shenzhen, China) (2018 Nov.). https://doi.org/10.1145/3274783.3275188.

[155] S. Klockgether and S. van de Par, "Just Noticeable Differences of Spatial Cues in Echoic and Anechoic Acoustical Environments," *J. Acoust. Soc. Am.*, vol. 140, no. 4, pp. EL352–EL357 (2016 Oct.). https://doi.org/10.1121/1.4964844.

[156] C. L. Christensen, G. Koutsouris, and J. H. Rindel, "The ISO 3382 Parameters: Can We Simulate Them? Can We Measure Them?" in *Proceedings of the International Symposium on Room Acoustics*, pp. 9–11 (Toronto, Canada) (2013 Jun.).

[157] Z. Meng, F. Zhao, and M. He, "The Just Noticeable Difference of Noise Length and Reverberation Perception," in *Proceedings of the International Symposium on Communications and Information Technologies*, pp. 418–421 (Bangkok, Thailand) (2006 Oct.). https://doi.org/10.1109/ISCIT.2006.339980.

[158] W. Bailey and B. Fazenda, "The Effect of Visual Cues and Binaural Rendering Method on Plausibility in Virtual Environments," presented at the *144th Convention of the Audio Engineering Society* (2018 May), paper 9921.

[159] 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 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4473–4477 (Florence, Italy) (2014 May). https://doi.org/10.1109/ICASSP.2014.6854448.

[160] A. Meshram, R. Mehra, H. Yang, et al., "P-HRTF: Efficient Personalized HRTF Computation for High-Fidelity Spatial Sound," in *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, pp. 53–61 (Munich, Germany) (2014 Oct.). https://doi.org/10.1109/ISMAR.2014.6948409.

[161] M. T. Islam and I. J. Tashev, "Anthropometric Features Estimation Using Integrated Sensors on a Headphone for HRTF Personalization," in *Proceedings of the AES International Conference on Audio for Virtual and Augmented Reality* (2020 Aug.), paper 1-7.

[162] S. Kaneko, T. Suenaga, and S. Sekine, "DeepEarNet: Individualizing Spatial Audio With Photography, Ear Shape Modeling, and Neural Networks," in *Proceedings of the AES International Conference on Audio for Virtual and Augmented Reality* (2016 Sep.), paper 6-3.

[163] F. Shahid, N. Javeri, K. Jain, and S. Badhwar, "AI DevOps for Large-Scale HRTF Prediction and Evaluation: An End to End Pipeline," in *Proceedings of the AES International Conference on Audio for Virtual and Augmented Reality* (2018 Aug.), paper P9-4.

[164] G. W. Lee, J. H. Lee, S. J. Kim, and H. K. Kim, "Directional Audio Rendering Using a Neural Network Based Personalized HRTF," in *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pp. 2364–2365 (Graz, Austria) (2019 Sep.).

[165] J.-M. Jot and K. S. Lee, "Augmented Reality Headphone Environment Rendering," in *Proceedings of the AES*

International Conference on Audio for Virtual and Augmented Reality (2016 Aug.), paper 8-2.

[166] R. Behringer, S. Chen, V. Sundareswaran, K. Wang, and M. Vassiliou, "A Distributed Device Diagnostics System Utilizing Augmented Reality and 3D Audio," in M. Gervautz, D. Schmalstieg, and A. Hildebrand (Eds.), *Virtual Environments '99*, Eurographics, pp. 105–114 (Springer, Vienna, Austria, 1999). https://doi.org/10.1007/978-3-7091-6805-9_11.

[167] R. Behringer, S. Chen, V. Sundareswaran, K. Wang, and M. Vassiliou, "A Novel Interface for Device Diagnostics Using Speech Recognition, Augmented Reality Visualization, and 3D Audio Auralization," in *Proceedings of the IEEE International Conference on Multimedia Computing and Systems*, vol. 1, pp. 427–432 (Florence, Italy) (1999 Jun.). https://doi.org/10.1109/MMCS.1999.779240.

[168] C. M. Tomaino, "Music Therapy and the Brain," in B. L. Wheeler (Ed.), *Music Therapy Handbook*, pp. 40–50 (Guilford Press, New York, NY, 2015).

[169] U. Chong and S. Alimardanov, "Audio Augmented Reality Using Unity for Marine Tourism," in M. Singh, D. K. Kang, J. H. Lee, U. S. Tiwary, D. Singh, W. Y. Chung (Eds.), *Intelligent Human Computer Interaction*, Lecture Notes in Computer Science, vol. 12616, pp. 303–311 (Springer, Cham, Switzerland, 2020). http://doi.org/10.1007/978-3-030-68452-5_31.

THE AUTHORS

Jing Yang

Amit Barde

Mark Billinghurst

Jing Yang was a research assistant at ETH Zurich and is now a senior researcher at Huawei. She received her Ph.D. in 2021 from ETH Zurich. Her research interests lie in the intersection of augmented reality and human-computer interaction with specific focus on spatial audio, room acoustics, music style transfer, and related applications. She has been active in academic and industrial activities in several companies and institutes, including Idiap Research Institute, The University of Auckland, and Nokia Bell Labs.

•

Amit Barde is a Research Fellow at the Empathic Computing Laboratory, University of Auckland. He received his Ph.D. in Human Interface Technology from the HIT-Lab NZ, where he explored the use of spatialized auditory cues for information delivery on wearable devices. His research interests include information delivery using spatialized auditory cues, interactive audio, the effects of sound on empathy, and its use in the management and treatment of tinnitus.

•

Mark Billinghurst is Director of the Empathic Computing Laboratory and Professor at the University of South Australia in Adelaide, Australia, and at the University of Auckland in Auckland, New Zealand. He earned a Ph.D. in 2002 from the University of Washington and conducts research on how virtual and real worlds can be merged, publishing over 650 papers on augmented reality (AR), virtual reality, remote collaboration, empathic computing, and related topics. In 2013, he was elected as a Fellow of the Royal Society of New Zealand and, in 2019, was given the IEEE International Symposium on Mixed and Augmented Reality (ISMAR) Career Impact Award in recognition for lifetime contribution to AR research and commercialization.