Extracting a Definition and Taxonomy for Audio Augmented Reality (AAR) Using Grounded Theory

Author(s) , Affiliation(s)

The concept of Augmented Reality (AR) has evolved over the years since its inception in 1957. However, as of today, AR is mostly realized using a visual medium. Almost all applications of AR rely on the user's ability to see. A few recent works have attempted to explore a new modality for AR–e.g., audio. However, the concept of using audio to augment reality has been considered debatable and there is no specific definition of the concept. To better understand this new concept, we launched a study using Grounded Theory to develop a definition and taxonomy for the concept of Audio Augmented Reality (AAR). This paper shares the preliminary results based on the activities conducted thus far, and hopes to generate discussion on using audio to augment reality within the Human Factors community.

INTRODUCTION

The ability to modify our reality to produce novel experiences has been a meaningful and novel endeavor since the late 20th century. These pursuits have led to the development of Augmented Reality (AR) and Virtual Reality (VR) technology. The technology for these concepts is still relatively young, given that these concepts were without a clear definition up until 1994 when Milgram's Reality-Virtuality Continuum was introduced (Milgram & Kishino, 1994). Their continuum described the virtual world and the real world to lay on either ends of a spectrum, whereas the space between was defined as Mixed Reality (MR).

Within this MR space and closer to the real-world end, lies AR. According to Drascic et al. (1993), AR is the essence of computer graphic enhancement of video images of real scenes. Sikora et al. (2018) defined AR as the combination virtual and real contents such that the virtual content is associated with real-world information, and it runs interactively in 3D in real time. Yet another definition is the superimposing computer-generated content, such as interactive 2D and 3D multimedia objects, in real time on a view of real objects (Rumiński, 2015). A study from 2021 (Skarbez et al., 2021) revisited the Reality Virtuality Continuum in an attempt to provide an updated taxonomy. This was necessary because the lines between AR, VR, and MR began to blur given the technological developments and applications. The updated taxonomy included three dimensions – Extent of World Knowledge, Immersion, and Coherence. Extent of World Knowledge describes the extent of real-world awareness by the system; Immersion describes the sense of illusion provided by the system, and Coherence describes the plausibility of the system as experienced by the user. According to the authors, these three dimensions together describe true Mixed Reality experience. Within the space of AR, Skarbez et al. (2021) postulates that the system should meet the following requirements: it should provide a high extent of world knowledge to allow the user to maintain above average situation awareness; it should provide immersion to the extent that the user's experience is mediated sufficiently, and the interaction must be coherent to maintain the authenticity of the experience.

As of the 21st century, AR applications truly have been a vision-based experience. Placing advertisements virtually, using mobile phone-based AR navigation, heads up displays in vehicles, cultural heritage tours of sites augmented with visual elements, and virtual safaris are all forms of vision-based AR; the applications also expand to using AR to develop and modify car designs, study vehicle crash performance, view- related documentation for hardware, as well as surgeons previewing simulations of their actions before acting (Krevelen & Poelman, 2010). There has been research of using audio to produce AR experiences, but these endeavors have not been realized to the extent visual AR has been. Human hearing holds two key advantages. First, since hearing is a separate modality of perception than vision, auditory information is processed using mental resources that are not shared with the visual system, according to Multiple Resource Theory (Wickens, 2008). This allows for better dual-task performance. AR experiences are defined by a higher sense of the real world (Extent of World Knowledge); visual overlays can be more distracting from the current task than auditory overlays. Auditory information allows for parallel processing with minimal interruption of the primary visual task, and increases knowledge of the real world. The second advantage lies in human hearing being omnidirectional. The human ear can detect sounds from all directions on the azimuth and altitude planes. Localization is the ability to perceive the direction of a sound source. Our ability to localize sounds has given rise to multiple localization studies especially in the field of military combat and occupational safety (Clasing & Casali, 2014; Casali, 2009; Cave et al., 2019; Khaled A. Alali, 2011; Alali & Casali, 2012), where knowing the source of the sound can be a matter of life or death.

To provide a new taxonomy specifically addressing the use of audio-based AR applications, and generate a definition for this concept, we are conducting this study using Grounded Theory. We hope to provide a theoretical understanding of AAR, identify how well it fits within the existing dimensions of AR, and lastly, develop a definition for it. This paper discusses the activities that have been done towards achieving this goal, shares preliminary findings, and discusses the next planned steps.

METHOD

Given the novel nature of this work, we hope to generate a theoretical understanding of Audio Augmented Reality (AAR) from data that are systematically generated and analyzed using comparative analysis (Tie et al., 2019). Our method to systematically generate this data followed a Grounded Theory approach involving two stages – firstly, we conducted a workshop at the International Conference on Auditory Display (ICAD) 2021 to hear from audio researchers and practitioners about their general understanding of this concept; secondly, we combed through the literature to develop a basic taxonomy and necessary aspects. We also conducted a preliminary focus group with non-experts to hear their thoughts on the concept of AAR. Using this iterative process, we hope to be able to refine our initial taxonomy on the different aspects of AAR, and finally produce a definition, grounded in data, that would define the AAR space on Milgram's continuum.

Grounded Theory

Grounded Theory is a form of qualitative research method that is excellent for developing taxonomy or creating frameworks. It involves all stages of research including getting participants, collecting data, analyzing data, employing literature, and writing up the paper itself (Hoda et al., 2011). It seeks a very data-driven approach to create theories and gives great ability to the researchers in being able to connect their data together to form conclusions (Furniss et al., 2011). There are two types of grounded theory, Straussian Grounded Theory (SGT) and Glaserian Grounded Theory (GGT). The main difference between the two is that SGT asks, "what if?" for each word in the data, while GGT asks "what do we have here?" for the data per se (Niekerk & Roode, 2009). The current study followed GGT where we allowed the concepts to emerge from data collected. Our literature review showed that AAR experiences shared common aspects such as the auditory display design, the form factors, and use cases. For the collection of data from participants, Grounded Theory suggests that we conduct "semi-structured interviews with open-ended questions", where the questions are iterative (develops as a result of previous questions) (Hoda et al., 2011). We used the concepts generated from the workshop to drive the following data collection strategies. During the workshop hosted at ICAD 2021, we asked attendees to brainstorm and rate what audio AR was, who might benefit from it, contexts of applications, and form factors in general. Next, we used the general ideas from the workshop to study the literature to further develop different aspects of AAR. The next activity, a preliminary focus group, used semi-structured questions as a guide on the same aspects but expanded to include broader questions on the concept. These data generation activities were part of the Purposive Sampling stage of Grounded Theory application.
To analyze our data, we applied the "Open Coding procedure" and the "Constant Comparison Method" to catch patterns across data. Effectively, the overall process is "interviewing a sample population about a topic, transcribing the interviews, coding parts of the transcript, and relating these codes to one another" (Furniss et al., 2011). For example, through the coding process we were able to generate an additional aspect, "Context". However, given that the Purposive Sampling stage has not been completed, further aspects are expected to emerge. Commonalities or differences between the generated aspects of AAR as identified by constant comparison of the data at each stage allowed us to identify dimensions of AAR. Once no new dimensions emerged through the process of constant comparison, we were able to string together the different dimensions by tying them together with the main concept of augmenting reality through audio to produce a definition grounded in data.

AAR Workshop

We conducted a two-hour workshop session with attendees from different countries who were both experts and non-experts at the virtual ICAD 2021. The main purpose was to create a discussion on the use of audio AR applications, gauge people's general understanding of the concept, and identify use cases and form factors for this technology. In total 28 audio researchers and practitioners participated in our workshop, but demographic information was available from only 24 attendees. 76% of attendees worked in academia, 21% worked in industry, and 3% in government; there were 8 attendees from USA, 4 from UK, 2 each from Sweden, France, and Canada, and 1 each from Netherlands, Italy, Germany, India, Finland, and South Korea.
The workshop was hosted virtually using Zoom, during which attendees took part in a brainstorming session using Miro, a virtual white board. They posted virtual notes against considerations such as "what is AAR", "who would benefit the most from AAR", "what form factors would AAR take" or "best context of AAR use cases". These questions were provided by the organizers as a general starting point of discussion. Following the brainstorming session, attendees worked in groups to generate more detailed use cases and a persona for each use case; these were presented after a coffee break. At the end of the workshop, attendees answered a questionnaire about their new understanding of AAR and also rated potential use cases on a scale from1 to 5 with 1 being a weak use case and 5 being a strong use case.

Literature Review

A review of the literature was conducted using keywords such as '(augmented reality)' OR '(auditory displays)' OR '(augmented reality AND audio)' OR '(audio in AR)'. The keywords were fed into databases such as EBSCOhost, IEEE Xplore, ACM Digital Library, and Google Scholar website. Papers were identified by reviewing the abstract. Eligible papers employed the use of audio and auditory displays to enhance experiences for users in different contexts. As part of the effort to use Grounded Theory, we looked for keywords and themes that matched with the ones that were generated in the workshop.

Preliminary Focus Group

We conducted a preliminary focus group with non-experts (those without a PhD or work experience of over 10 years in the field of audio or AR technology) to gauge their input towards defining AAR. We had two participants who were undergraduate engineering students. The focus group followed a guide with open-ended questions. Some of the questions included, “Do you think listening to music, through loudspeakers or earphones, can be considered an example of audio augmented reality?” or “As an emerging field, what future directions do you envision for AAR?”. Towards the end, we showed a video demonstration of a head worn prototype device developed by London based startup, ARx Vision, showing the different capabilities of the device; the participants were then asked to rate them on a scale of 1 to 7, and provide their comments.

PRELIMINARY RESULTS

Given that the current study is a work in progress, the results shared in this paper are preliminary results from the data generation activities that have been completed.

AAR Workshop

Table 2: User and Form Factors Popularity

Users	% Of attendees
Visually impaired, Tourists	100
Mentally challenged, High mental workload, Athletes	93.8
Surgeons, Context specific	6.3
Form Factors	**% Of attendees**
Air conduction earphones	87.5
Bone conduction earphones	81.3
Wrist wearables	75
Smart glasses	68.8
VR headsets	56.3
External devices (not on person)	6.3

Table 3: Use Case Ratings

Use Case	Rating Mode	% Of attendees
Sonifying historical events	5	56.3
Sonifying visual artwork	5	43.8
Sonifying graphical data	4	37.5
Wayfinding in the real world	4	37.5
Warning signals	4, 5	31.3
Feedback signals	4	31.3
3D sound with 3D movies	3	37.5
Listening to music through earphones	1	56.3
Listening to music through speakers	1	50
Navigating a visual menu	1	43.8
Navigating an auditory menu	1	31.3

Based on the discussion during the workshop, attendees generated the following data on plausible AAR users, form factors, and use cases. Each use case was rated on a scale of 1 to 5 with 1 being a weak example and 5 being a strong example. Tables 2 shows the percentage of attendees agreeing on the following items, and Table 3 shows the mode of ratings for each use case along with the percentage of attendees agreeing on it.

Literature Review

Our literature review, driven by the results of the workshop, was the next step in identifying use cases, targeted user population, form factors, and auditory display designs that are currently being used in research to investigate audio AR applications. Not surprisingly, commonalties were noted between the two that led to the emergence of certain aspects of audio AR. Figure 2 shows the initial taxonomy of the dimensions of AAR.

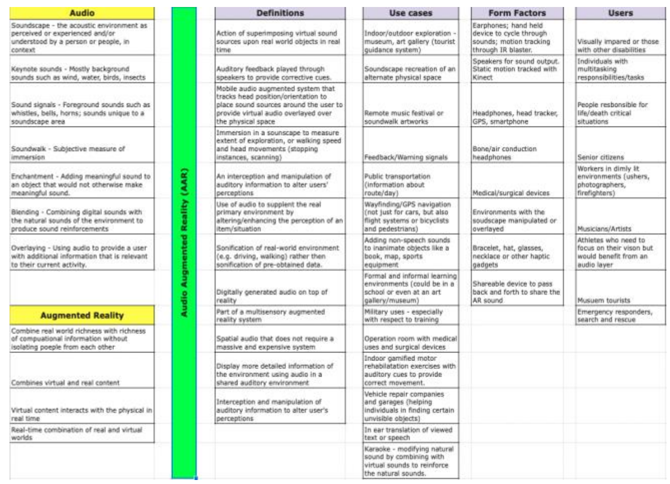

Figure 2: Initial taxonomy of AAR showing its different aspects

Preliminary Focus Group

The purpose of our preliminary focus group was to collect the input of non-experts on the data generated thus far, and identify the need to make any changes to our focus group guide for the future. Inputs from the current participants allowed us to refine the taxonomy and identify opportunities to refine our AAR aspects. Table 4 shows the unique keywords we extracted to coalesce the AR aspects further.

Table 4: Unique keywords from the preliminary focus group

Definition	Use Cases	Form Factors	Users
Personalized	Emotion regulation	Light weight	All working adults
Could be shared	Educational context	Long term use	Remote workers
Digitally generated	GPS guidance	Bone conduction earphones	With hearing disability
Remote source	AR games	Pass through features	With vision disability

DISCUSSION

This work-in-progress study aims to provide a structured theoretical foundation to the concept of audio AR in order to support the development of the technology for it.
Our literature review showed that indoor and outdoor exploration activities were of the highest importance. Most research towards improving indoor exploration has been

applied to locations such as museums, art galleries, or for virtual search tasks in a physical space (Bederson, 1995; Heller et al., 2016; Rumiński, 2015; Vazquez-Alvarez et al., 2016). Outdoor exploration studies focusing on providing enhanced experiences of events from the past, recreating soundscapes to supplement an existing experience such as that of a safari or providing information about their surroundings to tourists in a meaningful way, have also been conducted in attempts to improve the effectiveness of the auditory information being provided (Boletsis & Chasanidou, 2018; Lawton et al., 2020; Vazquez-Alvarez et al., 2012). A few niche applications have been using auditory cues to provide feedback to help maintain pace while exercising or to correct motor rehabilitation movements (Cavalcanti et al., 2019; Maculewicz & Serafin, 2015). Similar use cases were reported during our workshop with sonifying historical events and visual artwork, wayfinding, and feedback sounds being the most highly rated use cases. The focus group participants corroborated these results by affirming the notion that the auditory input in each use case needed to provide the appropriate information and be capable of providing an increased sense of immersion within their immediate or connected environment. In essence, the sense of immersion is a valuable dimension for AAR. Sikora et al. (2018) described the different elements of sound design necessary to create an immersive soundscape. Vazquez-Alvarez et al. (2016) took these considerations a step further to test a multi-level auditory display. An exocentric top layer of sound was designed to attract the user's attention, while a secondary layer (exocentric or egocentric) was made available to the user as they approach the object of interest. This distance-based approach to design auditory information was used in outdoor exploration studies to engage the user in an increased capacity with the environment as well (Vazquez-Alvarez et al., 2012). Furthermore, audio design requirements were provided by Krzyzaniak et al. (2019) to demonstrate how developers can implement exocentric sounds to enrich the environment; for example, "enchantment" to provide meaningful sounds to static inanimate objects, and "blending" to combining digital and natural sounds of the environment to produce sound reinforcement. These show that the existing immersion dimension, as defined by Skarbez et al. (2021), also applies to AAR.

Our data reflected the opinion that AAR could potentially impact all users, irrespective of their demographic information. For instance, museums may be visited by visitors of varying backgrounds, but AAR-based interventions can improve the experience regardless of their background (Bederson, 1995; Yang & Chan, 2019). AAR design should be context driven such unique user groups such as office workers as well as drone pilots can benefit from its application. As almost all our participants identified, visually impaired individuals can benefit more from AAR experiences. A study on visually impaired users showed that the requirements could vary appreciably from the requirements for those with normal vision, especially in regards to sound spatialization (Blum et al., 2013). At this stage we can thus conclude that the different users can benefit from AAR depending on the context or goal. Wayfinding was touted as a highly beneficial application of AAR, but its implementation for sighted and blind users would mean different levels of world knowledge and coherence for different users. Thus, for audio specific AR applications, we believe that these should be considered together as a user context.

A couple of other points naturally emerged during the workshop as well as in the focus group. First, whether AAR would require the sound to be digitally generated, and second, if AAR could be a shared experience. For example, participants pondered over the question, "Would listening to an electric guitar played live be an example of AAR?", and "Would attending a concert be an example of shared AAR?". Most of the responses opined that AAR can only be personalized experiences, and if it became a 'shared' experience, it would just become part of the overall real world. However, counter responses argued that sounds provided to specific individuals for their specific purposes at the same time would still be considered AAR. Therefore, we consider the customization of these sounds for an individual's specific task to be a defining factor for AAR. In comparison to visual or traditional AR, the visual elements are only made available to the single user and are usually done so for a specific purpose such as heads-up-displays while driving. For AAR, however, auditory cues can be available on a personal as well as environmental level. Museum goers might benefit from a general soundscape playing through speakers to enrich an installation, but individual users, using personal listening devices could opt to listen to sounds from a specific time period, thus creating a more customized experience. The distinction lies in the fact that customization of the general auditory information can create meaningful AAR experiences.

Relying on our generated data and the Grounded Theory approach, we postulate the following three dimensions for AAR. One, 'Immersion' to ensure that the sounds lead to enhanced perceptions or augmented experiences; two, 'User Context' meaning the information must be applicable and assistive to the user's primary task, and three, 'Personalization' meaning sounds may be audible to more than one user in the environment, but they are meaningful and unique only for the intended user. Figure 3 shows the three dimensions of AAR that we have been able to identify. Based on this, we can thus propose the following definition for AAR: "Auditory information, customized for the intended user that is capable of sufficiently immersing yet retaining awareness of their environment and designed to provide appropriate assistance in the user's primary task".

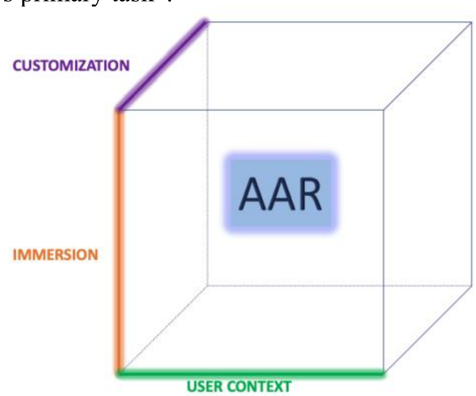

Figure 3: Dimensions of AAR

FUTURE WORK

Further data generation activities are yet to be completed. We will conduct three additional focus groups. Using the method of Constant Comparison, we plan to update our taxonomy, dimensions of AAR, and refine our definition.

After the focus groups, we will conduct one on one expert interviews with audio and AR experts. We identify experts as those with a PhD in the field of audio or AR technology, or having had worked in industry for at least 10 years. The expert interviews will follow the same semi-structured questions from the focus group. We hope that the current findings and questions raised in this paper will raise interest and inspire discussions on the concept of AAR within the Human Factors community.

REFERENCES

Bederson, B. B. (1995). Audio augmented reality: A prototype automated tour guide. *Conference Companion on Human Factors in Computing Systems - CHI '95*, 210–211. https://doi.org/10.1145/223355.223526

Blum, J. R., Bouchard, M., & Cooperstock, J. R. (2013). Spatialized Audio Environmental Awareness for Blind Users with a Smartphone. *Mobile Networks and Applications*, *18*(3), 295–309. https://doi.org/10.1007/s11036-012-0425-8

Boletsis, C., & Chasanidou, D. (2018). Audio augmented reality in public transport for exploring tourist sites. *Proceedings of the 10th Nordic Conference on Human-Computer Interaction*, 721–725. https://doi.org/10.1145/3240167.3240243

Cavalcanti, V. C., Ferreira, M. I. de S., Teichrieb, V., Barioni, R. R., Correia, W. F. M., & Da Gama, A. E. F. (2019). Usability and effects of text, image and audio feedback on exercise correction during augmented reality based motor rehabilitation. *Computers & Graphics*, *85*, 100–110. https://doi.org/10.1016/j.cag.2019.10.001

Chun Tie, Y., Birks, M., & Francis, K. (2019). Grounded theory research: A design framework for novice researchers. *SAGE Open Medicine*, *7*, 205031211882292. https://doi.org/10.1177/2050312118822927

Clasing, J. E., & Casali, J. G. (2014). Warfighter auditory situation awareness: Effects of augmented hearing protection/enhancement devices and TCAPS for military ground combat applications. *International Journal of Audiology*, *53*, 43–52.

Drascic D, Grodski J, Milgram P, Ruffo K, Wong P, & Zhai S. (1993). A Display System for Augmenting Reality. *InterCHI '93 Conf on Human Factors in Computing Systems*, *88*, 521.

Furniss, D., Blandford, A., & Curzon, P. (2011). Confessions from a grounded theory PhD: Experiences and lessons learnt. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 113–122. https://doi.org/10.1145/1978942.1978960

Heller, F., Jevanesan, J., Dietrich, P., & Borchers, J. (2016). Where are we?: Evaluating the current rendering fidelity of mobile audio augmented reality systems. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 278–282. https://doi.org/10.1145/2935334.2935365

Hoda, R., Noble, J., & Marshall, S. (2011). Grounded theory for geeks. *Proceedings of the 18th Conference on Pattern Languages of Programs - PLoP '11*, 1–17. https://doi.org/10.1145/2578903.2579162

John G. Casali, W. A. A., Jeff A. Lancaster, & Auditory Systems Laboratory, V. T., Blacksburg, VA, U. S. Army Aeromedical Research Laboratory (USAARL), Fort Rucker, AL, USA. (2009). A field investigation of hearing protection and hearing enhancement in one device: For soldiers whose ears and lives depend upon it. *Noise & Health*, *11*(42), 69–90.

Kara M. Cave, B. T., Kichol Lee and John G. Casali. (2019). Optimisation of an auditory azimuth localisation training protocol for military service members. *INTERNATIONAL JOURNAL OF AUDIOLOGY*. https://doi.org/10.1080/14992027.2019.1691747

Khaled A. Alali, J. G. C. (n.d.). The challenge of localizing vehicle backup alarms: Effects of passive and electronic hearing protectors, ambient noise level, and backup alarm spectral content. *Noise & Health*, *13*(51), 99–112. https://doi.org/10.4103/1463-1741.77202

Khaled Alali, and John G. Casali. (2012). Auditory backup alarms: Distance-at-first- detection via in-situ experimentation on alarm design and hearing protection effects. *Work 41*, 3599–3607. https://doi.org/10.3233/WOR-2012-0671-3599

Krzyzaniak, M., Frohlich, D., & Jackson, P. J. B. (2019). Six types of audio that DEFY reality!: A taxonomy of audio augmented reality with examples. *Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound*, 160–167. https://doi.org/10.1145/3356590.3356615

Lawton, M., Cunningham, S., & Convery, I. (2020). Nature soundscapes: An audio augmented reality experience. *Proceedings of the 15th International Conference on Audio Mostly*, 85–92. https://doi.org/10.1145/3411109.3411142

Maculewicz, J., & Serafin, S. (2015). A Stationary Bike in Augmented Audio Reality. *Proceedings of the 3rd 2015 Workshop on ICTs for Improving Patients Rehabilitation Research Techniques*, 164–166. https://doi.org/10.1145/2838944.2838984

Milgram, P. & Kishino, Fumio. (1994). A TAXONOMY OF MIXED REALITY VISUAL DISPLAYS. *IEICE Transactions on Information Systems, Vol E77-D*(No.12), 16.

Rumiński, D. (2015). An experimental study of spatial sound usefulness in searching and navigating through AR environments. *Virtual Reality*, *19*(3–4), 223–233. https://doi.org/10.1007/s10055-015-0274-4

Sikora, M., Russo, M., Đerek, J., & Jurčević, A. (2018). Soundscape of an Archaeological Site Recreated with Audio Augmented Reality. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *14*(3), 1–22. https://doi.org/10.1145/3230652

Skarbez, R., Smith, M., & Whitton, M. C. (2021). Revisiting Milgram and Kishino's Reality-Virtuality Continuum. *Frontiers in Virtual Reality*, *2*, 647997. https://doi.org/10.3389/frvir.2021.647997

Van Krevelen, D. W. F., & Poelman, R. (2010). A Survey of Augmented Reality Technologies, Applications and Limitations. *International Journal of Virtual Reality*, *9*(2), 1–20. https://doi.org/10.20870/IJVR.2010.9.2.2767

van Niekerk, J. C., & Roode, J. (2009). *Glaserian and Straussian grounded theory: Similar or completely different?* 8.

Vazquez-Alvarez, Y., Aylett, M. P., Brewster, S. A., Jungenfeld, R. V., & Virolainen, A. (2016). Designing Interactions with Multilevel Auditory Displays in Mobile Audio-Augmented Reality. *ACM Transactions on Computer-Human Interaction*, *23*(1), 1–30. https://doi.org/10.1145/2829944

Vazquez-Alvarez, Y., Oakley, I., & Brewster, S. A. (2012). Auditory display design for exploration in mobile audio-augmented reality. *Personal and Ubiquitous Computing*, *16*(8), 987–999. https://doi.org/10.1007/s00779-011-0459-0

Wickens, C. D. (2008). Multiple Resources and Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 449–455. https://doi.org/10.1518/001872008X288394

Yang, J., & Chan, C. Y. (2019). Audio-augmented museum experiences with gaze tracking. *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*, 1–5. https://doi.org/10.1145/3365610.3368415