

# Interactive Auditory Mediated Reality: Towards User-defined Personal Soundscapes

Gabriel Haas<sup>\*</sup>, Evgeny Stemasov<sup>\*</sup>, Michael Rietzler and Enrico Rukzio

Institute of Media Informatics  
Ulm University, Ulm  
<firstname>.<lastname>@uni-ulm.de

## ABSTRACT

Many people utilize audio equipment to escape from noises around them, leading to the desired isolation but also dangerously reduced awareness. Mediation of sounds through smarter headphones (e.g., hearables) could address this by providing non-uniform interaction with sounds while retaining a comfortable, yet informative soundscape. In a week-long event sampling study ( $n = 12$ ), we found that users mostly desire muting or a distinct "quiet-but-audible" volume for sound sources. A follow-up study ( $n = 12$ ) compared a reduced interaction granularity with a continuous one in VR. Usability and workload did not differ significantly for the two granularities but a set of four states can be considered sufficient for most scenarios, namely: "muted", "quieter", "louder" and "unchanged", allowing for smoother interaction flows. We provide implications for the design of interactive auditory mediated reality systems enabling users to be safe, comfortable and less isolated from their surroundings, while re-gaining agency over their sense of hearing.

## Author Keywords

Auditory Mediated Reality, Auditory Augmented Reality, Hearables, Augmented Hearing, Soundscapes

## CCS Concepts

•**Human-centered computing** → **Human computer interaction (HCI)**; *Ubiquitous and mobile computing*; •**Applied computing** → Sound and music computing;

## INTRODUCTION

Today's urban settings are populated by rich acoustic environments, composed of various sound sources [104]. Parts of these soundscapes are aesthetically pleasing to the population. However, there are also bothersome and irrelevant sounds which cause discomfort [120] and are linked to severe health issues [12]. Unlike vision, hearing is not directed and can hardly be averted. To avert it

<sup>\*</sup>Both authors contributed equally

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*DIS '20, July 6–10, 2020, Eindhoven, Netherlands.*

© 2020 Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-6974-9/20/07 ...\$15.00.

<http://dx.doi.org/10.1145/3357236.3395493>

like gaze, one has to essentially close it off entirely – which is what users attempt to replicate with personal audio technology [44, 26]. With the rise of ubiquitous personal audio devices these characteristics of hearing are reflected in the increasing use of headphones and similar devices. The use of personal audio technology is essentially an interaction with the surrounding soundscape on a personal level (i.e., the acoustic effect is only perceptible to the user). However, this interaction is uniform: all sounds in the environment are essentially affected to the same degree and in the same fashion. Therefore, all sounds are dampened or masked virtually alike. Current devices add a static degree of filtering (passive, like physical dampening e.g. earplugs or active, like active noise cancelling headphones) and augmentation via playback of media. Those augmentations can also act as masking sounds, supporting the already present dampening. Putting on headphones therefore drowns out all sound sources in the environment the same way, independently of aesthetics and relevance. While this allows users to curate what they hear [26], it also impacts situational awareness [67] and has distinct safety risks [28] when used in mobile contexts. In the current state of headphone technology, users have to cope with the drawbacks of this augmentation.

Emerging devices, called *hearables*, claim to be able to alter users' auditory perception through a permanently worn digital device, providing an auditory mediation of the perceived reality. In theory, they allow to filter particular unwanted sounds and set up a specific hearing profile, comparable to an equalizer for real world sounds. While this is a compelling vision and direction, practical requirements for such devices and interactions have not been determined yet.

In an initial step towards understanding the requirements for interactive auditory mediated reality (*IAMR*), two studies were conducted: 1) A week-long event sampling study with 12 participants, followed up by semi-structured interviews. 2) An interaction study comparing audio-only and screen-based interaction concepts conducted in virtual reality (VR) with  $n = 12$  participants. In the event sampling study participants recorded and rated 225 sources in their environment which they would like to alter the volume of. The semi-structured interviews aimed to further explore the concept of interacting with the surrounding soundscape and corresponding usage scenarios. Based on the insights of this study, we developed and implemented two interaction concepts for real-world sound source manipulation. In the second study those concepts were evaluated in terms of usability and workload.

The results of the first study indicate that besides muting, a distinct, "quiet-but-audible" volume exists, which caters to two requirements at the same time: aesthetics and information acquisition. For the second study, in terms of usability and workload, no significant differences were found for the two vastly different conceptual modalities of audio-only and screen-based interaction. However, we identified a set of four volume states to be sufficient for most scenarios: "muted", "quieter", "louder" and "unchanged". We propose interaction flows allowing better device-independent application of sound.

To summarize, the main contributions of this work are:

1. Usage requirements and patterns of potential users, which include a possible reduction of states, and usage contexts.
2. Specific insights for the system design of *IAMR*, such as workload of audio-only and screen-based implementations.
3. Implications for the design of future *IAMR* systems.

### INTERACTIVE AUDITORY MEDIATED REALITY

The concept of augmented, interactive hearing relies on three fundamental concepts: mediated reality, human augmentation, and hearables. The following sections introduce and define the relevant aspects of these concepts and ground them in previous research and products, concluding with a definition of *IAMR* which result from the basic concepts discussed. This concept of interacting with surrounding sounds on a per-source level was introduced and discussed in few other works [102, 119, 110]. We aim to build upon these works and further formalize the underlying concepts and nomenclature, along with applicable interaction patterns.

*Mediated reality* is the result of relaying and/or manipulating human sensation via a device. Usually, this involves real or virtual content being recorded, processed and played back to a user. The term *mediated reality* was coined by Steve Mann [75, 78]. He describes it as a framework which "combines augmented and diminished reality" [77], allowing deliberate addition as in augmented reality (AR) but also alteration and removal of content from any human sense. As a "smart" device is inserted between world and user, the device receives control over the user's sensation and perception. Conceptually, it allows *any* kind of sensation for the user whether they are somehow grounded in reality or not.

Inserting a device between world and user is a necessary prerequisite for *mediated reality* [62, 41, 75]. This applies to vision, as in AR glasses, as well as in hearing. As in vision, mediating sounds does not only include the addition of content. Through digital mediation, every single aspect of sounds and soundscapes can be altered or added, potentially creating a substitutional virtual world. Sound is recorded with a microphone and relayed to the user via headphones (i.e., via hearables or any other personal audio device). Similarly to visual AR, auditory AR primarily deals with the addition of content to users' perception. It does not intend to replace the real world, but to enrich it with information [14]. When mapped to two axes (Mediation and Virtuality) as done by Mann in [75], four clusters can be identified (see Figure 1) and presented using exemplary systems.

*Augmenting human* intellect and abilities has been a recurring topic in general literature and research [36, 84, 76]. Douglas Engelbart was among the first ones to frame this process and concept towards computers as tools or a way to augment

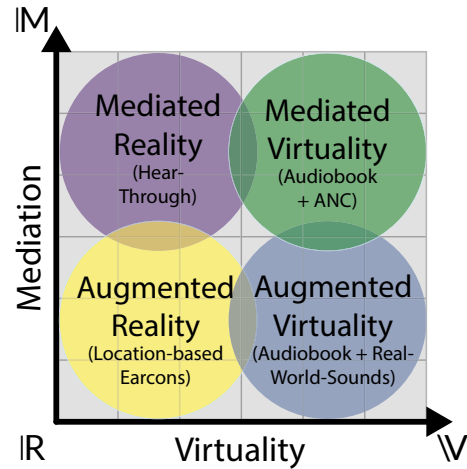


Figure 1: Depiction of mediation and virtuality axis adapted to audio, based on Steve Mann [75, 78]

human intellect [36]. With the help of devices, physical and cognitive capabilities of users can be restored or expanded. It is even possible to add new capabilities, either through bodily integration/implantation [23], re-mapping senses [51], adding entirely new perceptual dimensions or changing the body's morphology [64, 63].

Nowadays, augmenting is often conflated with the sole addition and superimposing of information. In a sense, added information or easier access supports and leads to better cognitive abilities and is therefore an augmentation of human abilities. However, removal or control over real and virtual information is equally augmentative. Augmentation can also happen through diminished reality, which removes content from the real world [76] and may make way for increased cognitive performance.

*Hearables* as the third concept are a device class of wireless earbuds providing functionality like voice assistance and fitness tracking [40, 96]. With *hearables*, auditory perception can be mediated by passing the real world sounds through a microphone and signal processor before playing back to the user. It allows the device to add to or alter a percept in any fashion and therefore enables personal soundscape interaction as opposed to global soundscape interaction (e.g., adding soundproof barriers along a highway). This concept is already implemented in some products [21, 33] in a rudimentary way (e.g., frequency based audio-filters).

Mediation, in turn, opens up a large design space for addition [19, 86, 71] and alteration [75, 111]. Additive mediation is known as audio/auditory AR, whereas alteration can be considered modulated reality [75, 78]. This fundamental concept is the prerequisite for any kind of change to users' abilities and perception. Ultimately, users would re-gain full agency and control over their sense of hearing by being able to choose and filter what they hear. This type of ability enhancement was proposed for vision in earlier research [75, 77, 76]. Hearing, being an important but less dominant human sense, was focus of research to a lesser extent. Nevertheless, the sheer usage and widespread adoption of personal audio technology indicates that there is a need to control the sense of hearing. This kind of control may go beyond established means

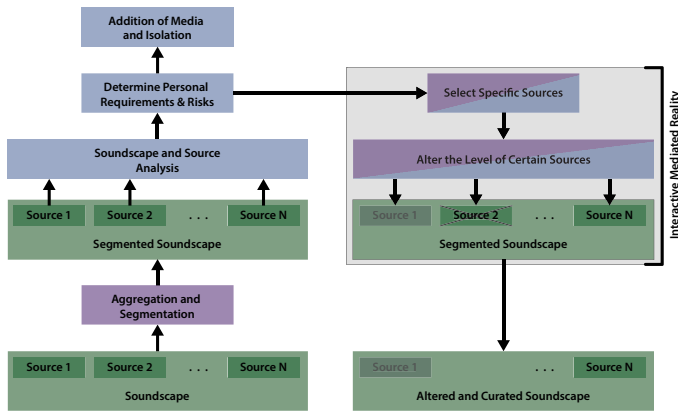


Figure 2: The process of mediating the acoustic environment, including the stages of understanding, analysis, source selection and alteration.

of restoring hearing with hearing aids. Current coping strategies like wearing headphones are flawed and, in part, dangerous: they drown out relevant and irrelevant sounds alike [44]. Risks arising from the use of personal audio devices include social isolation [50, 27], dangers in traffic [67], and acquired hearing loss [28, 61], all originating from the desire to curate what one is able to hear. Ideally, *IAMR* is able to alleviate the risks of personal audio while giving users an entire gradient of isolation which is not uniform, but relies on a more selective and fully interactive mediation. To conclude we define the concept of *IAMR* based on the definition of mediated reality by Steve Mann: *IAMR* refers to a general framework for user-driven artificial modification of human auditory perception via devices for augmenting, diminishing, and generally altering the sensory input of the auditory system.

A conceptual process for *IAMR* with a device that provides (auditory) scene understanding and interactive alteration of hearing is depicted in Figure 2. The system *segments* the soundscape to sound sources, which the users may initially *analyse* and understand further. Afterwards, *risks* and *requirements* can be weighted. With this estimation in mind, users may start to *add media* or *isolate*. These two steps are possible with established personal audio devices, like headphones [44]. If conducted with an *IAMR* system, users have further options for personal human-soundscape interaction: they may start to *selectively alter* specific sound sources, for instance reducing their perceived volume or removing them from their personal soundscape entirely. The result of the *IAMR* process is an *altered and curated soundscape*, which may include any mixture of real-world and virtual sounds.

## SCENARIOS FOR IAMR

Personal audio technology allows users to choose what they hear on a **personal** level [44], but is limited to uniform, almost binary interaction with the soundscape. With devices like hearables, the curation of the auditory environment is made possible to all users without the drawbacks of common personal audio devices. Personal soundscape curation then includes not only adding media and drowning out the world, but covers an entire gradient of level and other dimensions. Just like an aesthetic soundmark<sup>1</sup> [104, 107] (e.g., a fountain) can be added to a physical soundscape,

<sup>1</sup>Equivalent to a landmark for audition

it can also be added to a virtual one on demand. Notably, the addition of a fountain to the physical environment (i.e., altering the soundscape on a **global** level) is not easily possible and requires lengthy processes in city planning. With the rise of various machinery, sound levels rose, displacing the sound of nature [95] and silence itself. However, in biological and physiological terms, humans did not adapt as quickly – their sense of hearing is largely the same as before. They were given a tool to drown out sounds on a **personal** level with devices like headphones. In the following, we present scenarios for *IAMR* systems, where *comfort*, *media playback* and *information acquisition* are weighted differently, depending on the users' requirements and context. We consider established devices like headphones to be a precursor to *IAMR*, which are able to fulfill some, but not all potential user requirements. This is reflected in the presented scenarios, and follows the interaction flow of *IAMR* seen in Figure 2.

### Scenario I: Open Plan Office

In today's work culture, open plan offices are a common occurrence, requiring workers to cope with unwanted sounds [1, 43]. At the same time, communication between colleagues remains a relevant task. If a worker requires *auditory comfort*, she may isolate herself by using active noise cancellation (ANC) headphones and adding calm background noises. Alternatively, she may selectively remove the sound sources that are irrelevant and disturbing. If she desires *media* as an addition to her personal soundscape, she may use headphones, potentially drowning out sounds around her. Lastly, if she desires to retain *information* in her personal soundscape, she may leverage *IAMR* to exclude specific sound sources (e.g., colleagues not on her team) from the noise filter to be able to get information about their state without specifically devoting attention to them.

### Scenario II: Commuting

As with open plan offices, commuting is a relevant part of today's work culture, and is likewise a context where mobile audio devices are common [68]. If a user requires *comfort*, he may similarly isolate himself by using noise cancellation and drown out remaining background sounds with personal media. Alternatively, earplugs may suffice for ensuring comfortable levels of noise. If he desires to add *media* to his personal soundscape, the user may consume videos, music, audio books on the way home with the help of a mobile device and headphones. If the goal is to isolate while retaining the informative content of the auditory environment, the user may add speaker announcements to a list of desired sounds to avoid missing the correct station.

## RELATED WORK

The concept of *IAMR*, which was also discussed in [102, 119, 110], relies on theoretical and practical advancements in research and industry. Apart from ongoing research in academia, the space around augmented and altered hearing is populated by industry players and crowdfunding platforms. Not all products achieve widespread adoption or reach a state of market-readiness. However, many reflect the vision of human-soundscape interaction and *IAMR* or parts thereof.

### Human Perception and Sensation

Users' perception of (spatial) sound is crucial when attempting to create a method that mediates and alters it. Research on sensation

and perception is particularly relevant for filtering methods and helping users understand their soundscape [22]. The field can be separated into research on spatial perception and localization, spatial awareness and segmentation. Additionally, effects of noise and the perception thereof are related to the motivation and development of mediating devices.

Spatial perception and localization are relevant as the interface in world-fixed interactions fully surrounds the user. In 1967, Batteau showed that the auricle of the human ear is relevant for localization of sounds as it introduces delays to the signal propagation [13]. Furthermore, hearing is an omnidirectional sense in humans and therefore able to compensate for missing visual sensations. Sound augmentations like warning tones have been applied in vehicles [114] or general workspaces [29, 42, 4]. On a psychophysical level, humans are already able to segment and cluster the acoustic environments surrounding them [22]. They can also leverage phenomena like the cocktail party effect to be able to selectively perceive parts of the soundscape [9] which can in turn be supported and augmented with devices [116]. Effects like "gestalt laws" for audio or the precedence effect [59] can also be leveraged to support soundscape understanding.

### Interacting with Audio

A minority of previous research incorporated sounds as interaction targets, instead of it being a supporting feedback modality. In this case, the users' goal is to alter sound and they interact *with* sound, instead of *with the help* of sound. Work in this field is related to tasks like music composition, mixing and editing [38, 88, 16]. Adams et al. presented a system that allows exploration and alteration of arbitrary parameters of sound on a touchscreen [2]. In contrast, Mueller et al. presented a system for spatial interaction with sound in the shape of an experimental mixing interface or room [88]. Comparable, mid-air interaction with sound was presented by Alroe et al. [5], Dicke et al. [32] and Gelineck et al. [39]. In contrast, interfaces that allow exploration, search and organization of sound, are not necessarily sound-based or sound-focused [105, 16, 99, 112].

### Personal Audio Technology

Apart from industry-driven advances, there is ongoing research concerning personal audio technology and devices, particularly with regards to sociological and societal aspects. Mamuji et al. presented "attentive headphones" to manage interruptions and mediate communication [74]. A similar approach was used by Danninger et al. in an office scenario [31]. Tyler Bickford observed children's behaviour around headphones and earbuds like sharing a pair of earbuds or using them as speakers, transforming them to non-personal audio technology [17]. Social behaviour around sound and sharing thereof (e.g., music recommendations) was investigated by McGookin et al. [83, 82] and Hakansson et al. [45] on a larger scale. Michael Bull described modern use and implications of personal audio [25, 26], followed by Liikkanen et al. [68], Haas et al. [44] and Nettamo et al. [89] among others. Technical aspects like performance of noise-cancellation [7, 66], signal processing [10] or new form factors [40, 96, 52] are part of the progress in research. While there are hardware platforms available that originate from and are aimed at academic research [56], much work is primarily industry-driven and is described in more detail below.

### Auditory Augmented Reality

Beside the well-established field of visual AR [11, 18], projects focusing on hearing also exist. Core drivers are additive systems attempting to augment and superimpose virtuality over the real soundscape. While the concept envisioned in this work is focused on modulation and alteration instead of addition, the technical side remains the same. Auditory AR is most relevant to the proposed work, as methods and insights from systems can be transferred to the prototypes developed. It can further provide detailed reasoning on directions to prioritize or discard early in the process. Technical fundamentals include the use of hear-through technology [85, 79, 69, 87], bone conduction [69, 81] and the implementation of spatialization [80]. Auditory AR was also a focus in games research, with co-located [60] and location-based [94] games. Application-centred research was conducted in guide [14, 71, 113] and navigation [118] systems.

### Products and Industry

Personal audio technology and its advances are nowadays primarily driven by industry [52]. It encompasses new developments in terms of headphones, hearables and earplugs, along with progress in filtering and noise-cancellation. Some headphone-like devices are already equipped with spatial filtering [20, 92], whereas others attempt to blend real and virtual worlds: for example with the Xperia Ear Duo, users are meant to "Stay in tune with the world" [106]. As with other recent products, voice assistants are a selling point [117].

The market of hearables at its core consists of mostly crowd-funded startups. Doppler Labs' "Here One" comes closest to the vision of interactive hearing, providing presets, equalizers, and blending functionality with their headset [33]. A similar approach is taken by IqBuds [90], with other vendors focusing on fitness tracking [21].

Hearing aids are already an established way to recover and restoratively augment hearing [65, 58]. While they improve auditory perception, it is important to differentiate "audibility and intelligibility" [65], as not all of them are able to restore the ability to perceive speech as before. However, current devices provide situative presets [93], are controllable by the wearer and may even include smart adaptive functions that require no intervention from the user [49]. Some allow the streaming of media to them and have dedicated apps for controls [58, 8]. The intersection between personal audio and hearing aids is the core domain for future developments in *IAMR*.

### SOUNDWALK STUDY

To develop a system that is able to manipulate perceived ambient sounds, we first considered the needs of users. Therefore, we conducted a week-long event sampling study with 12 participants. They were instructed to record and rate sounds that they would like to alter via a mobile application. After this week, semi-structured interviews aimed to further explore the concept of interacting with the surrounding soundscape and possible usage scenarios.

### Methods and Apparatus

To gain insights into user requirements in real-world environments, a mobile application was created which allowed users to record and rate sound sources which they would like to make louder or

quieter. As this ability, at most, is currently only part of niche products, users are currently not actively aware of use cases. This is particularly the case for alterations that can not be executed with common personal audio devices, like fine-grained, selective alterations. The methodology is comparable to crowd-sourced research on noise-pollution [73, 98] and is also inspired by Schafer’s concept of “soundwalks” [104]. However, when surveying relevant literature, a core problem became evident: most soundscape research focuses on the soundscape as a whole, not on specific parts of it.

Choosing this method allows involving users in a design process for a future artifact with the requirements and chances being largely unknown to users and developers. Furthermore, it is likely to yield more natural results as users record data whenever they deem it necessary and do not have to remember specific situations. Users would also survey their surroundings primed towards the *ability* and *chance* to selectively change sounds, something they usually do not have. Additionally, there is no comparable data set as earlier research with similar methodology [3, 57, 34, 103] had a different framing. When attempting to gather data in-situ, a trade-off need to be made between the amount of data to gather and the number of samples received. The general goal was to enforce brief interactions with the app, nudging participants to collect many entries instead of few, detailed entries.

With the mobile application we provided, participants were able to record a brief sample of the sound they wanted to alter and report at which level they want it to be. The “desired level” ranged from 0 (“inaudible”) to 10 (“as loud as possible”). Furthermore, the application provided a user interface to label and rate those sounds on 7-point scales, derived and adapted from [55]: *Naturalness*, *Pleasantness*, *Regularity*, *Proximity* and *Directionality*. We also provided multiple context flags such as *working*, *leisure*, *social*, *solitary*, *public*, *private*, *commuting*, *at home*, *relaxing* and *focusing* that could be selected via checkboxes.

As users could freely decide to enter data, the application was meant to nudge users gently while not being annoying or overly present. A persistent notification was therefore shown to the user. Tapping it immediately opened the activity for data entry.

## Procedure

After the recruiting process, participants were invited for the introductory session consisting of application installation, explaining the target vision of *IAMR*, and giving instructions for the following week-long data acquisition. To verify the applications’ functionality, participants were asked to record a sound with the examiner present, listen to it, and delete it afterwards. The final meeting, approximately a week after the introduction, consisted of multiple parts. Participants were given the chance to listen to and delete sensitive recordings. This was followed by a semi-structured interview, which required it’s own informed consent. The reward given to participants was partially scaled with their “performance”, which essentially was the number of completely entered data points. The base reward was set to 7 *currency*, as it was estimated that the initial and final meetings would take roughly 40 to 50 minutes. For each recorded source, the participant would receive 0.4 *currency*, up to a maximum of 6 *currency* of added reward.

## Results

The participants were aged 27.83 years on average ( $SD = 2.7, min = 22, max = 31$ ), with exactly half identifying as female, the other half identifying as male. 5 were employed, 6 students (either undergraduate, graduate or PhD) and 1 reported being currently out of work. 7 held a master’s degree, 3 a bachelor’s degree and 2 a high school degree. 10 participants reported having had no diagnosed auditory impairments. The 2 remaining ones reported having had a sudden temporary hearing loss and having hearing loss in a specific frequency band which did not affect them in daily life. In total, 225 samples were collected by the participants.

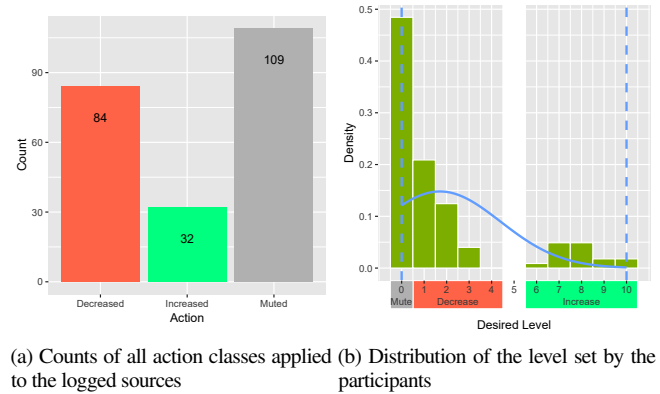


Figure 3: Action classes and the distribution of chosen volumes they are composed of.

For each source or source cluster, users set a volume level they want it to have. The full distribution can be seen in Figure 3, with estimated densities overlaid over a histogram. The histograms bins are equal to the possible level values users could select. A large cluster is found at the mute value. However, a large amount of sources was made quieter, without fully removing it from the soundscape. Additionally, the level below the “unchanged” value was not used at all. Increased values are more sparse and are mostly found in the middle between the maximum level and the unchanged state.

For all data points ( $n = 225$ ), the desired level/volume ranged from 0 to 10, covering the entire value range ( $M = 1.7, SD = 2.7, SE = 0.18$ ) with a median absolute deviation of 1.48. Far more sounds were decreased ( $n = 193$ ) in volume than increased ( $n = 32$ ). However, in terms of contribution to a soundscape, decreased volume is vastly different from completely silent sounds. When splitting the group of decreases in “decreased” and “muted”, a more balanced separation arises. The number of increases is unchanged ( $n = 32$ ), with muting ( $n = 109$ ) exhibiting a slightly higher sample count than decreasing ( $n = 84$ ), as can be seen in Figure 3.

The sounds users labelled were overwhelmingly associated with objects. We therefore applied an established coding and classification scheme to them which was introduced by Schafer. It covers natural sounds, human sounds, sounds and society, sounds as indicators, mechanical sounds, and quiet and silence [104]. Distribution of labeled events can be found in Figure 4.

As mentioned before, users had to rate the source or sources they want to alter on 5 scales: pleasantness, naturalness, proximity,

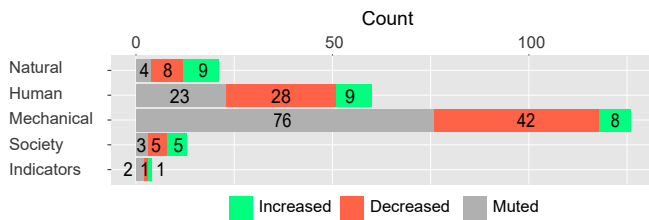


Figure 4: Occurrences of classes and alterations for Schafer's scheme [104] applied to the labeled data.

directionality and regularity. The rating on the unpleasantness-scale ranged from 0 to 6, covering the entire value range ( $M = 4.3, SD = 1.74, SE = 0.12$ ) with a median absolute deviation of 1.48. The ratings on the unnaturalness-scale also covered all possible values ( $M = 3.62, SD = 2.44, SE = 0.16$ ) with a median absolute deviation of 1.48. The distance-ratings also ranged from 0 to 6 ( $M = 3.65, SD = 1.89, SE = 0.13$ ) with a median absolute deviation of 2.97. Directionality ranged from 0 to 6, again covering the entire value range ( $M = 2.12, SD = 2.12, SE = 0.14$ ) with a median absolute deviation of 1.48. The rating on the regularity-scale covered all possible values, too ( $M = 2.67, SD = 2.04, SE = 0.14$ ) with a median absolute deviation of 2.97.

### Interviews

The interviews took 13:19 minutes on average ( $min = 6:56$  min,  $max = 28:31$  min), dependent on the participant's experience and desire to discuss the concept. After selective coding (thematic analysis), specific themes could be identified, which reoccurred for almost every participant. The following paragraphs summarize them.

**Dangers and Risks** The illusion of an intact auditory perception mediated through such devices was a criticism of the concept. This means that changes made to the auditory perception are no longer conscious for the user and therefore be forgotten. Furthermore, missed information, as already the case with personal audio devices, was linked to this aspect. Additionally, adaptation and overreliance emerged as issues.

**Dynamic Requirements** Participants repeatedly mentioned that besides a changing soundscape, their own requirements may change, depending on mood or context. For a future device, it is reasonable to assume that it can accommodate for some of these changes, but not necessarily all.

**Interaction and Interactivity** With the aforementioned aspects of dangers and changing requirements, a degree of interactivity was required by participants. Generic personal audio devices are already interactive, allowing alteration of media playback, but also physical interaction, like taking them off. This would not be the case for *permanently* worn hearables. Being able to interact also covers cases, where automation is insufficient or certain requirements change.

**Agency Over One's Own Perception** Considering that mediating devices may automatically alter and mediate perception itself, users reaffirmed their scepticism against automation. While headphones are an alteration of hearing, this alteration is deliberate and conscious. No other device currently mediates perception in an automated fashion, especially not in a subtractive or destructive fashion.

**Subtlety and Social Aspects** Most participants referred to the alteration of human speech in one way or another. Increased volume for better understanding is considered a noble goal but can be used for malicious eavesdropping. Similarly, muting people, especially without them knowing, is unfavourable on a social level. For some participants, these problems called for very subtle and unobtrusive interactions. For others, this required an enforced, system-side ruleset.

### Limitations and Discussion

Apart from the sole fact that per-source soundscape curation has its uses and potential, additional requirements can be derived from the interviews and data. Whenever a source is altered, the effect has to be clear to the user. This allows them to weigh advantages and risks appropriately. Furthermore, systems may leverage inherent abilities of users, as humans are already able to segment a soundscape into sources. Leveraging selective attention to define filters would make *IAMR* open to users. While *IAMR* may provide more comfort, filtering relevant sounds is a source of danger. Muting has to be treated carefully, especially if auditory comfort can be reached without total removal of sounds.

If given the chance to change the volume of any sound source, users mostly reduce or mute and rarely increase the level. This is a chance for simplification of all future interfaces, as continuous volume control may be reduced to few "states", like muting. Additionally, it can be assumed that there is a level between "muted" and "unaltered", which retains awareness and completeness of the soundscape, but caters to users' desire for silence and acoustical comfort – another chance for simplification. Simplification is ultimately a procedure to reduce interactions needed and therefore reduce the time required to interact and make it applicable for a wider range of devices.

Users identified sound sources predominantly as objects, rarely as textures and never by their acoustical properties. This makes frequency-based interfaces like equalizers unfeasible for everyday use. Therefore, interfaces should rely on understandable labels, as these may already describe the level of granularity required for alterations.

### INTERACTION STUDY

Based on the insights of the first study, a more concise evaluation was designed covering concrete aspects of future *IAMR* systems. A study in which participants actually interact with a system and are able to perceive the effects is likely to yield valuable feedback concerning the concept and vision, independently of the specific implementation. The first study conducted yielded a set of requirements and suggestions by users, as induced with their active examination of their everyday soundscapes. Most suggested screen-based interaction, but also required subtlety for it. Additionally, due to the way the desired volume was distributed and with the requirement for brief interactions, it was hypothesized that the granularity of interaction can be reduced or compressed, without a loss in usability. Based on the interviews and the above findings, the evaluation was set to study two interaction concepts (screen-based and head-tilt-based) and two manipulation granularities (discretized and continuous). The *screen-based* mode represent interactions with touchscreen devices like smartphones or tablets, whereas the *head-tilt-based* mode represented spatial

interaction. Tactile interfaces such as a mixing desk were not considered as they are not applicable for mobile scenarios.

### Application Design

Conducting an evaluation of interactions in VR is valid, as long as these interactions are either meant for VR, or are reproduced faithfully. Additionally simulated auditory scenes are a relevant topic in research, especially concerning their faithful reproduction [15, 37, 100, 109] and usage for analysis [101, 35, 72, 54]. Interaction with a small screen, like a smartphone or even a smart-watch, is challenging to reproduce faithfully in VR. It requires the users' fingers to have a dexterity close to their real abilities. Furthermore, small text, as used on such devices, requires high resolutions to remain legible. Generally speaking, each interaction tested in VR suffers from some degree of indirection. This applies to (touch-)screen-based interaction in particular, but also influences gestural interfaces to some degree. However, an evaluation in VR has distinct benefits other methods can not provide such as full control over the world, which was important for this investigation.

### Apparatus

To execute the study in VR, an appropriate environment was developed and designed. A top view of the scenery with sound sources is shown in Figure 5. It was meant to be compelling and visually consistent. Therefore, a pre-made asset was used. To avoid overloading the user with a large amount of sounds, yet encourage users to interact with the system, all sounds were introduced following a scripted timeline (Figure 5). To ensure proper spatialization and room effects, the Resonance Audio library was used. It supports spatialization by interaural time and level differences, head-related transfer functions (HRTFs), frequency-dependent effects, reverberation and occlusion. To ensure believable reflections, the reverberation baking functionality was used, which mapped visual materials to acoustic materials and properties.

### Variables and Study Design

The potential design space around *IAMR* covers a large set of dimensions, ranging from device use over interaction modality to treatment of specific sound source types. The first variable, **Concept / Paradigm** was chosen to cover two extremes in terms of modality: **screen-based** (close to visual-only) compared to **head-tilt-based** (close to audio-only). The second variable is **Granularity / Complexity**, which again covers two extremes in terms of manipulation. A **simplified** granularity provides minimal discretized steps or states for alteration. This subset of volume consists of a muted state, a quieter state, an unchanged state and a louder state, as derived from the first study. In contrast to this reduction, a method of **continuous** alteration allows for a potentially infinite amount of states.

#### Screen-based

Screen-based modes are meant to represent interaction via a surrogate device, like a smartphone or a tablet. For each sound source that is present in the scene, an element is added to a tablet-like interface the user is holding with its hand. As all elements are listed, interaction happens on an ordered aggregate of sources (Figure 6). Whenever a new source is added to the soundscape, it is placed above all others in the list, along with a visual highlight (blinking). The continuous mode (Figure 6, d), consists of 3 interactive modules per source. A slider covers the volume range from

0 (mute) over 0.5 (unchanged) to 1 (maximal), with the numeric value hidden from the user, as done by most other interfaces for sound. Additionally, 2 shortcuts to specific states are present, as suggested by users in the first study: reset the volume to its original state and mute the source. The on-screen discrete or simplified mode (Figure 6, e) operates on state. Each source offers 4 buttons, representing a muted, a quieter, an unchanged and a louder state.

#### Head-Based

The two modes for head-based interaction are implemented as audio-only systems. While this is not necessarily optimal, this depicts a fixed extreme point in the possible design space from which other prototypes may be derived. The design itself was derived from earlier prototypes mapping audio from or to sight [46, 47, 115] and interaction techniques relying on head roll or yaw [30, 53, 108, 91]. Sources are targeted with head rotation and a sphere-cast along the users' head orientation. Alteration is mapped to the head's tilt or roll, with a knob-like metaphor. This virtual knob then either has a continuous mapping, or has an ordered set of states covering the heads tilt range. Figure 6 a-c depicts the angle ranges and the functions mapped to them. The discrete head-based mode functioned similarly, but discretized input to a set of four states, as can be seen in Figure 6 b. The edit mode is reached in a similar fashion as the screen-based variant. To ensure comparability, both modes were enriched with a set of supporting functionality. For the head-based modes, the "labels" of targeted sources were read out to the user via text to speech (TTS). Similarly, the discrete states of each mode (e.g., "mute") were also read out to the user.

### Procedure

The study took place in a separate room of our institution. Participants received 8 *currency* and a bar of chocolate as compensation with the entire study taking roughly 60 minutes.

For each of the 4 study iterations, the participant received a brief explanation of the mode they are about to use. The 4 iterations were balanced in a Latin Square. The examiner relied on 4 cards, depicting the mode, with the help of which he showed the core part of the mode. Additionally, any VR-specifics were explained and demonstrated such as the activation gestures and which buttons are to be pressed for them. Participants wore Sony WH-1000XM2 ANC headphones. The VR-headset used was an Oculus Rift CV1. Users first entered a training scene, where 4 sound sources were placed in a neutral environment. The 4 sources were visually represented by grey cubes. 3 of the sources emitted irregular tones played on a piano, while the remaining one constantly played a 440 Hz sine wave. Whenever participants felt comfortable, the examiner loaded the real environment. The participants then received 5 minutes to configure the soundscape to their liking, while discovering the method. Additionally, they were encouraged to think aloud.

Each iteration was followed by a questionnaire, containing the NASA TLX [48], system usability scale (SUS) [24] and selected single-item questions on 5-point Likert scales. Additionally, users were provided with 3 comment fields to express likes, dislikes and general comments concerning the method they just used. After finishing all 4 trials, users received a final questionnaire with rankings, general questions and demographic data.



(a) Source placement (arrows indicate whether the source is slightly above or below the users' head height or above). User position is indicated through the axis gizmo.

Name	Delay	
1	Reading Person	150
2	Microwave	225
3	Cooking noises	50
4	Dripping faucet	25
5	Refrigerator	150
6	Buzzing lamp	0
7	Fan	0
8	Wind Chimes	75
9	Thunder	100
10	Traffic	50
11	Rain	0
12	Snoring dog	125
13	Howling wind	125
14	Keyboard	100
15	Clicking Mouse	100
16	Bathroom noises	175
17	Footsteps	200
18	Construction site	250

(b) Sound sources and their respective appearance delays.

Figure 5: Sound sources as they were placed and appeared for each participant

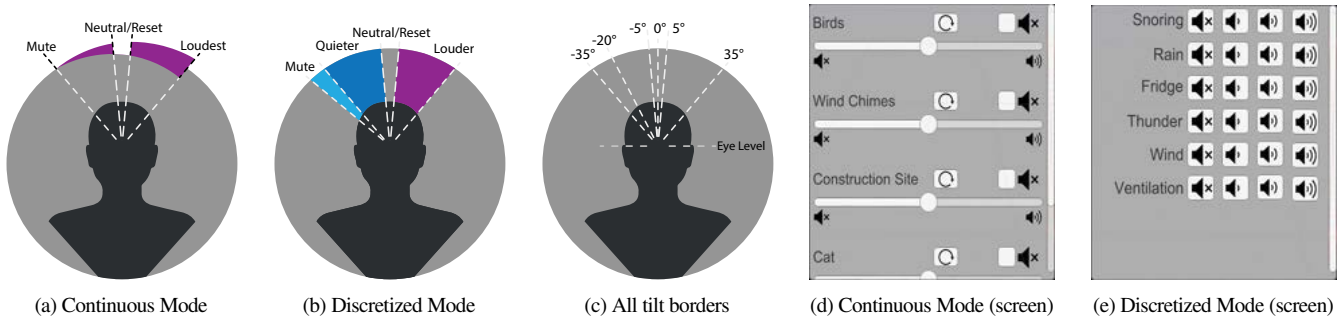


Figure 6: (a)-(c): Angle ranges for the systems based on head-tilt, as mapped for the user's view direction representing spatial interaction. Derived and adapted from [30] and [70]. (d)-(e): Virtual tablet interfaces representing aggregate interaction.

### Sample

The 12 participants were aged 28.58 years on average ( $SD = 2.55, min = 25, max = 32$ ), with 5 identifying as female, the remaining 7 identified as male. 10 participants reported having no diagnosed auditory impairments. One participant reported having had a temporary sudden hearing loss and one reported having hearing deficiency in a specific frequency band which does not constrain him. Their self-assessed, estimated average usage time of personal audio devices per week was 24.33 hours on average ( $SD = 17.3, min = 2, max = 56$ ).

### Results

The following sections describe and evaluate the quantitative and qualitative results gathered in the study.

#### Load and Usability

To assess the users' mental load and usability we used two standardized questionnaires. The NASA TLX [48] was meant to assess task load, while the system usability scale (SUS) [24] was meant to assess the usability of the concept. As mentioned before, users were asked and instructed to abstract from the specific implementation and from constraints imposed by the VR environment.

Core takeaways from the SUS scales are that granularity has no great influence on the perceived usability of the concepts presented to the users. While the median value was the same (both  $Mdn = 71$ ), the continuous granularity exhibits a larger variance in values, especially towards the lower end of the scale. This may be interpreted as an indicator that a metaphor of a continuous slider is overly complex for the use-case of sound alteration in most everyday scenarios. When comparing the SUS scores by paradigm, screen-based modes yield better results, with the interquartile range remaining above the marked score of 68. This can be ascribed to multiple aspects: aggregating sources (on a screen) makes them easy to interact with and it simplifies the task of finding them. The localisation task, especially when relying on audition only, adds another step of indirection, while being able to spatially understand and segment a soundscape becomes less relevant.

Results of the TLX are in line with the answers provided to the SUS questionnaire. High effort needed for the head-tilt modes can be primarily traced back to the difficulties users had localizing and selecting sources. This is likewise reflected in the performance item. Physical demand was less differentiated between the modes. Furthermore, the simplified head tilt was slightly less mentally demanding. This may be traced back to the fact that dealing



with an invisible state was easier in this mode: a selection of a source did not necessarily immediately alter the sound. For all other items, comparing granularity yields minimal differences. This indicates that reducing interaction complexity by reducing manipulation complexity is a viable path for practical systems.

As the SUS questionnaire yields a comparable score and each subscale is actually related to the system's operation, it was used for further quantitative analysis. When comparing each combination, differences are visible, primarily between modes and not between granularities. When assessing all gathered SUS scores, the mean score is found at 70.05 while the scores cover the range from 55 to 85 ( $SD = 7.21, SE = 1.04$ ). The Shapiro-Wilk normality test indicates that the data does not significantly deviate from a normal distribution, with  $W = 0.97$  and  $p = 0.26$ . Ratings gathered from the "Head Tilt with continuous change" mode had a mean SUS score of 65.63, ranging from 55.0 to 82.5 ( $SD = 8.19, SE = 2.37$ ). For "Head Tilt with discrete change", ratings had a mean SUS score of 67.71 and ranged from 57.5 to 82.5 ( $SD = 6.44, SE = 1.86$ ). "On-Screen with continuous change" had a mean SUS score of 72.08, which ranged from 57.5 to 80 ( $SD = 5.92, SE = 1.71$ ). Similarly, "On-Screen with discrete change" had a mean score of 74.79, ranging from 67.5 to 85 ( $SD = 4.70, SE = 1.36$ ).

To determine whether the differences in mean and range are significant, a Friedman rank sum test was conducted. The Friedman rank sum test (Friedman's ANOVA) determined that there were significant differences between the scores of the conditions, with  $\chi^2(3) = 15.7$  and  $p = 0.001306 < 0.005$ . As a follow up, multiple comparisons were executed, which indicated significant differences in the SUS score between the *On-Screen with discrete change* and the *Head Tilt with continuous change* condition and between the *On-Screen with discrete change* and the *Head Tilt with discrete change* condition. Additionally, pairwise comparisons using the Nemenyi test with  $q$  approximation were executed [97], yielding similar results with  $p = 0.0016 < 0.005$  and  $p = 0.0142 < 0.05$  respectively.

#### *Preferences by Context*

Four contexts were chosen to cover most application scenarios in an abstract fashion. A private context reduces social interaction and conveys a sense of security and safety, whereas a public context may have social implications and requirements. A static context exhibits no or limited changes in the soundscape and does not involve movement of the user. A dynamic context, in contrast, may involve movement and change of the environmental soundscape and the users' themselves. Having experienced all concepts and granularities, users were asked to rank the desired usage for each of these contexts, with 2 ranked items being mandatory.

A clear pattern over all 4 rankings is the preference of screen-based interfaces. In a private context, continuous change is preferred over discrete (5 and 7 mentions respectively). This is reversed for the public context, where the simplified version has nearly twice as many votes (8 and 3 respectively). A similar pattern emerges when comparing the static and dynamic contexts, where the static context has 6 mentions of the continuous change, compared to 4 of discrete change. The dynamic context has 8 mentions for the discrete change and only 4 for the continuous mode. This can be traced back to the desired brevity and speed of the interaction. With a simplified user interface, a desired state can be reached

quickly, albeit not as precisely as with a fine grained control. This aspect of speed and precision explains the differences found in the rankings. While a fine-grained alteration is necessary to retain, a simplified may suffice in many contexts, where neither the time nor the desire to configure sounds precisely is present.

#### *Source Ordering and Priority*

Users were additionally asked which criteria they would employ to order and prioritise sources on aggregated interfaces (i.e., screen-based lists). The criterion most commonly ranked at the first position, was volume (6x). This is in line with the dimension for acoustic comfort but also indicator for safety and urgency [6]. It was followed by distance (3x), an indicator for the feasibility of spatial filtering, as it is again an indicator for relevance and potential danger. The second rank is lead by annoyance (4x), a subjective criterion relevant for acoustic comfort and followed by relevance (3x). The third and last mandatory rank is lead by relevance (4x), followed by distance and pleasantness (3x each). The remaining criteria, including recency, as implemented for the screen interfaces in the study, were mentioned sparsely across all ranks. Notably, these preferences may be transferable to spatial user interfaces where occluding and intersecting sources require some sort of arbitration when they are being targeted and selected. Likewise, these features may serve as additional input for automated or partially automated systems if consistent rules can be derived and modelled.

#### *Interaction Concept and Granularity*

Most importantly, 100% of the participants agreed with the statement that 4 states are sufficient for most use-cases. This again confirms the potential of reduced manipulation granularity for optimisation for faster and dynamic contexts and environments. Additionally, 67% of the users tended towards agreement when offered that a fine-grained alteration is still possible, with 33% remaining undecided.

Furthermore, users predominantly agreed with the notion that they would configure their entire surrounding (67%) while disagreeing with the idea that they would alter only few, select sources (67%). The latter is in line with the process of soundscape interaction via personal audio devices, like headphones. The concept itself was met with acceptance, with 83% disagreeing with the statement that they are not interested in changing the sounds around them.

#### **Limitations**

The study design focused on two specific dimensions future systems can have and did not incorporate nuances or mixtures of these extremes. Users were additionally put in an environment they were not familiar with. This impacted their ability to locate sounds, as contextual information was not available to them. In a real home-scenario, occurrences and mappings of sounds would likely be known to the users, simplifying their process of locating and understanding. Despite these limitations, it was possible to receive valuable insights on a conceptual level, invaluable for future iterations.

#### **Discussion**

The core result of the evaluation is that a reduction of alteration complexity is not only accepted by users, but also preferred and considered more appropriate for dynamic scenarios. This insight is likely applicable independently of modality or device, as it is

based on the interaction between human and soundscape and less in the interaction between human and device.

Another takeaway is the aspect that an audio-only system, despite its theoretical advantages, is hard to develop and requires thorough testing and iterations. Despite the lower performance of both head-tilt prototypes, the idea of a natural and rather subtle mapping was accepted and understood by the participants. It is also reasonable to assume that visual AR can serve as supporting or core modality, if it is desired to retain the spatial aspect of interaction with a soundscape.

If a spatial reference is needed or desired, metaphors aggregating sources along with spatial information may be a reasonable direction for future evaluations.

## **IMPLICATIONS FOR DESIGN**

The two studies yielded a large set of insights for future systems that build upon the concept of *IAMR*. Currently, personal soundscape curation is made possible via personal audio technology. Yet, their functionality is limited and the way they let users interact with the soundscape is uniform and almost binary. Having gained more selective control over sounds opens up possibilities of human-soundscape interaction on a personal level. Based on both conducted studies and the quantitative and qualitative insights gained, we present a set of implications for such systems.

***Reclaiming and Retaining Agency*** The concept of *IAMR* is meant to support users to *re-gain* some agency over their sense of hearing, without suffering from the disadvantages of established personal audio technology. An ever-increasing amount of noise in urban soundscapes can be regarded as a coercion to hear. Automating this process may be conceptually a similar coercion of the user to hear or miss specific noises. Likewise, automated selective alterations to the users' acoustic environment may convey an illusion of an intact soundscape, whereas manual alterations may not. Interactivity, either based on manual mixing or the definition or selection of presets, is therefore an appropriate path to agency.

***Soundscape Understanding Precedes Soundscape Alteration*** Any system aiming to allow users to alter their auditory perception has likewise to support the users in their task of *understanding* their acoustic environment. This is particularly relevant for dynamic environments or environments unknown to the users. Apart from filtering sounds alone, systems should provide labelling and other functions to the users to aid them in understanding the current soundscape and the consequences of their alterations. This improves the judgment of risks, but also the expression of the users' requirements towards the soundscape.

***Spatial Interaction vs. Aggregated Interaction*** Conceptually, interacting with sound sources in space may help users to segment and understand the soundscape, leveraging spatial and selective hearing. However, this introduces some degree of reliance on visual information. In the case of occluded sources or sources without a precise location, spatial interaction may either be inefficient or introduce additional issues. Interacting on an aggregated set of representations for sources therefore is applicable but may require the use of visual interfaces, like smartphones or AR-headsets.

***Appropriate Interaction Fidelity*** Real-world sound does not have to be treated equal to media sounds. Treating a soundscape as a sound mixing system used in the industry is rarely appropriate and deters users from soundscape interaction and alteration. A

large set of use-cases for interactive hearing can be covered by a greatly reduced set of states for volume: "quiet-but-audible" for aesthetic requirements and information acquisition alike, lightly increased volume for better understanding and muted volume for the removal of irrelevant sources from the soundscape. Long configuration is not desirable for dynamic environments. This reduced interaction fidelity is likely to suffice, while allowing for more brief and unobtrusive interactions.

## ***Meaningful Sound Source Granularities and Decompositions***

It is also important to consider the granularity of sound sources to be interacted with. One can consider "traffic" to be a sound source. Alternatively, one can decompose it into its parts like cars, pedestrians. Likewise, it is possible to group sparsely occurring sounds into ambience as an interaction target, like the voices of different bystanders into "speech" as an aggregate. These aggregations or decompositions can be defined by semantics, spatial relation to the user or the relevance to the user.

***Harmless and Harmful Use*** By providing heightened or even superhuman senses, issues of acceptance and privacy arise. Distinguishing sounds and their sources through selective playback in situations where it would not be possible otherwise is problematic. Furthermore, it is not intended that the system allows for eavesdropping on people beyond users' natural ability for selective hearing. As it is the case with many intrusive technologies, acceptance must be discussed and weighted thoroughly.

## **CONCLUSION AND FUTURE WORK**

Given appropriate tools, users may regain agency over their auditory perception. The intersection and interface between user and soundscape is currently non-interactive. For interactive mediated hearing, requirements are different from current soundscape design approaches.

The first study was able to uncover real-world requirements for augmented hearing. It yielded a large set of sources users wanted to alter, with various mechanical sounds irrelevant to users. Nearly all were referenced by object and not by high-level clustering or physical properties. Similarly, sources were made louder, to increase aesthetic qualities or to acquire additional information. Core insight however was that users reduce volume instead of muting sources to account for both acoustical comfort and information acquisition. The second study showed that a reduced manipulation granularity had no significant impact on usability and acceptance. In general, the development of specific systems and the evaluation of their effects in real-world scenarios is a goal to progress *IAMR* systems.

This work focused on one dimension of sound: volume. While it is the most graspable one to users, spatial position, frequency and various other factors are relevant for future research. Additionally, more specific interactions and visualizations can be implemented and evaluated with the proposed guidelines. Audio-only, visual-only and mixed-modality systems can be compared on a more refined level, based on the results gathered in this work.

## **ACKNOWLEDGEMENTS**

This work has been supported by the DFG project "Empirical Assessment of Presence and Immersion in Augmented and Virtual Realities" (RU 1605/4-1).

## REFERENCES

- [1] Volkan Acun and Semiha Yilmazer. 2018. A Grounded Theory Approach to Investigate the Perceived Soundscape of Open-Plan Offices. *Applied Acoustics* 131 (Feb. 2018), 28–37. DOI: <http://dx.doi.org/10.1016/j.apacoust.2017.09.018>
- [2] Alexander Travis Adams, Berto Gonzalez, and Celine Latulipe. 2014. SonicExplorer: Fluid Exploration of Audio Parameters. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 237–246. DOI: <http://dx.doi.org/10.1145/2556288.2557206>
- [3] Luca Maria Aiello, Rossano Schifanella, Daniele Quercia, and Francesco Aletta. 2016. Chatty Maps: Constructing Sound Maps of Urban Areas from Social Media Data. *Open Science* 3, 3 (March 2016), 150690. DOI: <http://dx.doi.org/10.1098/rsos.150690>
- [4] Petter Alexanderson. 2004. Peripheral Awareness and Smooth Notification: The Use of Natural Sounds in Process Control Work. In *Proceedings of the Third Nordic Conference on Human-Computer Interaction (NordCHI '04)*. ACM, New York, NY, USA, 281–284. DOI: <http://dx.doi.org/10.1145/1028014.1028057>
- [5] Tobias Alrøe, Jonas Grann, Erik Grönvall, Marianne Graves Petersen, and Jesper L. Rasmussen. 2012. Aerial Tunes: Exploring Interaction Qualities of Mid-Air Displays. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design (NordCHI '12)*. ACM, New York, NY, USA, 514–523. DOI: <http://dx.doi.org/10.1145/2399016.2399095>
- [6] Tjeerd C Andringa. 2013. Core Affect and Soundscape Assessment: Fore- and Background Soundscape Design for Quality of Life. *InterNoise 2013* (Sept. 2013), 11.
- [7] Linus Yinn Leng Ang, Yong Khiang Koh, and Heow Pueh Lee. 2017. The Performance of Active Noise-Canceling Headphones in Different Noise Environments. *Applied Acoustics* 122 (July 2017), 16–22. DOI: <http://dx.doi.org/10.1016/j.apacoust.2017.02.005>
- [8] Apple Inc. 2018. Use Live Listen with Made for iPhone Hearing Aids. (2018). <https://support.apple.com/en-us/HT203990> Accessed: 09.08.2019.
- [9] Barry Arons. 1992. A Review of The Cocktail Party Effect. *Journal of the American Voice I/O Society* (1992), 16.
- [10] Joshua Atkins and Daniele Giacobello. 2014. Trends and Perspectives for Signal Processing in Consumer Audio. *arXiv:1405.4843 [cs]* (May 2014). <http://arxiv.org/abs/1405.4843>
- [11] Ronald T. Azuma. 1997. A Survey of Augmented Reality. *Presence: Teleoperators & Virtual Environments* 6, 4 (1997), 355–385.
- [12] Mathias Basner, Wolfgang Babisch, Adrian Davis, Mark Brink, Charlotte Clark, Sabine Janssen, and Stephen Stansfeld. 2014. Auditory and Non-Auditory Effects of Noise on Health. *The Lancet* 383, 9925 (2014), 1325–1332. <http://www.sciencedirect.com/science/article/pii/S014067361361613X>
- [13] Dwight W. Batteau. 1967. The Role of the Pinna in Human Localization. *Proceedings of the Royal Society of London B: Biological Sciences* 168, 1011 (1967), 158–180. <http://rspb.royalsocietypublishing.org/content/168/1011/158.short>
- [14] Benjamin B. Bederson. 1995. Audio Augmented Reality: A Prototype Automated Tour Guide. In *Conference Companion on Human Factors in Computing Systems (CHI '95)*. ACM, New York, NY, USA, 210–211. DOI: <http://dx.doi.org/10.1145/223355.223526>
- [15] Wolfgang Beinhauer and Cornelia Hipp. 2009. Using Acoustic Landscapes for the Evaluation of Multimodal Mobile Applications. In *Proceedings of the 13th International Conference on Human-Computer Interaction. Part II: Novel Interaction Methods and Techniques*. Springer-Verlag, Berlin, Heidelberg, 3–11. DOI: [http://dx.doi.org/10.1007/978-3-642-02577-8\\_1](http://dx.doi.org/10.1007/978-3-642-02577-8_1)
- [16] Janne Bergman, Jarmo Kauko, and Jaakko Keränen. 2009. Hands on Music: Physical Approach to Interaction with Digital Music. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '09)*. ACM, New York, NY, USA, 33:1–33:11. DOI: <http://dx.doi.org/10.1145/1613858.1613901>
- [17] Tyler Bickford, Jason Stanyek, and Sumanth Gopinath. 2014. Earbuds Are Good for Sharing: Children’s Headphones as Social Media at a Vermont School. In *The Oxford Handbook of Mobile Music Studies*. The Oxford Handbook of Mobile Music Studies, Vol. 1. Oxford University Press, Oxford. [dx.doi.org/10.1093/oxfordhb/9780195375725.013.015](http://dx.doi.org/10.1093/oxfordhb/9780195375725.013.015)
- [18] Mark Billinghurst, Adrian Clark, and Gun Lee. 2015. A Survey of Augmented Reality. *Foundations and Trends® in Human-Computer Interaction* 8, 2-3 (March 2015), 73–272. DOI: <http://dx.doi.org/10.1561/11000000049>
- [19] Andrew Blakney and Sudhir Mudur. 2011. Sounds in Space: 3D Audio Experiences through Tangible Navigation. In *HCI International 2011 – Posters’ Extended Abstracts (Communications in Computer and Information Science)*. Springer, Berlin, Heidelberg, 217–221. DOI: [http://dx.doi.org/10.1007/978-3-642-22095-1\\_45](http://dx.doi.org/10.1007/978-3-642-22095-1_45)
- [20] Bose Corporation. 2018. Bose Hearphones | Conversation-Enhancing Headphones. (2018). [https://www.bose.com/en\\_us/products/headphones/conversation\\_enhancing\\_headphones/earphones.html](https://www.bose.com/en_us/products/headphones/conversation_enhancing_headphones/earphones.html) Accessed: 09.08.2019.
- [21] BRAGI GmbH. 2018. Wireless Earphones - The Dash Pro. (2018). <https://web.archive.org/web/20190331120847/https://bragi.com/products/thedashpro> Accessed: 09.08.2019.

- [22] Albert S. Bregman. 1994. *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press.
- [23] Lauren M. Britton and Bryan Semaan. 2017. Manifesting the Cyborg Through Techno-Body Modification: From Human-Computer Interaction to Integration. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 2499–2510. DOI: <http://dx.doi.org/10.1145/3025453.3025629>
- [24] John Brooke. 1996. SUS-A Quick and Dirty Usability Scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.
- [25] Michael Bull. 2001. The World According to Sound: Investigating the World of Walkman Users. *New Media & Society* 3, 2 (June 2001), 179–197. DOI: <http://dx.doi.org/10.1177/14614440122226047>
- [26] Michael Bull. 2008. Sounding Out The City: Personal Stereos and The Management of Everyday Life; Music in Everyday Life. In *Sounding Out the City: Personal Stereos and the Management of Everyday Life (Materializing Culture)*. Berg 3PL.
- [27] Todd A. Campbell. 2018. iPads, iPods and Technology-Enabled Isolation: If We're So Connected, Why Are We so Alone? *GSTF Journal of Music (JMusic)* 1, 1 (May 2018). <http://dl16.globalstf.org/index.php/JMusic/article/view/1375>
- [28] Peter J. Catalano and Stephen M. Levin. 1985. Noise-Induced Hearing Loss and Portable Radios with Headphones. *International Journal of Pediatric Otorhinolaryngology* 9, 1 (June 1985), 59–67. DOI: [http://dx.doi.org/10.1016/S0165-5876\(85\)80004-5](http://dx.doi.org/10.1016/S0165-5876(85)80004-5)
- [29] Marina Cidota, Stephan Lukosch, Dragos Datcu, and Heide Lukosch. 2016. Workspace Awareness in Collaborative AR Using HMDs: A User Study Comparing Audio and Visual Notifications. In *Proceedings of the 7th Augmented Human International Conference 2016 (AH '16)*. ACM, New York, NY, USA, 3:1–3:8. DOI: <http://dx.doi.org/10.1145/2875194.2875204>
- [30] Andrew Crossan, Mark McGill, Stephen Brewster, and Roderick Murray-Smith. 2009. Head Tilting for Interaction in Mobile Contexts. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '09)*. ACM, New York, NY, USA, 6:1–6:10. DOI: <http://dx.doi.org/10.1145/1613858.1613866>
- [31] Maria Danninger, Roel Vertegaal, Daniel P. Siewiorek, and Aadil Mamuji. 2005. Using Social Geometry to Manage Interruptions and Co-Worker Attention in Office Environments. In *Proceedings of Graphics Interface 2005 (GI '05)*. Canadian Human-Computer Communications Society, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, 211–218. <http://dl.acm.org/citation.cfm?id=1089508.1089543>
- [32] Christina Dicke and Jörg Müller. 2015. Evaluating Mid-Air List Interaction for Spatial Audio Interfaces. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction (SUI '15)*. ACM, New York, NY, USA, 24–33. DOI: <http://dx.doi.org/10.1145/2788940.2788945>
- [33] Doppler Labs Inc. 2018. Here One - Wireless Smart Earbuds. (2018). <https://hereplus.me/products/here-one/> Accessed: 09.08.2019.
- [34] Danièle Dubois, Catherine Guastavino, and Manon Raimbault. 2006. A Cognitive Approach to Urban Soundscapes: Using Verbal Data to Access Everyday Life Auditory Categories. *Acta Acustica united with Acustica* 92, 6 (Nov. 2006), 865–874.
- [35] Gemma Maria Echevarria Sanchez, Timothy Van Renterghem, Kang Sun, Bert De Coensel, and Dick Botteldooren. 2017. Using Virtual Reality for Assessing the Role of Noise in the Audio-Visual Design of an Urban Public Space. *Landscape and Urban Planning* 167 (Nov. 2017), 98–107. DOI: <http://dx.doi.org/10.1016/j.landurbplan.2017.05.018>
- [36] Douglas C. Engelbart and William K. English. 1968. A Research Center for Augmenting Human Intellect. In *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I (AFIPS '68 (Fall, Part I))*. ACM, New York, NY, USA, 395–410. DOI: <http://dx.doi.org/10.1145/1476589.1476645>
- [37] Nathaniel Finney and Jordi Janer. 2010. Soundscape Generation for Virtual Environments Using Community-Provided Audio Databases. In *W3C Workshop: Augmented Reality on the Web*. Barcelona.
- [38] Jérémie Garcia, Xavier Favory, and Jean Bresson. 2016. Trajectoires: A Mobile Application for Controlling Sound Spatialization. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 3671–3674. DOI: <http://dx.doi.org/10.1145/2851581.2890246>
- [39] Steven Gelineck and Dan Overholt. 2015. Haptic and Visual Feedback in 3D Audio Mixing Interfaces. In *Proceedings of the Audio Mostly 2015 on Interaction With Sound (AM '15)*. ACM, New York, NY, USA, 14:1–14:6. DOI: <http://dx.doi.org/10.1145/2814895.2814918>
- [40] Valentin Goverdovsky, Wilhelm von Rosenberg, Takashi Nakamura, David Looney, David J. Sharp, Christos Papavassiliou, Mary J. Morrell, and Danilo P. Mandic. 2017. Hearables: Multimodal Physiological in-Ear Sensing. *Scientific reports* 7, 1 (2017), 6948.
- [41] Raphael Grasset, Laurence Boissieux, Jean D. Gascuel, and Dieter Schmalstieg. 2005. Interactive Mediated Reality. In *Proceedings of the Sixth Australasian Conference on User Interface - Volume 40 (AUIC '05)*. Australian Computer Society, Inc., Darlinghurst, Australia, Australia, 21–29. <http://dl.acm.org/citation.cfm?id=1082243.1082247>

- [42] Carl Gutwin, Oliver Schneider, Robert Xiao, and Stephen Brewster. 2011. Chalk Sounds: The Effects of Dynamic Synthesized Audio on Workspace Awareness in Distributed Groupware. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work (CSCW '11)*. ACM, New York, NY, USA, 85–94. DOI : <http://dx.doi.org/10.1145/1958824.1958838>
- [43] A. Haapakangas, E. Kankkunen, V. Hongisto, P. Virjonen, D. Oliva, and E. Keskinen. 2011. Effects of Five Speech Masking Sounds on Performance and Acoustic Satisfaction. Implications for Open-Plan Offices. *Acta Acustica united with Acustica* 97, 4 (July 2011), 641–655. DOI : <http://dx.doi.org/10.3813/AAA.918444>
- [44] Gabriel Haas, Evgeny Stemasov, and Enrico Rukzio. 2018. Can't You Hear Me?: Investigating Personal Soundscape Curation. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia (MUM 2018)*. ACM, New York, NY, USA, 59–69. DOI : <http://dx.doi.org/10.1145/3282894.3282897>
- [45] Maria Hakansson, Mattias Rost, Mattias Jacobsson, and Lars Erik Holmquist. 2007. Facilitating Mobile Music Sharing and Social Interaction with Push!Music. In *Proceedings of the 40th Annual Hawaii International Conference on System Sciences (HICSS '07)*. IEEE Computer Society, Washington, DC, USA, 87–. DOI : <http://dx.doi.org/10.1109/HICSS.2007.225>
- [46] Masatoshi Hamanaka and SuengHee Lee. 2009. Sound Scope Headphones. In *SIGGRAPH 2009: Talks (SIGGRAPH '09)*. ACM, New York, NY, USA, 43:1–43:1. DOI : <http://dx.doi.org/10.1145/1597990.1598033>
- [47] Masatoshi Hamanaka and Seunghee Lee. 2013. Concert Viewing Headphones. In *SIGGRAPH Asia 2013 Emerging Technologies (SA '13)*. ACM, New York, NY, USA, 4:1–4:3. DOI : <http://dx.doi.org/10.1145/2542284.2542288>
- [48] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*. Vol. 52. Elsevier, 139–183.
- [49] Chris Heddon. 2018. Why Artificial Intelligence Is Coming to a Hearing Aid (and Audiology Practice) Near You. (June 2018). <https://hearinghealthmatters.org/blog/2018/artificial-intelligence-audiology-hearing-aids/> Accessed: 09.08.2019.
- [50] Virginia Heffernan. 2011. The Argument Against Headphones. *The New York Times* (Jan. 2011). <https://www.nytimes.com/2011/01/09/magazine/09FOB-medium-t.html> Accessed: 09.08.2019.
- [51] Jochen Huber, Roy Shilkrot, Pattie Maes, and Suranga Nanayakkara. 2017. *Assistive Augmentation*. Springer Berlin Heidelberg, New York, NY.
- [52] Nick Hunn. 2016. The Market for Hearable Devices 2016-2020. (2016), 36.
- [53] Thibaut Jacob, Gilles Bailly, Eric Lecolinet, Géry Casiez, and Marc Teyssier. 2016. Desktop Orbital Camera Motions Using Rotational Head Movements. In *Proceedings of the 2016 Symposium on Spatial User Interaction (SUI '16)*. ACM, New York, NY, USA, 139–148. DOI : <http://dx.doi.org/10.1145/2983310.2985758>
- [54] Like Jiang, Massimiliano Masullo, Luigi Maffei, Fanyu Meng, and Michael Vorländer. 2018. A Demonstrator Tool of Web-Based Virtual Reality for Participatory Evaluation of Urban Sound Environment. *Landscape and Urban Planning* 170 (Feb. 2018), 276–282. DOI : <http://dx.doi.org/10.1016/j.landurbplan.2017.09.007>
- [55] Jian Kang and Mei Zhang. 2010. Semantic Differential Analysis of the Soundscape in Urban Open Public Spaces. *Building and environment* 45, 1 (2010), 150–157. <http://www.sciencedirect.com/science/article/pii/S0360132309001309>
- [56] Fahim Kawsar, Chulhong Min, Akhil Mathur, Alessandro Montanari, Utku Günay Acer, and Marc Van den Broeck. 2018. eSense: Open Earable Platform for Human Sensing. In *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems (SenSys '18)*. ACM, New York, NY, USA, 371–372. DOI : <http://dx.doi.org/10.1145/3274783.3275188>
- [57] Pablo Kogan, Bruno Turra, Jorge P. Arenas, and María Hinalaf. 2017. A Comprehensive Methodology for the Multidimensional and Synchronic Data Collecting in Soundscape. *Science of The Total Environment* 580 (Feb. 2017), 1068–1077. DOI : <http://dx.doi.org/10.1016/j.scitotenv.2016.12.061>
- [58] Birger Kollmeier and Jürgen Kiessling. 2016. Functionality of Hearing Aids: State-of-the-Art and Future Model-Based Solutions. *International Journal of Audiology* 57, sup3 (May 2016), S3–S28. DOI : <http://dx.doi.org/10.1080/14992027.2016.1256504>
- [59] David M. Krum, Evan A. Suma, and Mark Bolas. 2012. Spatial Misregistration of Virtual Human Audio: Implications of the Precedence Effect. In *Proceedings of the 12th International Conference on Intelligent Virtual Agents (IVA '12)*. Springer-Verlag, Berlin, Heidelberg, 139–145. DOI : [http://dx.doi.org/10.1007/978-3-642-33197-8\\_14](http://dx.doi.org/10.1007/978-3-642-33197-8_14)
- [60] Jennifer Lade and Jonathan Duckworth. 2013. Audio Arc: An Audio-Spatial Game Using Mobile Phone Ringtones. In *Proceedings of the 9th ACM Conference on Creativity & Cognition (C&C '13)*. ACM, New York, NY, USA, 401–402. DOI : <http://dx.doi.org/10.1145/2466627.2481228>
- [61] Phillip C. Lee, Craig W. Senders, Bruce J. Gantz, and Steven R. Otto. 2016. Transient Sensorineural Hearing Loss after Overuse of Portable Headphone Cassette Radios. *Otolaryngology—Head and Neck Surgery* (May 2016). DOI : <http://dx.doi.org/10.1177/019459988509300510>

- [62] Youngho Lee, Jongmyung Choi, and Young-Jae Ryoo. 2014. Socially Wise Mediated Reality for Hollistic Smart Environments. *IEEE*, 431–434. DOI: <http://dx.doi.org/10.1109/SCIS-ISIS.2014.7044885>
- [63] S. Leigh, H. Agrawal, and P. Maes. 2018. Robotic Symbionts: Interweaving Human and Machine Actions. *IEEE Pervasive Computing* 17, 2 (April 2018), 34–43. DOI: <http://dx.doi.org/10.1109/MPRV.2018.022511241>
- [64] Sang-won Leigh, Harpreet Sareen, Hsin-Liu (Cindy) Kao, Xin Liu, and Pattie Maes. 2017. Body-Borne Computers as Extensions of Self. *Computers* 6, 1 (March 2017), 12. DOI: <http://dx.doi.org/10.3390/computers6010012>
- [65] Nicholas A. Lesica. 2018. Hearing Aids: Limitations and Opportunities. *The Hearing Journal* 71, 5 (May 2018), 43. DOI: <http://dx.doi.org/10.1097/01.HJ.0000533807.87095.2a>
- [66] Maojin Liang, Fei Zhao, David French, and Yiqing Zheng. 2012. Characteristics of Noise-Canceling Headphones to Reduce the Hearing Hazard for MP3 Users. *The Journal of the Acoustical Society of America* 131, 6 (June 2012), 4526–4534. DOI: <http://dx.doi.org/10.1121/1.4707457>
- [67] Richard Lichenstein, Daniel Clarence Smith, Jordan Lynne Ambrose, and Laurel Anne Moody. 2012. Headphone Use and Pedestrian Injury and Death in the United States: 2004–2011. *Injury prevention* (2012), injuryprev–2011.
- [68] Lassi A. Liikkanen and Mikko Lahdensuo. 2010. Observing the Mobile Music Phenomenon: One in Nine Commuters Is Wired. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems (DIS '10)*. ACM, New York, NY, USA, 348–351. DOI: <http://dx.doi.org/10.1145/1858171.1858234>
- [69] Robert W. Lindeman, Haruo Noma, and Paulo Goncalves de Barros. 2007. Hear-Through and Mic-Through Augmented Reality: Using Bone Conduction to Display Spatialized Audio. In *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '07)*. IEEE Computer Society, Washington, DC, USA, 1–4. DOI: <http://dx.doi.org/10.1109/ISMAR.2007.4538843>
- [70] Edmund LoPresti, David M. Brienza, Jennifer Angelo, Lars Gilbertson, and Jonathan Sakai. 2000. Neck Range of Motion and Use of Computer Head Controls. In *Proceedings of the Fourth International ACM Conference on Assistive Technologies (Assets '00)*. ACM, New York, NY, USA, 121–128. DOI: <http://dx.doi.org/10.1145/354324.354352>
- [71] Kent Lyons, Maribeth Gandy, and Thad Starner. 2000. Guided by Voices: An Audio Augmented Reality System. In *International Conference on Auditory Display*. Georgia Institute of Technology. <https://aragorn.library.gatech.edu/handle/1853/50672>
- [72] Luigi Maffei, Massimiliano Masullo, Aniello Pascale, Gennaro Ruggiero, and Virginia Puyana Romero. 2016. Immersive Virtual Reality in Community Planning: Acoustic and Visual Congruence of Simulated vs Real World. *Sustainable Cities and Society* 27 (Nov. 2016), 338–345. DOI: <http://dx.doi.org/10.1016/j.scs.2016.06.022>
- [73] Nicolas Maisonneuve, Matthias Stevens, Maria E. Niessen, and Luc Steels. 2009. NoiseTube: Measuring and Mapping Noise Pollution with Mobile Phones. In *Information Technologies in Environmental Engineering*. Springer, Berlin, Heidelberg, 215–228. DOI: [http://dx.doi.org/10.1007/978-3-540-88351-7\\_16](http://dx.doi.org/10.1007/978-3-540-88351-7_16)
- [74] Aadil Mamuji, Roel Vertegaal, Changuk Sohn, and Daniel Cheng. 2005. Attentive Headphones: Augmenting Conversational Attention with a Real World TiVo. In *Extended Abstracts of CHI*, Vol. 5. ACM, Portland, Oregon, USA, 2223–.
- [75] Steve Mann. 1999. Mediated Reality. *Linux Journal* 1999, 59es (March 1999), 5. <http://dl.acm.org/citation.cfm?id=327697.327702>
- [76] S. Mann. 2001. Wearable Computing: Toward Humanistic Intelligence. *IEEE Intelligent Systems* 16, 3 (May 2001), 10–15. DOI: <http://dx.doi.org/10.1109/5254.940020>
- [77] Steve Mann and Woodrow Barfield. 2003. Introduction to Mediated Reality. *International Journal of Human-Computer Interaction* 15, 2 (April 2003), 205–208. DOI: [http://dx.doi.org/10.1207/S15327590IJHC1502\\_1](http://dx.doi.org/10.1207/S15327590IJHC1502_1)
- [78] Steve Mann, Tom Furness, Yu Yuan, Jay Iorio, and Zixin Wang. 2018. All Reality: Virtual, Augmented, Mixed (X), Mediated (X,Y), and Multimeditated Reality. *arXiv:1804.08386 [cs]* (April 2018). <http://arxiv.org/abs/1804.08386>
- [79] Georgios Marentakis and Rudolfs Liepins. 2014. Evaluation of Hear-through Sound Localization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 267–270. DOI: <http://dx.doi.org/10.1145/2556288.2557168>
- [80] Aengus Martin, Craig Jin, and André van Schaik. 2009. Psychoacoustic Evaluation of Systems for Delivering Spatialized Augmented-Reality Audio. *Journal of the Audio Engineering Society* 57, 12 (Dec. 2009), 1016–1027. <http://www.aes.org/e-lib/browse.cfm?elib=15234>
- [81] Keenan R. May and Bruce N. Walker. 2017. The Effects of Distractor Sounds Presented through Bone Conduction Headphones on the Localization of Critical Environmental Sounds. *Applied ergonomics* 61 (2017), 144–158.
- [82] David McGookin and Stephen Brewster. 2011. PULSE: An Auditory Display to Provide a Social Vibe. In *Proceedings of Interacting with Sound Workshop: Exploring Context-Aware, Local and Social Audio Applications (IwS '11)*. ACM, New York, NY, USA, 12–15. DOI: <http://dx.doi.org/10.1145/2019335.2019338>
- [83] David McGookin and Stephen Brewster. 2012. PULSE: The Design and Evaluation of an Auditory Display to Provide a Social Vibe. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 1263–1272. DOI: <http://dx.doi.org/10.1145/2207676.2208580>

- [84] J. L. Mey and B. Gorayska. 1995. *Cognitive Technology: In Search of a Humane Interface*. Elsevier.
- [85] Anders Kalsgaard Møller, Pablo Faundez Hoffmann, Flemming Christensen, and Dorte Hammershøi. 2014. Evaluation of a Hear-Through Device. In *HCI International 2014 - Posters' Extended Abstracts (Communications in Computer and Information Science)*. Springer, Cham, 440–445. DOI: [http://dx.doi.org/10.1007/978-3-319-07857-1\\_77](http://dx.doi.org/10.1007/978-3-319-07857-1_77)
- [86] Nikos Moustakas, Andreas Floros, and Nicolas Grigoriou. 2011. Interactive Audio Realities: An Augmented / Mixed Reality Audio Game Prototype. In *Audio Engineering Society Convention 130*. <http://www.aes.org/e-lib/browse.cfm?elib=15826>
- [87] Florian Mueller and Matthew Karau. 2002. Transparent Hearing. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems (CHI EA '02)*. ACM, New York, NY, USA, 730–731. DOI: <http://dx.doi.org/10.1145/506443.506569>
- [88] Jörg Müller, Matthias Geier, Christina Dicke, and Sascha Spors. 2014. The boomRoom: Mid-Air Direct Interaction with Virtual Sound Sources. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 247–256. DOI: <http://dx.doi.org/10.1145/2556288.2557000>
- [89] Esa Nettamo, Mikko Nirhamo, and Jonna Häkkinen. 2006. A Cross-Cultural Study of Mobile Music: Retrieval, Management and Consumption. In *Proceedings of the 18th Australia Conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments (OZCHI '06)*. ACM, New York, NY, USA, 87–94. DOI: <http://dx.doi.org/10.1145/1228175.1228193>
- [90] Nuheara Limited. 2018. Nuheara: Intelligent Truly Wireless Earbuds. (2018). <https://www.nuheara.com/> Accessed: 09.02.2019.
- [91] Tomi Nukarinen, Jari Kangas, Oleg Špakov, Poika Isokoski, Deepak Akkil, Jussi Rantala, and Roope Raisamo. 2016. Evaluation of HeadTurn: An Interaction Technique Using the Gaze and Head Turns. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*. ACM, New York, NY, USA, 43:1–43:8. DOI: <http://dx.doi.org/10.1145/2971485.2971490>
- [92] Orosound. 2018. Tilde Earphones - Experience Noise-Filtering Technology. (2018). <https://www.orosound.com/tilde-earphones/> Accessed: 09.08.2018.
- [93] Oticon A/S. 2016. Oticon Hearing Aids | Rediscover the Sounds of Your Life. (2016). <https://www.oticon.co.uk/> Accessed: 09.08.2019.
- [94] Mikko Peltola, Tapio Lokki, and Lauri Savioja. 2009. Augmented Reality Audio for Location-Based Games. In *Audio Engineering Society Conference: 35th International Conference: Audio for Games*. Audio Engineering Society. <http://www.aes.org/e-lib/browse.cfm?elib=15176>
- [95] Bryan C. Pijanowski, Almo Farina, Stuart H. Gage, Sarah L. Dumyahn, and Bernie L. Krause. 2011. What Is Soundscape Ecology? An Introduction and Overview of an Emerging New Science. *Landscape Ecology* 26, 9 (Nov. 2011), 1213–1232. DOI: <http://dx.doi.org/10.1007/s10980-011-9600-8>
- [96] Joseph Plazak and Marta Kersten-Oertel. 2018. A Survey on the Affordances of “Hearables”. *Inventions* 3, 3 (July 2018), 48. DOI: <http://dx.doi.org/10.3390/inventions3030048>
- [97] Thorsten Pohlert. 2016. The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR). (Jan. 2016), 27.
- [98] Rajib Rana, Chun Tung Chou, Nirupama Bulusu, Salil Kanhere, and Wen Hu. 2013. Ear-Phone: A Context-Aware Noise Mapping Using Smart Phones. *arXiv:1310.4270 [cs]* (Oct. 2013). <http://arxiv.org/abs/1310.4270>
- [99] Abhishek Ranjan, Ravin Balakrishnan, and Mark Chignell. 2006. Searching in Audio: The Utility of Transcripts, Dichotic Presentation, and Time-Compression. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 721–730. DOI: <http://dx.doi.org/10.1145/1124772.1124879>
- [100] Mathias Rossignol, Gregoire Lafay, Mathieu Lagrange, and Nicolas Misdariis. 2015. SimScene : A Web-Based Acoustic Scenes Simulator. In *1st Web Audio Conference (WAC)*. IRCAM & Mozilla, Paris, France. <https://hal.archives-ouvertes.fr/hal-01078098>
- [101] Francesco Ruotolo, Luigi Maffei, Maria Di Gabriele, Tina Iachini, Massimiliano Masullo, Gennaro Ruggiero, and Vincenzo Paolo Senese. 2013. Immersive Virtual Reality and Environmental Noise Assessment: An Innovative Audio-Visual Approach. *Environmental Impact Assessment Review* 41 (July 2013), 10–20. DOI: <http://dx.doi.org/10.1016/j.eiar.2013.01.007>
- [102] Spencer Russell, Gershon Dublon, and Joseph A. Paradiso. 2016. HearThere: Networked Sensory Prosthetics Through Auditory Augmented Reality. In *Proceedings of the 7th Augmented Human International Conference 2016 (AH '16)*. ACM, New York, NY, USA, 20:1–20:8. DOI: <http://dx.doi.org/10.1145/2875194.2875247>
- [103] Justin Salamon, Christopher Jacoby, and Juan Pablo Bello. 2014. A Dataset and Taxonomy for Urban Sound Research. In *Proceedings of the 22Nd ACM International Conference on Multimedia (MM '14)*. ACM, New York, NY, USA, 1041–1044. DOI: <http://dx.doi.org/10.1145/2647868.2655045>
- [104] R. Murray Schafer. 1993. *The Soundscape: Our Sonic Environment and the Tuning of the World*. Simon and Schuster.

- [105] Chris Schmandt. 1998. Audio Hallway: A Virtual Acoustic Environment for Browsing. In *Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology (UIST '98)*. ACM, New York, NY, USA, 163–170. DOI: <http://dx.doi.org/10.1145/288392.288597>
- [106] Sony Mobile Communications Inc. 2018. Wireless Stereo Headset—Xperia® Ear Duo. (2018). <https://www.sonymobile.com/us/products/smart-products/xperia-ear-duo/> Accessed: 09.08.2019.
- [107] Michael Southworth. 1969. The Sonic Environment of Cities. *Environment and Behavior* 1, 1 (June 1969), 49–70. DOI: <http://dx.doi.org/10.1177/001391656900100104>
- [108] Oleg Špakov, Poika Isokoski, Jari Kangas, Jussi Rantala, Deepak Akkil, and Roope Raisamo. 2016. Comparison of Three Implementations of HeadTurn: A Multimodal Interaction Technique with Gaze and Head Turns. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI 2016)*. ACM, New York, NY, USA, 289–296. DOI: <http://dx.doi.org/10.1145/2993148.2993153>
- [109] G. Christopher Stecker, Steven Carter, Travis M. Moore, and Monica L. Folkerts. 2018. Validating Auditory Spatial Awareness with Virtual Reality and Vice-Versa. *The Journal of the Acoustical Society of America* 143, 3 (April 2018), 1828. DOI: <http://dx.doi.org/10.1121/1.5035998>
- [110] Evgeny Stemasov, Gabriel Haas, Michael Rietzler, and Enrico Rukzio. 2018. Augmenting Human Hearing Through Interactive Auditory Mediated Reality. In *The 31st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings (UIST '18 Adjunct)*. ACM, New York, NY, USA, 69–71. DOI: <http://dx.doi.org/10.1145/3266037.3266104>
- [111] Yuichiro Takeuchi. 2010. Weightless Walls and the Future Office. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 619–628. DOI: <http://dx.doi.org/10.1145/1753326.1753419>
- [112] Gabriel Urbain, Christian Frisson, Alexis Moinet, and Thierry Dutoit. 2016. A Semantic and Content-Based Search User Interface for Browsing Large Collections of Foley Sounds. In *Proceedings of the Audio Mostly 2016 (AM '16)*. ACM, New York, NY, USA, 272–277. DOI: <http://dx.doi.org/10.1145/2986416.2986436>
- [113] Yolanda Vazquez-Alvarez, Matthew P. Aylett, Stephen A. Brewster, Rocio von Jungendorf, and Antti Virolainen. 2014. Multilevel Auditory Displays for Mobile Eyes-Free Location-Based Interaction. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14)*. ACM, New York, NY, USA, 1567–1572. DOI: <http://dx.doi.org/10.1145/2559206.2581254>
- [114] MinJuan Wang, Sus Lundgren Lyckvi, Chenhui Chen, Pelle Dahlstedt, and Fang Chen. 2017. Using Advisory 3D Sound Cues to Improve Drivers' Performance and Situation Awareness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 2814–2825. DOI: <http://dx.doi.org/10.1145/3025453.3025634>
- [115] Junji Watanabe, Yuki Hashimoto, Hideaki Nii, and Masahiko Inami. 2006a. Visual Resonator: Sight-Based Auditory Experience. In *Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology (ACE '06)*. ACM, New York, NY, USA. DOI: <http://dx.doi.org/10.1145/1178823.1178868>
- [116] Junji Watanabe, Hideaki Nii, Yuki Hashimoto, and Masahiko Inami. 2006b. Visual Resonator: Interface for Interactive Cocktail Party Phenomenon. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems (CHI EA '06)*. ACM, New York, NY, USA, 1505–1510. DOI: <http://dx.doi.org/10.1145/1125451.1125727>
- [117] Chris Welch. 2018. Sony's Futuristic Earbuds Are a Pain in the Present. (June 2018). <https://www.theverge.com/2018/6/29/17511618/sony-xperia-ear-duo-wireless-earbuds-review> Accessed: 09.08.2019.
- [118] Jeff Wilson, Bruce N. Walker, Jeffrey Lindsay, Craig Cambias, and Frank Dellaert. 2007. SWAN: System for Wearable Audio Navigation. In *Proceedings of the 2007 11th IEEE International Symposium on Wearable Computers (ISWC '07)*. IEEE Computer Society, Washington, DC, USA, 1–8. DOI: <http://dx.doi.org/10.1109/ISWC.2007.4373786>
- [119] Selman Yüçetürk, Mohammad Obaid, and Astm Evren Yantaç. 2016. Probing Human-Soundscape Interaction Using Observational User Experience Methods. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*. ACM, New York, NY, USA, 33:1–33:4. DOI: <http://dx.doi.org/10.1145/2971485.2971495>
- [120] Xu Zhang, Meihui Ba, Jian Kang, and Qi Meng. 2018. Effect of Soundscape Dimensions on Acoustic Comfort in Urban Open Public Spaces. *Applied Acoustics* 133 (April 2018), 73–81. DOI: <http://dx.doi.org/10.1016/j.apacoust.2017.11.024>