Can’t You Hear Me? Investigating Personal Soundscape Curation

Gabriel Haas, Evgeny Stemasov, Enrico Rukzio
Institute of Media Informatics, Ulm University
Ulm, Germany
<*firstname*>.<*lastname*>@uni-ulm.de

ABSTRACT
Continuous advances in personal audio technology (e.g. headphones), led to efficient noise cancellation and allowed users to build and influence their personal acoustic environment. Despite the high adoption and ubiquitous character of the technology, we do not fully understand which particular factors influence and form usage patterns. As a step towards understanding the usage of personal audio technology, we conducted two focus groups ($n = 10$) to investigate current headphone usage and users’ wishes regarding current and future personal audio technology. Based on this data, we derive a model for what we call *personal soundscape curation*. This model was assessed with the data of a crowdsourced survey on Amazon Mechanical Turk ($n = 194$) on state of the art practices. *Personal soundscape curation* allows to describe usage strategies (curation, adaptation, renunciation) and break down influencing factors of context and environment as well as illustrate which consequences may arise from the users’ behavior.

CCS Concepts
•**Human-centered computing** → *Human computer interaction (HCI);*

Author Keywords
Personal Audio; Soundscapes; Usage Behavior; Headphones;

INTRODUCTION
The number of people who have access to the equipment necessary for private consumption of audio content is constantly increasing. Among other factors, this is due to always available smart devices and digital content. Smartphones and the included hands-free headsets alone contribute with more than 1.5 billion sales in 2017, according to Gartner Inc. [15].

The number of types and form factors of headphones are constantly increasing to adapt to changing requirements: fully untethered and wireless devices, microphone-based hear-through [12], bone conduction headphones and even intelligent functionalities, like translation are included [4]. Active noise-cancelling filters undesired noises, all driven by miniaturization and other technological advances leading to increasing adoption and sales, despite proven risks for user’s health and safety [22, 26]. Apart from being a tool for pure media consumption, headphones serve as a status symbol, emphasizing design and lifestyle [34, 35].

All of this leads to the conclusion that headphones are a genuinely ubiquitous technology. Developments such as noise-cancelling and completely wireless earphones show the enormous potential that still exists in the process of designing headphones. Personal audio technology could develop beyond traditional headphones and include novel and more technological advanced concepts such as Hearables which are able to change the auditory perception of users by applying selective filtering. Smart features could enrich the experience of headphones with sensors that are able to acquire the user context. Furthermore, problems regarding the acquisition of information can be reduced or eliminated. This includes direct interpersonal communication, undirected (i.e. broadcasted) communication and information that is intrinsically embedded within the soundscape. Those devices offer far more opportunities than audio playback. Lower- or higher-pitched voices could be modulated or increased in clarity, eventually providing augmented abilities [23]. Acquiring information in unfavourable, loud environments can be seen as such a relevant and desirable augmentation of hearing. These concepts can lead to the vision of a mixed reality as introduced by Milgram and Kishino [31], using only the sense of hearing.

Using a qualitative research approach, we aimed for an understanding of the ways in which people are using personal audio technology right now in order to influence and manage their own acoustic environment. By analyzing our focus groups, we developed a model that explains usage strategies such as the *curation* of personal soundscapes, *adaptation* of usage behavior or the outright *abandonment* of personal audio technology. We identified the influencing factors within contextual situations such as informativeness of environment and comfort of soundscape. There are also *intervening conditions* such as social norms, device characteristics and the need for information that can have an impact on usage behavior. Furthermore, we illustrate which *consequences* arise from the users’ cho-

MUM ’18, November 25–28, 2018, Cairo, Egypt

ISBN 978-1-4503-6594-9/18/11...$15.00

DOI: https://doi.org/10.1145/3282894.3282897

sen behavior. To back these findings with data from a larger population, we conducted a crowdsourced survey and discuss the results in respect to the set up model. In conclusion, we recapitulate important aspects regarding future devices which emerged from the focus groups. We believe that this work can help to create further improved personal audio technology that meets the dynamic and diverse requirements of modern-day users.

To summarize, this paper has the following contributions:

1. Focus group interviews which provide insights into usage behavior and wishes for personal audio technology.
2. A model that introduces and describes the phenomenon of *personal soundscape curation* with influencing factors and following consequences.
3. An online survey with a larger user population to support the proposed model.

RELATED WORK

Studies of Personal Audio Technology

Michael Bull, founder of the academic discipline of "sound studies", investigated the usage of portable media consumption devices extensively [5, 6, 7]. He described how users manage their cognition, interpersonal behavior and social space with the help of a Walkman or other portable music players. He stated that users play music with great volume to "drown out the industrialized sounds of the city" [5], and that "in a world filled with noise, rather than craving for silence, users demand their own noise" [5]. However, he mostly focused on personal music listening as a medium of prioritizing one's own requirements and preferences to disengage from forms of interaction.

Tyler Bickford investigated the use of earbuds in the context of a school. He states that personal audio technology (e.g. MP3 players) is used by children for co-located, social and shared media consumption [2]. Furthermore, he found that the children put less emphasis on audio quality metrics, but rather on social commitment by performing *"earbud-sharing"* (and therefore losing one audio channel) and room-volume-playback through earbuds [2]. Reinelt et al. [35] studied the perceived design aesthetics of headphones. They influence the enjoyment of use and therefore makes headphones an, at least partly, hedonic technology.

Soundscape Manipulation and Curation

The acoustic environment that surrounds us, is topic of ongoing, extensive research. Scopes differ from character and perception of city soundscapes [37] over automated classification of acoustic environments [28] to smartphone notifications [14]. Lichenstein et al. [26] showed that the absence of sound signals increases risks when walking in heavy city traffic, which implies an increased risk for headphone users. McFarlane investigated cognitive limitations of people in the context of task interruptions. He suggested to give users control over their acoustic environment to improve task performance, which could be achieved with noise-cancelling headphones [30].

Implications of such noise-cancelling technology in headphones was studied by Hagood [18]. He describes them as "soundscaping devices" that enable the "production of personal space" [18]. They allow for a shift of attention that reconfigures the relation of individuals to their surroundings by putting the focus on a virtual soundscape. However, according to Mamuji et al. [29], these headphones impede social interaction, as they lack intelligent filtering. As stated in their work, eye contact can serve as a call for attention and is used to mediate the process of interruption and therefore maintaining proxemic social protocols (e.g. the initiation of communication without physical contact) [19]. They built a prototype of attentive headphones that leverages eyetracking to allow users to participate in interpersonal communication. Not only social interaction is constrained by headphones, but also is information acquisition. To compensate for this, an assistive framework for headphone users in traffic situations was presented by Yoon et al. [41].

Human capabilities can not only be influenced by specific types of headphones, but also by the content that is rendered on them. The influence of nature-based masking sounds on performance and mood was investigated by Newbold et al. [32]. These soundscapes give users a "sense of control over the environment and self" as they stimulate cognition and help to drown out environmental sounds. These masking sounds are of particular interest in environments with many speakers, like in open-plan offices [40]. Effects on performance differ by type of sound and was shown to be optimal for nature sounds such as rippling water [17, 21]. Each of these papers cover a specific aspect of soundscape manipulation but we aimed to explore it as a whole.

Auditory Augmented Reality

Among traffic noise and many other sources of information, digital notification sounds enrich the informativeness of acoustic environments [16, 8]. Those computer generated sounds, which act as carriers of information, were investigated by Sara Bly [3]. She found that sounds can be a useful medium to convey digital information to a human user. Several attempts have already been made to add digital content to our listening experience and therefore augment our reality. Rijswijk and Strijbos presented a mobile application called *Walk With Me* that adds specific sounds based on GPS location "completing the soundscape composition" of a city [39] and create a musical experience. Additional information can also be embedded into the real soundscape. This was presented by Butz and Jung [8] and Chernyshov et al. [9] among others. Russel et al. [36] presented a novel wearable device that makes use of head tracking and bone conduction speakers to achieve a "seamless mixing of real and virtual sounds." [36]. While they all contribute to the field of auditory augmented reality none incorporated the subtraction of sounds.

FOCUS GROUPS

To gain deeper insights into the impact of personal audio technology on information acquisition, interpersonal relations and communication, we conducted two focus groups, consisting of 5 people each, to ensure that each participant has the opportunity to express his or her views. In larger groups, it is easy for

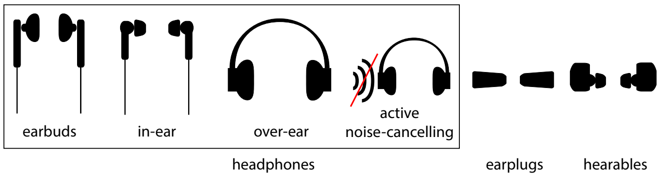

Figure 1. A categorization of personal audio technology that we introduce to form a common understanding and avoid miscommunication.

the more vocal participants to dominate the discussion [25]. We chose focus groups over interviews to identify topics that we as authors did not consider and to benefit from mutual inspiration.

Preliminary Considerations

To create a consistent understanding of personal audio devices, we grouped them into 6 categories as illustrated in Figure 1. This classification is used throughout the paper. Headphones with two earcups and a connecting headband that rest on the head are called *over-ears*. Notably, the class of On-ear headphones exists. These often exhibit a lower degree of occlusion than *over-ears*, but are largely similar in form factor to them. As such, they are treated interchangeably in the following. The headphones that sit on the outer ear are called *earbuds* as long as they do not make use of a silicon tip which would make them *in-ears*. It is important to separate them, because *in-ears* create a much tighter sealing and block out external noise more effectively. Blocking out external noise is also very effective with *active noise-cancelling*. Those headphones were invented to enhance comfort while travelling by cancelling out external waves of sound [18]. This technology can be incorporated into many types of headphones and heavily influences how ambient noise is perceived. Therefore, we made it a category of its own. A completely passive device that is used for personal soundscape curation are *earplugs* which are also treated as an own category. The last category included are so-called *hearables*. This emerging category utilizes advanced audio technology and sensors to create truly wireless headphones that borrow features from hearing aids and offer physiological sensing [4, 12, 42].

Procedure

The focus groups took place in a meeting room of our institution. We prepared a script containing the discussion topics that are requirements, usage behavior, perception, expectations, interaction and form factor. Participants were shown a video consisting of selected scenes from a futuristic vision in which a user wearing hearables manipulates his own soundscape by substituting sound sources such as traffic or construction work with more pleasing sounds, like seagulls and ocean waves [20]. This served as an illustration of which problems are present in the context of headphone usage at the moment and to get the discussion going in the first place.

After the discussion round, the participants were administered a questionnaire to acquire demographic data and data about the usage of personal audio devices. Both sessions had a duration of about 70 minutes and were recorded and transcribed by the authors.

Demographics

The total sample consisted of 10 researchers, 2 of them were female. Their age ranged from 23 to 31 years. 7 have an HCI background, the other three work in the fields of computer graphics, visualization and automotive. All reported a high general interest in technology. All 10 use headphones and, on average, spend 14.5 hours per week wearing them (σ = 7.24), irregardless of type. Altogether, 5 of the participants use *earplugs*, 6 use *active noise-cancelling* headphones and none owns or uses *hearables*. 4 participants reported that they wear headphones without media playback, too (on average 0.82 hours per week, σ = 1.54).

Analysis

We chose a qualitative, exploratory and open-ended approach to analyze the transcribed recordings, aiming to create a concise understanding of the insufficiently documented phenomena of *personal soundscape curation* that people are performing frequently when deciding if and how they use their personal audio technology.

Initial Coding

The authors conducted an initial round of open coding to retrieve a set of concise quotations and wording used for certain issues and aspects (*in-vivo* codes). In total, 270 quotations were extracted from the transcripts. Those were spread out and organized to find clusters of similar statements that were then converted into codes and subcodes.

This initial codebook was then applied to a transcript of a single focus group. Two of the authors independently coded the discussions, marking each passage with an appropriate code. Missing concepts and suggested additions to the codebook were noted for discussion.

Subsequently, the codebook was rewritten during a joint discussion. Several codes were slightly changed in scope and additional subcodes were introduced. With the adapted codebook, the authors independently attempted to apply a coding to the second transcript. Again, suggested additions and changes to the codebook were discussed and incorporated in the final codebook if necessary.

For the final stage of coding, the two transcripts were broken down to a fixed line length of 80 characters. Furthermore, timestamps were removed entirely, while indicators of changing speakers were reduced to a single unique character. The presenter, who guided the focus groups and intervened occasionally to keep the discussion on track, was not removed but coded with an own code, to retain the discussion's context. The two restructured transcripts served as a foundation for *line-by-line-coding* with the developed codebook which was carried out by both coders in two sessions, each lasting approximately 5 hours. In case of conflicting codings, the authors resolved the issue through a structured debate. Any occurrences of off-topic discussion were marked with an own code, which, at 942 lines resulted in 176 lines of 'unrelated' content.

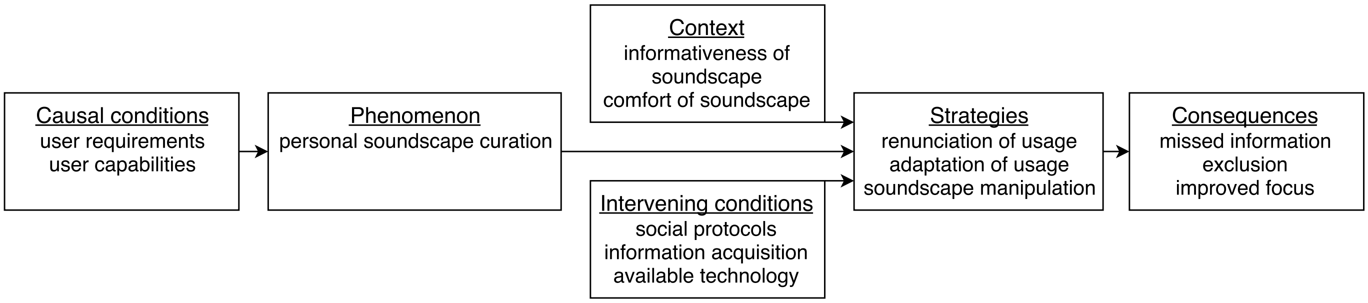

Figure 2. The model we propose, derived from our focus group analysis. Causal conditions trigger the phenomenon of soundscape curation, strategies thereof are influenced by the situational context and possible intervening conditions. Last component are consequences resulting from actions of strategies.

Axial Coding and Model Building

The next stage of qualitative data analysis consisted of axial coding, to uncover and solidify relations between the codes of the final codebook, namely: usage behaviour, information acquisition, social protocols, soundscape manipulation, human capabilities, form factors and future systems. For this purpose, we used the procedure described by Corbin et al. [10]. The focus group discussions yielded a large number of specific system requirements, considering their priming on *hearables* and comparable technology. However, the participants reported many experiences and issues concerning current device usage.

Following the procedure, a core phenomenon was chosen for axial coding: *personal soundscape curation*. The concept encompasses all types of manipulation to the perceived acoustical environment.

Results

The derived model as depicted in figure 2 is centered around the core *phenomenon* of *personal soundscape curation*. Curation of soundscapes is primarily triggered by the *causal conditions*, which consist of a user's personal abilities and requirements. The *strategies* employed, include renunciation of usage, its adaptation and manipulation. They are influenced by the user's *context*, which includes traits of the physical, unaltered soundscape (e.g. the perceived comfort and potentially accessible information). The use and execution of the *strategies* is further shaped by *intervening conditions*. These contain social aspects, availability of devices, and therefore technology, as well as potential dependencies on information. The application of chosen *strategies* has intended and unintended *consequences* for the user (e.g. improved focus or social exclusion).

Causal Conditions

In our model, personal requirements and capabilities serve as the causal condition for soundscape curation. This indicates that many users feel a need to alter their acoustic environment which was repeatedly mentioned in our focus groups. As P4 stated:

> "There are also situations in which, for goodness sake, I do not want to hear conversations. And not only quieter, but not at all."

The omni-directional nature of sound does not allow people to steer their focus in the way it is possible for the visual sense. Accordingly, users try to block out all distracting noises to be more focused on a specific task. Unfortunately, blocking out all sounds, as made possible by *active noise-cancelling*, feels unfamiliar and unnatural at first. An example of this effect in combination with personal audio technology was stated by P1:

> "When you put on noise-cancelling headphones for the first time, it is sometimes a weird impression. [...] Because suddenly you don't hear anything. But [usually] you always hear a thing. And hearing nothing is weird, too. So I can imagine that sometimes it's better [...], to make it quieter in general but not to fade out completely."

Most people end up with music, spoken content or natural soundscapes rendered on headphones to mask and drown out sounds of their environment.

Phenomenon

The chosen phenomenon of *personal soundscape curation* covers all forms of manipulation that users may apply to their very own perception of sounds. This includes the partial or full blocking of noise as well as the addition of user selected content. For example it is performed if a user decides to use *earplugs* or headphones with music playback enabled to escape from unpleasant environmental noises. In a general fashion, this was stated by P2:

> "I often wear headphones and that is pretty good to mask the environmental sounds like that construction site or when people are chatting in the hallway [...]"

Although it is not yet available, the phenomenon could also include manipulation of other sound properties besides volume. For instance, the frequency spectrum of sounds could be shifted and adjusted to create a more pleasant impression of voices or sounds emitted by machines (e.g. the screeching of a drill).

Context

The acoustic environment a user resides in, can take on many forms. In almost every case, it is full of various information sources. While a certain set of messages and notifications is received through digital means (i.e. notifications of personal devices), an even larger array of information is emitted by the physical world itself. The importance of certain information may be assessed through evaluating the costs and consequences of missing it. While being a participant in traffic (e.g. a pedestrian or a cyclist), a missed acoustic signal of a car can lead to harm. This means that the information relayed by engine noises, car horns or rolling noise is potentially crucial to a user's well-being. It can even lead to a full renunciation of usage. P5 put it the following way:

> "So far I don't have any headphones on while riding my bike because I don't want to die. [...] because I did not hear something."

The context is not only characterized through its informativeness. The factor of comfort is important, too. Usually loud noises and high frequencies i.e. screaming children are perceived as disturbing but low, monotonous sounds can also be troublesome. P3 stated the following aspect in communication:

> "[...] you always have those peaks [in volume] from someone [...] who does not possess this 'indoor-voice'."

Intervening Conditions

The participants also recurrently mentioned social relations and hierarchies. Some contexts allow a full isolation through headphones. Nevertheless, it is often perceived as impolite. Furthermore, hierarchies at work may imply a forced communication. The communication aspect is more important in the context of work, compared to other scenarios, e.g. in public transport, where users usually have no close relationship to the people around them. P2 expressed this by saying:

> "I just want to work, usually for myself, and maybe I want to have some peace of mind, but it's not realistic that I just do this and ignore everything else. Usually people visit throughout the day to talk [...]"

Headphones do not only isolate the user at a level of sound but also socially. A user wearing headphones is less likely to be approached by fellow people, because of the signalling effect of the headphones [5]. Users that put on their headphones are perceived as being already occupied. For instance, P9 reported:

> "So when I see someone in the train putting on their headphones or something, I don't talk to them [...]"

While wearing headphones, the prioritization of information is entrusted to others and may lead to missed opportunities, for instance social encounters or events.

The availability and the capabilities of technology or devices may also interfere with the user-chosen strategy, as mentioned by P6:

> "I guess I can listen to some music, but I will still be hearing the construction site somehow [...]"

Strategies of Soundscape Curation

The previously described factors influence the strategies of soundscape manipulation. If information embedded in the environment is prioritized sufficiently high (e.g. while participating in traffic), users completely refrain from using personal audio technology. The expectancy of somewhat less important information may lead to an adapted usage. One example of such a strategy was reported by P3:

> "If you are sitting at the airport, listening to content, then every time you hear the 'DingDong', in theory I still have one hour until my flight, but sometimes they'll relocate the gate, maybe they do this and that, and then you constantly take off the headphones."

However, if users do not have to rely on acoustic information or even want to block out some forms of information they choose to alter their personal soundscape completely. Users rely on these mechanisms and suffer if these do not work as intended:

> "Yesterday I had to listen to, I guess, twenty minutes of [discussions about] hair dyeing products because my headphones were broken. That was terrible." (P10)

Consequences of Soundscape Curation

The prioritization of sounds can lead to missed information, in case the user decides for the addition of selected content or purely blocking the surrounding sounds. The resulting impacts can vary in severity. P9 mentioned the following example:

> "While riding the train I most often listen to audio books and then you have the announcements. And 90% of the railroad announcements are 'I welcome you on the trip ...' [...] but sometimes there are important things [...] I'd like to know [...]"

However, users tend to accept this kind of information loss if the benefits are significant. This includes media consumption, relaxation, less distraction or the ability to communicate with distant people. Even interpersonal communication can be affected as expressed by P4:

> "[...] if I am sitting, focused, in the office and do some work and turn on music to isolate, then I don't give a shit if someone comes in."

Bystanders either interpret the state of the wearer, or poll it by attempting to communicate, as stated by P10:

> "You have surely [tried to] talk to someone who wore headphones and simply called them, and then you'll see whether [they] react or not."

If interpreted incorrectly, a bystander may either interrupt the wearer (and therefore disrupt focus or flow) or falsely refrain from communicating at all – a negative consequence. Such interruptions may disrespect proxemics [19], as a bystander may feel the need for physical contact to catch the wearer's attention (e.g. through tapping on the shoulder).

ONLINE QUESTIONNAIRE

To back aspects of the derived model with data from a larger population, we conducted an online survey. We recruited 201 participants via Amazons Mechanical Turk (mTurk) platform. While mTurk does not necessarily draw a representative sample of population, the people that participate on this platform have been shown to be able to generate meaningful results if appropriate precautions are taken [13, 24]. Only participants from English speaking countries (US, CA, UK, AU, NZ) were allowed to participate to ensure language comprehensibility. To verify sufficient attention of the workers, multiple checks were present in the questionnaire, ranging from simple attention monitoring (Please type the word 'cat' below.) to exclusions based on numerical ranges and mutually exclusive options. In total, 7 participants were removed from the data set due to contradictions in their answers, which left us with responses from 194 participants. Participants were given 60 minutes to complete the task after accepting it but it only took them 4:55 minutes on average to complete and submit the survey. Successful participation was rewarded with 0.50 $.

Demographics

Our participants were almost exclusively located in the US with only two participants from Great Britain, two participants from Canada and a single participant from Australia. We achieved an almost equal gender distribution with 96 participants reporting being male (49.5%) and 98 reporting being female. Ages ranged from 21 to 68 with a median of 32 (σ = 9.3). Participants reported high educational levels with the biggest group being Bachelor's degrees (46.9%) followed by high school degrees (36.1%), Master's degrees (10.3%), few professional degrees (2.1%) and doctorate degrees (0.5%). The remaining 4.1% mostly has some college education but did not graduate. 62.37% of our participants reported being employed for wages, while 22.16% reported being self-employed. Students, being out of work, homemakers, retired and being unable to work accounted for the remaining 15.46%. To add to the understanding of the population we included further questions regarding profession. We used the existing classification based on color of collar historically worn at work. According to that 56.2% of our participants identified as white collar (Office work, either employed, managerial or administrative). Second biggest group was pink collar (Service work, like healthcare, retail or sales) accounting for 20.1% followed by blue collar (Manual work, like construction, assembly or maintenance) at 6.2% and green collar (Environmental work, like farming, waste management or landscaping) at 1.0%. 16.2% choose 'Other' or did not report.

Results

Device Usage

Only 7 out of our 194 participants (3.6%) stated that they do not own or use headphones. The usage per week ranged from 1 to 100 hours (M = 6, σ = 12.1), same as duration of wear per week (M = 8.5, σ = 14.2). It is to be noted that our participants tend to wear headphones even if they are not actively in use which means there is no media playback enabled. When asked for which purpose they are using headphones, the most named answer was to 'block external noise' (57.2%), followed by 'do not want to disturb others' (53.6%), to 'concentrate better' (42.8%), 'privacy of the content listened to' (37.6%) and better sound quality (4.2%). We asked our participants what kind of personal audio technology they use in different situations (home alone, home with others, at work, during walking, in public transport (e.g. bus, train), sports (e.g. running, swimming), individual transport (e.g. car, bike) and events (stadium, concert)). Options were the types of headphones as described in preliminary considerations in the *Focus Groups* section and illustrated in figure 1. The results were ordered descending by number of entries and are shown in figure 3.

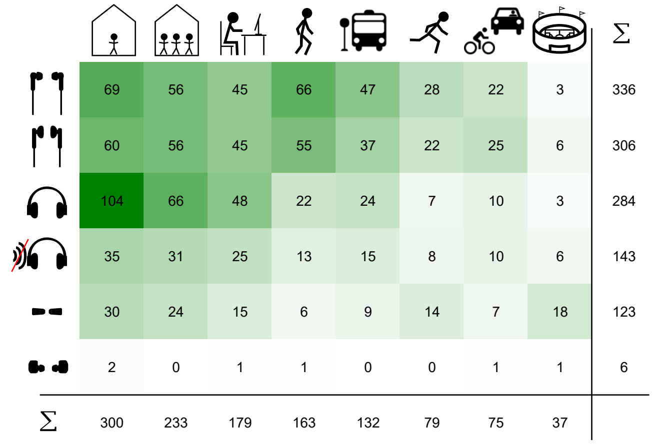

Figure 3. The reported usage of different audio technology (*in-ears*, *earbuds*, *over-ears*, *active noise-cancelling*, *earplugs*, *hearables*) in selected environments: home alone, home with others, work, walking, public transport, sports, personal transport (e.g. car and bike), events (e.g. stadium, concert, clubbing).

The second part of our questionnaire dealt with problems mentioned in the focus group discussions that may arise while using headphones. A problem that more than half of our participants reported to have already faced, is the 'accidental ignoring of an acquaintance trying to talk to them' (52.0%). 42.2% reported that they 'missed a relevant signal indoors', 32.0% were 'accused of not listening/paying attention'. Most other problems such as being 'accused of being antisocial' (23.7%), 'unable to classify a sound' (13.9%), 'unable to locate a sound' (12.4%) and 'missed a relevant signal outdoors' (11.3%) were stated less often. The additional comment field was used to report few problems regarding form factor of devices such as comfort and tangled cords. 24.2% of the participants stated that they did not encounter any of these problems.

Environment Assessment and Personal Preferences

Beside plain usage of devices, we wanted to get insights into the acoustical environment of our participants. On a 7-point Likert scale they had to express how their home and workplace is assessed by them in terms of quietness. Two thirds stated that they agree with the statement 'I consider my home to be quiet'. Workplaces are considered to be quiet by 60% of our participants. The rating of the statement 'I require silence to be able to concentrate' was more ambivalent with 54% agreement and 33% disagreement. This is also the case for being 'easily annoyed by sounds' (53% agree, 31% disagree) and 'easily distracted by sounds' (53% agree, 34% disagree).

Headphones are most frequently used for the consumption of media. The questionnaire contained two ranking-questions to coarsely determine the type of media people consume while relaxing or trying to relax and while working/trying to focus. These two categories were chosen as a representation

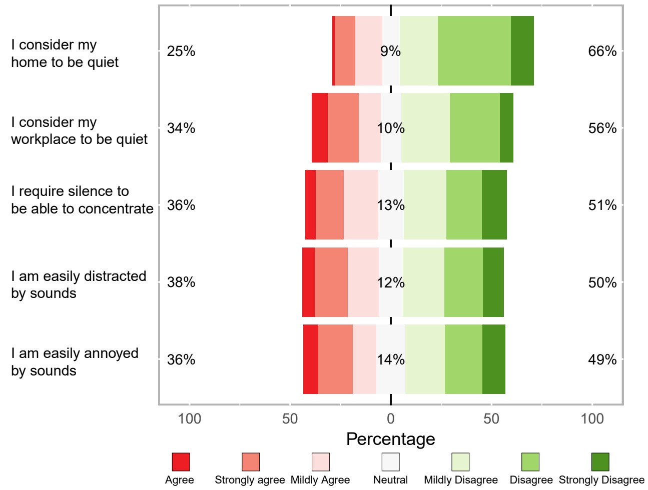

Figure 4. Participants' assessment of acoustic environment and their personal capabilities. Left-hand side percentages include all disagreeing answers, center percentages the neutral ones and right-hand side percentages include all agreeing ones.

for possibly endless sets of environments and personal goals for content consumption. For the rank one and two of media consumption while doing work/focusing, most frequently chosen category was music (81 and 44 mentions), followed by silence/no media (32 and 10 mentions) and spoken content (32 and 36 mentions). The most frequently mentioned category for the third rank was sounds of nature (24 mentions). For 'relaxing', the most frequently chosen category for the first rank was also music with 92 mentions, followed by spoken content with 35 mentions. The second rank is led by spoken content (56 mentions), closely followed by music (50 mentions).

Discussion

The low percentage of users that do not own or use headphones regularly and the high periods of use reported by respondents show us that headphones can be considered a fitting example of an ubiquitous technology which is deeply integrated into the life of many. This is furthermore one of the necessary *causal conditions* for soundscape curation.

Overall, headphones that are *in-ears* are most commonly used, followed by *earbuds*. This could at least partly be due to the fact that these are the types of headphones that are often included when buying a smartphone, music player or other equipment that is used for mobile media playback. The situational distribution of usage is noticeably different for *over-ears* headphones. Apparently, users tend to use these types of headphones in stationary settings, like home and work more often than other types of headphones. This is most likely due to higher wearing comfort but also due to worse portability which is an example of an *intervening condition* as proposed in our model. Divergent usage behavior can also be observed with *earplugs*. The *context* completely forms the usage in this case. They are most often used at home where the soundscape does not provide important cues. Situations in which the acoustic environment is rich in information such as walking and other forms of transportation show the lowest reported usage. Events like concerts or sport games are an exception to the rule. In this context, it is most likely attempted to protect the health of the ears by using this equipment – an *intervening condition* paired with an uncomfortable *context*.

Various *causal conditions* are documented in answers to the questions regarding environment assessment and personal preferences. Almost a third of the participants did not consider their workplace to be quiet. Therefore, a need for soundscape manipulation is given. This is also indicated by the ranking questions were users reported music as the most commonly used masking sound while trying to focus followed by silence/no media. Despite their positive effects [32], nature sounds are rarely used.

Most dominant *strategy* reported is the manipulation of the soundscape through addition of content, as music and spoken content are preferred for both work and relaxation. The *context* influencing the *strategies* used, is reflected in some of the reported issues, where users miss signals or consider their environment to be loud or unfavourable. *Consequences*, as described in our model, were also reported by many participants though we could not dig deep into interpersonal communication and social aspects due to its complexity. Only a quarter of our respondents stated to never have experienced the listed problems.

PROSPECT

The widespread use of headphones and similar devices in mobile scenarios shows the acceptance and ubiquity of *personal soundscape curation*. Even the use of simple earplugs provides a curated soundscape by attempting to block out signals. However, subtracting from or drowning out the environmental sounds has a core disadvantage: the potential to miss important signals. Despite this fundamental limitation, users still employ a diverse set of devices and tools to specifically regain their personal space [7, 18]. This is attempted by excluding an entire dimension of information from their perception.

Unlike the sense of hearing, the visual sense is directed and the human body possesses the ability to instantly shield the eyes from unwanted visual sensations. Attention can instead be managed with the support of tools (e.g. headphones or earplugs). Users also already leverage inherent traits of devices (i.e. their degree of transparency to external sounds) to manipulate how they perceive the world around them and to add content to their personal soundscape.

Recent Developments

The abilities of personal audio as an ubiquitous and mobile technology will certainly continue to expand and evolve. However, their co-evolution with the functionality of other, similarly ubiquitous devices like smartphones, has attained consumer availability only recently. Smarter devices attempt to go beyond filtering or drowning out unwanted sound: they incorporate additional sensors [4] and provide selective filtering methods [12, 33].

Notably, attempts to alleviate some *intervening conditions* are already being made by current products which go beyond active noise-cancelling. *SoundBrake* [38] is able to detect abnormal events occurring in a soundscape, to briefly interrupt media playback. Main issue of its implementation is that it is unable to judge the relevance of a sound event – something human users aware of their context usually are able to. The *Orosound Tilde* headphones [33] transfer a feature of the visual sense to the sense of hearing: directionality. By applying noise-cancellation to sounds outside of a fixed cone in front of the user, an artificial "field-of-hearing" is created, at the expense of situational awareness. As it is the case with most such devices, this is beneficial in some situations, where situational awareness can be sacrificed in favor of increased focus. *Here One* [12] provided presets and equalizer-like modifications to human hearing. The former is limited in scope, the latter poorly accessible to humans. As an alternative, bone-conduction and weakly-shielding earphones can be employed, where awareness is required [27]. Such bone-conduction headphones like *AfterShokz* [1] do not obstruct the user's hearing, but also do not provide any shielding against unwanted real-world sounds. This directly contradicts the causal condition and requirement of isolation.

All of these devices are certainly a step in the right direction, but exhibit flaws. As per the proposed model, they account for a specific, mostly *static* context (e.g. traffic or offices) and are not able to account for *dynamic* intervening conditions. Furthermore, each device and most usage scenarios have to balance information influx and the user's desire to filter it on social and environmental levels. Currently, users are applying a set of strategies, when they require an altered personal soundscape:

1. They *renounce the usage* if potential consequences are too severe (e.g. forgoing music playback while riding a bicycle).

2. They *alter and curate their soundscape*, defining a ratio between digital media and real world signals (e.g. drowning out construction noise with music).

3. They *adapt their usage*, by applying dynamic changes to their devices or media (e.g. repeatedly removing one or both earbuds).

All strategies are currently handled manually and require thorough analysis of context and potential intervening conditions by the user.

Considerations for Future Audio Technology

The context itself consists of the user's surrounding soundscape. Said soundscape is formed by the sounds emitted by physical objects. This includes the *states* and *events* of objects in their surrounding (e.g a whistling teapot). Only a subset of these signals conveys relevant information for the user. In contrast, the acquisition of this particular subset of signals, may be crucial for the users' wellbeing and safety [26].

A more specialized aspect of information acquisition and interaction are social protocols. Today's social protocols are already influenced by the use of personal audio technology. This includes communication, interruptions and social signalling. Currently, the use of headphones changes [2] and even impedes [11] social interaction.

Just as the state of the environment is dynamic, the social context and its role as information source changes. A person's voice, filtered out by the user may eventually try to relay critical information (e.g. danger). Alternatively, an "important" person, like a user's principal, may not always transmit important information. The relevance and significance of a signal, can often not be determined automatically. This excludes most static metaphors and filtering methods from being practicable.

The use of and the interaction with personal audio devices, is an information source in itself. Putting on or wearing *visible* personal audio devices is seen as an established signal to bystanders: it implies an ongoing activity and a certain aversion to interpersonal contact and disruption. In contrast to the process of interrupting and being interrupted, the causal conditions of deliberate isolation (of oneself) and exclusion (of bystanders) remain one of the core use cases of personal audio, apart from media consumption.

Users regularly weigh advantages and consequences of soundscape curation and accept potential consequences *willingly*. This decision is founded on two possible assumptions about the current environment: 1) The soundscape's informational relevance and sound levels are static, which is not the case for most environments. 2) The negative effects are either seen as unlikely or considered to be tolerable and are consciously accepted. Judging the necessity of interruption and the relevance of information is then, at least partly, offloaded to bystanders.

To conclude, current personal audio technology is not designed in a way that it allows users to handle their social context satisfactorily. Furthermore, information acquisition is made more difficult and users often have to make a choice between the surrounding acoustic environment and their own content and media.

LIMITATIONS AND FUTURE WORK

Our work has a few limitations. For instance, the initial framing of the focus groups aimed to create a foundation for the development of future mobile audio technologies. However, this was the exact motivation for the participants to elaborate on their current practices and devices in detail.

One could argue that we may not have reached proper saturation while conducting the focus groups. It should be noted that we did rely on groups of five researchers each, who digressed rarely and provided us with a broad set of insights, experiences and issues. Despite the lower sample size, we were able to derive a meaningful model but it may be more accurate having more participants from diverse backgrounds. Even if we do not consider it being likely, we cannot rule out the possibility that our model is only valid for a certain cultural area.

In addition, the crowdsourced survey relied on self-reporting and therefore may be subject to bias. It can only give an idea of practices and usage behavior of respondents but should be treated with caution. The sample gained from mTurk could be considered limited. However, the participants which passed the validation steps turned out to be from diverse educational and professional backgrounds.

To entirely cover users needs, more people from diverse backgrounds should be considered in additional focus groups. This could help to tailor devices that adapt to other, more specific target groups. Going beyond that, novel, adaptive systems may account for dynamic social and environmental factors, instead of offloading this task to users or bystanders.

CONCLUSION

In this work we presented the phenomenon of *personal soundscape curation* and a model that covers user centered aspects surrounding personal audio technology. It takes *context* and *intervening conditions* into account to explain the *strategies* that users evolve and explains resulting *consequences*. We conducted two focus groups and analyzed the transcripts of the resulting discussion by going through several stages of coding were the model resulted from. We furthermore conducted an online survey via mTurk to back our model with data from a larger and diverse population.

The current state of personal audio can be described as a widespread, but flawed attempt to gain full control over a human sense. Users try to balance curation and awareness, depending on a set of dynamic environmental factors, which should be taken into account in future endeavours to improve personal audio technology.

REFERENCES

1. AfterShokz. 2018. Technology - AfterShokz. (2018). https://aftershokz.com/pages/technology Accessed: 8.2.2018.
2. Tyler Bickford, Jason Stanyek, and Sumanth Gopinath. 2014. *Earbuds Are Good for Sharing: Children's Headphones as Social Media at a Vermont School*. The Oxford Handbook of Mobile Music Studies, Vol. 1. Oxford University Press, Oxford.
3. Sara Bly. 1982. Presenting Information in Sound. In *Proceedings of the 1982 Conference on Human Factors in Computing Systems (CHI '82)*. ACM, New York, NY, USA, 371–375. DOI: http://dx.doi.org/10.1145/800049.801814
4. BRAGI GmbH. 2017. The Dash Pro - True Wireless Intelligent Earphones. (2017). https://www.bragi.com/thedashpro/ Accessed: 17.9.2017.
5. 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
6. Michael Bull. 2008a. *Sound Moves: IPod Culture and Urban Experience*. Routledge, New York, NY, 10001.
7. Michael Bull. 2008b. 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, Oxford, UK.
8. Andreas Butz and Ralf Jung. 2005. Seamless User Notification in Ambient Soundscapes. In *Proceedings of the 10th International Conference on Intelligent User Interfaces (IUI '05)*. ACM, New York, NY, USA, 320–322. DOI: http://dx.doi.org/10.1145/1040830.1040914
9. George Chernyshov, Benjamin Tag, Jiajun Chen, Vontin Noriyasu, Paul Lukowicz, and Kai Kunze. 2016. Wearable Ambient Sound Display: Embedding Information in Personal Music. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers (ISWC '16)*. ACM, New York, NY, USA, 58–59. DOI: http://dx.doi.org/10.1145/2971763.2971789
10. Juliet Corbin and Anselm Strauss. 1990. Grounded Theory Research: Procedures, Canons and Evaluative Criteria. *Zeitschrift fuer Soziologie* 19, 6 (1990), 418–427.
11. Robert Morrison Crane. 2005. *Social distance and loneliness as they relate to headphones used with portable audio technology*. Ph.D. Dissertation. Humboldt State University.
12. Doppler Labs Inc. 2017. Here One - Wireless Smart Earbuds. (2017). https://hereplus.me/products/here-one Accessed: 17.9.2017.
13. Julie S. Downs, Mandy B. Holbrook, Steve Sheng, and Lorrie Faith Cranor. 2010. Are your participants gaming the system?: screening mechanical turk workers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Atlanta, GA, USA, 2399–2402. DOI: http://dx.doi.org/10.1145/1753326.1753688
14. Jose A. Gallud and Ricardo Tesoriero. 2015. Smartphone Notifications: A Study on the Sound to Soundless Tendency. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '15)*. ACM, New York, NY, USA, 819–824. DOI: http://dx.doi.org/10.1145/2786567.2793706
15. Gartner Inc. 2018. Gartner Says Worldwide Sales of Smartphones Recorded First Ever Decline During the Fourth Quarter of 2017. (Feb. 2018). https://www.gartner.com/en/newsroom/press-releases/2018-02-22-gartner-says-worldwide-sales-of-smartphones-recorded-first-ever-decline-during-the-fourth-quarter-of-2017 Accessed: 3.10.2018.
16. Stavros Garzonis, Chris Bevan, and Eamonn O'Neill. 2008. Mobile Service Audio Notifications: Intuitive Semantics and Noises. In *Proceedings of the 20th Australasian Conference on Computer-Human Interaction: Designing for Habitus and Habitat (OZCHI '08)*. ACM, New York, NY, USA, 156–163. DOI: http://dx.doi.org/10.1145/1517744.1517793
17. 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

18. Mack Hagood. 2011. Quiet comfort: Noise, otherness, and the mobile production of personal space. *American Quarterly* 63, 3 (2011), 573–589.

19. Edward T. Hall, Ray L. Birdwhistell, Bernhard Bock, Paul Bohannan, A. Richard Diebold, Marshall Durbin, Munro S. Edmonson, J. L. Fischer, Dell Hymes, Solon T. Kimball, Weston La Barre, J. E. McClellan, Donald S. Marshall, G. B. Milner, Harvey B. Sarles, George L Trager, and Andrew P. Vayda. 1968. Proxemics. *Current Anthropology* 9, 2/3 (April 1968), 83–108. DOI: http://dx.doi.org/10.1086/200975

20. Harman and Silencio Films. 2015. HARMAN | Auditory Augmented Reality. (Jan. 2015). https://vimeo.com/115933849 Accessed: 10.9.2017.

21. Helena Jahncke, Patrik Björkeholm, John E. Marsh, Johan Odelius, and Patrik Sörqvist. 2016. Office noise: Can headphones and masking sound attenuate distraction by background speech? *Work* 55, 3 (2016), 505–513.

22. Wen Jiang, Fei Zhao, Nicola Guderley, and Vinaya Manchaiah. 2016. Daily music exposure dose and hearing problems using personal listening devices in adolescents and young adults: A systematic review. *International Journal of Audiology* 55, 4 (April 2016), 197–205. DOI: http://dx.doi.org/10.3109/14992027.2015.1122237

23. Katherine Kenna and Alex Ryan. 2016. Superhearo: Sensory Augmentation for Your Friendly Neighborhood Vigilante. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (UbiComp '16)*. ACM, New York, NY, USA, 481–486. DOI: http://dx.doi.org/10.1145/2968219.2971342

24. Aniket Kittur, Ed H. Chi, and Bongwon Suh. 2008. Crowdsourcing user studies with Mechanical Turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, Florence, Italy, 453–456. DOI:http://dx.doi.org/10.1145/1357054.1357127

25. Richard A. Krueger and Mary Anne Casey. 2014. *Focus groups: A practical guide for applied research*. Sage Publications, Thousand Oaks, California.

26. 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* 18, 5 (Oct. 2012), 287–290. DOI: http://dx.doi.org/10.1136/injuryprev-2011-040161

27. 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 *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*. IEEE Computer Society, Washington, DC, USA, 173–176. DOI: http://dx.doi.org/10.1109/ISMAR.2007.4538843

28. Ling Ma, Ben Milner, and Dan Smith. 2006. Acoustic Environment Classification. *ACM Trans. Speech Lang. Process.* 3, 2 (July 2006), 1–22. DOI: http://dx.doi.org/10.1145/1149290.1149292

29. 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–.

30. Daniel McFarlane. 1999. Coordinating the interruption of people in human-computer interaction. In *Human-computer interaction, INTERACT*, Vol. 99. IOS Press, Edinburgh, UK, 295.

31. Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems* 77, 12 (1994), 1321–1329.

32. Joseph W. Newbold, Jacob Luton, Anna L. Cox, and Sandy J. J. Gould. 2017. Using Nature-based Soundscapes to Support Task Performance and Mood. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. ACM, New York, NY, USA, 2802–2809. DOI: http://dx.doi.org/10.1145/3027063.3053214

33. Orosound SAS. 2018. Tilde Earphones - Experience Noise-Filtering Technology. (2018). https://www.orosound.com/tilde-earphones/ Accessed: 3.10.2018.

34. Amanda Petrusich. 2016. Headphones Everywhere. (July 2016). https://www.newyorker.com/culture/cultural-comment/headphones-everywhere Accessed: 18.9.2017, ISSN: 0028-792X.

35. Patrick Reinelt, Shewit Hadish, and Claus-Peter H. Ernst. 2016. How Design Influences Headphone Usage. In *The Drivers of Wearable Device Usage*. Springer, Cham, Switzerland, 59–68. https://link.springer.com/chapter/10.1007/978-3-319-30376-5_6 DOI: 10.1007/978-3-319-30376-5_6.

36. 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, Article 20, 8 pages. DOI: http://dx.doi.org/10.1145/2875194.2875247

37. 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

38. Sundae Electronics LLC. 2018. SoundBrake - Make Your Headphones Smarter. (2018). https://www.sundaelectronics.com/ Accessed: 7.2.2018.

39. Rob Van Rijswijk and Jeroen Strijbos. 2013. Sounds in Your Pocket: Composing Live Soundscapes with an App. *Leonardo Music Journal* 23 (2013), 27–29. DOI: http://dx.doi.org/10.1162/LMJ_a_00149

40. Ken Vassie and Miles Richardson. 2017. Effect of self-adjustable masking noise on open-plan office worker's concentration, task performance and attitudes. *Applied Acoustics* 119 (2017), 119–127.

41. Sungwook Yoon, Yangwon Lim, Hankyu Lim, and Hyenki Kim. 2012. Architecture of Automatic Warning System on Urgent Traffic situation for Headphone Users. *International Journal of Multimedia and Ubiquitous Engineering* 7, 2 (2012), 421–426.

42. Fan-Gang Zeng. 2016. Goodbye Google Glass, Hello Smart Earphones. *The Hearing Journal* 69, 12 (Dec. 2016), 6. DOI: http://dx.doi.org/10.1097/01.HJ.0000511124.64048.c1