Institute for Visualization and Interactive Systems

University of Stuttgart
Universitätsstraße 38
D–70569 Stuttgart

Bachelorarbeit

# Elicitation of Modalities for Controlling Augmented Hearing

Maike Ernst

| | |
|---|---|
| **Course of Study:** | Medieninformatik |
| **Examiner:** | Jun.-Prof. Michael Sedlmair |
| **Supervisor:** | Francisco Kiss M.Sc.,<br>Dipl.-Inf. Alexandra Voit |
| **Commenced:** | February 7, 2019 |
| **Completed:** | August 7, 2019 |

## Abstract

Augmented Reality technologies are continually evolving and changing our interaction with the environment. While visual extensions of the environment are prevalent in leisure time as well as at the workplace, hardly any auditory changes are applied to the environment. The best-known application of Augmented Sound is active noise-canceling headphones, which enable the filtering of distracting background noise. Currently this approach of changing our soundscape is being pursued and expanded. Nowadays, we are capable of amplifying the sound coming from our point of view and integrating virtual sounds into the environment, but it is currently not possible to completely adjust our soundscape. In this work, we investigate the expectations and needs of potential end-users regarding the control of Augmented Hearing. We have conducted three consecutive studies to identify not only desired actions but also control modalities. In an elicitation study, possible gestures/interactions for three different modalities were identified and verified through a large-scale validation study. Based on the study results, we can suggest applications for Augmented Sound and user-friendly control methods.

## Kurzfassung

Augmented Reality-Technologien entwickeln sich ständig weiter und verändern unsere Interaktion mit der Umwelt. Während visuelle Erweiterungen der Umgebung sowohl in der Freizeit als auch am Arbeitsplatz weit verbreitet sind, werden akustische Veränderungen an der Umgebung kaum vorgenommen. Die bekannteste Anwendung von Augmented Sound sind aktive geräuschunterdrückende Kopfhörer, welche die Filterung von störenden Hintergrundgeräuschen ermöglichen. Derzeit wird dieser Ansatz, unsere Geräuschkulisse zu verändern verfolgt und ausgebaut. Heutzutage sind wir in der Lage, den Sound in unserer Blickrichtung zu verstärken und virtuelle Klänge in die Umgebung zu integrieren. Es ist derzeit aber nicht möglich, unsere Klanglandschaft vollständig an unsere Bedürfnisse anzupassen. In dieser Arbeit untersuchen wir die Erwartungen und Bedürfnisse potenzieller Endverbraucher hinsichtlich der Kontrolle von Augmented Hearing. Wir haben drei aufeinander folgende Studien durchgeführt, um nicht nur die gewünschten Maßnahmen, sondern auch die Kontrollmodalitäten zu identifizieren. In einer Gesten-Elicitation-Studie wurden mögliche Befehle für drei verschiedene Modalitäten identifiziert und durch eine groß angelegte Validierungsstudie verifiziert. Basierend auf den Studienergebnissen können wir Anwendungen für Augmented Sound und benutzerfreundliche Steuerungsmethoden vorschlagen.

# Contents

# List of Tables

# 1 Introduction

Interaction with virtual content is no longer unusual in everyday life. While 3 billion people worldwide already own a smartphone[1], more and more technologies and devices are being developed to equip the user's environment with digital information. This is why we are interacting more and more with the so-called Augmented Reality (AR). This technology enriches the real world with additional digital information, such as graphics, messages, or sounds, which in reality are not there. The possibilities in this area are just being exploited, and the potential is almost infinite. Many applications designed to simplify and improve everyday life are based on the Augmented Vision approach: additional information is visually perceptible, for example through glasses, such as the HoloLens[2], providing the user with insights into his environment. Generally, AR technologies are widely represented in research and everyday life nowadays. The possible applications range from the workplace to gaming in almost all areas of life. Primarily visual data, such as images or informational text, is used to enrich reality with virtual data, although any sensory perception can be affected by adding information. While a human being is able to freely decide what he will and will not see by looking away or even closing his eyes, this is not possible with hearing. We may be able to focus on individual sound sources and almost blend out background noises, but we cannot turn away from a sound. Since we are at the mercy of constant noise, computer-assisted hearing could improve our everyday lives. An example for this technology are headphones that minimize the noise level around us and have been around for several years now. With the help of directional microphones, current technologies are even capable of recognizing the direction we are looking in, and intensifying the sounds from that direction. More than just changing the soundscape, it is also possible to enrich it, with virtual sound. Technologies exist that on the one hand, can record sounds and on the other hand, realistically place them in the environment. More precisely, 360° recordings have to be made so not only the sound but also its position and intensity in relation to the audience can be recorded and played. This 3D sound allows not only to transmit auditory information but also to create realistic ambient sounds without anyone else being aware of them. So there are already many approaches to enable complete customization of the users' soundscape. The required technologies are not the main issue in realization but a user-friendly design, both in terms of the functions provided and the control system. In order to accomplish that, it is possible to apply a user-based design. Therefore, we conducted an elicitation study as part of this work. We followed the approach of Wobbrock et al. [1, 2] as it has already been carried out successfully in a variety of application areas. In order to provide a first approach for developing such systems, we conducted three successive user studies. According to this methodology, participants are confronted with possible actions to be executed with a given control modality. First started with focus groups to identify the actions wanted by potential users. The participants considered which soundscapes they experience in everyday life and how they would like to modify each one.

[1] https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/

[2] https://www.microsoft.com/en-us/hololens

We used the information regarding the actions needed for this customization to define referents for an elicitation study. We also asked the participants what control options they could imagine and developed three different approaches based on their answers: Mid-air gestures, touch gestures, and key assignments for the three-button control element used by headphones. We then presented the different actions and control modalities to the participants. We asked the potential users to suggest suitable control actions for each referent for the three different control systems. The level of consensus between the participants was determined using an agreement rate analysis to evaluate the proposed commands. Finally, the gestures were compared and rated by many participants using an online survey. Our work comprises (1) a classification of gestures for controlling augmented hearing using a taxonomy, (2) three gesture sets for different control modalities (mid-air gestures, touch gestures, key assignments) which reveal user preferences according to specific use cases. Besides, we discuss how these insights can affect the design of control options for augmented hearing. The goal of this thesis is to enable user-friendly control of future AR technologies designed to enhance the hearing experience.

## Outline

This work is structured as follows:

**Chapter 2 - Background and Related Work** Presentation of related work and technologies.

**Chapter 3 - Concept** Potentials of Augmented Sound Technology

**Chapter 4 - Focus Groups** Procedure and results of the focus groups.

**Chapter 5 - Elicitation Study** Finding gestures for identified actions and modalities.

**Chapter 6 - Validation Study** Procedure and results of an online survey to validate previous results.

**Chapter 7 - Discussion** Analysis of the results of all three studies.

**Chapter 8 - Conclusion and outlook** Summary of results and outlook on future work.

# 2 Background and Related Work

Our work is based on the following foundations:
(1) Gesture elicitation studies and
(2) The current state of research and technologies in the field of Augmented Sound

## 2.1 Elicitation Studies

A gesture elicitation study invites users into the design process, known as participatory design [3]. Their opinions and ideas are then taken into account when creating gesture controls. The participants of an elicitation study are presented with several actions for one application area. In gestures elicitation, an action is called a referent. For each of these referents, the participant proposes one or (ideally) several gestures which he considers suitable. Conducting an elicitation study is very appreciated and user-friendly since potential users tend to prefer gestures created by other users [4]. This is explained by the fact that experts tend to create more complex gestures [4] and technology-based approaches achieve non-intuitive results [5]. After all suggested gestures have been collected, a gesture set is defined for the examined application area. For each referent, we search for the most frequently mentioned gesture and then include it in the gesture set for the corresponding action. Finally, the degree of agreement between the participants is calculated to verify the set. The required metrics have already been established: agreement rate, disagreement rate, and co-agreement rate by Wobbrock et al. [1]. Their methodology was already successfully performed in a variety of fields of application [6–11], including the field of AR [12–14].
The purpose of several of these studies is to make it easier to control existing (everyday) technologies. For instance, Vatavu et al. [6] searched for mid-air gestures for television control. Ruiz et al.[7] collected gestures performed with mobile phones to facilitate everyday situations such as answering phone calls. Since gestures for different systems are also required to meet varying requirements, it makes sense and is necessary for new systems to conduct its own study. It is possible to use the same or similar gestures for related technologies, but small differences, such as the display size, influence the participants' suggestions. Therefore various other studies are focusing on gesture control of display systems. This included both common displays of different sizes [8, 9] and surface computing technologies [10]. Another study investigated how to design consistent gestures for different input devices as consumers consistently (can) use increasingly more devices to trigger the same actions or control the same device [11]. As already mentioned, elicitation studies were also successfully conducted in the AR domain. T. Piumsomboon et al. [12] conducted a study to identify natural hand gestures for applications in the AR domain to gain a better understanding of users' thought and behavior. They were interested in gestures for a set of 40 different referents to edit a virtual 3D object, e.g., rotate or zoom, or interact with an interface like opening a menu. Although a tabletop construction was used, researchers stated that the user-designed gestures could be performed mid-air without a surface in any AR environment.

Other studies gathered user commands for an AR interface that allows multi-modal input [13, 14]. Voida et al. [13] asked their participants to suggest gestures and voice commands to manipulate objects in a visually enhanced office environment. The results show that users generally prefer voice commands, but the distance to the object is essential in the choice of modality. The closer they are to the object, the more likely it is, they will use hand gestures rather than speech. The researches also discovered that gestures are preferably performed sitting and with minor movements. Lee et al.[14] observed the influence of display type and different display conditions on user instructions. Their participants modified visual objects on two different display types by hand gestures or speech commands. The results reveal that the task influences the selection of the input type, and most users tend to combine both modalities. Although some of the studies covered sound-based actions such as making the sound louder or quieter, the current state of research is not sufficient to provide suitable gestures for controlling augmented hearing.

## 2.2 Augmented Reality

Augmented Reality (AR) is the computer-aided enrichment of reality by adding virtual information [16, 17]. It has evolved into a well-known technology in the last years The most common form of AR in everyday life is Augmented Vision - the display of digital visual data, such as images or information. Over 100,000,000 Android users have downloaded the mobile game Pokémon Go[1] since its release in 2016, making it one of the most successful mobile games of recent years and one of the best-known applications of AR. The game environment is mapped to the real world, so the player has to walk around in real life to find and interact with game content like virtual monsters. Also widespread are the so-called Head-Up Displays in the automotive sector, which project information onto the windshield in front of the driver, such as speed and navigation signals. Driver safety is improved because the driver no longer has to take his eyes off the road for reading information like speed.

Many approaches focusing on AR in the industry aiming to relieve and support workers [18, 19] or to develop technologies in educational environments [15, 20]. While learning the application of AR can increase efficiency and effectiveness as it makes education interactive. Furthermore, learning is simplified when information is presented according to context and exactly where it is needed [15]. Issues like technical difficulties, as well as learning and pedagogical problems, have to be considered when using such systems[20]. Jost et al. [18] developed the system SmARPro, which preprocesses information based on the user and the context and displays it using AR. For instance, it provides the information required for the logistics process at the appropriate location. Work instructions could also become more understandable by using AR [19]. For this purpose, the researchers compared different ways of displaying instructions: paper-based, head-mounted display (HMD), tablet, and in-situ. HMD and in-situ based views allowed for faster task completion than paper-based and via a tablet. In both cases, the worker is shown the required information exactly where he needs it. AR offers great potential not only at the workplace but also to improve and simplify everyday life. Especially in the travel sector, AR provides many possibilities to optimize vacations and trips. Zhou et al. [21] created an app that provides information about tourist attractions using an image retrieval algorithm. An image is loaded into the app, the algorithm compares it with an existing data set and returns the image with the highest similarity.

[1] https://play.google.com/store/apps/details?id=com.nianticlabs.pokemongo&hl=de

Now the user has several possibilities to get more information about the location, including a video. The researchers believe that such an algorithm can be used not only for travel but also for other everyday activities, such as shopping. Other approaches were also explored in order to help users with purchasing groceries [22, 23]. Colored tags will appear on the image of the aisles to indicate whether a product is healthy and meets the user's chosen restrictions (e.g., lactose-free) [22]. Besides, color tags can also identify (un)suitable products immediately, for example, if they contain allergens [23].

### 2.2.1 Augmented Sound

On the contrary, Augmented Sound has so far been a comparatively small area in research and everyday life. The technology encompasses various approaches. Augmented sound can be understood not only as the extension of the soundscape by acoustic information but also as the manipulation of ambient sounds. Headphones[2], which can filter ambient noise to enable the wearer more peace of mind, are now widely used. So-called active noise-canceling (ANC) headphones can be equipped with sound-insulating material that prevents ambient noise from reaching the user's ear. Alternatively, some of those headphones can additionally suppress sound waves. They mirror the incoming sound waves to create absolute silence by adding up the real, and the artificially generated sound waves.[24]
Moreover, some manufacturers are taking this idea a step further. For example, there are headphones[3] that, in addition to an ANC function, they also amplify the sound the user wants to hear. They use directional microphones to detect the direction in which the wearer is looking and amplify the sound source(s) coming from this direction.
Google Pixel Buds[4] are headphones having a real-time translation function implemented. Users can communicate directly with foreign-language people without the need for (technical) translation assistance. Sennheiser is also active in the Augmented Sound sector with its AMBEO segment. The company offers headphones[5] that enable transparent hearing. This technology allows the user to hear ambient sounds even while using the headphones [25] to, for example, listen to music. It can continuously adjust the degree of transparency to adapt its soundscape for the best possible quality. They also developed a microphone which is capable of making 360° sound recordings[6]. Such records can be used to map virtual sound to the location of real-world objects, meaning the user perceives the pre-recorded sound as if the object was actually producing it. Virtual 360° sounds can be used in a variety of ways. Zhou et al. [26] discovered that integrating 3D sound helps complete AR environments. Additionally, this study reveals that immersion in visual AR environments is improved and a more realistic experience can be created by applying 360° sound. It simplifies depth perception and task performance, and the cooperation between users. *Dramagic* is a table designed to encourage the creativity of children by using virtual sound[27]. It is an effective method of stimulating children's imagination and inspiring them. This technology can enhance not only virtual and augmented reality with realistic sound, but also in everyday life, users can adjust their soundscapes through it.

[2]https://www.audiophileon.com/news/best-noise-cancelling-earbuds-earphones
[3]https://www.gadget-rausch.de/iqbuds-in-ears-mit-hoergeraet-funktion-und-noise-cancelling/
[4]https://store.google.com/product/google_pixel_buds
[5]https://www.youtube.com/watch?v=8j4X-uN83I4
[6]https://en-de.sennheiser.com/microphone-3d-audio-ambeo-vr-mic

Bose follows a somewhat similar idea with its Bose AR frames[7]. These are sunglasses with built-in speakers around the ears. In contrast to previous technologies which manipulate the soundscape of the user, additional information can be played via the loudspeakers, e.g., about the surroundings. This functionality is also used for audio guides in museums, which is becoming more common. The visitor does not need a guided tour done by another person to adapt his visit independently and individually to his interests. Hatala et al. [28, 29] applied this approach and developed a more feature-rich model of such an audio guide. It not only provides information but also recognizes which background music or narrative style appeals to the visitor and dynamically adapts to it. Systems that are supposed to work using location-related functions must know where the user is positioned and their location in relation to the object. Therefore, technologies, such as the patents [30, 31], have been developed to virtual position sound correctly in AR in order to offer the most realistic possible experience. Larsen et al. [32] developed a system to consider the delay of sound waves. Throughout this study, concert-goers wore a headset which played the sound with a delay, depending on the distance between the wearer and the stage. The study shows that the participants perceived this Augmented Sound approach as an improvement of their sound perception. However, they also felt it was a disadvantage to be isolated by the required equipment. These concerns should be considered when designing and realizing similar designs.

## 2.3 Summary

Augmented Sound offers great potential for improving everyday life. Current technologies enable the suppression of background noise, the amplification of sounds the user wants to hear and providing information; for instance, about the environment. It is also possible to record and play 3D-sound, meaning sounds are recorded and played back with the correct position and intensity in relation to the listener. Although there is already research and technology that can modify the soundscape of humans, there is still no system that can ensure full freedom of choice. The main problem in the implementation of such systems is not necessarily the technical realization, but to enable a user-friendly interface. We learned that several gesture elicitation studies have already been conducted in a variety of areas and can ensure user-friendly systems. For this reason, we conducted an elicitation study to explore how people would like to adjust their soundscape in order to find suitable control options. So we will take user opinions into account, already during the development of such systems, by finding out beforehand what the device has to be capable of and how potential users would like to control it. We believe the results of this investigation will enable a user-friendly use of Augmented Hearing.

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

# 3 Concept

By integrating Augmented Sound based technologies into everyday life, users can customize their soundscape to their needs and thereby improving their perception of sound. This technology creates new opportunities for interaction with the environment. We can access visual information at any time in the form of images or text on our smart devices, such as the smartphone. However, this kind of information transfer between humans and devices always requires a display and can, therefore distract. As already mentioned in the previous Chapter 2.3, existing technologies such as glasses can deliver auditory information about buildings or restaurants to the wearer. This virtual (360°) sound can also be integrated into Virtual Reality (VR) for an even more realistic experience. But not only the provision of information is an application for augmented sound. The user can remove unwanted sounds or amplify important sounds to hear exactly what he wants. Today, removing noise through active noise-canceling headphones is an established application of augmented sound, but the possibilities of modifying the user's sound environment remain unexploited. The integration of virtual sound, the removal of sound sources, the changing of attributes of sound sources, such as volume, can support and improve people's perception of sound. Above all, it is essential to enable user-friendly control so the user can quickly and easily adapt his environment to his needs.

# 4 Focus Groups

Since as far as to our knowledge no research yet explored which actions potential users could need when using Augmented Sound, we started with Focus Groups to identify them.

## 4.1 Participants

Following Onwuegbuzies et al. [33] recommendations, we planned two focus groups to obtain more significant results. Based on the experience of previous researchers [34], we aimed for relatively small groups of six to eight participants. We recruited a total of 16 participants (6 female) with an average age of 22.31 (SD = 1.9) years using university mailing lists, flyers, and social networks. Our participants were students of different faculties (medical-technical computer science, cybernetics, software engineering, computer science, simulation technology, media computer science, mechanical language processing, mechanical engineering, and economics). Each of the participants had little or no experience with AR and no experience at all in the field of Augmented Sound.

### 4.1.1 Session One

In the first session, we had eight participants, three of whom identified themselves as female. The age ranged from 18 to 24, with an average age of 21.75 (SD = 1.9).

### 4.1.2 Session Two

In the second session, eight subjects participated, three of them identified as female. The participants' age ranged from 18 to 29 years and had an average age of 22.87 (SD = 1.9). Our participants discussed their daily soundscapes and how they would like to modify them.

## 4.2 Procedure

The schedule of the two meetings was identical. Since the conversation between the individual participants in focus groups has a significant influence on the results, we started with an introduction round and a short icebreaker game. We introduced the participants afterward to AR, especially in the area of audio and were able to gain an initial overview of possible applications through a video [1]. The video demonstrates the BOSE glasses already mentioned in Chapter 2.3, which can provide the wearer with information about his environment. Since none of the participants used a comparable technique or even thought about the use of Augmented Sound, the participants should consider soundscapes they are exposed to in their everyday lives. Afterward, they should consider modifications they want to apply to their soundscape and which functions should be integrated into the device. We asked the participants to use sticky notes to write down the actions important to them, and to compare and structure them with each other, see Figure 4.1. In addition to one person who was instructed to take notes, both sessions were audio and video recorded with the consent of the participants.

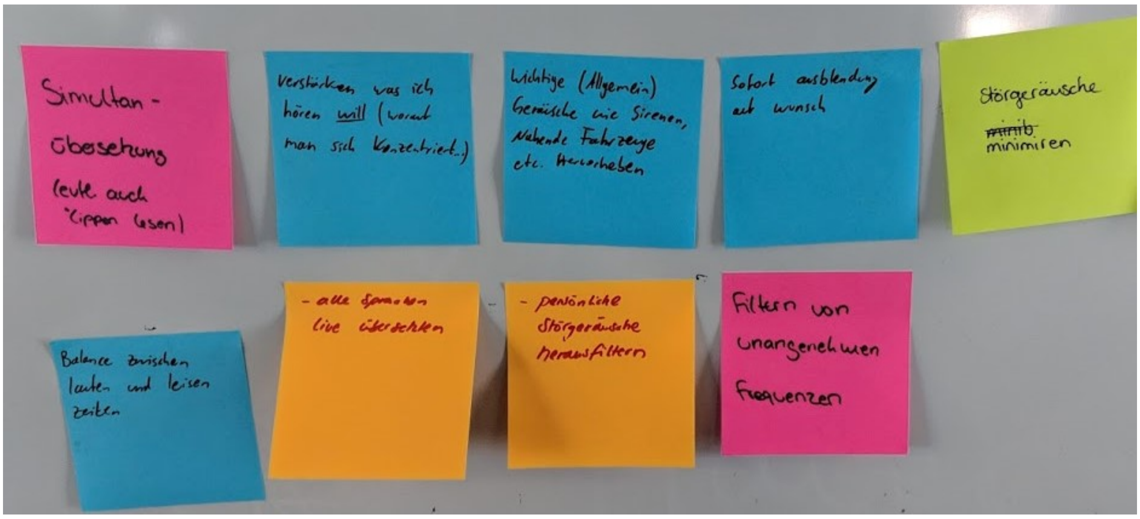



**Figure 4.1:** Brainstorming during the focus group

[1] https://www.youtube.com/watch?v=5_dCrJjRGT8

## 4.3 Findings

With the help of the recordings and notes, the results of the focus groups were analyzed after the second session to define actions that are relevant to users.

### 4.3.1 Actions

To identify potential actions for controlling Augmented Sound, we asked the participants to name modifications they consider as meaningful and relevant. Afterward, they examined and discussed the distinctions between each other. In the end, each participant had to choose the 7 to 10 most important actions. Based on these discussions and lists, we gathered all mentioned actions. We were able to identify a total of sixteen actions as relevant during both sessions, using the audio recordings and notes.

| | **Action** | **Explanation** |
|---|---|---|
| 1. | Augmentation on | Turn on augmented sense |
| 2. | Augmentation off | Turn of augmented sense |
| 3. | Volume up | Increase the volume of the current soundscape |
| 4. | Volume down | Decrease the volume of the current soundscape |
| 5. | Add sound source | Add an individual sound to current soundscape |
| 6. | Remove sound source | Remove an individual sound to current soundscape |
| 7. | Focus on sound source | Increase the volume of a specific sound source while making the remaining sound sources quieter |
| 8. | Listen from here | Listen to sound coming from a reduced area multiple sources simultaneously |
| 9. | Active Noise Cancelling higher | Reduce the environmental sound volume |
| 10. | Active Noise Cancelling lower | Listen to environmental sounds closer to their real intensity |
| 11. | Frequency higher | Raise the frequency of a sound source |
| 12. | Frequency lower | Decrease the frequency of a sound source |
| 13. | Translation | Translate foreign languages or unclear pronunciation |
| 14. | Replace soundsource | Replace an unpleasant sound with a more pleasant one |
| 15. | Identify soundsource | Get information about the soundsource (Location, type of sound, etc.) |
| 16. | Virtual Sound | sounds that do not occur in the soundscape (e.g. information, reminders, news, phone ringing) |

**Table 4.1:** List of all actions identified during the focus groups with their explanations

### 4.3.2 Modalities

During the sessions, we also asked the participants about possibilities and devices they could imagine for controlling augmented sound. In the first sessions, two participants brought up futuristic ideas like mind control or plaster-like patches on the skin, which can perceive touch. The remaining participants, both in the first and second session, focused on ordinary options. In addition to mid-air and touch gestures, the control element for headphones has also established itself. Participants considered it convenient that using headphones to improve their sound perception usually provides such a controller.

### 4.3.3 Further issues

During the conversation about the different soundscapes, the participants discussed other aspects besides possible actions. Many of them consider Augmented Hearing Technology to be useful, as they often feel annoyed by noise in public. While they see the opportunity to adapt their soundscape to their needs and have the chance to improve their (working) environment through this technology, they also look at some aspects more critically. Privacy could be violated by recording functions, fading out other people creates distance between people, and carries great risks. For the majority of participants, however, the benefits outweigh the risks, as each is responsible for their own usage. A popular request of the participants is intelligent software capable of making the necessary modifications autonomously.

## 4.4 Discussion

Based on these actions, we worked out the relevant referents for the follow-up study. We excluded actions that were only listed in one group or by very few participants (action 14 and 15) as these may be important for individuals, but most likely do not reflect the needs of everyone. Actions beyond basic options were also not considered for further studies (action 13 and 16). While selecting the control modalities, we did not consider futuristic approaches, as a technical realization would currently not be possible. Mid-air gestures and headphone buttons were mentioned or confirmed as meaningful by most participants. The participants considered both the smartphone and the smartwatch for the touch gestures. They criticized the phone for not always being nearby, unlike the watch, so we decided to elicit the touch gestures based on a smartwatch, especially since these gestures can also be performed on a smartphone.

# 5 Elicitation Study

We conducted the elicitation study based on the relevant actions and modalities we identified in the focus groups.

## 5.1 Participants

Twenty participants took part in our gesture elicitation study, four of them identified as female. The age ranged from 20 to 28, with an average age of 23.45 (SD=1.80). Each of the participants had little or no experience with AR and no experience at all in the field of Augmented Sound. Our participants were students of different faculties (software engineering, computer science, media computer science, mechanical language processing, real estate technology and management and pedagogy).

| **ID** | **gender** | **age** | **ID** | **gender** | **age** | **ID** | **gender** | **age** |
|---|---|---|---|---|---|---|---|---|
| P1 | male | 23 | P8 | male | 25 | P15 | male | 20 |
| P2 | male | 26 | P9 | male | 22 | P16 | male | 21 |
| P3 | female | 25 | P10 | male | 23 | P17 | male | 24 |
| P4 | female | 24 | P11 | male | 24 | P18 | male | 24 |
| P5 | male | 28 | P12 | male | 23 | P19 | male | 23 |
| P6 | female | 22 | P13 | male | 25 | P20 | female | 21 |
| P7 | male | 23 | P14 | male | 23 | | | |

**Table 5.1:** Participants of the gesture elicitation

## 5.2 Referents

As mentioned in the previous chapter, we decided not to take some actions as a referent for the gestures elicitation. We devided the remaining twelve referents into six groups of pairs to randomize the order throughout the study: (A) *augmentation_on* and *augmentation_off*, (B) *volume_up* and *volume_down* (C) *add_source* and *remove_source*, (D) *focus_on_source* and *listen_from_here*, (E) *ANC_up* and *ANC_down*, (F) *frequency_higher* and *frequency_lower*. (explanations see Table 4.1 Referents in a group are either actions that have an opposite effect (volume_up - volume_down) or are related on some level (focus_on_source - listen_from_here). For each action, we have created a short video to demonstrate its effect. Each video is edited with Adobe Photoshop, Premiere Pro, and Audition and shows a modification of these five sound sources according to the action shown (see Figure 5.1).

**Figure 5.1:** This figure shows screenshots of the footage we created. Each of these sources produces sounds, but only those highlighted are audible to the user. The brightness of this mark indicates the volume of the sound source.

## 5.3 Modalities

Besides **mid-air gestures**, which can be performed with the whole body, we searched for user commands for two additional control modalities.
We were looking for **touch gestures** for the small display of a smartwatch. In this case, the user can only operate with one hand, but several fingers, the whole hand, and motions that can be tracked by sensors. Furthermore, the participants had to define **key assignments** for headphones, precisely for a built-in control element that can be found on almost every Bluetooth headphones. They usually have three keys: plus (+), minus (-) and a multi-function key (MF).

## 5.4 Method

We have two independent variables in our study. First, the control modality (mid-air gestures, touch gestures, and key assignments) and second, the referents (see Section5.2). The participants' proposals depend on this, so the dependent variable is the proposed gesture. Since each of our participants attends all conditions, i.e., command finding for all actions and all modalities, it is a within-group design. In addition to the list and videos of the referent, our apparatus includes a laptop to play the videos. We also used a dictating device for audio recording and a camcorder with a tripod for video recording.

## 5.5 Procedure

We briefed the participants about the goals of the experiment and the associated risks. We also informed them about their privacy and the right to withdraw their participation and data at any point. If the participants agreed to take part in the experiment, they read and signed a consent form and filled a demographics questionnaire. First, we showed the subjects short videos of all twelve actions and asked if he understood not only the video but also the demonstrated action and its effect. We asked them to suggest gestures for these actions for the first modality and evaluate each suggestion according to intuitiveness, the simplicity of performance and social acceptability. If they wanted to see the corresponding video again, they were allowed to do so. For the two control modalities that require devices (smartwatch gestures and key assignment), exemplary equipment was distributed to the participants so that they could familiarize themselves with it. After they repeated this for the remaining two modalities, we conducted a short final interview with the participants. We asked them to name actions they missed, would use or not use and which of the three control options they prefer. We also wanted to know if the participants could imagine using such a system himself and which modality they would prefer for the control. The study took approximately one hour per participant. In the end, we paid each participant a compensation of 10 EUR.

**Figure 5.2:** Exemplary representation of the three modalities: mid-air gestures (top left), key assignments (top right) and touch gestures (bottom).

## 5.6 Findings

The participants proposed a total of 341 unique mid-air gestures, 295 unique touch gestures, and 262 unique key assignments. Below, these 898 gestures are classified using a taxonomy. Furthermore, we calculate the consensus between the participants and report on their statements regarding the different modalities, actions, and Augmented Hearing.

### 5.6.1 Gesture Taxonomy

We classified the mid-air gestures and touch gestures based on existing taxonomies. We used the dimensions *nature*, *flow* and *form* of Wobbrock et al. [10] and added the dimension *complexity* of Ruiz et al. [7]. In addition, added the dimension *interaction* from Bader et al. [8] with an additional category *whole hand* We considered *form* only for the mid-air gestures and *interaction* only for the touch gestures. Since the key assignments for the headphone control are non-gestic, the taxonomy is not applicable here. We have classified the assignments according to (1) the number of buttons used, (2) whether the buttons used were briefly pressed or held down, (3) whether the buttons were pressed one after the other or simultaneously and (4) whether pressed once or several times. Using the *nature* dimension, we can understand how the gesture was selected. *Symbolic* gestures represent a symbol; for example, an X is drawn on the display to remove a selected source from the soundscape. While *metaphorical* gestures have figurative characteristics, *abstract* gestures are chosen without a meaningful context.

The dimension *flow* for mid-air and touch gestures indicates whether the desired effect already occurs during the execution of the gesture (*continuous*) or afterwards (*discrete*). An example of a continuous gesture would be adjusting the volume by slowly pinching two fingers in the air or on display. Switching the augmentation on or off with a clap, on the other hand, is a discrete gesture. When defining the key mappings, the *flow* indicates whether the buttons are pressed *simultaneously* or *sequentially*. This specification is *not needed* if a button is pressed only once.

The *complexity* of a gesture is either *simple*, such as a tap on the display, or *compound*, when two or more simple gestures are combined.

Mid-air gestures can vary in form. If the thumb is held down to reduce the volume until it reaches the desired level, it is a *static pose*. A *dynamic pose* is changing the volume by turning an imaginary knob (counter)clockwise. When pointing to a sound source in order to add it to the soundscape, the gesture consists of a *static pose and path*. If something else is performed afterward, such as clenching the hand to a fist, this gesture has the form *dynamic pose and path*.

The *interaction* dimension of the headphone buttons indicates whether the buttons used are only pressed once or several times. For touch gestures, the dimension *interaction* states the number of fingers used. Either one or multi fingers are needed. We needed to add the *whole hand* category for gestures performed with the palm, not the fingers.

The *button count* is the number of different buttons used for the mapping, which in our case can be a maximum of three (plus (+), minus (-) and multi-function (MF)).

*Mode* determines how the keys are used. This is either a short *press*, a long *hold* or a combination of both.

For each dimension, we calculated the relative occurrence of each category as shown in Figure 5.3.

| General | | |
|---|---|---|
| Nature | abstract | Gesture is arbitrarily chosen |
| | metaphorical | Gesture is based on a metaphor |
| | symbolic | Gesture represents a symbol |
| | physical | Gesture is a manipulation on the object |
| Flow | discrete | Effect is triggered after gesture |
| | continuous | Effect is triggered during gesture |
| Complexity | simple | Gesture consists only of a single action |
| | compund | Gesture consists of several simple gestures |

| Gesture | | |
|---|---|---|
| Form | static pose | The pose remains unchanged during the gesture |
| | dynamic pose | The pose will be changed during the gesture |
| | static pose and path | The pose is not changed during the gesture, but its position will |
| | dynamic pose and path | Both the pose and its position will be changed during the gesture |

| Smartwatch | | |
|---|---|---|
| Interaction | one-finger | Gesture can be performed with one finger |
| | multi-finger | Gesture can be performed with two or more fingers |
| | without touch | Gesture can be performed without touching the display |
| | whole hand | Gesture can be performed with the entire hand |

| Headphone keys | | |
|---|---|---|
| button count | zero | no button is used |
| | one | one button is used |
| | two | two different buttons are used |
| | three | three different buttons are used |
| flow | simultaneously | the buttons are used at the same time |
| | sequentially | the buttons are used one after another |
| | not needed | no definition of flow is possible |
| mean | press | the button(s) are only pressed briefly |
| | hold | the button(s) are held for some time |
| | combined | the button(s) are both pressed and hold. |
| interaction | once | the buttons(s) are used one time |
| | repeatedly | the button(s) are used several times |

**Table 5.2:** Gesture taxonomy for mid-air gestures and touch gestures for smartwatch. Classification of key mappings.

### Taxonomy breakdown

| | | mid-air gestures | touch gestures |
|---|---|---|---|
| nature | abstract | 0.27 | 0.39 |
| | metaphorical | 0.38 | 0.46 |
| | physical | 0.31 | 0.11 |
| | symbolic | 0.04 | 0.04 |
| flow | continuous | 0.41 | 0.40 |
| | dicrete | 0.59 | 0.60 |
| complexity | compound | 0.31 | 0.30 |
| | simple | 0.69 | 0.70 |
| form | dynamic pose | 0.13 | |
| | static pose | 0.05 | |
| | dynamic pose and path | 0.31 | |
| | static pose and path | 0.51 | |
| interaction | one-finger | | 0.58 |
| | multi-finger | | 0.24 |
| | without touch | | 0.09 |
| | hand | | 0.09 |

| | | key assignments |
|---|---|---|
| buttons | zero | 0.02 |
| | one | 0.47 |
| | two | 0.49 |
| | three | 0.02 |
| flow | sequentially | 0.21 |
| | simultaneously | 0.30 |
| | not needed | 0.49 |
| mode | press | 0.50 |
| | hold | 0.26 |
| | combined | 0.24 |
| interaction | once | 0.59 |
| | repeatedly | 0.41 |

**Figure 5.3:** Taxonomy breakdown

As one can see in figure 5.3 mid-air and touch gestures scored same for *flow* and *complexity*. About 60% of proposed gestures are discrete, so the desired effect occurs after their execution. Nearly 70% of gestures is only one movement. The rest consists of several movements. The participants proposed metaphorical gestures most frequently and symbolic gestures least frequently, although with different proportions per modality. While in mid-air gestures, physical gestures used to be more common than abstract gestures, in touch gestures, it is reversed. With mid-air gestures, the position of the pose changes in over 80% of gestures during execution, but the pose itself stays the same in over 60% of these gestures. Almost 60% of touch gestures are performed with only one finger, a quarter of gestures with two fingers. 10% of the gestures require the use of the whole hand, while another 10% do not require any touch interaction with the display.
For most key assignments, the participants used one or two keys. Only in very few cases none or all three keys were used. The buttons used are pressed only once for 60% of the assignments. In 20% of the cases, the buttons used were pressed sequentially, in 30% at the same time. If only one key is used and pressed only once, information about the flow is not needed. The buttons are pressed briefly for half of the assignments and hold down for 25%. The remaining quarter uses both means combined.

### 5.6.2 User-defined gesture sets

In order to create a gesture set for each modality, we have included the best, i.e., the most frequently mentioned gesture, gesture per referent in the corresponding set. Sometimes one gesture mentioned the most for more than one referent. In those cases, we compared the participants' review for these actions and added the gesture to the set referring to the action that achieved a higher degree of agreement. The only exception is a so-called toggle action, meaning a binary feature that can either be enabled or disabled, in our case switching augmentation on or off. With these actions, it is possible to use the same gesture for both referents. When an action had several equally frequent mentioned gestures, we selected the one which was more consistent concerning the existing set.

#### Evaluation

A personal review should accompany each gesture proposed by a participant. Therefore, they should rate on a scale from 1 to 7 (1) the intuitiveness of the gesture, (2) how easy it is to perform, and (3) how suitable it is for the public. In the following Table 5.4, we will review how the gestures of the three sets were rated on average by the participants.

***Intuitiveness*** The evaluation of intuitiveness of the gestures, with respect to the action, is very consistent between the sets, but within the sets it is variable. The intuitiveness of a gesture does, therefore, not depend on the input method but the respective action. The participants rated the frequency-changing actions least intuitively for each modality. The most intuitive features are the change of volume and the switching on and off of the augmentation.

***Simpleness*** Both within a set and between the sets, there are no significant differences in the simplicity of the gestures.

***Acceptance*** Compared to the other two modalities, the proposals of the mid-air gestures have been rated significantly less suitable for the public. Especially problematic are the gestures of the actions, which focus on the manipulation of selected sound sources, for example, removing or adding sources. Within the sets of the other two modalities, there are no significant differences in acceptance between the gestures.

| | mid-air gestures | key assignments | touch gestures |
|---|---|---|---|
| augmentation_off | Tap your ear with one hand | Hold MF | Double-tap with one finger° |
| augmentation_on | Tap your ear with one hand | Hold MF | Double-tap with one finger° |
| volume_down | Turn imaginary knob clockwise with one hand | Adjust gradually with - | Swipe down with one finger |
| volume_up | Turn imaginary knob counterclockwise with one hand | Adjust gradually with + | Swipe up with one finger |
| add_source | Point to the source you want to add | Look at the source you want to add while holding + | Look at the source you want to add and swipe to the right |
| remove_source | Point to the source you want to remove then wipe left△ | Look at the source you want to remove while holding - | Look at the source you want to remove and swipe to the left |
| focus_on_source | Point to the source you want to focus on for a long time° | Look at the source you want to focus on and click MF twice | Look at the source you want to focus on and hold with one finger |
| listen_from_here | Circle the area with one hand | Gaze roams the area while holding MF | Gaze roams the area while holding one finger on the display△ |
| anc_down | Raise one arm in front of your body, palm points up | Hold + until NC is as desired | Turn arm 90° away from body |
| anc_up | Lower one arm in front of your body, palm points down | Hold - until NC is as desired | Turn arm 90° towards body△ |
| frequency_lower | Flick then lower one arm in front of your body, palm points down° | Click - twice | Circle the edge of the display counterclockwise with one finger until frequency is as desired |
| frequency_higher | Flick then raise one arm in front of your body, palm points up° | Click + twice | Circle the edge of the display clockwise with one finger until frequency is as desired |

**Table 5.3:** User-defined gesture sets for all three modalities, based on findings of the interview.
°second best gesture, since the superior one is used for another referent (with higher agreement)
△ between equally favored gestures, the one leading to a consistent set was chosen

| | **Mid-air gestures** | | | | | |
|---|---|---|---|---|---|---|
| | Intuitiveness | | Simpleness | | Acceptance | |
| | mean | SD | mean | SD | mean | SD |
| augmentation_off | 6.00 | 1.22 | 7.00 | 0.00 | 6.50 | 0.50 |
| augmentation_on | 6.20 | 1.17 | 7.00 | 0.00 | 6.60 | 0.49 |
| volume_down | 6.57 | 0.73 | 5.86 | 1.25 | 6.43 | 0.73 |
| volume_up | 6.57 | 0.73 | 5.86 | 1.25 | 6.42 | 0.73 |
| add_source | 5.67 | 0.75 | 6.84 | 0.37 | 5.17 | 1.34 |
| remove_source | 5.50 | 0.50 | 7.00 | 0.00 | 5.00 | 1.00 |
| focus_on_source | 5.00 | 1.51 | 5.50 | 1.08 | 4.00 | 1.53 |
| listen_from_here | 6.00 | 0.94 | 6.67 | 0.67 | 4.67 | 1.41 |
| anc_down | 6.00 | 0.00 | 6.00 | 1.00 | 6.50 | 0.50 |
| anc_up | 5.00 | 1.00 | 7.00 | 0.00 | 5.50 | 1.50 |
| frequency_lower | 3.50 | 0.50 | 7.00 | 0.00 | 5.00 | 0.00 |
| frequency_higher | 3.50 | 0.50 | 7.00 | 0.00 | 5.00 | 0.00 |
| | **Touch gestures** | | | | | |
| augmentation_off | 6.00 | 1.55 | 6.80 | 0.40 | 6.60 | 0.80 |
| augmentation_on | 6.00 | 1.55 | 6.80 | 0.40 | 6.60 | 0.80 |
| volume_down | 6.60 | 0.80 | 6.60 | 0.49 | 7.00 | 0.00 |
| volume_up | 6.60 | 0.80 | 6.60 | 0.49 | 7.00 | 0.00 |
| add_source | 5.75 | 0.83 | 7.00 | 0.00 | 7.00 | 0.00 |
| remove_source | 5.75 | 0.83 | 7.00 | 0.00 | 7.00 | 0.00 |
| focus_on_source | 5.34 | 0.47 | 7.00 | 0.00 | 6.34 | 0.94 |
| listen_from_here | 5.40 | 1.02 | 6.40 | 1.17 | 6.60 | 0.80 |
| anc_down | 5.00 | 1.41 | 6.34 | 0.94 | 7.00 | 0.00 |
| anc_up | 5.00 | 1.41 | 6.34 | 0.94 | 7.00 | 0.00 |
| frequency_lower | 4.75 | 0.83 | 6.00 | 1.22 | 7.00 | 0.00 |
| frequency_higher | 4.75 | 0.83 | 6.00 | 1.22 | 7.00 | 0.00 |
| | **Key assignments** | | | | | |
| augmentation_off | 6.64 | 0.64 | 7.00 | 0.00 | 7.00 | 0.00 |
| augmentation_on | 6.64 | 0.64 | 7.00 | 0.00 | 7.00 | 0.00 |
| volume_down | 6.94 | 0.23 | 6.83 | 0.69 | 7.00 | 0.00 |
| volume_up | 6.94 | 0.23 | 6.83 | 0.69 | 7.00 | 0.00 |
| add_source | 5.00 | 2.08 | 6.67 | 0.47 | 6.83 | 0.37 |
| remove_source | 5.00 | 2.08 | 6.67 | 0.47 | 6.83 | 0.37 |
| focus_on_source | 5.22 | 0.40 | 6.60 | 0.49 | 6.60 | 0.49 |
| listen_from_here | 4.20 | 2.64 | 5.60 | 1.02 | 6.20 | 1.17 |
| anc_down | 4.75 | 1.09 | 6.50 | 0.50 | 7.00 | 0.00 |
| anc_up | 4.75 | 1.09 | 6.50 | 0.50 | 7.00 | 0.00 |
| frequency_lower | 3.75 | 0.83 | 6.75 | 0.43 | 5.75 | 1.64 |
| frequency_higher | 3.75 | 0.83 | 6.75 | 0.43 | 5.75 | 1.64 |

**Table 5.4:** Intuitiveness, Simpleness of execution and Suitability for the public of our key assignments set

### 5.6.3 Agreement Rate

We calculated the agreement rate for each referent, according to the methodology of Wobbrock et al.[1, 2] to determine the consensus between the participants. The agreement rate can take a value between 0 (no agreement) and 1 (total agreement) and is calculated as follows.

$$AR(r) = \frac{|P|}{|P|-1} \sum_{P_i \subseteq P} \left(\frac{|P_i|}{|P|}\right)^2 - \frac{1}{|P|-1}$$

$|P|$ is the total number of gestures proposed for referent $r$. $P_i$ refers to a set of identical proposals for r. $|P_i|$ is the size of this set. In other words, how often this specific gesture was proposed for referent r.

As one can see in figure 5.4, our agreement scores range from .01 to .14 for mid-air gestures, from .03 to 0.1 for touch gestures and from .04 to 0.62 for key mappings. The average agreement rate is .06 (mid-air gestures), .06 (touch gestures), and .19 (key mappings). The highest rates were achieved by *volume down/up* (mid-air gesture, key assignment) and *listen from here* (touch gestures). The actions *remove source* (mid-air gestures), *ANC up* (touch gestures) and *frequency lower* (key assignment) reached the smallest score.

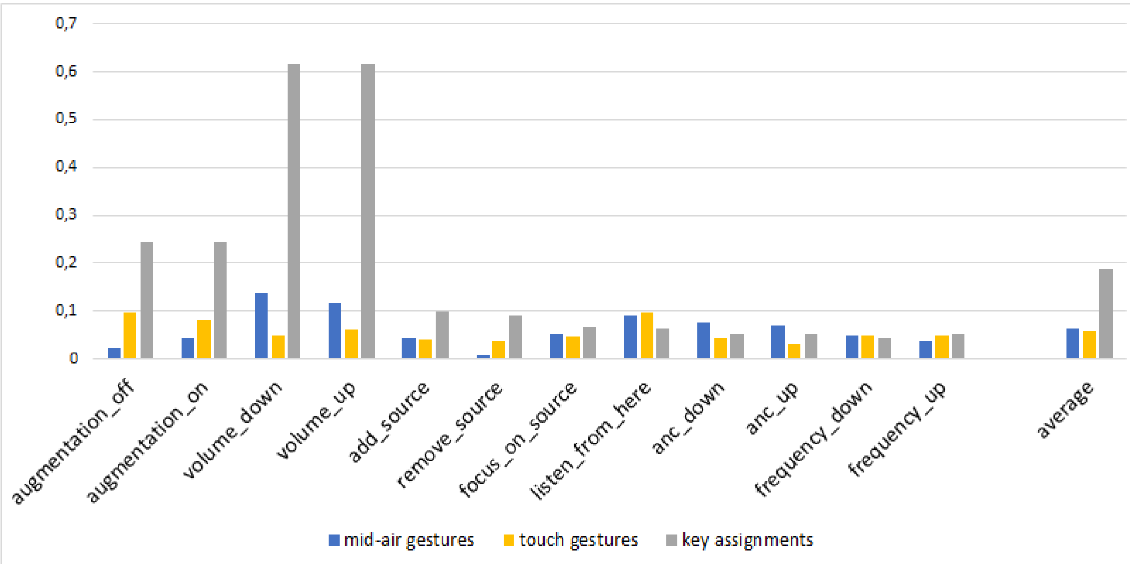



**Figure 5.4:** Overview of the agreement rate for all referents and control modalities

### 5.6.4 User review

As mentioned before, we audio-recorded the thoughts the participants expressed during the session and their answers to the few questions we asked before and after the gestures elicitation. We transcribed the records and used atlas.ti [1] to analyze the data. For this purpose, we made a thematic analysis with open coding, according to Blandford et al. [35]. While coding, we discovered three themes (modality evaluation, feature evaluation, Augmented Hearing evaluation) that we partially divided into subsections.

| **Modality evaluation** | **Feature evaluation** | **Augmented Hearing evaluation** |
|---|---|---|
| ∘ Mid-air gestures | ∘ Unneeded action | ∘ Chances |
| ∘ Touch gestures | ∘ Missing Features | ∘ Concerns |
| ∘ Key assignments | | |

**Table 5.5:** Coding Themes with Subsections

Each of the following quotations has been translated from German into English.

#### Modality evaluation

**Mid-air gestures** The mid-air gestures are the most disliked option; eleven participants rated it as the most unsuitable modality. The main criticism is that most users would not feel comfortable using mid-air gestures in public. Besides, gestures are too tiring for the participants compared the other modalities.

> *"I think gestures are really elaborate."* (P04, mid-air gestures)

> *"Gestures, in general, have the problem, it looks a bit strange when you gesticulate in the air. That's why I chose discrete gestures."* (P10, mid-air gestures)

Although the majority of participants consider the gestures unsuitable, many participants perceive the modality itself is intuitive for the application.

> *"It looks terrible, but it's intuitive."* (P05, mid-air gestures)

***Touch gestures*** Twelve participants named touch gestures as their favorite, six as second best and two as least choice. They like the visual feedback that a display can provide and that touch gestures are suitable for the public. The execution of touch gestures is also easy for users, as devices with displays such as smartphones are commonplace.

> *"I think the Smartwatch is great, good for public use as well"* (P11, touch gestures)

> *"I think it's better if I have something visual."* (P13, touch gestures)

[1] https://atlasti.com/

The participants furthermore see potential in extending and improving this modality. The additional use of an interface so the Augmented Hearing System can be controlled without gestures is considered valuable by most participants.

> *"(...) an interface complements the gestures. So I don't only have the interface or the gestures, but both."* (P13, touch gestures)

> *"You would be so fast in a menu, without having to do anything time-consuming, that it is not worth it for me to make gestures."* (P04, touch gestures)

***Key assignments*** Seven participants considered the headphones to be the least, nine to be the second and only four to be the first choice. While the headphone buttons were judged quite intuitively, easy to use, and suitable for the public, many felt very limited by the number of buttons. Almost all participants stated that it was not only difficult to find assignments covering all twelve actions, but that the time and effort required to learn is too high.

> *"I felt there were too many functions for three buttons"* (P02, key assignments)

> *"The learning effort is relatively high because it is not intuitive."* (P11, key assignments)

> *"I thought the headphone was actually quite cool, it's also subtle, so it's very suitable for the public."* (P12, key assignments)

#### Feature evaluation

***Unneeded action*** Except for two participants, everyone stated they would not need to change the frequency. While a few consider these actions as entirely unnecessary, the majority were able to find use cases but believed that they were not relevant.

> *"I wouldn't use the frequency, no, because objects have their frequency and that would just be surreal to me."* (P08, final interview)

> *"When you think of it as a hearing aid, it is important, of course, because the older you are, the harder it is to hear high frequencies. (...) I personally wouldn't use it (yet). Someday I'll get old, and then I'll need it."* (P17, final interview)

All other actions were perceived as rather useful by the participants.

***Missing features*** Almost all participants stated that the actions available are sufficient. Participant P04 stated that he felt that the gradual adjustment of frequency, volume, and NC was unnecessary and would prefer pre-selected levels (off, low, medium, high). Three participants were unsure whether they would use*listen from here*.

> *"I don't think the area is necessary, because I can add individual sources."* (P13, final interview)

Four missed the opportunity to undo the focusing of a source. Several participants wish to preset profiles, e.g., work, in order to avoid manipulating the same sound sources each time but to ensure that this happens automatically.

> *"I'd like to have the option of setting up some kind of template (...) so I can say I'm at work.* (P11, final interview)

Individual participants wished for more in-depth possibilities, such as a connection to the internet, a recording function, voice control, or translation.

#### Augmented Hearing

***Chances*** Out of twenty participants twelve are interested in using augmented hearing. Most subjects can imagine using such a system mainly at work.

> *"It makes the most sense for working."* (P16, final interview)

> *"Yeah, I think so, that would be kind of cool."* (P18, final interview)

***Concerns*** Three participants see no necessity in this technology, and five are unsure.

> *"I don't know. Personally, I just think that's nonsense."* (P13, final interview)

Participants critical about the usage shared concerns about privacy and social life.

> *"I think once you can add and remove sources, it would be bad for social life, so this (...) shouldn't exist"* (P04, final interview)

> *"That's just a strong separation from the environment."* (P10, final interview)

# 6 Validation Study

To validate our gesture sets, we finally conducted a large-scale online study. Participants should compare the four best suggestions for each action and modality and rank them according to their preferences. Besides, we asked them to evaluate the modalities and actions, as well as their suggestions. In the end, the participants assessed whether and why they would use Augmented Hearing or not.

## 6.1 Preparation

For the validation of the previous study, we launched an extensive online survey. For each action, we selected the four most frequently mentioned gestures. If fewer gestures then four were mentioned several times in context with the action, we considered only these gestures. For cases with more than four gestures, we reviewed the participants' ratings and included only the four best-rated gestures in the survey. After we selected gestures for all actions, we created an animated GIF for each gesture and a survey with Qualtrics[1].

## 6.2 Participants

Since Qualtrics surveys can be linked to Amazon Mechanical Turk[2], we used this platform to recruit participants. A total of 155 participants took part in the study, but we evaluated only the responses from those 94 participants who completed the survey. Forty-one identify themselves as female; one participant preferred not to specify his gender. The participants' age ranged from 21 to 57 years, with an average of 33.07 years (SD = 8.13). Approximately 78% of the participants are employed, 83% full-time, and the rest part-time. 14% are unemployed, 7% are still in education, and one participant is unable to work. The participants came from 10 different countries, most of them from North America (USA: 42, Canada:10) or Europe (UK: 17, Italy: 12, Germany: 4, Poland: 3, France: 2). Two more participants are from India, one each from Australia and Bahrain.

[1]https://www.qualtrics.com
[2]https://www.mturk.com/

## 6.3 Procedure

Initially, the participants had to agree to participate. They were informed about the purpose of the study, the data collected, and their compensation for participation. If a participant did not agree, the survey terminated. After the participants gave their consent, they were first requested to provide demographic information, such as age and gender. After a short introduction in Augmented Hearing, supported by a video, and the participants' task, the primary survey began. We showed a short video per referent to demonstrate the desired effect. Afterward, the participants should rank the selected gestures for each modality and the modalities as well. We randomized not only the order of the actions but also the modality order due to avoid learn-ability and fatigue effects.
At the end of the survey, we asked the participants for overall valuations. They had to evaluate the relevance of each action on a scale from 1 (not important at all) to 5 (very important). The participants were also able to suggest action(s) they were missing in an open text question. Besides, they should rank the three modalities and give reasons for their choice in an open text. The last two questions allowed the participants to say whether they wanted to use Augmented Hearing or not and to explain their decision in an open text question.

## 6.4 Findings

In order to validate the gesture sets, we developed the favorite gestures of the online study and compared them with the sets. We also summarized the evaluation of the actions and modalities, as well as the reviews of augmented hearing.

### 6.4.1 Gesture set validation

To verify the gesture set from 5.3, we presented the participants the most frequently mentioned gestures per referent and for each modality. As already explained, they established a ranking of these gestures for us to find out which gesture satisfies most users. Since we do not want to create a gesture set unlike the elicitation study bu to validate the determined sets, the selection of the most chosen gesture is sufficient, and the same gesture can be used for several actions. In Table 6.1, we have listed which gesture was most frequently mentioned for which action and modality. For clarification, we highlighted the gesture in green if it matches the one from the interviews. Mid-air gestures have the most matching gestures (6). The key assignments have five matching gestures, and the touch gestures have the fewest matches (4).

| | mid-air gestures | key assignments | touch gestures |
|---|---|---|---|
| augmentation_off | Tap your ear with one hand | Hold MF | Cover the display with your entire hand until augmentation is disabled |
| augmentation_on | Flick once | Hold MF | Double-tap with one finger |
| volume_down | Lower arm in front of your body, palm points down | Adjust gradually with - | Drag 1 finger downwards until volume is as desired |
| volume_up | Raise one arm in front of your body, palm points down | Adjust gradually with + | Drag 1 finger upwards until volume is as desired |
| add_source | Point to the source you want to add | Look at the source you want to add while holding MF | Look at the source you want to add and tap once |
| remove_source | Point to the source you want to remove then wipe left | Double-click - | Look at the source you want to remove and swipe to the left |
| focus_on_source | Point to source you want to focus on for a long time | Look at the source you want to focus on and hold MF for a few seconds | Look at the source you want to focus while holding 2 fingers in the middle of the display |
| listen_from_here | Circle the area with one hand | Gaze roams the area while holding + | Gaze roams the area while holding 1 finger on the display |
| anc_down | Turn imaginary knob backwards with one hand in front of your ear | Hold MF while adjusting NC gradually with + | Tap in the middle of the display and circle clockwise at its edge until NC is as desired |
| anc_up | Lower arm in front of body, palm points down | Hold - until NC is as you desired | Drag one finger down until NC is as desired |
| frequency_lower | Flick then lower arm in front of your body, palm points down | Hold - until the frequency is as desired | Circle the edge of the display counterclockwise with 1 finger until frequency is as desired |
| frequency_higher | Turn imaginary knob clockwise with one hand | Click MF twice then adjust gradually with + | Tap in the middle of the display and circle clockwise at its edge until the frequency is as desired |

**Table 6.1:** Overview of actions rated according to the importance for participants. Cells highlighted in green indicate the most frequently selected gesture per referent.

### 6.4.2 Relevance

On a scale from 1 (not important at all) to 5 (very important), each participant had to assess how relevant each action is to him. Although participants were able to make recommendations for additional features in an open text question, they would like to see implemented in the system no participant made any suggestions.

| | 1 | 2 | 3 | 4 | 5 | median |
|---|---|---|---|---|---|---|
| augmentation_off | 10.11% | 13.48% | 28.09% | 25.84% | 22.47% | 3 |
| augmentation_on | 7.95% | 9.09% | 30.68% | 29.55% | 22.73% | 4 |
| volume_down | 5.62% | 4.49% | 8.99% | 23.60% | 57.30% | 5 |
| volume_up | 6.74% | 2.25% | 11.24% | 24.72% | 55.06% | 5 |
| add_source | 5.62% | 17.98% | 21.35% | 39.33% | 15.73% | 3 |
| remove_source | 4.49% | 8.99% | 23.60% | 40.45% | 22.47% | 4 |
| focus_on_source | 3.37% | 15.73% | 23.60% | 34.83% | 22.47% | 4 |
| listen_from_here | 6.98% | 19.77% | 30.23% | 29.07% | 13.95% | 3 |
| ANC_down | 6.74% | 20.22% | 43.82% | 16.85% | 12.36% | 3 |
| ANC_up | 6.74% | 20.22% | 42.70% | 19.10% | 11.24% | 3 |
| frequency_lower | 19.10% | 24.72% | 30.34% | 14.61% | 11.24% | 3 |
| frequency_higher | 21.59% | 29.55% | 23.86% | 19.32% | 5.68% | 2 |

**Table 6.2:** Overview of actions rated according to the importance for participants. Cells highlighted in green indicate the most frequently selected score per referent.

### 6.4.3 Preferred modality

The participants were asked to specify which modality of control they prefer for each referent. They preferably performed six of the twelve actions with the use of a smartwatch, three with the headphone buttons and only two by gestures. In one action no favorite among the modalities could be established. This result is approximately a reflection of the overall evaluation of the modalities at the end of the questionnaire. The participants prefer touch gestures, followed by mid-air gestures and finally, headphones.

> *"Touch gestures on a watch are subtle and easy to do."* (P04, online survey)

> *"It is most effortless (in contrast to gestures) and precise (in contrast to headphones' buttons) at the same time."* (P89, online survey)

Many think mid-air gestures are suitable for the application but do not feel comfortable doing them in public.

> *"Gestures have a lot of potential, but I feel they would look quite silly from a third party."* (P05, online survey)

> *"I feel like it would seem more natural to be able to gesture"* (P70, online survey)

*"I feel uncomfortable doing big gestures in the middle of the street."* (P13, online survey)

The participants ranked the key assignments as the least option in general.

*"Because some things were just too hard to do with headphone buttons."* (P11, online survey)

*"The headphone option seems like it would be inconvenient for more cases because of the location of controls"* (P05, online survey)

### 6.4.4 Augmented Hearing

As the participants of the elicitation study, we also asked the online survey respondents whether they would use augmented hearing and why. Accordingly, 37% of the participants would use augmented hearing, 10% said they would not use it, and the rest is unsure. The ones who are willing to use it proposed reasonable applications, such as removing unwanted noises, safety aspects, and health benefits.

*"It lets me concentrate on the things I want to hear whilst also cutting out external noise"* (P88, online study)

*"I sometimes want to adjust how I'm hearing outside noises relatively to my headphones for safety."* (P71, online survey)

*"It allows you to (...) be cut [out] from noise pollution"* (P59, online survey)

The participants, who are uncertain about the use, often struggled to imagine the technical implementation.

*"I am not sure how exactly would it work so I'd like to try it and then decide or use cases"* (P80, online study)

*"I can't think of a specific situation where I might use it, but I wouldn't be opposed to using it either"* (P33, online survey)

Those who say they are unwilling to use augmented hearing cannot see any value in using such systems

*"I don't think that it's something really necessary"* (P72, online survey)

or they simply do not want to use the features.

*"I only want to hear the real world."* (P68, online survey)

# 7 Discussion

## 7.1 Modalities and their gesture sets

In the following, we discuss how our results can be taken into account in the development of Augmented Hearing systems.

#### Implications for gesture design

The accordance between study results (see Table 6.1) is not sufficient to verify the gesture sets based on the elicitation. It follows that these should not be used as a standard set for augmented hearing systems, but rather as an inspiration for the design. The different application contexts and locations had an essential impact on the choice and rating of gestures among the participants of both studies. These different needs may be one reason why we did not identify a valid set for each modality. Another problem could be the difficulties many participants faced when trying to imagine the technical implementation of the systems. Morris et al. [36] addressed how legacy bias affects the results of elicitation studies. According to them, participants do not always suggest the gestures they personally prefer, but gestures inspired by metaphors or other external influences. We followed their approach by asking our participants to propose more than one gesture per referent (procedure) and to show the different ways to utilize each modality(priming). In contrast to earlier work [9, 12, 37], our gestures are mostly metaphorical and not physical. This could be an indication that our elicitation study is limited by legacy bias since the participants seemed to need metaphors to come up with suitable gestures. The calculation of the agreement rate showed that the key assignments on average reached the highest consensus between the participants. The two referents *volume_down* and *volume_up* achieved the highest rates by far. The actions affecting the volume were also rated most intuitively, as users are already familiar with these settings from other devices or use cases. This is another indication that metaphors or already known control mechanisms affected the participants' suggestions. Overall, we achieved rather low agreement scores, but in the same range as other studies [11, 12]. We believe legacy bias and low agreement rates are due to the fact that we carried out the elicitation study without prototypes or existing technologies. Our participants faced problems to imagine the technical implementation, have hardly any experience with augmented reality and never experienced augmented hearing. Especially when adjusting the volume or removing and adding sound sources, many participants of the elicitation study searched for applicable metaphors in the music and computer context. Respondents to the online study only have to indicate which gesture is most suitable, without having to recognize or find metaphors leading to different preferences in the selection of the gesture. On the other hand, the majority of our actions are not related to the manipulation of individual objects, as known from other studies.

Only four of twelve actions (*remove_source*, *add_source*, *focus_on_source*, *listen_from_here*) applied to objects, in this case, sound sources. The other functions are changing the whole sound environment (volume, frequency, ANC) or switching the augmentation on and off. It is therefore not surprising that the participants proposed fewer physical gestures in relation to other studies.

### 7.1.1 Implications for control modalities

Among the three modalities, most participants found touch gestures to be the most appropriate. Several stated by using the smartphone on a daily basis, they are used to this kind of control and therefore perceived touch gestures as most intuitive and most comfortable to perform. During the elicitation study, many people suggested using an (additional) interface. Gestures for functions such as switching augmentation on and off, or changing the frequency, are unnecessary for many participants, as in these cases accessing a menu would be faster. A graphical user interface, in addition to the gestures, is recommended when implementing an augmented hearing system. Both the interface and the gestures should be available with a smartphone because the participants criticized that an additional device is needed. Compared with the other two modalities, the right execution of a touch gesture is the easiest for users to monitor. They often worry about being misinterpreted when gesturing, and they are not able to see the buttons on the headphones. Touch gestures have to be easy to perform with as few fingers as possible, as users prefer simple movements with one or two fingers. For mid-air gestures, they also favor simple small movements in order to reduce the effort. The majority perceived the gestures as inadequate for the public. Although they were judged to be quite intuitive and easy to perform, since most subjects would not feel comfortable performing these gestures anywhere, they do not seem suitable for such a system. Earlier work [38, 39] observed that gestures are classified as socially acceptable if bystanders can understand their intention. We believe that among other issues, here lies the main problem. The participants considered the gestures not intuitive and had difficulties imagining gestures suitable for the actions. So they possibly assume that external or unrelated bystanders are not able to interpret the purpose of the gesture. Since most participants wanting to use it for work rather than for their daily life, gesture-based control is still an option for applications in the office or outside the public, as gestures already exist for similar use cases [6, 13]. Since all previous elicitation studies only collected gestures, touch gestures, and speech commands, we have no reference material for the key assignments. This input modality is also application-specific and therefore can hardly be used in other contexts. Many of the participants believed that the use of the control element makes sense because it does not cause any interference if it is attached to the headphones. Nonetheless, there was criticism that three buttons were not enough to cover so many actions. Such a control element works only as an additional input device. For simple and frequently used functions, such as changing the volume, the buttons can be used to offer users the best possible control modality. Since the favored modality depends on the referent and no modality satisfies all participants, it is worth considering multi-modal input.

## 7.2 Functionality

In addition to the actions that potential users want, our work also presents in-depth features for Augmented Hearing systems.

### 7.2.1 Actions

The actions presented were, in general, described as useful and sufficient. Only frequency-changing seems to be the function that most people would not use. A few participants mentioned one could change the frequency of an unpleasant voice, but they personally would not do so. Based on table 6.2 increasing the frequency is more relevant than lowering. During the elicitation study, one participant recognized that lowering frequency is useful for the elderly because they may have problems perceiving higher frequencies. Although we had a 57-year-old participant, the average age was about 23 (gesture elicitation) and 33 (online study). Impairments with the perception of higher frequencies are therefore rather unlikely in our samples, which is why changing the frequency is not (yet) required.
Some participants complained about a missing reset function without needing to shut down the device. Adding a reset function for any settings is therefore useful and should be considered in the implementation.

### 7.2.2 In-depth functions

In addition to adjusting the soundscape using various actions, the participants specified further functional requirements for the system. Among other things, a connection to the internet for retrieving and playing information about the surroundings. The participants also want to add virtual sound, which is already in use with the Sennheiser AMBEO headphones[1]. Besides, they asked for real-time translation feature as known from the Google Pixel Buds[2] in the focus groups, but not taken into account in the elicitation study as it is beyond the scope of adjusting the soundscape. Since this feature also came up in the interviews, it is advisable to consider adding it. Whether there needs to be a gesture for this is questionable, as the participants would not use the function frequently, but would like to have the chance to do so. As already mentioned, users want visual feedback that simplifies usage. Besides a graphical user interface a combination of Augmented Sound and Augmented Vision is also possible. This linkage could also allow modalities not supporting a GUI to provide visual feedback, possibly resulting in even superior control options. During the focus groups as well as during the interviews, almost all participants stated that the system should be intelligent. Meaning they want to be able to make presets so they can select specific scenarios (at work) and the system adapts to it autonomously. Besides, the system should be able to recognize through individual use, which sound sources are desired and which are not.

[1] https://www.youtube.com/watch?v=8j4X-uN83I4
[2] https://store.google.com/product/google_pixel_buds

## 7.3 Concerns

The participants raised privacy concerns during the focus groups as well as during the elicitation study. Especially regarding the possibility of a recording function, it must be ensured that no personal rights will be violated. There has to be a legal arrangement, like with photographing or video recording individuals. A further concern with augmented hearing is the separation from the environment. Although users can decide by themselves if and what kind of sounds they want to suppress, they are not able to prevent being blocked out by someone else. New possibilities arise to isolate people or to isolate oneself, which is why many participants consider computer-aided hearing as surreal and critical. As the participants of Larsen et al. [32] a few of our subjects feared to feel isolated from the environment due to the equipment. Despite this, more than half of the participants would use it or at least consider using it.

# 8 Conclusion and Outlook

In this thesis, we conducted three successive studies concerning augmented hearing. We started with focus groups to explore how users want to customize their sound environment. According to these preferences, we identified twelve actions, and in the following elicitation study, we found user commands to execute them. The participants suggested mid-air gestures, touch gestures on a smartwatch and key assignments for a headphone controller, leading to gesture sets for the different control modalities. We presented a classification of these gestures using a taxonomy to replicate the mental models of potential users. Finally, we wanted to verify our gesture sets with an online survey. We observed some differences between the results of the elicitation study and the survey. As possible causes for those outcomes, we see the different needs based on place and context of usage, as well as the difficulties experienced by the participants in imagining the technical realization. We discussed how these insights could affect the design of control options for augmented hearing, leading to user-friendly control of future AR technologies designed to enhance the hearing experience.

## Outlook

Future work on augmented hearing should identify location and context-based needs of potential users. The usage in leisure time, travel or at work is entirely different and therefore has an influence on the choice of modality as well as on the user commands themselves. Since many participants had difficulties imagining the technical realization and were therefore possibly limited in identifying gestures and key assignments, various prototypes should be developed based on these results. Those prototypes can be evaluated in follow-up user studies aiming to design a user-friendly system. Another possibility would be to explore how personal rights and social coexistence can be maintained.

# Bibliography


[1] R.-D. Vatavu, J. O. Wobbrock. "Formalizing Agreement Analysis for Elicitation Studies: New Measures, Significance Test, and Toolkit". en. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. Seoul, Republic of Korea: ACM Press, 2015, pp. 1325–1334. ISBN: 978-1-4503-3145-6. DOI: 10.1145/2702123.2702223. URL: http://dl.acm.org/citation.cfm?doid=2702123.2702223 (visited on 01/20/2019) (cit. on pp. 11, 13, 32).

[2] R.-D. Vatavu, J. O. Wobbrock. "Between-Subjects Elicitation Studies: Formalization and Tool Support". en. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. Santa Clara, California, USA: ACM Press, 2016, pp. 3390–3402. ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858228. URL: http://dl.acm.org/citation.cfm?doid=2858036.2858228 (visited on 01/20/2019) (cit. on pp. 11, 32).

[3] J. Blomberg, J. Giacomi, A. Mosher, P. Swenton-Wall. "Ethnographic Field Methods and Their Relation to Design". In: *Participatory Design*. CRC Press, Dec. 2017, pp. 123–155. DOI: 10.1201/9780203744338-7 (cit. on p. 13).

[4] M. R. Morris, J. O. Wobbrock, A. D. Wilson. "Understanding Users' Preferences for Surface Gestures". In: *Proceedings of Graphics Interface 2010*. GI '10. event-place: Ottawa, Ontario, Canada. Toronto, Ont., Canada, Canada: Canadian Information Processing Society, 2010, pp. 261–268. ISBN: 978-1-56881-712-5. URL: http://dl.acm.org/citation.cfm?id=1839214.1839260 (visited on 05/09/2019) (cit. on p. 13).

[5] M. Nielsen, M. Störring, T. B. Moeslund, E. Granum. "A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI". en. In: *Gesture-Based Communication in Human-Computer Interaction*. Ed. by A. Camurri, G. Volpe. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2004, pp. 409–420. ISBN: 978-3-540-24598-8 (cit. on p. 13).

[6] R.-D. Vatavu, I.-A. Zaiti. "Leap gestures for TV: insights from an elicitation study". en. In: *Proceedings of the 2014 ACM international conference on Interactive experiences for TV and online video - TVX '14*. Newcastle Upon Tyne, United Kingdom: ACM Press, 2014, pp. 131–138. ISBN: 978-1-4503-2838-8. DOI: 10.1145/2602299.2602316. URL: http://dl.acm.org/citation.cfm?doid=2602299.2602316 (visited on 01/20/2019) (cit. on pp. 13, 44).

[7] J. Ruiz, Y. Li, E. Lank. "User-defined Motion Gestures for Mobile Interaction". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '11. event-place: Vancouver, BC, Canada. New York, NY, USA: ACM, 2011, pp. 197–206. ISBN: 978-1-4503-0228-9. DOI: 10.1145/1978942.1978971. URL: http://doi.acm.org/10.1145/1978942.1978971 (visited on 06/13/2019) (cit. on pp. 13, 26).

[8] P. Bader, A. Voit, H. V. Le, P. Woźniak, N. Henze, A. Schmidt. "WindowWall: Towards Adaptive Buildings with Interactive Windows As Ubiquitous Displays". In: *ACM Trans. Comput.-Hum. Interact.* 26.2 (Mar. 2019), 11:1–11:42. ISSN: 1073-0516. DOI: 10.1145/3310275. URL: http://doi.acm.org/10.1145/3310275 (visited on 05/09/2019) (cit. on pp. 13, 26).

[9] M. L. Wittorf, M. R. Jakobsen. "Eliciting Mid-Air Gestures for Wall-Display Interaction". en. In: *Proceedings of the 9th Nordic Conference on Human-Computer Interaction - NordiCHI '16*. Gothenburg, Sweden: ACM Press, 2016, pp. 1–4. ISBN: 978-1-4503-4763-1. DOI: 10.1145/2971485.2971503. URL: http://dl.acm.org/citation.cfm?doid=2971485.2971503 (visited on 01/20/2019) (cit. on pp. 13, 43).

[10] J. O. Wobbrock, M. R. Morris, A. D. Wilson. "User-defined Gestures for Surface Computing". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '09. New York, NY, USA: ACM, 2009, pp. 1083–1092. ISBN: 978-1-60558-246-7. DOI: 10.1145/1518701.1518866. URL: http://doi.acm.org/10.1145/1518701.1518866 (visited on 01/20/2019) (cit. on pp. 13, 26).

[11] T. Dingler, R. Rzayev, A. S. Shirazi, N. Henze. "Designing Consistent Gestures Across Device Types: Eliciting RSVP Controls for Phone, Watch, and Glasses". en. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. Montreal QC, Canada: ACM Press, 2018, pp. 1–12. ISBN: 978-1-4503-5620-6. DOI: 10.1145/3173574.3173993. URL: http://dl.acm.org/citation.cfm?doid=3173574.3173993 (visited on 01/20/2019) (cit. on pp. 13, 43).

[12] T. Piumsomboon, A. Clark, M. Billinghurst, A. Cockburn. "User-defined Gestures for Augmented Reality". In: *CHI '13 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '13. event-place: Paris, France. New York, NY, USA: ACM, 2013, pp. 955–960. ISBN: 978-1-4503-1952-2. DOI: 10.1145/2468356.2468527. URL: http://doi.acm.org/10.1145/2468356.2468527 (visited on 06/13/2019) (cit. on pp. 13, 43).

[13] S. Voida, M. Podlaseck, R. Kjeldsen, C. Pinhanez. "A study on the manipulation of 2D objects in a projector/camera-based augmented reality environment". en. In: *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '05*. Portland, Oregon, USA: ACM Press, 2005, p. 611. ISBN: 978-1-58113-998-3. DOI: 10.1145/1054972.1055056. URL: http://portal.acm.org/citation.cfm?doid=1054972.1055056 (visited on 07/06/2019) (cit. on pp. 13, 14, 44).

[14] M. Lee, M. Billinghurst. "A Wizard of Oz Study for an AR Multimodal Interface". In: *Proceedings of the 10th International Conference on Multimodal Interfaces*. ICMI '08. event-place: Chania, Crete, Greece. New York, NY, USA: ACM, 2008, pp. 249–256. ISBN: 978-1-60558-198-9. DOI: 10.1145/1452392.1452444. URL: http://doi.acm.org/10.1145/1452392.1452444 (visited on 07/20/2019) (cit. on pp. 13, 14).

[15] K. Lee. "Augmented Reality in Education and Training". en. In: *TechTrends* 56.2 (Mar. 2012), pp. 13–21. ISSN: 1559-7075. DOI: 10.1007/s11528-012-0559-3. URL: https://doi.org/10.1007/s11528-012-0559-3 (visited on 07/21/2019) (cit. on p. 14).

[16] P. Milgram, H. Takemura, A. Utsumi, F. Kishino. "Augmented reality: a class of displays on the reality-virtuality continuum". In: *Telemanipulator and Telepresence Technologies*. Vol. 2351. International Society for Optics and Photonics, Dec. 1995, pp. 282–292. DOI:

10.1117/12.197321. URL: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/2351/0000/Augmented-reality--a-class-of-displays-on-the-reality/10.1117/12.197321.short (visited on 07/19/2019) (cit. on p. 14).

[17] R. V. Krevelen. *Augmented Reality: Technologies, Applications, and Limitations*. en. 2007. DOI: 10.13140/rg.2.1.1874.7929. URL: http://rgdoi.net/10.13140/RG.2.1.1874.7929 (visited on 07/21/2019) (cit. on p. 14).

[18] J. Jost, T. Kirks, B. Mättig, A. Sinsel, T. U. Trapp. "Der Mensch in der Industrie – Innovative Unterstützung durch Augmented Reality". de. In: *Handbuch Industrie 4.0 Bd.1: Produktion*. Ed. by B. Vogel-Heuser, T. Bauernhansl, M. ten Hompel. Springer Reference Technik. Berlin, Heidelberg: Springer Berlin Heidelberg, 2017, pp. 153–174. ISBN: 978-3-662-45279-0. DOI: 10.1007/978-3-662-45279-0_86. URL: https://doi.org/10.1007/978-3-662-45279-0_86 (visited on 05/09/2019) (cit. on p. 14).

[19] M. Funk, T. Kosch, A. Schmidt. "Interactive Worker Assistance: Comparing the Effects of In-situ Projection, Head-mounted Displays, Tablet, and Paper Instructions". In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. UbiComp '16. event-place: Heidelberg, Germany. New York, NY, USA: ACM, 2016, pp. 934–939. ISBN: 978-1-4503-4461-6. DOI: 10.1145/2971648.2971706. URL: http://doi.acm.org/10.1145/2971648.2971706 (visited on 07/23/2019) (cit. on p. 14).

[20] H.-K. Wu, S. W.-Y. Lee, H.-Y. Chang, J.-C. Liang. "Current status, opportunities and challenges of augmented reality in education". In: *Computers & Education* 62 (Mar. 2013), pp. 41–49. ISSN: 0360-1315. DOI: 10.1016/j.compedu.2012.10.024. URL: http://www.sciencedirect.com/science/article/pii/S0360131512002527 (visited on 07/21/2019) (cit. on p. 14).

[21] Y. Zhou, Y. Tie, L. Qi. "Augmented Reality Travel App Based on Image Retrieval Algorithm". In: *Proceedings of the 2018 International Conference on Computing and Pattern Recognition*. ICCPR '18. event-place: Shenzhen, China. New York, NY, USA: ACM, 2018, pp. 73–77. ISBN: 978-1-4503-6471-3. DOI: 10.1145/3232829.3232834. URL: http://doi.acm.org/10.1145/3232829.3232834 (visited on 07/22/2019) (cit. on p. 14).

[22] J. Ahn, J. Williamson, M. Gartrell, R. Han, Q. Lv, S. Mishra. "Supporting Healthy Grocery Shopping via Mobile Augmented Reality". In: *ACM Trans. Multimedia Comput. Commun. Appl.* 12.1s (Oct. 2015), 16:1–16:24. ISSN: 1551-6857. DOI: 10.1145/2808207. URL: http://doi.acm.org/10.1145/2808207 (visited on 07/22/2019) (cit. on p. 15).

[23] T. Röddiger, D. Doerner, M. Beigl. "ARMart: AR-Based Shopping Assistant to Choose and Find Store Items". In: *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. UbiComp '18. event-place: Singapore, Singapore. New York, NY, USA: ACM, 2018, pp. 440–443. ISBN: 978-1-4503-5966-5. DOI: 10.1145/3267305.3267631. URL: http://doi.acm.org/10.1145/3267305.3267631 (visited on 07/22/2019) (cit. on p. 15).

[24] *Wayback Machine*. Apr. 2012. URL: https://web.archive.org/web/20120426050529/http://www.medialab.ch/archiv/pdf_studien_diplomarbeiten/da05/da2005-104_ActiveNoiseControl.pdf (visited on 06/18/2019) (cit. on p. 15).

[25] F. Mueller, M. Karau. "Transparent Hearing". In: *CHI '02 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '02. event-place: Minneapolis, Minnesota, USA. New York, NY, USA: ACM, 2002, pp. 730–731. ISBN: 978-1-58113-454-4. DOI: 10.1145/506443.506569. URL: http://doi.acm.org/10.1145/506443.506569 (visited on 05/09/2019) (cit. on p. 15).

[26] Z. Zhou, A. D. Cheok, Y. Qiu, X. Yang. "The Role of 3-D Sound in Human Reaction and Performance in Augmented Reality Environments". In: *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 37.2 (Mar. 2007), pp. 262–272. DOI: 10.1109/tsmca.2006.886376 (cit. on p. 15).

[27] M. Yamamoto, N. Hanamitsu, H. Kim, K. Minamizawa, M. Inami. "Dramagic: A Sound Animation Table to Enhance Imagination". In: *Proceedings of the 2016 Virtual Reality International Conference*. VRIC '16. event-place: Laval, France. New York, NY, USA: ACM, 2016, 11:1–11:4. ISBN: 978-1-4503-4180-6. DOI: 10.1145/2927929.2927950. URL: http://doi.acm.org/10.1145/2927929.2927950 (visited on 07/19/2019) (cit. on p. 15).

[28] M. Hatala, R. Wakkary. "Ontology-Based User Modeling in an Augmented Audio Reality System for Museums". en. In: *User Modeling and User-Adapted Interaction* 15.3 (Aug. 2005), pp. 339–380. ISSN: 1573-1391. DOI: 10.1007/s11257-005-2304-5. URL: https://doi.org/10.1007/s11257-005-2304-5 (visited on 05/09/2019) (cit. on p. 16).

[29] L. Kalantari, M. Hatala, J. Willms. "Using Semantic Web Approach in Augmented Audio Reality System for Museum Visitors". In: *Proceedings of the 13th International World Wide Web Conference on Alternate Track Papers & Posters*. WWW Alt. '04. event-place: New York, NY, USA. New York, NY, USA: ACM, 2004, pp. 386–387. ISBN: 978-1-58113-912-9. DOI: 10.1145/1013367.1013488. URL: http://doi.acm.org/10.1145/1013367.1013488 (visited on 05/09/2019) (cit. on p. 16).

[30] G. A. Dunko. "Augmented reality enhanced audio". en. US8170222B2. May 2012. URL: https://patents.google.com/patent/US8170222B2/en (visited on 05/09/2019) (cit. on p. 16).

[31] G. Arnold, D. Bates. "Discrete dynamic positioning of audio signals in a 360° environment". US5850455A. Dec. 1998. URL: https://patents.google.com/patent/US5850455A/en (visited on 07/21/2019) (cit. on p. 16).

[32] J. E. Larsen, A. Stopczynski, J. Larsen, C. Vesterskov, P. Krogsgaard, T. Sondrup. "Augmenting the Sound Experience at Music Festivals Using Mobile Phones". In: *Proceedings of the 16th International Conference on Intelligent User Interfaces*. IUI '11. event-place: Palo Alto, CA, USA. New York, NY, USA: ACM, 2011, pp. 383–386. ISBN: 978-1-4503-0419-1. DOI: 10.1145/1943403.1943471. URL: http://doi.acm.org/10.1145/1943403.1943471 (visited on 05/09/2019) (cit. on pp. 16, 46).

[33] A. J. Onwuegbuzie, W. B. Dickinson, N. L. Leech, A. G. Zoran. "A Qualitative Framework for Collecting and Analyzing Data in Focus Group Research". en. In: *International Journal of Qualitative Methods* 8.3 (Sept. 2009), pp. 1–21. ISSN: 1609-4069, 1609-4069. DOI: 10.1177/160940690900800301. URL: http://journals.sagepub.com/doi/10.1177/160940690900800301 (visited on 03/22/2019) (cit. on p. 19).

[34] J. Lazar, J. H. Feng, H. Hochheiser. “Chapter 8 - Interviews and focus groups”. In: *Research Methods in Human Computer Interaction (Second Edition)*. Ed. by J. Lazar, J. H. Feng, H. Hochheiser. Boston: Morgan Kaufmann, Jan. 2017, pp. 187–228. ISBN: 978-0-12-805390-4. DOI: 10.1016/B978-0-12-805390-4.00008-X. URL: http://www.sciencedirect.com/science/article/pii/B978012805390400008X (visited on 02/01/2019) (cit. on p. 19).

[35] A. Blandford, D. Furniss, S. Makri. *Qualitative HCI Research: Going Behind the Scenes*. en. Google-Books-ID: zoz2CwAAQBAJ. Morgan & Claypool Publishers, Apr. 2016. ISBN: 978-1-62705-760-8 (cit. on p. 33).

[36] M. R. Morris, A. Danielescu, S. Drucker, D. Fisher, B. Lee, m. c. schraefel m. c., J. O. Wobbrock. “Reducing Legacy Bias in Gesture Elicitation Studies”. In: *interactions* 21.3 (May 2014), pp. 40–45. ISSN: 1072-5520. DOI: 10.1145/2591689. URL: http://doi.acm.org/10.1145/2591689 (visited on 03/13/2019) (cit. on p. 43).

[37] R.-D. Vatavu. “A Comparative Study of User-defined Handheld vs. Freehand Gestures for Home Entertainment Environments”. In: *J. Ambient Intell. Smart Environ.* 5.2 (Mar. 2013), pp. 187–211. ISSN: 1876-1364. URL: http://dl.acm.org/citation.cfm?id=2594684.2594688 (visited on 01/20/2019) (cit. on p. 43).

[38] J. Rico, S. Brewster. “Usable Gestures for Mobile Interfaces: Evaluating Social Acceptability”. en. In: *Proceedings of the 28th international conference on Human factors in computing systems - CHI ’10*. ACM Press, 2010, p. 10. DOI: 10.1145/1753326.1753458 (cit. on p. 44).

[39] C. S. Montero, J. Alexander, M. T. Marshall, S. Subramanian. “Would You Do That?: Understanding Social Acceptance of Gestural Interfaces”. In: *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services*. MobileHCI ’10. event-place: Lisbon, Portugal. New York, NY, USA: ACM, 2010, pp. 275–278. ISBN: 978-1-60558-835-3. DOI: 10.1145/1851600.1851647. URL: http://doi.acm.org/10.1145/1851600.1851647 (visited on 07/24/2019) (cit. on p. 44).

All links were last followed on August 05, 2019.

# A Appendix

## A.1 Consent Form Focus Groups

## A.2 Consent Form Elicitation Study

## A.3 Interview Script Elicitation Study

## A.4 Questionnaire Online Study

# Informed Consent of Participation

You are invited to participate in the focus group **Actions and Use Cases For Augmented Hearing** initiated and conducted by Francisco Kiss. The research is supervised by Prof. Michael Sedlmair at the University of Stuttgart. Please note:

- Your participation is enitrely voluntary
- The focus group will last approximately 120 minutes
- We will record videos during the session
- We will record audio during the session
- We may publish our results from this and other sessions in our reports, but all such reports will be confidential and will not include your name

If you have any questions or complaints about the whole informed consent process of this research study or your rights as a human research subject, please contact Francisco Kiss (E-Mail: francisco.kiss@vis.uni-stuttgart.de) or Jun-Prof. Michael Sedlmair. You should carefully read the information below. Please take as much time as you need to read the consent form.

## 1. Purpose and Goal of this Research

The study is part of my bachelor thesis. The goal of this focus group is to identify relevant interactions and application cases for the integration of augmented hearing in everyday life. Your participation will help us achieve this goal. The results of this research may be presented at scientific or professional meetings or published in scientific proceedings and journals.

## 2. Participation and Compensation

Your participation in this user study is completely voluntary. You will be one of approximately 16 people being studied for this research. You will receive 20 EUR as compensation for your participation. You may withdraw and discontinue participation at any time without penalty or losing the compensation. If you decline to participate or withdraw from the user study, no one on the campus will be told. You may refuse to answer any questions you do not want to answer.

## 3. Procedure

After confirming your informed consent, you will:

1. Fill out a consent form and a demographic questionnaire
2. Take part in group discussions and contribute your ideas
3. Summarize and evaluate the results of the group session
4. Get paid

The complete procedure of this user study will last approximately 120 minutes.

## 4. Risks and Benefits

There are no risks associated with this user study. Discomforts or inconveniences will be minor and are not likely to happen. If any discomforts become a problem, you may discontinue your participation. Your benefit in participating is your compensation of 20 EUR.

## 5. Data Protection and Confidentiality

The researcher will not identify you by your real name in any reports using information obtained from this user study and that your confidentiality as a participant in this user study will remain secure. All data you provide in this user study will be published anonymized and treated confidentially in compliance with the General Data Protection Regulation (GDPR) of the European Union (EU). In all cases uses of records and data will be subject to the GDPR. Faculty and administrators from the campus will not have access to raw data or transcripts. This precaution will prevent your individual comments from having any negative repercussions. We will record videos, record audio, and will potentially take notes during the user study. Any recordings cannot be viewed by anyone outside this research project unless we have you sign a separate permission form allowing us to use them (see below). As with any publication or online related activity, the risk of a breach of confidentiality is always possible. According to the GDPR, the researchers will inform the participant if a breach of confidential data was detected. We will record videos, record audio, and will potentially take notes during the user study.

## 6. Identification of Investigators

If you have any questions or concerns about the research, please feel free to contact:

**Francisco Kiss**
**Principal Investigator**
**Human-Computer Interaction**
**Pfaffenwaldring 5A**
**70569 Stuttgart, Germany**

**Jun. - Prof. Michael Sedlmair**
**Professor/Head of Department**
**Human-Computer Interaction**
**Allmandring 19**
**70569 Stuttgart, Germany**

## 7. Informed Consent and Agreement

☐ **I understand the explanation provided to me. I have been given a copy of this form. I have had all my questions answered to my satisfaction, and voluntarily agree to participate in this user study**

______________________________
Printed Name of Subject

______________________________ ______________________________
Signature of Subject | Location, Date

# Informed Consent of Participation

You are invited to participate in the elicitation study **Augmented Hearing Controls** initiated and conducted by Francisco Kiss. The research is supervised by Prof. Michael Sedlmair at the University of Stuttgart. Please note:

- Your participation is enitrely voluntary
- The study will last approximately 60 minutes
- We will record videos during the session
- We will record audio during the session
- We may publish our results from this and other sessions in our reports, but all such reports will be confidential and will not include your name

If you have any questions or complaints about the whole informed consent process of this research study or your rights as a human research subject, please contact Francisco Kiss (E-Mail: francisco.kiss@vis.uni-stuttgart.de) or Jun-Prof. Michael Sedlmair. You should carefully read the information below. Please take as much time as you need to read the consent form.

## 1. Purpose and Goal of this Research

The study is part of my bachelor thesis. The goal of this study is to identify relevant gestures for several control modalities for augmented hearing. Your participation will help us achieve this goal. The results of this research may be presented at scientific or professional meetings or published in scientific proceedings and journals.

## 2. Participation and Compensation

Your participation in this user study is completely voluntary. You will be one of approximately 16 people being studied for this research. You will receive 10 EUR as compensation for your participation. You may withdraw and discontinue participation at any time without penalty or losing the compensation. If you decline to participate or withdraw from the user study, no one on the campus will be told. You may refuse to answer any questions you do not want to answer.

## 3. Procedure

After confirming your informed consent, you will:

1. Fill out a consent form and a demographic questionnaire
2. Conduct the study with different settings, propose gestures for each and answer questionnaires after each one.
3. Answer final questionnaire at the end
4. Get paid

The complete procedure of this user study will last approximately 60 minutes.

## 4. Risks and Benefits

There are no risks associated with this user study. Discomforts or inconveniences will be minor and are not likely to happen. If any discomforts become a problem, you may discontinue your participation. Your benefit in participating is your compensation of 20 EUR.

## 5. Data Protection and Confidentiality

The researcher will not identify you by your real name in any reports using information obtained from this user study and that your confidentiality as a participant in this user study will remain secure. All data you provide in this user study will be published anonymized and treated confidentially in compliance with the General Data Protection Regulation (GDPR) of the European Union (EU). In all cases uses of records and data will be subject to the GDPR. Faculty and administrators from the campus will not have access to raw data or transcripts. This precaution will prevent your individual comments from having any negative repercussions. We will record videos, record audio, and will potentially take notes during the user study. Any recordings cannot be viewed by anyone outside this research project unless we have you sign a separate permission form allowing us to use them (see below). As with any publication or online related activity, the risk of a breach of confidentiality is always possible. According to the GDPR, the researchers will inform the participant if a breach of confidential data was detected. We will record videos, record audio, and will potentially take notes during the user study.

## 6. Identification of Investigators

If you have any questions or concerns about the research, please feel free to contact:

**Francisco Kiss**
**Principal Investigator**
**Human-Computer Interaction**
**Pfaffenwaldring 5A**
**70569 Stuttgart, Germany**

**Jun. - Prof. Michael Sedlmair**
**Professor/Head of Department**
**Human-Computer Interaction**
**Allmandring 19**
**70569 Stuttgart, Germany**

## 7. Informed Consent and Agreement

☐ **I understand the explanation provided to me. I have been given a copy of this form. I have had all my questions answered to my satisfaction, and voluntarily agree to participate in this user study**

______________________________
Printed Name of Subject

______________________________ ______________________________
Signature of Subject | Location, Date

# Interview Script Elicitation Study

## Vor dem eigentlichen Interview:

1. Consent Form ausfüllen
2. Demographics ausfüllen
   - Geschlecht
   - Alter
   - Beruf

## Intro

1. Was weißt du über Augmented Reality?
   - Was sind deine eigenen Erfahrungen ?
   - Was hast du von Augmented Reality schon gehört, aber noch nie selbst ausprobiert oder gesehen?
2. Auf vorherige Antwort eingehen
   - **Augmented Reality beschreiben** (falls wenig Vorwissen) :
     Augmented Reality ist die Erweiterung der Umgebung, also der Realität, durch den Einsatz von Technik. Also zum Beispiel das Einblenden von visuellen Informationen (Head-Up Display) oder andere Daten
   - **Augmented Sound/Hearing erklären:**
     Augmented Sound kann verschieden umgesetzt werden. Für die Studie stellst du dir vor, dass du Köpfhörer trägst, denen es möglich ist deine Geräuschkulisse zu manipulieren. Die bekannteste Form ist Noise Canceling: Man reduziert bzw. eliminiert die Umgebungsgeräusche, indem die Schallwellen invertiert werden, um einen Zustand der absoluten Ruhe erreichen zu können.
   - **Erklärung der Aufgabe, Überleitung zur ersten Modalität**:
     Wir haben einige Aktionen vorbereitet, die solche Kopfhörer in Zukunft noch ausführen könnten. Deine Aufgabe ist es nur zu verstehen, welchen Effekt diese Aktion hat und für verschiedene Geräte, mögliche Gesten oder Interaktionen zu finden, um die Aktion durchzuführen. Wir haben 3 verschiedene Steuermöglichkeiten ausgesucht: (1) Gesten (2) ein Touchinterface einer SmartWatch (3) Kopfhörer Kontroll Buttons. Wir zeigen dir für jedes dieser 3 Geräte, das gleiche Set an Aktionen und du musst dir geeignete Gesten oder Steuermöglichkeiten dafür überlegen. Im Idealfall fallen dir mehrere Gesten für eine einzelne Aktion ein.

## Gesten (G)

Wir suchen passende Gesten für die einzelnen Aktionen. Dabei bist du vollkommen frei ob du nur einzelne Körperteile, den ganzen Körper, Mimik, ein oder zwei Hände für die Geste benutzen willst.
(Zoomen - zwei finger)
(klatschen - licht an/aus)
Ganz körper geste
(Zwinkern - screenshot machen)
(kopfnicken - akzeptieren)

## TouchInterface (T)

Für die SmartWatch suchen wir Gesten für die Durchführung der einzelnen Aktionen. Dabei kannst du eine Hand benutzen, aber so viele Finger wie du möchtest.
(Benachrichtigungen ansehen - nach unten wischen)
Auch die ganze Hand zu verwenden ist eine Möglichkeit (Kinomodus - Hand darauf legen)

## Köpfhörer Tasten (K)

Viele besitzen Kopfhörer, die am Kabel ein Steuerelement mit meist 3 Tasten haben, um die Musik lauter oder leiser zu machen, Gespräche anzunehmen und so weiter. Du kannst die für die Aktionen Tasten überlegen, die du verwenden würdest. Dabei darfst du die Tasten mehrfach, mehrere Tasten gleichzeitig drücken oder neue Tasten hinzufügen, wenn du das Gefühl hast, das 3 nicht ausreichen.
(Lauter - Plus Taste)
(Lied Pausieren - Multifunktionstaste)
(Lied skippen - 2 mal plus Taste)

### Zusammenfassung

1. War es bei den Kopfhörern schwer geeignete Belegungen zu finden?
2. War es bei der Smartwatch schwer geeignete Gesten zu finden?
3. War es bei den Körpergesten schwer geeignete Gesten zu finden?
4. Kannst du dir eine (oder mehrere) Aktion(en) vorstellen, die du gerne nutzen würdest, die nicht erwähnt wurde?
5. Gibt es eine (oder mehrere) Aktion(en) die du nicht nutzen würdest? Aus welchen Gründen?
6. Wenn du eine Rangordnung der Modalitäten erstellen müsstes, wie sähe diese aus und warum?

## Aktionen

**Gruppe A**
Augmentation an
Augmentation aus

**Gruppe B**
Lauter
Leiser

**Gruppe C**
Soundquelle entfernen
Soundquelle hinzufügen

**Gruppe D**
Soundquelle fokussieren
Soundquellen gruppieren

**Gruppe E**
ACN stärker
ACN schwächer

**Gruppe F**
Frequenz höher
Frequenz tiefer

# Online Survey

**Start of Block: Informed Consent**

Q1.1

Welcome to the research study!

We are interested in understanding user-friendly gesture design of actions in the Augmented Sound area. You will be presented user-defined gestures for various Augmented Sound actions and asked to answer some questions about it. Please be assured that your responses will be kept completely confidential.

The study should take you around **20** minutes to complete. Your participation in this research is voluntary. You have the right to withdraw at any point during the study, for any reason, and without any prejudice. If you would like to contact the Principal Investigator in the study to discuss this research, please e-mail Maike Ernst (st143200@stud.uni-stuttgart.de).

Please be aware that the study **cannot be performed in the Safari web browser** due to technical reasons. All other major browsers are supported.

By clicking the button below, you acknowledge that your participation in the study is voluntary, you are 18 years of age, and that you are aware that you may choose to terminate your participation in the study at any time and for any reason.

Please note that this survey will be best displayed on a laptop or desktop computer. Some features may be less compatible for use on a mobile device.

- ◯ I consent. Begin the study (1)
- ◯ I do not consent. I do not wish to participate in this study (2)

*Skip To: End of Survey If Welcome to the research study! We are interested in understanding user-friendly gesture desig... = I do not consent. I do not wish to participate in this study*

**End of Block: Informed Consent**

**Start of Block: Demographics**

Q2.1 Age

________________________________________________________________



Q2.2 I identify as

- ◯ male (1)
- ◯ female (2)
- ◯ other (3)
- ◯ i prefer not to say (4)

Q2.3 What is your highest level of education you have completed?

- ◯ No formal education (1)
- ◯ Primary school (2)
- ◯ High school (3)
- ◯ Secondary vocational education (MBO) (4)
- ◯ Bachelor's' degree (5)
- ◯ Master's degree (6)
- ◯ PHD (7)

Q2.4 Which of the following categories best describes your employment status?

- ◯ Employed, working full time (1)
- ◯ Employed, working part time (2)
- ◯ Not employed (3)
- ◯ Student, trainee (4)
- ◯ Retired (5)
- ◯ Disable, not able to work (6)

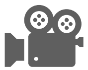
Q2.5 Place of residence

▼ Please select below... (1) ... Other (195)

End of Block: Demographics

Start of Block: intro

Q3.1
Please watch the following video and carefully review the survey instructions.



In the course of this study you will be shown 12 short videos demonstrating the effect of an action in Augmented Sound.
Imagine wearing headphones that allow all these actions. While each of the 5 sources actually makes a noise, through the headphones you only perceive those that are colored.

There are 3 different control modalities:
1) Gestures
2) Touch gestures on a smartwatch
3) Headphone with a common control element with 3 keys (+, - and multi-function)

For each action, you compare several gestures for each modality, as well as the modalities among themselves.

Please TURN ON SOUND!

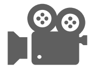
At the end of this survey, we provide you a unique ID. You need to paste this ID into MTurk, to get paid.

End of Block: intro



Start of Block: volume up

Q4.1



This is "volume up". This action increases the volume of all currently active sound sources.

Q4.2 Which touch gesture is best suited for this action? Give your ranking.
- Swipe up with one finger. (1)
- Tap in the middle of the display once and circle clockwise at the edge of the display until the volume is as you prefer. (2)
- Drag one finger upwards until the volume is as you prefer. (3)
- Circle clockwise at the edge of the display until the volume is as you prefer. (4)

Q4.3 Which gesture is best suited for this action? Give your ranking.
- Turn an imaginary knob clockwise with one hand until the volume is as you prefer. (1)
- Fold one hand up with palms pointing up. (2)
- Raise one arm in front of your body with palms pointing up until the volume is as you prefer. (3)
- Raise both arms in front of your body with palms pointing up until the volume is as you prefer. (4)

Q4.4 Which key mapping is best suited for this action? Give your ranking.
- Hold + button until the volume is as you prefer. (1)
- Adjust the volume gradually with + button. (2)

Q4.5 Which control modality is best suited for this action? Give your ranking.
Gestures (1)
Touch gestures (2)
Headphone buttons (3)

End of Block: volume up

Start of Block: volume down

Q5.1

This is "volume down". This action reduces the volume of all currently active sound sources.

Q5.2 Which gesture is best suited for this action? Give your ranking.
- Turn imaginary knob counterclockwise with one hand until the volume is as you prefer. (1)
- Lower arm in front of your body with palms pointing down until the volume is as you prefer. (2)
- Fold both hands downwards two times with palms pointing down. (3)
- Lower both arms in front of your body with palms pointing down until the volume is as you prefer. (4)

Q5.3 Which touch gesture is best suited for this action? Give your ranking.
- Drag down with one finger until the volume is as you prefer. (1)
- Tap in the middle of the display once and circle counterclockwise at the edge of the display until the volume is as you prefer. (2)
- Swipe down with one finger. (3)
- Circle counterclockwise at the edge of the display until the volume is as you prefer. (4)



Q5.4 Which key mapping is best suited for this action? Give your ranking.
- Hold - button until the volume is as you prefer. (1)
- Adjust the volume gradually with - button. (2)

Q5.5 Which control modality is best suited for this action? Give your ranking.
Gestures (1)
Touch gestures (2)
Headphone buttons (3)

End of Block: volume down

Start of Block: augOff

Q6.1

This is "augmentation off". This action turns off the augmented hearing function, so one hears normally again. Here, the user has previously chosen to augment their hearing by removing several sound sources. By turning off augmentation he can hear the sources again.

Q6.2 Which gesture is best suited for this action? Give your ranking.
- Tap on one ear with one hand. (1)
- Flick once. (2)
- Clap twice. (3)
- Clench a fist with one hand. (4)

Q6.3 Which touch gesture is best suited for this action? Give your ranking.
- Cover the display with your whole hand. (1)
- Double tap with one finger. (2)
- Hold one finger on the display. (3)

Q6.4 Which key mapping is best suited for this action? Give your ranking.

- Hold multi-function key until augmentation is disabled. (1)
- Click multi-function button once. (2)

Q6.5 Which control modality is best suited for this action? Give your ranking.
Gestures (1)
Touch gestures (2)
Headphone buttons (3)

End of Block: augOff

Start of Block: augOn

Q7.1

This is "augmentation on". This action turns on the augmented hearing function and applies previous settings. Here, the user chosen to augment their hearing by removing several sound sources, before turning off augmentation. By turning on augmentation again, these sound sources will be instantly removed from his soundscape.

Q7.2 Which gesture is best suited for this action? Give your ranking.

- Tap on one ear with one hand. (1)
- Flick once. (2)
- Pinch once with one hand. (3)
- Put your flat hand on one ear and move it away. (4)



Q7.3 Which touch gesture is best suited for this action? Give your ranking.

- Cover the display with your whole hand. (1)
- Double tap with one finger. (2)
- Hold one finger on the display. (3)

Q7.4 Which key mapping is best suited for this action? Give your ranking.

- Hold multi-function key until augmentation is disabled. (1)
- Click multi-function button once. (2)

Q7.5 Which control modality is best suited for this action? Give your ranking.
Gestures (1)
Touch gestures (2)
Headphone buttons (3)

End of Block: augOn

Start of Block: add source

Q8.1

This is "add sound source". This action adds a sound source to soundscape (currently active sound sources).

Q8.2 Which gesture is best suited for this action? Give your ranking.

- Point to the source you want to add. (1)
- Point to the source you want to add with a clenched fist and open it. (2)

Q8.3 Which touch gesture is best suited for this action? Give your ranking.
- Focus on a sound source you want to add with your gaze then swipe right. (1)
- Focus on a sound source you want to add with your gaze then tap once. (2)
- Focus on a sound source you want to add with your gaze then tap twice. (3)
- Point your watch to a sound source you want to add then tap once. (4)

Q8.4 Which key mapping is best suited for this action? Give your ranking.
- Focus on a sound source you want to add with your gaze and hold + button until the source is added. (1)
- Focus on a sound source you want to add with your gaze and hold + and multi-function button simultaneously until the source is added. (2)
- Focus on a sound source you want to add with your gaze and click + and multi-function button once simultaneously. (3)
- Focus on a sound source you want to add with your gaze and hold multi-function key until the source is added. (4)

Q8.5 Which control modality is best suited for this action? Give your ranking.
Gestures (1)
Touch gestures (2)
Headphone buttons (3)

End of Block: add source

Start of Block: remove source

Q9.1

This is "remove sound source". This action removes a sound source from soundscape (currently active sound sources).



Q9.2 Which gesture is best suited for this action? Give your ranking.
- Point to the source you want to remove. (1)
- Point to the source you want to remove and wipe left. (2)
- Lower your arm in the direction of the source you want to remove. (3)

- Q9.3 Which touch gesture is best suited for this action? Give your ranking.
- Focus on a sound source you want to remove with your gaze then swipe left. (1)
- Point your watch to a sound source you want to remove then tap twice. (2)
- Point your watch to a sound source you want to remove then tap once. (3)

Q9.4 Which key mapping is best suited for this action? Give your ranking.
- Focus on a sound source with your gaze and hold - button until the source is removed. (1)
- Focus on a sound source with your gaze and hold - and multi-function button until the source is removed. (2)
- Focus on a sound source with your gaze and click - twice. (3)
- Focus on a sound source with your gaze and hold multi-function button until the source is removed. (4)

Q9.5 Which control modality is best suited for this action? Give your ranking.
Gestures (1)
Touch gestures (2)
Headphone buttons (3)

End of Block: remove source

Start of Block: Focus

Q10.1

This is "focus sound source". This action makes a selected sound source louder, while the other active sound sources become quieter.

---

Q10.2 Which gesture is best suited for this action? Give your ranking.

- Point to the source you want to focus on. (1)
- Focus on a sound source with your gaze and keep your eyes closed for a moment. (2)
- Point to the source you want to focus on for a moment. (3)

Q10.3 Which touch gesture is best suited for this action? Give your ranking.

- Focus on the sound source with your gaze then hold one finger on the display. (1)
- Focus on the sound source with your gaze then tap once. (2)
- Point your watch to the sound source then tap once. (3)
- Focus on the sound source with your gaze then place two fingers in the middle of the display and drag them outwards. (4)

---

Q10.4 Which key mapping is best suited for this action? Give your ranking.

- Focus on the sound source with your gaze and click multi-function button twice. (1)
- Focus on the sound source with your gaze and hold + button a few seconds. (2)
- Focus on the sound source with your gaze and click + button twice. (3)
- Focus on the sound source with your gaze hold multi-function button a few seconds. (4)



---

Q10.5 Which control modality is best suited for this action? Give your ranking.
Gestures (1)
Touch gestures (2)
Headphone buttons (3)

End of Block: Focus

---

Start of Block: LFH

Q11.1

This is "listen from here". Through this action, only sound sources within a selected area are heard.

---

Q11.2 Which gesture is best suited for this action? Give your ranking.

- Circle the area you want to listen to with one arm. (1)
- Frame the area you want to listen to with two fingers of each hand. (2)
- Pull a box with both arms around the area you want to listen to. (3)
- Extend both arms in front of you. The area between your arms is the area you want to listen to. (4)

---

Q11.3 Which touch gesture is best suited for this action? Give your ranking.

- Let your gaze roam over the area while holding one finger on the display. (1)
- Let your watch roam over the area while holding one finger on the display. (2)
- Let your gaze roam over the area while holding one finger on the display. (3)
- The edge of the display reflects the environment around you. You can select the area by drawing the corresponding section at the edge. (4)

Q11.4 Which key mapping is best suited for this action? Give your ranking.
- Let your gaze roam over the area while holding multi-function button. (1)
- Let your gaze roam over the area while holding + button. (2)
- Focus on the center of the area with your gaze and click multi-function button twice. (3)
- Let your gaze roam over the area while holding all three buttons. (4)

Q11.5 Which control modality is best suited for this action? Give your ranking.
Gestures (1)
Touch gestures (2)
Headphone buttons (3)

**End of Block: LFH**

**Start of Block: frequency higher**

Q12.1

This is "frequency higher". This action increases the pitch of all currently active sound sources.

Q12.2 Which gesture is best suited for this action? Give your ranking.
- Turn an imaginary knob clockwise with one hand until the frequency is as you prefer. (1)
- Flick once then raise arm in front of your body with palms pointing up until the frequency is as you prefer. (2)



Q12.3 Which touch gesture is best suited for this action? Give your ranking.
- Tap in the middle of the display once and circle clockwise at the edge of the display until the frequency is as you prefer. (1)
- Pinch slowly with two fingers until the frequency is as you prefer. (2)

Q12.4 Which key mapping is best suited for this action? Give your ranking.
- Click + button twice. (1)
- Hold + and multi-function simultaneously until the frequency is as you prefer. (2)
- Hold multi-function button while adjusting gradually with + button. (3)
- Click multi-function button twice then adjust gradually with + button. (4)

Q12.5 Which control modality is best suited for this action? Give your ranking.
Gestures (1)
Touch gestures (2)
Headphone buttons (3)

**End of Block: frequency higher**

**Start of Block: frequency lower**

Q13.1

This is "frequency lower". This action lowers the pitch of all currently active sound sources.

Q11.4 Which key mapping is best suited for this action? Give your ranking.
- Let your gaze roam over the area while holding multi-function button. (1)
- Let your gaze roam over the area while holding + button. (2)
- Focus on the center of the area with your gaze and click multi-function button twice. (3)
- Let your gaze roam over the area while holding all three buttons. (4)

Q11.5 Which control modality is best suited for this action? Give your ranking.
Gestures (1)
Touch gestures (2)
Headphone buttons (3)

**End of Block: LFH**

**Start of Block: frequency higher**

Q12.1

This is "frequency higher". This action increases the pitch of all currently active sound sources.

Q12.2 Which gesture is best suited for this action? Give your ranking.
- Turn an imaginary knob clockwise with one hand until the frequency is as you prefer. (1)
- Flick once then raise arm in front of your body with palms pointing up until the frequency is as you prefer. (2)



Q12.3 Which touch gesture is best suited for this action? Give your ranking.
- Tap in the middle of the display once and circle clockwise at the edge of the display until the frequency is as you prefer. (1)
- Pinch slowly with two fingers until the frequency is as you prefer. (2)

Q12.4 Which key mapping is best suited for this action? Give your ranking.
- Click + button twice. (1)
- Hold + and multi-function simultaneously until the frequency is as you prefer. (2)
- Hold multi-function button while adjusting gradually with + button. (3)
- Click multi-function button twice then adjust gradually with + button. (4)

Q12.5 Which control modality is best suited for this action? Give your ranking.
Gestures (1)
Touch gestures (2)
Headphone buttons (3)

**End of Block: frequency higher**

**Start of Block: frequency lower**

Q13.1

This is "frequency lower". This action lowers the pitch of all currently active sound sources.

Q14.5 Which control modality is best suited for this action? Give your ranking.
Gestures (1)
Touch gestures (2)
Headphone buttons (3)

End of Block: ANC up

Start of Block: ANC down

Q15.1

This is "ANC down". This action **increases** the environmental sound volume by **reducing** activ noise cancelling.

Q15.2 Which gesture is best suited for this action? Give your ranking.
- Raise one arm in front of your body with palms pointing up until the noise cancelling is as you prefer. (1)
- Raise both arms in front of your body with palms pointing up until the noise cancelling is as you prefer. (2)
- Turn imaginary knob backwards in front of the ear. (3)

Q15.3 Which touch gesture is best suited for this action? Give your ranking.
- Turn that arm carrying the watch 90° away from your body. (1)
- Tap in the middle of the display once and circle clockwise at the edge of the display until the noise cancelling is as you prefer. (2)
- Tap in the middle of the display once and circle counterclockwise at the edge of the display with two fingers until the noise cancelling is as you prefer. (3)



Q15.4 Which key mapping is best suited for this action? Give your ranking.
- Hold + button until the noise cancelling is as you prefer. (1)
- Click multi-function button once then adjust the noise cancelling gradually with + button. (2)
- Hold multi-function button while adjusting the noise cancelling gradually with + button. (3)
- Click - button twice. (4)

Q15.5 Which control modality is best suited for this action? Give your ranking.
Gestures (1)
Touch gestures (2)
Headphone buttons (3)

End of Block: ANC down

Start of Block: actions_end

Q16.1 Please rate how important each action is for you on a scale from 1 (not important at all) to 5 (very important).

| | 1 (1) | 2 (2) | 3 (3) | 4 (4) | 5 (5) |
|---|---|---|---|---|---|
| Augmentation On (1) | ○ | ○ | ○ | ○ | ○ |
| Augmentation Off (2) | ○ | ○ | ○ | ○ | ○ |
| Volume down (3) | ○ | ○ | ○ | ○ | ○ |
| Volume up (4) | ○ | ○ | ○ | ○ | ○ |
| Frequency higher (5) | ○ | ○ | ○ | ○ | ○ |
| Frequency lower (6) | ○ | ○ | ○ | ○ | ○ |
| AC up (7) | ○ | ○ | ○ | ○ | ○ |
| ANC down (8) | ○ | ○ | ○ | ○ | ○ |
| Listen from here (9) | ○ | ○ | ○ | ○ | ○ |
| Focus on sound source (10) | ○ | ○ | ○ | ○ | ○ |
| Add sound source (11) | ○ | ○ | ○ | ○ | ○ |
| Remove sound source (12) | ○ | ○ | ○ | ○ | ○ |

Q16.2 Can you imagine an action you would use that wasn't shown?

________________________________

End of Block: actions_end



Start of Block: modalities_end

Q17.1 Which modality would you prefer to use? Give your ranking
Gestures (1)
Touch gestures (2)
Headphone buttons (3)

Q17.2 Why did you choose this ranking?

________________________________

End of Block: modalities_end

Start of Block: end

Q18.1 Would you use augmented hearing?

○ Yes (1)

○ Maybe (2)

○ No (3)

Q18.2 Why?

________________________________

End of Block: end

Start of Block: Random ID

Q76 If you are participating in this survey through MTruk, please paste your worker ID here.

________________________________

Q75
Here is your ID:  ${e://Field/RandomID}

In order to receive your payment, please copy it to paste into the MTurk window.

When you have copied this ID, please click the next button to submit your survey.

End of Block: Random ID

---

The answers to the questions marked with this icon appear randomized

This symbol signals a video is being shown at this point.

○ This symbol indicates a GIF is inserted at this point.

**Declaration**

I hereby declare that the work presented in this thesis is entirely my own and that I did not use any other sources and references than the listed ones. I have marked all direct or indirect statements from other sources contained therein as quotations. Neither this work nor significant parts of it were part of another examination procedure. I have not published this work in whole or in part before. The electronic copy is consistent with all submitted copies.

place, date, signature