S5: Selective Sensing of Single Sound Sources

Francisco Kiss
University of Stuttgart
Stuttgart, Germany
francisco.kiss@vis.uni-stuttgart.de

Romina Poguntke
VIS-MCI, University of Stuttgart
Stuttgart, Germany
romina.poguntke@vis.uni-stuttgart.de

Albrecht Schmidt
Human Centered Ubiquitous Media Group, LMU Munich
Munich, Germany
albrecht.schmidt@ifi.lmu.de

Paweł W. Woźniak
VIS-MCI, University of Stuttgart
Stuttgart, Germany
pawel.wozniak@vis.uni-stuttgart.de

MobileHCI '18 Adjunct: 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, September 3–6, 2018, Barcelona, Spain

ACM ISBN 978-1-4503-5941-2/18/09.
http://dx.doi.org/10.1145/3236112.3236118

Abstract
The sense of hearing provides humans with information about their surroundings and is the primary means of communication, yet it is limited in its ability to focus on particular stimuli. To provide this ability, we designed and built *S5*, a mobile proof-of-concept prototype that allows *Selective Sensing of Single Sound Sources*. Our design consists of a head-mounted directional microphone attached to a smartphone, which acts as controller, filter and amplifier. Users hear the selective signal through headphones and activate the device by touching their ear. To evaluate this sensory augmentation, we conducted a study with 16 participants that showed the system was appealing and perceived as useful. Based on our findings, we conclude that the proposed augmentation is feasible and we provide insights for further development of the concept.

Author Keywords
Augmented hearing; interfaces

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous

Introduction and Background
Focusing our sight on individual targets is an intuitive ability in humans, which allows us to differentiate objects and

events without effort. Although hearing enables humans to perceive spatial properties, such as distance and relative position of a source, focusing hearing on single sources is challenging. This becomes evident when trying to focus on a conversation in noisy venues or identifying the source of a noise in an industrial machine.

Mobile devices such as Microsoft's HoloLens[1] or Bose AR[2] already provide a series of sensory augmentation. Even if dedicated devices allow for more complex and versatile solutions, common-use smartphones can support sensory augmentation.

In this paper, we investigate the feasibility of focusing hearing based on a mobile device and propose a system capable of providing this ability. To assess user expectations about this technology, we ran a *future techonology workshop*, presenting the idea to a group of participants and collecting feedback through structured tasks. Based on the gathered insights, we implemented a prototype and ran a study with 16 participants, to investigate the benefits and limitations of the concept and its practical implementation. Based on the data obtained from the study, we propose further steps to enhance and investigate the design concept.

Related work

The use of mobile devices to augment the sense of hearing has been a subject of exploration in HCI and a large corpus of research exists. The use of directional microphones for augmentation was investigated by Heller and Borchers. *AudioTorch* [2] and *AudioScope* [3] used the sensors present in smartphones to calculate their position and orientation and simulated sounds coming from particular directions.

[1] https://www.microsoft.com/de-de/hololens
[2] https://developer.bose.com/bose-ar

Albrecht et al. proposed a "microphone-hear-through augmented reality hardware system" (sic) [1], adding a microphone to the outer side of each speaker of a headphone set. This way, they forwarded the signal of each microphone to the speaker, and using equalization and filtering, they reproduced the sounds of the environment as if no headphones were being used. The system allowed overlaying sounds within the environment, which proved not only useful for AR applications, but also for listening to music without blocking external sounds.

In the field of AR, Shen et al. combined visualization and hearing to provide feedback on the sources of sounds [6]. Their system recognized the origin of sounds by means of sound processing and presented that information to the user with visual markers on a head mounted display. Although the main use of this concept was to aid people with hearing impairments, this technology could be useful in many different application scenarios, such as machinery diagnostics, fire fighting or remote monitoring. Hussaini's *Mobile SoundAR* is a system for realistic augmented hearing. The system uses the sensors on an iPhone 4 to track the orientation of the user's head, thus enabling his system to render environmental sounds in a realistic way [4].

The above mentioned work explores different approaches to auditory augmentation, putting an emphasis in giving people abilities beyond human capability. It also provides a base to explore the relations between spacial dimensions and hearing, selecting hearing and isolating sound sources, although the combination of this three concepts remains yet to be investigated.

Selective Sensing of Sound

In contrast to existing work, we propose to enable humans to focus hearing at will and listen to selected sound sources.

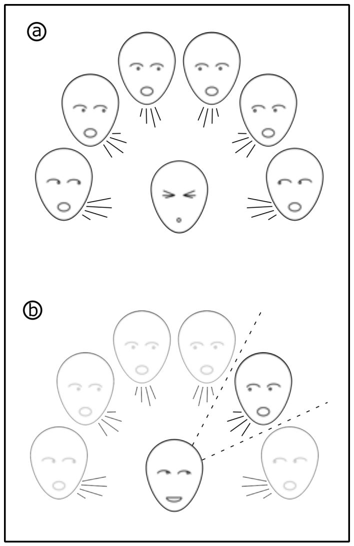

Figure 1: (a) Multiple competing sound signals can be overwhelming. (b) The ability to focus hearing on a narrow direction improves perception and cognition.

Our goal is to augment human hearing with a mobile device that permits using this ability in the real world. We believe this ability can enhance human perception and provide benefits in better cognition in multiple daily life circumstances.

Our hypothesis is that providing humans with a system that supports focused hearing will result in (a) adaptation to this new sense (even if learning is required), and (b) decrease cognitive effort by improving the signal-to-noise ratio between information of interest and environmental sound, and aiding to rapidly identify the source of a sound.

The contribution of this paper is twofold: first, we introduce the concept of selective sensing of single sound sources (S5). In a workshop, we explored user requirements linked to the prototypical implementation and then present the first prototype of our system. Second, we test our concept with 16 users. Based on the gained insights we derive initial findings about building augmented audio systems.

Design

The goal of our design is to empower people to focus their hearing as they would do with sight. Thus, users should be enabled to control their sense of hearing and spatially select sources of sound from their environment. They should also toggle this sensory augmentation, alternating between normal hearing and directional hearing at will. From the user perspective, this would result in the ability to isolate the sound signal from a single source, while phasing out, or even muting, all other environmental sound (see Figure 1).

The current prototype is based on generic stereo headphones, augmented with a directional microphone and a control unit. This provides the desired functionality in a wearable device and allows us to investigate the relevant challenges for HCI. We envision that S5 will be embedded in consumer-grade headphones, in particular in active noise-cancelling headsets since they are already equipped with some of the required electronics.

We ran two sessions of a *future technology workshop* [5], each time with six participants (a total of 5 males and 7 females, aged between 20 and 25). During each session, we explained our concept to participants, showed them a video portraying the main functionalities of an early prototype, and prompted them to ideate different control modalities for such a system. In addition, we motivated discussions about the application fields of this technology and the implications of its use in daily life.

The collected insights suggested a strong preference for gesture-based controls, especially the gesture that mimics positioning an open hand next to the ear, as when trying to improve hearing. Most application cases reported by the participants were accessibility oriented, and the remaining ones for very noisy environments, such as crowded venues or industrial installations. An issue of concern for the workshop participants were the implications for privacy, in particular for those subject to eavesdropping, who would likely be unaware of it.

Our design consists of a system that enables users to focus their hearing on single sound sources. We based our first prototype on the insights provided during the workshop. The system amplifies sound coming from a narrow area in the direction towards which the user is facing. By moving the head, the user can control the area in focus. For controlling our system, we designed a hand-based gesture mechanism: when users touch their ear, the sound focusing system is activated, amplifying the sound coming from directly in front of the user. The gesture was identified as convenient during the workshop. It is a cultural metaphor that adds transparency to the activity: bystanders will likely realize that the user is trying to listen to some-

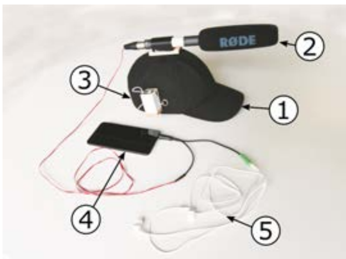

Figure 2: First prototype: (1) rigid cap, (2) shotgun microphone, (3) IR distance sensor, (4) smartphone and (5) headphones

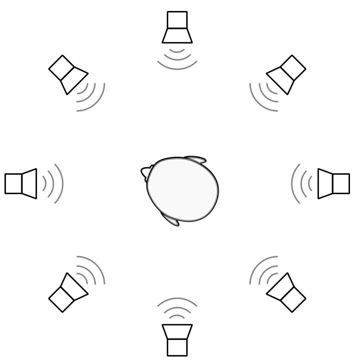

Figure 3: Experiment setup: an array of 8 speakers on an octagonal disposition. The distance between opposite speakers amounts to 3 meters.

thing. Combining target selection and gesture control, users can switch on and off the functionality and focus on single sound sources. Further, these characteristics make the system mobile and functional in most situations.

Prototype

To evaluate our concept, we implemented a prototype (see Figure 2) that provides the main functionality and control of the system as described above. We used generic off-the-shelf in-ear headphones for playing the sound that is sensed with a directional shotgun microphone (RØDE NTG-2 [3]). The signal received by the microphone is amplified and output to the headphones by a smartphone (LG Nexus 5X) running Android. The *mono* signal received by the microphone is played in both *stereo* channels unaltered. This is correct from a *binaural* perspective, since both ears perceive exactly the same primary signal when the source is in front of the user.

The gesture detection is performed with an infra-red distance sensor controlled by an RFduino[4], which sends a signal to the smartphone over a Bluetooth Low Energy connection when the hand of the user is closer than 5 cm. We implemented an Android application that forwards the audio signal from the microphone to the headphones only while the signal indicating hand in the proximity of the sensor is received. For this first prototype, we chose not to process the signal in any way, but to amplify it and control the output intensity with the volume buttons of the phone. This way we ensure that the observed effects are not altered in any way by sound filters.

We attached the microphone on top of a reinforced hat, pointing forward. The infra-red sensor was fixed to the right side of the hat, in a way that it detected the hand of the wearer when it was next to the ear. The sensor can be removed and re-attached on the left side of the hat for left-handed users. The smartphone is stored by the user in a pocket.

[3]http://www.rode.com/microphones/ntg-2
[4]http://www.rfduino.com

Evaluating the Concept of Selective Hearing

We evaluated our concept in an experiment involving 16 participants. We assessed the performance, usability and the benefits S5 offered to users when focusing on particular information from noisy environments.

We used a large, quiet room in a cellar, where we placed an array of 16 speakers (8W RMS, Trust REMO[5]), as shown in Figure 3. The speakers were connected to an eight-channel audio interface (Behringer FIREPOWER FCA610[6]), controlled with a Windows computer running Reaper v5.40, a software application that allows controlling and reproducing multiple audio tracks simultaneously. The result was that a participant standing in the middle of the array would hear different audio tracks coming from many directions, thus generating a real-like environmental soundscape.

Experiment Design

The experiment consisted of two different tasks, designed to evaluate two different scenarios: (a) the information is constantly repeated and (b) the information is presented just once.

$Task\ A$ consisted of the identification and comprehension of a particular repetitive sound signal within several competing ones. Eight audio tracks generated with the Ivona [7]

[5]https://www.trust.com/en/product/17595-remo-2-0-speaker-set
[6]https://www.music-group.com/Categories/Behringer/Computer-Audio/Audio-Interfaces/FCA610/p/P0A3B
[7]https://www.ivona.com

package for Python constantly repeated 4-digit series. During the first minute, all tracks repeated their respective series and indicated that the participant should ignore each number (i.e. "Ignore four, ignore seven..."). After the first minute was over, one of the tracks kept repeating the four-digit series, but indicated the participant to annotate the numbers (i.e. "Write four, write seven..."). The task of the participant was to detect this series as soon as possible and input the four digits on an tablet, starting at any digit but respecting the order.

$Task\ B$ consisted of the identification and comprehension of information within a sound signal that is presented just once. We recorded environmental noise from the local train station and generated train departure announcements with the Ivona package for Python. We combined the recordings to generate a noisy realistic train station sound environment. The announcements overlapped each other and the ambient sound made it difficult to discern messages. In this setup, participants were asked to determine the departure time and platform of a train heading to a given destination, as well as the delay.

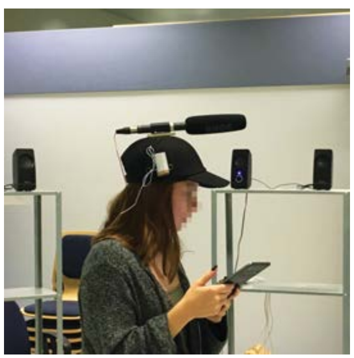

Figure 4: A participant during the experiment.

Participants and Procedure
We recruited 16 participants (11 male and 5 female) aged between 22 and 31 years (M=24.93, SD=2.57) to take part in our experiment. All participants stated to be right-handed and to had no hearing, sight or cognition impairments. Participants were compensated for their participation with 10 Euros.

Participants were informed about the experiment details and their right to interrupt their participation or deny the use of the collected data. After expressing their consent for participation and data collection, they performed each task twice, once with and once without the use of our prototype. To minimize any possible bias, we generated two sets of audio tracks for each task and balanced among participants the order of the conditions (with and without the aid of the device) and which set of audio tracks was used for each condition. To minimize the effect of wearing headphones, participants were asked to wear them during the control condition, even if the device was deactivated.

For each task and condition, the completion time was recorded, as well as the amount of errors in the answers provided by the participants. After all tasks were completed, participants provided feedback on their user experience and general perception of our design in a semi structured interview.

Results
All participants completed the first task correctly within the minimal necessary time, while no participant could get any correct information during the second task.

During the interviews, participants stated that the first task was too easy and the second one too difficult, which is consistent with the collected results. All participants describe the system as attractive, innovative and potentially useful. Some participants wished for an alternative control method, suggesting that a simple button would be easier and more comfortable to use. One of the participants expressed that she did not feel sure about performing the gesture correctly, but “there is no wrong way to press a button”.

Some participants suggested that the reason for failing the second task is due to the need of first locating the sound source with the system deactivated and then toggle it to extract the data. This works only if the information is repeated or consists of a relatively long signal, since the information broadcasted during the time required to locate the direction of the signal is lost to the user. Some participants claimed that not being used to their “new sense” added difficulty to the task. They described that trying to find a particular sound source while thinking of gesturing and solving the task required a high cognitive effort.

Some participants reported perceiving a latency between the sound played on the speakers and the signal amplified by the prototype. This delay in sound was described as distracting and making the tasks more difficult.

Findings

We derived findings both from the workshop and from the experiment. The findings from the workshop refer to the abstract concept of enabling humans to focus hearing on single sound sources. It was suggested that here are two general application scenarios for our concept: focused hearing can help to focus in noisy environments and it can benefit people with hearing impairments.

There was general consensus among workshop participants about using a gesture to control the system in an intuitive and simple way. Additionally, participants considered the potential violation of privacy to be an important issue.

From the experiment using the prototype in a experimental setting we have learned about users' impressions on the concept and implementation. All users perceived the concept as appealing and useful. Some of the participants considered the control method to be complicated and suggested to replace it by a simpler or more conventional approach (i.e. a button). Participants found the device especially useful for listening to repetitive signals. Transient signals, played just once, were portrayed as challenging, both with and without the device.

During the interviews after the experiment, most participants showed a bias towards associating augmented hearing exclusively with accessibility technologies. When asked to suggest application scenarios, most suggestions were related to tackling hearing impairments instead of augmenting hearing beyond natural capabilities.

Discussion

During both the workshop and the study, participants consistently expressed a positive impression of concept. Interestingly, when asked to suggest application scenarios, most experiment participants only suggested accessibility applications.

The privacy issue was discussed deeply in the workshop but barely mentioned by the experiment participants. Eavesdropping was suggested as the problem but participants could not think of a practical solution. It was proposed to add a signal method on the device, that makes it clear to others when it is active (this resulted in the hand-to-ear gesture).

From a technical point of view, we showed the feasibility of implementing such a technology, although some shortcomings of this first approach made themselves evident. Namely, using a smartphone as filter and amplifier introduces perceivable latency in the sound signal that participants found annoying. Some participants disliked the control method, preferring a simpler approach and suggesting to provide some unambiguous feedback about the system being active. This can be caused by legacy bias, suggesting a question for further exploration.

Further, understanding information that is presented just once was very challenging, since participants needed to find the source of the sound. During the time needed to achieve this, the presented information is lost. This points to a problem that escapes the scope of this sensory augmentation but might suggest a new line of work.

Figure 5: Analogue audio power amplifier based on the LM386 chip. The output intensity of this circuit can me controlled with a potentiometer.

Future work

Based on the gained findings, we identified a set of challenges and opportunities that will guide out future work on augmented sound.

The first challenge is the control method. We plan to evaluate how different alternatives can benefit focused hearing in separate experiments. This will allow us to isolate the mutual effect of both variables. For this, the first step will be the evaluation of an improved prototype of the system, controlled with a simple button.

Another challenge is to reduce the latency of the sound. We plan to replace the smartphone with a dedicated amplifier based on an LM386[8] analogue amplifier (see Figure 5). This will not only reduce the size of the prototype but also minimize the latency of the audio output. Additionally, the control method based on a button will benefit from this different approach, since its implementation consists of a switch in the power circuit.

A further improvement could be the use of active noise cancelling headphones. This will result in improving the isolation of the target sound signal from the ambient sound and also eliminate any remaining latency effect. In this case, it will be necessary to link the noise cancelling function to the button, to allow sound to be heard while the aid is deactivated. Alternatively, it might be useful to modify the headphones, so instead of deactivating the noise cancelling, the device works as hear-through headphones. This can be achieved by bridging the inverted buffer in the noise cancelling circuit.

Finally, we believe that this kind of augmentation creates opportunities in different areas of HCI: selective hearing can potentially increase the throughput of information humans can process in loud environments. It also opens a new space for interaction with public displays, enabling computer and humans to selectively interact with an increased degree of individuality in public venues. The metaphor of selecting sound sources with gaze can be extended to other use cases and implementations, such as a combination of eye-tracking and 3D sound. Additionally, we hope our concept prompts discussions about privacy when working with augmented senses, and measures to tackle the arising threats, not only from technical point of view, but also from legal and social perspectives.

[8]www.ti.com/lit/ds/symlink/lm386.pdf

Conclusion

In this work, we introduced and proofed the concept of selective sensing of single sound sources. We presented user requirements based on findings from a workshop involving a total of 12 participants. We implemented the basic functionality of the proposed design in a proof-of-concept prototype. Further, we reported on the findings from an experiment asking 16 participants to use and control S5 to gather qualitative feedback. The results reveal that augmenting human hearing to allow focusing into single sound sources is technically feasible and perceived as a valuable aid for its users. Finally, we derive a set of findings that yield key insights about improvement and enhancement possibilities.

Acknowledgements

Part of this research was supported by the European Union's Horizon 2020 Programme under ERCEA grant no. 683008 AMPLIFY. This work is also part of Be-Greifen and was supported by the German Federal Ministry of Education and Research, grant no. 16SV7527.

REFERENCES

1. Robert Albrecht, Tapio Lokki, and Lauri Savioja. 2011. A Mobile Augmented Reality Audio System with Binaural Microphones. In *Proceedings of Interacting with Sound Workshop: Exploring Context-Aware, Local and Social Audio Applications (IwS '11)*. ACM, New York, NY, USA, 7–11. DOI: http://dx.doi.org/10.1145/2019335.2019337

2. Florian Heller and Jan Borchers. 2014. AudioTorch: Using a Smartphone As Directional Microphone in Virtual Audio Spaces. In *Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices & Services (MobileHCI '14)*. ACM, New York, NY, USA, 483–488. DOI: http://dx.doi.org/10.1145/2628363.2634220

3. Florian Heller and Jan Borchers. 2015. AudioScope: Smartphones As Directional Microphones in Mobile Audio Augmented Reality Systems. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 949–952. DOI: http://dx.doi.org/10.1145/2702123.2702159

4. Syed Naseh Hussaini. 2011. Mobile SoundAR: Your Phone on Your Head. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems (CHI EA '11)*. ACM, New York, NY, USA, 1777–1782. DOI: http://dx.doi.org/10.1145/1979742.1979844

5. Helen Sharp, Yvonne Rogers, and Jenny Preece. 2007. Interaction design: beyond human-computer interaction. (2007).

6. Ruiwei Shen, Tsutomu Terada, and Masahiko Tsukamoto. 2012. A System for Visualizing Sound Source Using Augmented Reality. In *Proceedings of the 10th International Conference on Advances in Mobile Computing & Multimedia (MoMM '12)*. ACM, New York, NY, USA, 97–102. DOI: http://dx.doi.org/10.1145/2428955.2428979