A HEARING AID TO VISUALIZE THE DIRECTION OF SOUND

Udo Gebelein, Stefan Rapp

University of Applied Sciences Darmstadt
udo.gebelein@h-da.de

Abstract: Loss of directional hearing is not only related to one-sided deafness, but may also occur when using hearing aids or cochlear implants. Loss of directional hearing does not only mean a significant reduction in quality of life. It becomes a risk factor when it prevents, for example, reacting quickly to a potentially dangerous situation in traffic. The aim of the project is to process auditory signals so that they are suitable for visual sensory perception in order to help hearing impaired people. Two prototypes are presented to graphically display sound. Despite their limited functionality, they demonstrate the technical feasibility of such a device consisting of commercially available components at a reasonable price. The first prototype is based on a Raspberry Pi with a 3.2 inch display. The sound acquisition is done with 4 lavalier microphones and USB sound cards. The second prototype is also based on a Raspberry Pi. However, sound processing is moved to an enhanced computing platform with an array of 8 SMD microphones, to ensure a synchronous sound sampling, and a sound source location framework is used. Visualization options based on the requirements of hearing impaired people are introduced. Lessons learned are discussed and future work is presented.

1 Motivation

Loss of directional hearing is not only related to one-sided deafness, but may also occur when using hearing aids or cochlear implants [1]. Loss of directional hearing means a significant reduction in quality of life:

- Aggravation in everyday life since the ability to determine the direction of sound is essential for orientation.
- "Party Deafness" – spatial separation of speech signals helps to understand speech. This is impaired in groups, since the identification of the speaker is difficult.
- Difficult assessment of threatening situations since hearing impaired people can not estimate how close a vehicle is and from which direction it comes. This is a serious risk factor when it is about reacting quickly to a potentially dangerous traffic situation.

The aim of the project is the development of a hearing aid to indicate the direction of sound. Sound sources are graphically displayed and can be visually perceived. The regained ability to perceive the direction of sound facilitates the participation of hearing-impaired people in road traffic, in social and professional life and improves protection in and against threatening situations.

2 Binaural Directional Hearing and Related Work

Directional hearing is done by comparing the sound waves hitting the ears and is therefore called binaural directional hearing. Both intensity and time differences are used. The anatomical structure of the outer ear has a significant influence on the quality of directional hearing [1, 2]. The basic mechanisms and terms are briefly outlined below:

Interaural level difference (ILD): The head acts as an obstacle for frequencies with a wavelength smaller than the head dimensions. The other ear is reached by diffraction around the head. The attenuation of the wave amplitude at the other ear leads to a level difference. Higher frequencies are attenuated more than lower ones.

Interaural time difference (ITD): The lateral incidence of the acoustic waves leads to different path lengths, which results in a time difference of the signals when hitting the ears. The interaural time difference applies only for wavelengths larger than twice the head diameter, since the phase differences become ambiguous from half of the period.

Duplex theory of binaural direction perception: The duplex theory postulates that for frequencies lower than approximately 1 kHz primarily interaural time differences are used, whilst for higher frequencies primarily interaural level differences are used.

Head Related Transfer Function (HRTF): The HRTF defines the filter behavior that head, body and outer ear impose on the incoming sound waves. The sound waves are distorted by reflections in the ear conch. The sound color changes depending on the direction of sound which facilitates the front / back and bottom / top location of sound sources. The distortion of the sound waves is used by the brain to estimate the direction and distance of the sound source and is the basis of monaural directional hearing [3].

Directional hearing is the subject of medical and technical research [2, 4]. In [5] the Digital Sound Display is presented which provides a visual representation of audio information as an overlay for games. Hearing aid manufacturers try to leave the ear canal as free as possible so that the natural directional perception is preserved. The company Audio Analytic (audioanalytic.com) deals with the classification and visualization of sound events of everyday life. Audio Analytic claims to have a data set of several million sound events, and advertizes it for applications such as surveillance of residential homes. The commercially available "Sound Shirt" from CuteCircuit (cutecircuit.com) transforms music into haptic impressions and thus enables an alternative perception of music for the hearing impaired [6]. The "ActiveBelt" [7] is a belt-type wearable tactile display that transmits directional information for navigational purposes. [8] deals with tactile displays such as torso displays and discusses the human factors.

3 First Prototype

A first prototype has been built to demonstrate the technical feasibility and to communicate the idea. It is based on a Raspberry Pi with a 3.2 inch touch display, 4 lavalier microphones and USB sound cards and is depicted in Figure 1. It shows the open prototype exposed to a synthetic sound source, and a possible future miniaturized platform (left).

The software is client-server based with a web front-end as presented in [9]. The main page on the screen shows four larger dots, placed on the middle of the display's edges. They are colored in blue and represent the spatial configuration of the four microphones as well as their measured sound level. The dots increase in size with higher sound levels, already hinting a direction to the user. A further larger dot, colored in red, indicates the estimated direction of a prominent sound. In Figure 1, it can be seen in the upper left corner of the display. In order to be useful for the task at hand, users need to be able to have an idea of the localization of past sound events. We use a special visualization for this in the middle of the display, with smaller dots, also

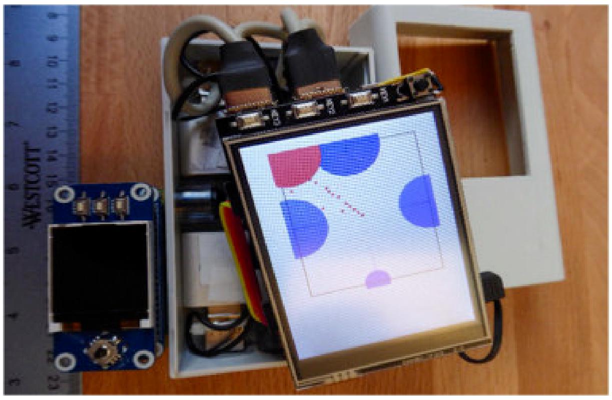

Figure 1 – A first prototype based on a Raspberry Pi with 4 USB microphones and a display.

colored red. They represent the history of past sound events. With each new localized source of sound, the outer red dot becomes small and moves inward until it disappears in the middle.

Sound processing in the first prototype has been done as follows. The algorithm is based on the calculation of level differences using classical signal processing algorithms. Audio chunks at a fixed length are retrieved from the sound cards. Various algorithms for smoothing and sound level and sound energy calculation were tested. A plain maximum value calculation proved to be useful, dropping the first and last samples of each audio chunk. The quadrant is determined where the maximum sound level is to be expected. Implausible quadruples are discarded. A geometrical consideration is used to determine the polar angle. Again, a plausibility test is used to decide about the validity of the result. Needless to say, the microphones have to be calibrated beforehand. The device only handles one active sound source and works best with a synthetic sound stream. No classification of sound types is performed.

Initially, USB microphones were used to avoid sound cards. One issue has been that the inexpensive USB microphones did not have very consistent frequency responses, despite being of the same make and model and being sourced simultaneously. Another difficulty comes from the fact that the four microphones do not necessarily sample in perfect synchronization: as all four are independent sound cards to the computer, they employ their own sample rate generation. Even if they would have the exact same sample rate, or rates similar enough for small time spans, they still may sample at different times. This behavior makes a good estimation of directions, for example via cross correlation or time of arrival methods, difficult if not impossible.

4 Second Prototype

An obvious solution is to connect several digital microphones directly to the Raspberry Pi, and have the sampling under tighter control. Unfortunately, the GPIO header only exposes a single inter-IC-sound (I2S) channel, the interface commonly used to attach sound hardware such as digital microphones. I2S is designed to only support two channels, usually left and right, with varying bit depths. The I2S interface to the Raspberry Pi's system-on-chip (SOC) accepts two channels, with 8 to 32 bits sample depth [10]. Monaural microphones typically feed just one of the channels on an I2S interface. In consequence, only one microphone can be directly connected to the SOC, or put the other way, four I2S interfaces would have been necessary to circumvent some of the problems from the first prototype. As there are 32 bits available for two input channels, some self-designed hardware could paste the output from several synchronously triggered microphones into one frame, for example by using an FPGA or micro-controller. Once received by the SOC, the individual channels would be split again by software. This way up to 4 microphones with 16 bit resolution (or 6 with 10 bit or 8 with 8 bit resolution) could be packed into the existing I2S of the Raspberry Pi. Another obvious solution would be to use more processing on the micro-controller side and use another interface such as USB to connect to the Raspberry Pi.

A search for already available multiple microphone solutions for the Raspberry Pi revealed that several products are on the market, generally in circular shape and primarily targeted for supporting voice assistants. We decided to go with hardware from Matrix Labs LLC, the Matrix Creator and the Matrix Voice. They both connect eight SMD Microphones via a Xilinx Spartan FPGA to a Raspberry Pi's GPIO port. In addition, the boards feature a ring of RGB LED that are suitable for easy instant visualization during development. While the Matrix Creator has a Cortex-M3 micro-controller and many sensors including an inertial measurement unit (IMU) connected as well, the Matrix Voice can optionally be equipped with an ESP32 micro controller with Bluetooth and WiFi, but lacks additional sensors. In essence, both may be operated also without the Raspberry Pi, but either miss Bluetooth/WiFi connectivity or an IMU. The IMU of the Creator is however also available as a breakout board and so may be added to the Matrix Voice externally, or, vice versa an ESP32 module with Bluetooth/WiFi capability could be added to the Matrix Creator. Among the multiple microphone products it is popular to use the Open embeddeD Audition System (ODAS) [11] for demonstration purposes, and so does Matrix Labs. It does processing of multiple microphone input including sound source tracking and is designed to run on embedded hardware. For our second prototype, we swap our own sound processing for the seemingly superior ODAS and concentrate on visualization and system aspects next.

5 System design

We identify three scenarios as being worthy of consideration in our research for the hearing impaired:

1. A personal device that users carry with them. When coming into an uncomfortable situation, they should be able to look at the visualization of sound in reasonable discretion.

2. When expecting dangerous situations that need instant reaction, such as when walking on the street, users may also almost continuously need to monitor the visualization in their field of view.

3. A display integrated in the vehicle that 'listens' on the vehicle's outside when users are driving in the car or riding on a bicycle. It shall differentiate emergency car signals, blowing of horns, ringings of bells from bicyclists and trams, or other approaching vehicles etc.

In the first scenario, smartphones come to mind first. Smartphones have typically at least two microphones, one to record the user's voice, and one for echo canceling in hands free mode (typically on the opposite side, close to the speaker), but not more. This is not sufficient for our purpose, as we would like to have more microphones for the detection of direction of sounds. With the omnipresence of smartphones, the display serves us well, however an application running on the device would need to be activated very quickly for the first use case. A microphone array could be strapped to the smartphone's back (taking into account the finger positions of the user holding the device) and connected to the phone for example via Bluetooth. Our prime candidate for the first use case however is a smart watch. It may show both time and the visual hearing aid simultaneously, and a quick glance at the watch is unobtrusive and socially acceptable behavior in most situations.

The first use case is difficult in so far as users are also interested in past sound events. They hear something and take out the phone or look at their watch at a time the event has already passed. Our visualization of past sound is designed to be suitable for exactly that. Attaching the microphones to the display is however not a good idea, as we are not interested in the sound inside a users pocket or arm cuffs. The microphones therefore must be separated from the display solution to offer continuous recording of the sound event history. We believe that the microphone should be placed naturally, i.e. roughly at the height of users' ears. In the first step, a baseball cap or other type of hat will serve as the base for our microphones. Later we may integrate microphones into the frame of eyeglasses or into clothing for a more unobtrusive solution.

The second scenario can possibly be covered with a solution that satisfies the first scenario. Another candidate for the constant monitoring is a head mounted display (HMD). HMD are socially less acceptable as the controversy around Google Glass has shown. For the hearing impaired, there may however be a solution that complements or replaces the solution for the first scenario. After Google has shut down the sales to consumers, the hardware for Google Glass is still available for industrial use cases through re-sellers. Also other manufacturers offer see-through HMD at comparably low prices, such as the Epson Moverio line of products. Obviously, using HMD has the benefit of integrating the microphones with the display solution – however, HMD tend to be less comfortable to wear over a longer period because of the weight, and it remains to be seen whether the displays disturb the view sufficiently to be a safety risk in themselves.

In the last scenario, one aspect to overcome is the placement of the microphones so that they are not affected by noises from the air stream. More futuristic designs will consider a display that covers the full windscreen or making use of HMD, but the visual hearing aid may also be integrated into existing display solutions. There are solutions on the market since around 2014 that visualize the outside of a car from a bird's-eye-view by stitching together four camera streams placed around the car such as Continental's ProViu ASL360. The so-called surround view helps drivers maneuver their vehicles when driving up to a loading ramp or getting into tight parking lots. The visualization of sound events could be integrated in such a display or be independently rendered without.

Tactile information has been previously used for giving directional or navigation information to users [8, 7]. One could imagine that users wear a belt equipped with about eight or ten vibrotactile actuators that indicate the sound direction in another modality. Obviously, direction information could also be presented multimodally, i.e. users could wear a belt with tactile actuators for a coarse and immediate directional 'feel' of a sound event, and consult the display for more detailed information or for consulting the history of events. We believe such a combination would be equally suitable for all three scenarios. Still, for our initial prototypes we do not plan to use other modalities for conveying the direction information to users.

All three scenarios would benefit from a classification of sound events so that users can also see different acoustic characteristics apart from their occurrence in time. Especially when multiple sound sources occur simultaneously, it may be difficult to connect a visual indication to the corresponding sound sources. We consider two solutions here: first, a classification into (semantic) classes such as speech / non-speech, male / female voice, blowing of the horn / ambulance signaling / ringing of a tram / of a bicyclist, barking of a dog etc. Different classes may be signaled by icons or by different colors, so that users will learn their meaning over time. Second, it may be useful to directly encode the spectral characteristic of sound into a color representation / the spectrum of color. Note that this approach is similar to the 1970s light organs found in discotheques of the time and therefore somehow already familiar to the elderly. The solution also relies on the fact that human beings are able to learn, so rather than training an artificial neural network to differentiate between a potentially unlimited number of classes, we could simply let the users learn to differentiate using representations in color. It has to be seen however whether the three color channels blue, green, red are sufficient to convey the complexity of a sound spectrum.

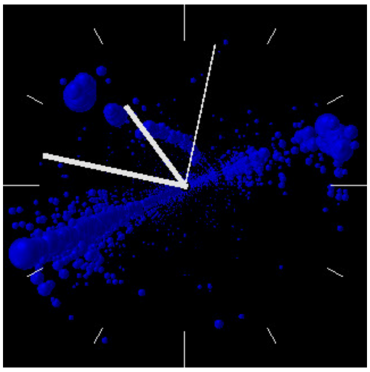

Figure 2 – Visualization of sound (music from left, a speaker to the right, and one moving front to left)

6 Visualization

Our current approach to visualization builds on the experience with the first prototype. We believe that a short memory of acoustic events is indispensible and that it should be accessible to users at first sight, not only after some interactions. As the nature of the surrounding sound suggests, a circular visualization of the current situation seems most natural for a bird's eye perspective. In a 2D plane, the historic information can therefore sensibly only extend towards the inner or outer areas while maintaining their directional information. As the most recent sound events seem to be more important than those about to be dismissed, it seems most natural to make recent sound events larger than those from the past. Therefore the most recent events are shown on the outside of the circular display and over time they drift towards the center while getting smaller and smaller. The most natural animation towards the center is not a linear movement but one that follows $1/t$, displaying a theoretically unlimited time span. This is the same way distance is handled in central projection, and as with distance in images, the age of past sound events is limited by the resolution in the display center. We found that keeping the history of sound events for one minute or even 30 seconds is probably sufficient for daily situations. It gives reasonable visibility for the more recent events also on smaller display sizes. As can be seen in Figure 2, the display shows the currently active sound sources as well as the direction (and even movement) of past sound sources.

As sounds do not have a physical manifestation, there is no obvious visual representation. It may be abstract (a dot, a disk, a sphere, a square, a line, etc.) or it may be iconic (representing also the type of sound). We go with the idea of a sound bubble for the time being, i.e. a sound event is represented by a sphere the size of which depends on its intensity (energy). Our visualization is implemented in C++ using the OpenGL ES 2.0 API for graphical rendering. We foresee that we will have to deploy the visualization on a range of embedded devices during our experiments, so we do not want to compromize on graphics performance or portability. OpenGL ES is still the most widespread hardware-accelerated graphics solution in mobile devices. Sound events are processed as point fragments through the vertex shader with position and size calculated from sound direction and age in seconds. They are rendered as point textures in the fragment shader by coloring a greyscale texture image of a sphere. Using this implementation it is also easy to experiment with different colors or with iconic representations for different sound types. The visualization runs easily in realtime on a Raspberry Pi on sensible display resolutions and should therefore also be directly usable on smartphones and smart watches. For our practical experiments, we have embedded the visualization of sound events into faces of analog and digital clocks, see Figure 3.

The illustrations in Figures 2 and 3 display data from the very first stage of ODAS, called sound source localization (SSL). It also contains erroneous detections that can come from early reflections or other reverberation, or are simply glitches from the algorithm or noise from the

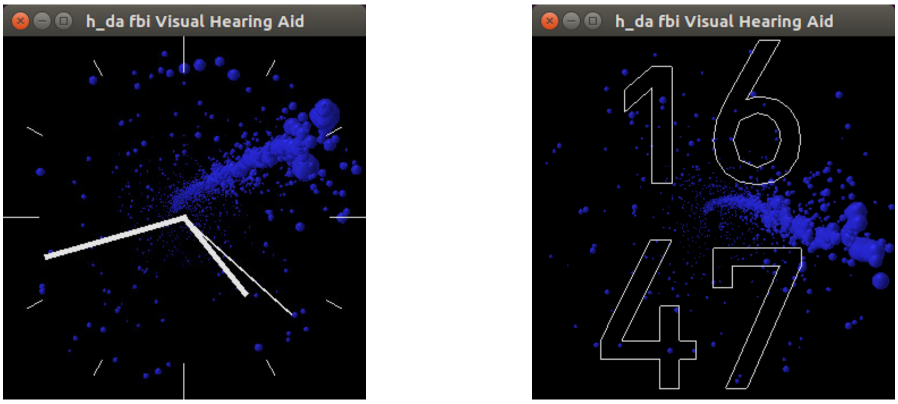

Figure 3 – Visualization of directional sound events embedded in watch faces

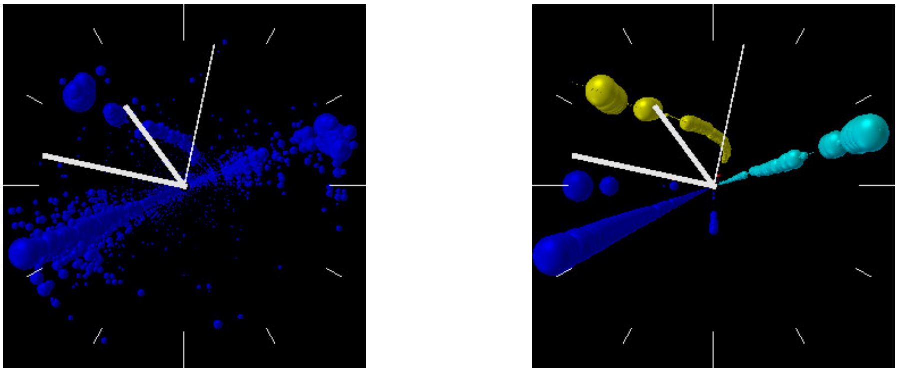

Figure 4 – Visualization of sound source localization versus sound source tracking

microphones. It would be desirable to have a clearer display where these reflections and reverberation do not contribute sound bubbles. ODAS also provides a Kalman filter stage called sound source tracking (SST) that takes the data from SSL, assigns an id to a detected sound source and tracks its movement with a given innertia [11]. Their implementation of a Kalman filter also takes into consideration that a sound source may be quiet for a short time (thus keeping the same id) or a new sound source may appear (receiving a new id). While we have not experimented exhaustively with the parameters of ODAS, we observe that this modeling is difficult with singular sound events, such as the sound from a dropping object. It works very well with human voices (for which ODAS is designed) or music from a speaker, but short sound events are filtered out completely, as they are not contributing energy long enough from their direction, see Figure 4. This is problematic for our application.

From a philosophical point of view, it is already interesting now to speculate about the role of a visual hearing aid: Shall it be a "visual microphone" that does not interpret at all what it hears? This would mean we should do as little processing and filtering as possible and just display the signals as directly as possible, like a measuring instrument does. It would then be left to the human brain to understand the information, make an interpretation of some noise as irrelevant, as reflections of sounds, but on the other hand also allow users to see traces of a fallen object's ting-a-ling in the noisy signal, or the weaker signals of more distant sounds. On the other hand, the visual hearing aid could aim to be an "intelligent companion" that truly understands what is going on around the user. It would clearly differentiate between different speakers (possibly even recognizing voices it had heard before, marking them with their initials or a photo of their faces), would classify sound events (such as the falling of a bunch of keys versus foot steps on a hard wood floor) and so on. The increased difficulty with processing by the visual hearing aid may pay back in reduced mental load for the user, who does not have to learn how to interpret the data. While it may be scientifically more interesting to pursue the more difficult task, even at the risk of making errors in the processing, we are largely undecided at this time and will try to answer the usefulness of different solutions and preferences in user tests. It may also turn out that the measurement instrument may be preferable for one user group while the intelligent companion approach may better fit another group of users.

7 Future Work

Our next steps will be integrating the second prototype in wearable devices and making user tests. We have plans for a partially automated user test setup, where the direction indication of test subjects is measured by our optical tracking system. Further experiments with various visualizations will yield most suitable visual representations of sounds and their history and show shortcomings of our solution so far. A classification of sound sources (human voices, cars, ring tones, horns and sirens) will help users to manage multiple simultaneous sound sources. Finally, we intend to deploy end-to-end classification for direct sound detection using machine learning methods. We are actively looking for industrial collaborators on the subject, as we see possible improvements in quality of live for patients suffering from limited directional hearing.

References

[1] SEEBER, B. U.: *Binaurales Hören mit Cochlea Implantaten. 36. Deutsche Jahrestagung für Akustik - DAGA 2010, Berlin*, pp. 27–35, 2010.

[2] EVERS, C.: *Richtungshören und Sprachverstehen im Störgeräusch bei Patienten mit unilateraler Atresia auris congenita*. Ph.D. thesis, Universität zu Lübeck, 2006.

[3] SAXENA, A. and A. NG: *Learning sound location from a single microphone*. In *IEEE International Conference on Robotics and Automation*, pp. 1737–1742. 2009.

[4] RASCON, C. and I. MEZA: *Localization of sound sources in robotics: A review*. *Robotics and Autonomous Systems*, 96, pp. 184–210, 2017.

[5] HOLLOWAY, A., R. DEARMOND, M. FRANCOEUR, D. SEAGAL, A. ZUILL, and S. KURNIAWAN: *Visualizing audio in a first-person shooter with directional sound display*. In *1st Workshop on Game Accessibility: Xtreme Interaction Design (GAXID'11) at Foundations of Digital Games, Bordeaux, France*. 2011.

[6] VOGLER, B.: *Alternative Soundempfindung für Gehörlose und Schwerhörige – Das Sound Shirt*. Bachelorarbeit, Hochschule der Medien Stuttgart, 2017.

[7] TSUKADA, K. and M. YASUMURA: *ActiveBelt: Belt-type wearable tactile display for directional navigation*. In *International Conference on Ubiquitous Computing*, pp. 384–399. Springer, 2004.

[8] VAN ERP, J. B.: *Tactile navigation display*. In *International Workshop on Haptic Human-Computer Interaction*, pp. 165–173. Springer, 2000.

[9] YAGER, N.: *LittleSleeper: A baby sleep monitor using a Raspberry Pi and Python*. Tech. Rep., AICBT Consulting Ltd., 2015.

[10] *BCM2835 ARM Peripherals*. Broadcom Europe Ltd., 2012.

[11] GRONDIN, F. and F. MICHAUD: *Lightweight and optimized sound source localization and tracking methods for open and closed microphone array configurations*. *Robotics and Autonomous Systems*, 113, pp. 63–80, 2018.