WEARABLE SOUND LOCALIZATION ASSISTIVE DEVICE FOR THE HEARING IMPAIRED

Mauricio Kugler*, Hiroyuki Sakamoto* and Masatoki Suto**

* Department of Computer Science & Engineering, Nagoya Institute of Technology, Nagoya, Japan
** Department of Architecture & Design, Nagoya Institute of Technology, Nagoya, Japan
e-mail: mauricio@kugler.com

Abstract: The sense of hearing can provide immediate information about remote events, even when outside of the field of vision and beyond obstacles, facilitating functioning in uncontrolled environments. Hearing impairment can thus have a huge disabling effect on an individual. This paper proposes a wearable self-contained dedicated device capable of full-plane sound localization. The system, shaped as a glass frame, uses only four microphones spaced by 10 mm, and is initially targeted at a resolution of 45°. The individual binaural angles are calculated by a process loosely based on the human hearing system. These angles are then combined in order to determine the final direction. A prototype of the proposed system was implemented using 3D printing and MEMS microphones. Experiments with the prototype in a reverberant environment show an error of 6.73° when it is tested standalone and 21.16° when tested in a dummy head.
Keywords: Wearable device, sound localization, hearing impairment, hardware prototype.

Introduction

Sense of hearing is one of the two senses that can provide immediate information about remote events, along with sight. Unlike the latter, which is limited to line-of-sight applications, hearing can sense events in any direction and it is good in bringing them to attention of an unaware person [1]. Hearing impairment can thus have a huge disabling effect on an individual. The degree of hearing loss can range from a minor decline to complete loss of the hearing ability. Although hearing impairment occurs primarily among seniors as a natural result of aging, it can occur in any age group.

The human hearing system is capable of localizing sound sources in three-dimensions, outside of the field of vision and beyond obstacles, thus providing a more complete situational awareness that facilitates functioning in uncontrolled environments. However, conventional hearing aids and implanted devices usually only improve sound recognition, but not sound localization [2]. An artificial system capable of localizing a sound source would hence be beneficial in many fields, but especially for the hearing impaired.

Wearable technologies have been receiving increasing attention, fueled by advances in battery technology and ever increasing computing power of mobile processors. Wearable systems capable of sound localization have been proposed in the literature, but usually depend on external processing, have high computational demands or poor frequency range.

For instance, Hwang *et al.* [3] used sound localization as complementary information for face recognition in their head-mount-display (HMD) based system. The system uses two microphones separated by 4 cm. Jain *et al.* [2] recently presented a detailed study about visual feedback for sound localization. Their motivation is that, although hearing aids and cochlea implants improve sound recognition, the user is still deprived of sound localization capabilities. The proposed system is implemented in a head-mount-display and relies on an external 64-microphone array for sound localization. Among their conclusions, most hearing-impaired participants reported that higher resolution in localization was strongly preferred, that currently available HMD are not appropriate for such application, and thus customized designs are needed.

This paper proposes a wearable self-contained dedicated device capable of full-plane sound localization. The system uses only four microphones, can localize high frequency sounds and is initially targeted at a resolution of 45°. A prototype of the proposed system was implemented and used for experiments.

Materials and methods

Sound localization methods can be divided into three groups, methods based on head-like structures and Head Related Transfer Functions (HRTF), methods using two or more microphones and utilizing binaural cues, and methods based on microphone arrays and beamforming techniques.

The HRTFs summarize the filtering effect of head, torso and pinna, for each direction and distance of a sound source [4]. Localization then becomes a search for the pair of left and right HRTF that produces the highest correlation value when applied to the incoming signals. Obtaining such functions, however, is a complex process. Beamforming based implementations require multi-microphone arrays with specific geometries and are notorious for their high computational complexity [5].

Human hearing is based on binary cues, which correspond to differences in a sound arriving to each ear. Interaural Level Difference (ILD) focuses on

differences in sound energy level as a cue for localization, while Interaural Time Difference (ITD) corresponds to the Time-Difference of Arrival (TDoA) of sounds between the two ears [6]. The latter is the most direct cue for estimating the direction of the sound source and does not impose so many restrictions to the structure of the system when compared to other methods.

Reducing dimensions of the device presents challenges for the sound localization method and the implementation of the device. As the distance between the microphones has to be small and no obstacle is present between them, differences in spectra and level between the signals in each microphone are neglectable. Hence, ILD and HRTF methods are not suitable for the intended implementation, and only ITD can be used.

Binaural sound localization – Among the most popular alternatives for estimating the TDoA is the Generalized Cross-Correlation (GCC) [7]. This method, however, is computationally expensive. Instead, the proposed system uses the well-know Jeffress model [8] to calculate the TDoA. In its basic form, the model contains two antiparallel delay lines that receive the spike trains corresponding to the sound coming from the left and right ears. The delay lines are connected to coincidence detectors along their lengths, which generate a spike if its two inputs receive a spike simultaneously. If there is no delay between the two sound signals, the majority of the spikes will coincide at the center of the delay lines. For instance, the more to the left the sound source is, the earlier the spikes from the left ear arrive at the delay line and the coincidence happens closer to the right side of the model. Thus, the temporal information is transformed to spatial information.

The space between the microphones directly affects the Jeffress model regarding bandwidth and resolution. Considering a Time-Difference Extractor (TDE) based on the Jeffress model with n coincidence detectors and n positions in each of the anti-parallel delay lines, the index z of the coincidence detector corresponding to a delay of Δt is given by:

$$z = \left\lceil \Delta t \frac{n f_s}{2n} + \frac{n}{2} \right\rceil = \left\lceil \frac{f_s \Delta t + n}{2} \right\rceil \quad (1)$$

where $\lceil \cdot \rceil$ represents the ceiling operator, f_s is the sampling frequency and n can be calculated as follows:

$$n = \left\lceil \frac{2 d f_s}{v} \right\rceil \quad (2)$$

The quantized angle θ_q represented by the index z is given by:

$$\theta_q = \arccos\left(\frac{2z}{n-1} - 1 \right) \quad (3)$$

If the signal wavelength λ is shorter than the distance between the microphones, spatial aliasing occurs [9]. The maximal frequency $f_{\max}$ for a given distance $2d$ can be roughly determined as:

$$v = \lambda f_{\max} \therefore f_{\max} = \frac{v}{2d} \quad (4)$$

Proposed system – The proposed method independently calculated the binary angles of the three pairs of adjacent microphones and combines them into a final direction value. The process to calculate each angle is loosely based on the human hearing system and is shown in Figure 1. The input signal (left and right channel) is applied to a bandpass filter bank which divides it in 128 frequency bands from 100 Hz to 8 kHz. Each channel is then upsampled in order to increase its spatial resolution, according to equation (2). For each channel, a spike train is generated using a zero-crossing spike generator and applied to the TDE. The output spikes are counted and the maximal value indicates the TDoA of a given frequency channel. Finally, the values are converted to angles using equation (3) and averaged.

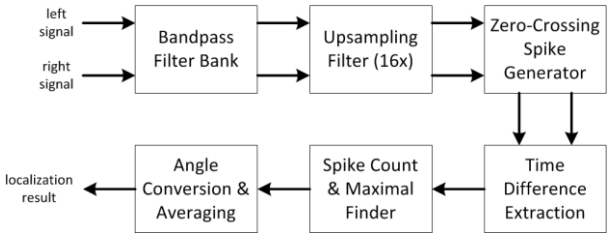

Figure 1: Summarized block diagram of the binaural angles calculation.

The final direction is obtained as follows. Let the four microphones be named M_1, M_2, M_3 and M_4, and let the points located at the middle of each connection between pairs of adjacent microphones be *A*, *B*, *C*, as shown in Figure 2. Also let angles α, β, γ, δ be the angles formed between the horizontal axis and the sound source *S* at points *A*, *B*, *C* and *D*, respectively. Finally, let point D be placed at the origin of the coordinate system. The objective is to find $\delta = f(\alpha, \beta, \gamma)$. Angle δ will be positioned at the crossing of the 3 angles, but, intuitively, only 2 angles are necessary for the calculation.

First, the tangents of angles α, β and δ are given by:

$$\tan\delta = \frac{S_y}{S_x} \quad (5)$$

$$\tan\alpha = \frac{S_y - A_y}{S_x - A_x} \quad (6)$$

$$\tan\beta = \frac{S_y - B_y}{S_x - B_x} \quad (7)$$

where P_x and P_y denote the Cartesian coordinates of a point *P*. In order to obtain δ, it is necessary to isolate S_x and S_y from equations (6) and (7):

$$S_x = \frac{A_x \tan\alpha - A_y - B_x \tan\beta + B_y}{\tan\alpha - \tan\beta} \quad (8)$$

$$S_y = \frac{A_y \tan\beta - A_x \tan\alpha \tan\beta - B_y \tan\alpha + B_x \tan\alpha \tan\beta}{\tan\beta - \tan\alpha} \quad (9)$$

Substituting (8) and (9) in (5) gives:

$$\delta = \arctan\left(\frac{A_y \tan\beta - B_y \tan\alpha + (B_x - A_x)\tan\alpha \tan\beta}{A_y - B_y - A_x \tan\alpha + B_x \tan\beta}\right) \quad (10)$$

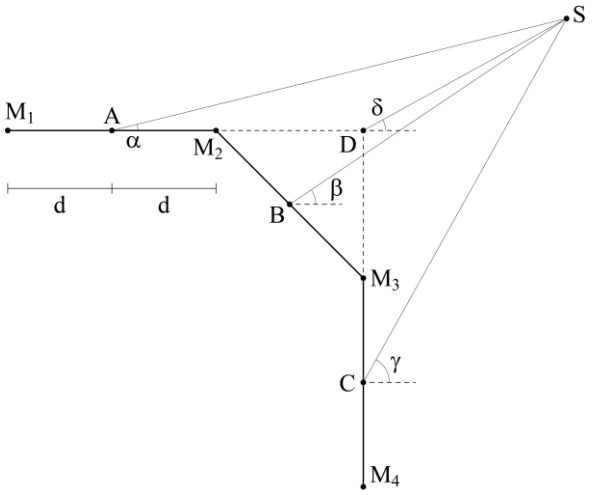

Figure 2: Position of the four microphones and binaural angles.

Equation (10) requires only two of the binaural angles to be used. As equation (3) presents a low resolution for angles close to -90° or +90°, the two angles further from these regions are the chosen ones, thus reducing the overall error of the system.

The proposed system sets the distance 2*d* between adjacent microphones to just 10 mm, considerably smaller than the majority of methods proposed in the literature. This has the advantage of enabling a compact system and increase the frequency range, according to equation (4). The downside is a reduction in resolution, as shown in equation (2). The adopted solution was to increase the sampling frequency, which compensates for the reduction of *d*. This is accomplished by the upsampling stage shown in Figure 1.

Implementation – Similar to the human hearing system, the performance of a sound localization device can be influence by its geometry, due to effects such as reflection and refraction of the sound on its parts. Thus, in order for the experimental results to be as realistic as possible, a prototype of the proposed system was created using 3D printing, as shown in Figure 3. It contains four INMP401 MEMS microphones, separated by 10 mm, positioned on the right-front corner of the frame, similar to the diagram in Figure 2. The microphones are powered by a CR2032 battery, placed on the left-front corner. This initial prototype does not include any embedded processing unit and it is connected to external audio recording equipment by cable coming from the back of the right temple.

Figure 3: Wearable sound localization prototype, placed on the dummy head.

Results

The experiments were performed using the prototype shown in Figure 3, connected to a Roland Octa-Capture Digital Audio Interface. Each recording consists of 3 s of white noise, in directions from 0° to 360° at steps of 5°. In all experiments, the sound source was stationary, and the prototype, placed 3 m away, was rotated by a Kohzu RA07AW rotating stage. The experiments were made in a non-anechoic standard meeting room, measuring 12.9×6.7×2.8 m, with a background noise of 32 dB and reverberation time RT_{60} of 0.55 s.

Localization accuracy – At first, the prototype was tested alone, without being placed on the dummy head. The localization results are shown in Figure 4. The average error was 6.73° and the maximal error was 20.41°.

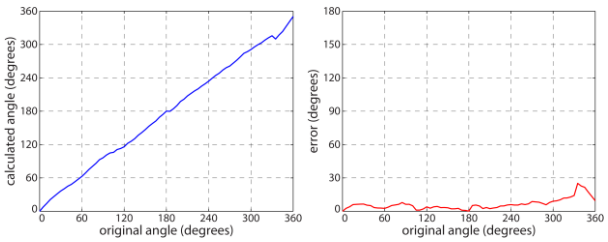

Figure 4: Localization results (blue) and angular error (red) of the stand-alone prototype.

After, the prototype was tested on the dummy head. The results are shown in Figure 5. The average error was 21.16° and the maximal error was 158.77°.

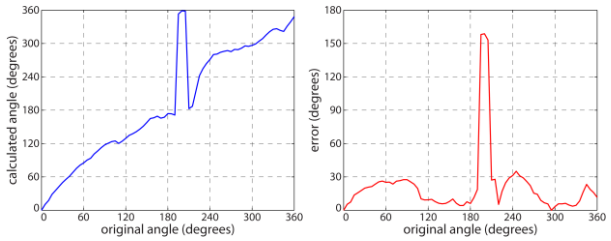

Figure 5: Localization results (blue) and angular error (red) of the prototype on the dummy head

Attitude Estimation – The objective of this experiment is to verify the maximal inclination of an individual's head (pitch and roll) while walking normally. If severe inclination values are observed, this could be a problem for the proposed wearable device, as the direction values would be invalid.

A 3-axis accelerometer MMA8451Q was placed in a cap, which was used by several individuals while walking a pre-determined path. The acceleration value was acquired 12.5 times per second, low-pass filtered with 0.5 Hz of cut-frequency and later converted to inclination angles, considering the earth's gravity. Five volunteers, unaware of the objective of the experiment, performed the experiment. The results are shown in Figure 6.

The average and standard deviation for the pitch angle were, respectively, -1.72° and 10.79°, while for the roll angle, average and standard deviation were -1.57° and 8.74°, respectively. As these values are considerably small, the resulting differences in the calculated horizontal directions are neglectable.

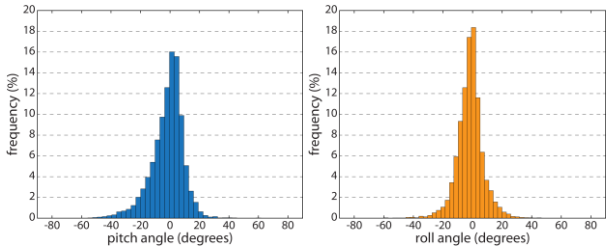

Figure 6: Pitch and roll histograms for the walking pattern of five subjects.

Discussion

This paper proposes a self-contained wearable sound localization device for hearing impaired people's assistance. A hardware prototype was implemented and used to conduct experiments.

The results shown that, when tested alone, the glass-frame prototype achieved high accuracy, more than enough for the intended 45° resolution, even in a reverberant environment. However, a large error was observed when the azimuth was around 200° in the case were the system was used on a dummy head. In order to try to prevent such error, several types of non-reflective materials in different shapes were added to the dummy-head, without success. Nonetheless, these experiments showed that, when the sound is coming from behind the dummy head, several paths of different distances are created for the sound to reach the microphones, resulting in inconsistent time-delays. Possible solutions for this problem include the use of the sound pressure of each microphone, or the difference in pressure between the microphones, as an extra clue for determining the final direction of the sound source.

Future works also include the implementation of a new prototype with the microphones actually embedded in the glass frame, as well as an FPGA-based processing unit also installed within the frame.

Acknowledgements

The authors would like to thank Jana Makovníková and Shohei Fujimura for their contributions to this research.

References

[1] Kim S-W, Lee J-Y, Kim D, You B-J, Doh NL. Human localization based on the fusion of vision and sound system. Proc. of the 8th International Conf. on Ubiquitous Robots and Ambient Intelligence; 2011. p. 495-498.

[2] Jain D, Findlater L, Gilkeson J, Holland B, Duraiswami R, Zotkin D, Vogler C, Froehlich JE. Head-Mounted Display Visualizations to Support Sound Awareness for the Deaf and Hard of Hearing. Proc. of the 33rd Annual ACM Conf. on Human Factors in Computing Systems; 2015. p. 241-250.

[3] Hwang B, Kim C-S, Park H-M, Lee Y-J, Kim M-Y, Lee M. Development of Visualizing Earphone and Hearing Glasses for Human Augmented Cognition. Proc. of the 18th International Conf. on Neural Information Processing; 2011. p. 342-349.

[4] Cheng CI, Wakefield GH. Introduction to Head-Related Transfer Functions HRTFs: Representations of HRTFs in Time, Frequency, and Space. Proc. of the 107th Audio Engineering Society Convention; 1999. p. 231-249.

[5] Valin J, Michaud F, Rouat J. Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering. Robotics and Autonomous Systems. 2007; 55 (3): 216-228.

[6] Keyrouz F. Advanced Binaural Sound Localization in 3-D for Humanoid Robots. IEEE Transactions on Instrumentation and Measurement. 2014; 63(9): 2098-2107.

[7] Knapp CH, Carter GC. The generalized correlation method for estimation of time delay. IEEE Transactions on Acoustics, Speech, and Signal Processing. 1976; 24(4): 320-327.

[8] Jeffress LA, A place theory of sound localization. Journal of Comparative and Physiological Psychology. 1948; 41: 35-39.

[9] Blauert J. Spatial Hearing: The Psychophysics of Human Sound Localization. Rev. ed. Cambridge: The MIT Press; 1983.