

Assessment of sound source localization of an intra-aural audio wearable device for audio augmented reality applications

Narimene Lezzoum and Jérémie Voix

École de technologie supérieure,

University of Québec

Montreal, Canada

Email: narimene.lezzoum@ens.etsmtl.ca, jeremie.voix@etsmtl.ca

Abstract—This paper presents a study on the effect of an intra-aural Audio Wearable Device (AWD) on sound source localization. The AWD used in this study is equipped with a miniature outer-ear microphone, a miniature digital signal processor and a miniature internal loudspeaker in addition to other electronics. It is aimed for audio augmented reality applications, for example to play 3D audio sounds and stimuli while keeping the wearer protected from loud or unwanted ambient noise. This AWD is evaluated in terms of ambient source localization using three localization cues computed using signals played from different positions in the horizontal and sagittal planes and recorded in the ear canals of an artificial head with and without a pair of AWDs. The localization cues are: the interaural time difference, the interaural level difference, and the head related transfer functions. Results showed that the used AWD does barely affect the localization of low frequency sounds with localization error around 2° , and only affects the localization of higher frequency sound depending on their position and frequency range.

I. INTRODUCTION

Recently, a significant research effort has been seen for the development of intra-aural AWDs that combine different functionalities such as audio navigation, hearing protection, music playback, face-to-face and long distance communication while maintaining awareness to ambient sounds [1], [2]. These AWDs are capable of protecting the ear from background noise while simultaneously transmitting warning signals to the wearer's ear [3] and enabling face-to-face communication by detecting and transmitting enhanced speech signals to the protected ear [4].

AWDs are designed to maintain ambient sound awareness, thereby enabling their wearers to hear and localize the position of specific sounds sources such as speech and alarms while being protected from noise or listening to music through earphones, an audio experience often referred to as *audio augmented reality*.

Sounds generated by multiple physical sources propagate in the air and arrive from different directions to the human ears. The human auditory system is capable of accurately localizing physical sound sources in 3-D environments using different acoustical cues resulting from the contact of the incident sound with the human head, pinna and ears [5]. In sound source localization, there are two primary cues used to determine

the direction of the sound source in the horizontal plane: the interaural time difference (ITD) and the interaural level difference (ILD). ITD and ILD are binaural cues and permit to determine the location of the sound source by comparing in the horizontal plane the signals received by the left and right ears. In early 1900, Lord Rayleigh introduced a principle of sound source localization that combines ITD and ILD cues depending on the frequency of the incoming sound called: the "duplex theory" [6]. This theory states that since low frequency sounds have long wavelengths and the head is very small compared to that wavelength, the sound is not very altered by the head, thus the auditory system uses the difference of time of arrival to localize the sound source. In addition, it states that since high frequency sounds have short wavelengths, the head acts as an acoustical shadow and attenuates these sounds, thus the auditory system uses the level difference between the two ears to localize the sound source [7], [5], [8]. Thus, ITD cues dominate localization of low frequency sounds and ILDs dominate the localization of high frequency sounds.

While ITD and ILD represent primary binaural cues for source localization in the horizontal plane, the head related transfer function (HRTF) represents the primary spectral cue for sound localization in the sagittal plane in addition to the horizontal. It permits to compare the signals received at the ears to a reference signal (the source signal) and provides information such as elevation and distance between the source and the ears [9].

When wearing an AWD capable of transmitting specific external sounds to the ear, the ITD, ILD and HRTF cues may be different from those of an open ear, as they are affected by the frequency response of the electroacoustic components used and these differences may yield to a non-accurate localization of the external sound sources in the horizontal and sagittal planes.

While a considerable research is conducted for sound source localization using multiple microphones and signal processing algorithms [10], [11], the less expensive solution for AWDs equipped with miniaturized electronics is the development of an acoustically "transparent" intra-aural AWD that affects the least the localization cues and provides audio augmented

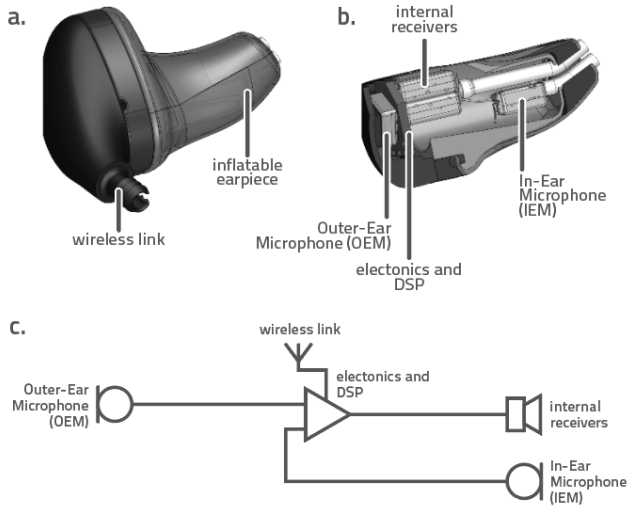


Fig. 1. The auditory research platform built on a custom molded earpiece (a), its electroacoustic components (b), and equivalent electrical schematic (c).

reality to the user.

To assess the acoustic transparency of an intra-aural AWD, acoustic localization cues need to be calculated and compared to those when no device is used. Therefore, the current paper presents a study on the effect of an intra-aural AWD on sound source localization using ITD, ILD and HRTF cues. This study is part of a project undertaken at the CRITIAS Research Chair in In-Ear Technologies for the development of a *bionic ear* [2].

This paper is organized as follows: Section II introduces the intra-aural AWD used in this study. Section III presents the localization cues namely the ITD, ILD and HRTF. Section IV presents the experimental setup, while Section V shows the results and Section VI discusses them and concludes the work.

II. THE AUDITORY RESEARCH PLATFORM (ARP)

The Auditory Research Platform (ARP) is used in this study as a representative AWD and is illustrated in Fig. 1. It is integrated with miniaturized electronics: an outer-ear microphone, a digital signal processor (DSP), an internal miniaturized loudspeaker, an internal microphone and a wireless radio link. In the current study, the external microphone is used to capture the surrounding sounds and the internal loudspeaker is used to transmit the captured sounds to the occluded ear. The DSP, internal microphone and the radio wireless link are not used for the present study.

For increased comfort and superior acoustics performance, this ARP is built on a custom molded earpiece obtained using the proprietary SonoFit™ instantaneous molded earpiece [12].

III. LOCALIZATION CUES

A. Interaural Time Difference (ITD)

The ITD corresponds to the time difference between a sound reaching one ear and the same sound reaching the other ear. Humans are able to detect changes in the ITD as small as

$20 \mu\text{s}$ [13] [14]. To determine the ITD, the delay between the signal perceived at one ear for example the right ear $x(n)_R$ and the signal perceived at the other ear (left ear) $x(n)_L$ is calculated as the lag of the maximum of the cross-correlation function as follows:

$$C_{LR}(k) = \sum_{n=0}^{N-1} x_L(n)x_R(n-k) \quad (1)$$

$$\text{ITD} = \arg \max(C_{LR}(k)) \quad (2)$$

In cases where the ITD is available, the sound source location can be estimated using the distance between the two microphones D placed inside the ear and the speed of sound c_{sound} as follows:

$$\theta = \arcsin \frac{c_{\text{sound}} \cdot \text{ITD}}{D} \quad (3)$$

B. Interaural Level Difference (ILD)

The ILD corresponds to the difference between the sound pressure level of a signal reaching one ear and the sound pressure level of the same signal reaching the other ear. It is calculated as follows:

$$\text{ILD} = 20 \log_{10} \left(\frac{x_L(n)}{x_R(n)} \right) \quad (4)$$

C. Head-Related Transfer Functions

HRTFs were first introduced by Wightman et al. in 1989 [15], and their goal is to measure the effect of the head and pinna's shape on the external sounds [16] [17]. HRTFs are measured as the ratio between the signal recorded in the ear canal of an artificial head $x(n)$ (or using a miniature microphone placed in the open ear-canal of a human subject) and the signal $y(n)$ recorded at a reference position that corresponds to the position of the center of the head. Thus the HRTF is calculated as follows:

$$\text{HRTF}(f) = \frac{X(f)}{Y(f)} \quad (5)$$

with $X(f)$ and $Y(f)$ the Fourier transforms of $x(n)$ and $y(n)$ respectively.

IV. EXPERIMENTAL SET UP

The experiments were conducted in a 95 m^3 semi-anechoic room equipped with an automated rotating arm that can move from -60° to 60° in the sagittal plane and from 0° to 360° on azimuth. At one of the end of the automated arm, a loudspeaker is placed for sound playback, as illustrated in Fig. 2.

Before starting the recordings with and without the ARP, pink noise was played through the loudspeaker and recorded using a reference microphone placed in the middle of the room at the center of the head location as illustrated in Fig. 2. The recordings were performed for each couple of azimuth/elevation position of the sound source: from 0° to

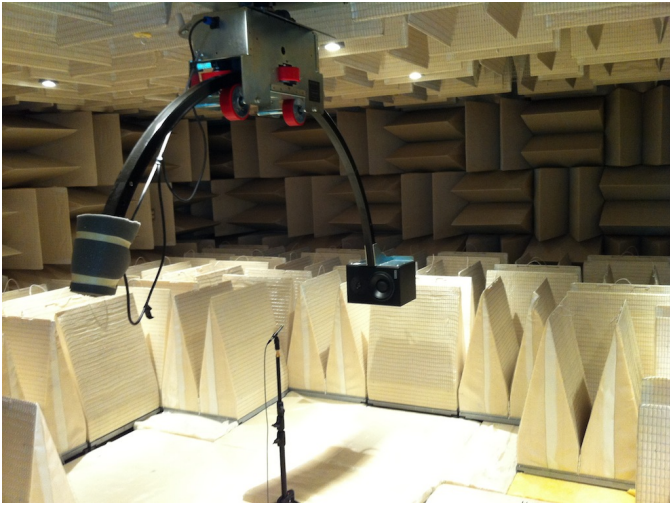


Fig. 2. The anechoic room set up showing the automated arm with a loudspeaker placed at one of its ends and a reference microphone placed in the middle of the room at the head center position.

360° on the horizontal plane with a step of 5°, and from -40° to 60° on the sagittal plane with a step of 10°.

After recording the reference signals, two recording experiments were conducted with a G.R.A.S 45CB (Holte, Denmark) artificial head. The head center was placed at the exact same position where the reference microphone was, facing 0° on the horizontal plane. In the first experiment, recordings were performed with open ears and in the second experiment the artificial head was fitted with a pair of ARPs.

A. Without the ARP (open ears)

Pink noise played through the loudspeaker was captured for each couple of azimuth/elevation position by the microphones located inside each ear canal of the artificial head. They were amplified and then recorded by the National Instruments PXI-6221 data acquisition (DAQ) board located outside of the room. Recordings were sampled at 44 kHz with 24 bits resolution.

B. With the ARP (occluded ears)

In this step, the artificial head was fitted with the pair of ARPs as illustrated in Fig. 3. Pink noise played through the loudspeaker was captured using the external miniature microphones of the ARP fitted in both ears of the artificial head. These signals were amplified and then sent directly to the internal miniature loudspeaker, transmitted to the ears, where they were captured by the microphones located inside each occluded ear canal of the artificial head and recorded with the data acquisition system mentioned above.

V. RESULTS

The recorded signals are analyzed in terms of ITD in the low frequencies (octave bands with center frequencies ≤ 1000 Hz), and ILD in the high frequencies (octave bands with center frequencies ≥ 2000 Hz). Since the ITD and ILD give information about the horizontal location of the sound

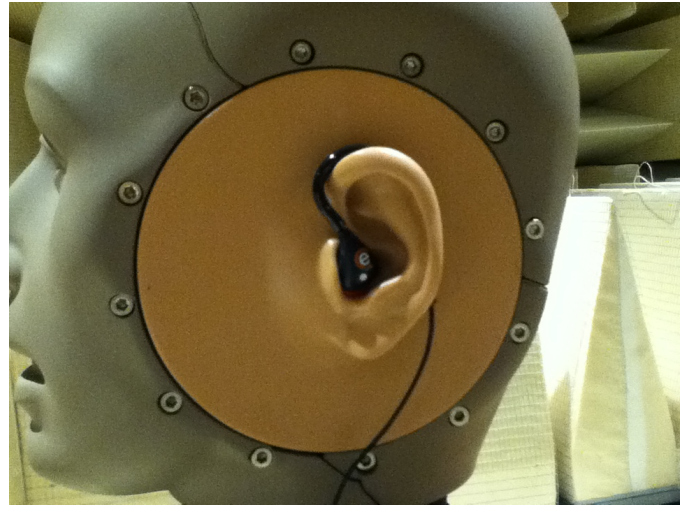


Fig. 3. G.R.A.S 45CB artificial head fitted with the ARP.

source, only signals recorded at 0° elevation are used. For HRTF analysis, recordings with varying elevation and azimuth angles are used and analyzed.

A. ITD

For each position of the sound source and for both scenarios (with and without the ARP), the ITD was calculated as showed in equations 1 and 2 for octave bands with center frequencies of 125, 250, 500 and 1000 Hz.

Fig.4 illustrates for all the positions of the sound source, the ITD when having an open ear and when wearing the device. This figure shows that the ITD when wearing the ARP is very similar to the ITD of an open ear across all positions of the sound source. The difference between the ITDs with and without the ARP is lower than 22 μ s which corresponds to a localization error around 2°. This localization error is calculated as the difference between the angle at which the sound was played and the estimated angle using equation 3, knowing that the distance between the two outer ear pinna of the artificial head D equals 138 mm.

The aforementioned small differences in ITD are explained by the fact that the ARP is an intra-aural device and the external microphone is very close to the ear-canal entrance.

B. ILD

ILD, measured with and without the ARP for all horizontal positions of the sound source are analyzed in octave bands with center frequency of 2000, 4000 and 8000 Hz. Fig.5 illustrates ILD results and shows that in both scenarios (with and without the ARP), the ILD increases when the frequency increases, which is due to the fact that the high frequencies are more affected by the head shadow. This figure also shows that the ILD in both scenarios depends on the position of the sound source, and ILD with the ARP has the same trends as the ILD without the ARP. In addition, from Fig.5 we observe that the maximum ILD occurs at different positions depending on the frequency: in some cases it occurs at 90° and

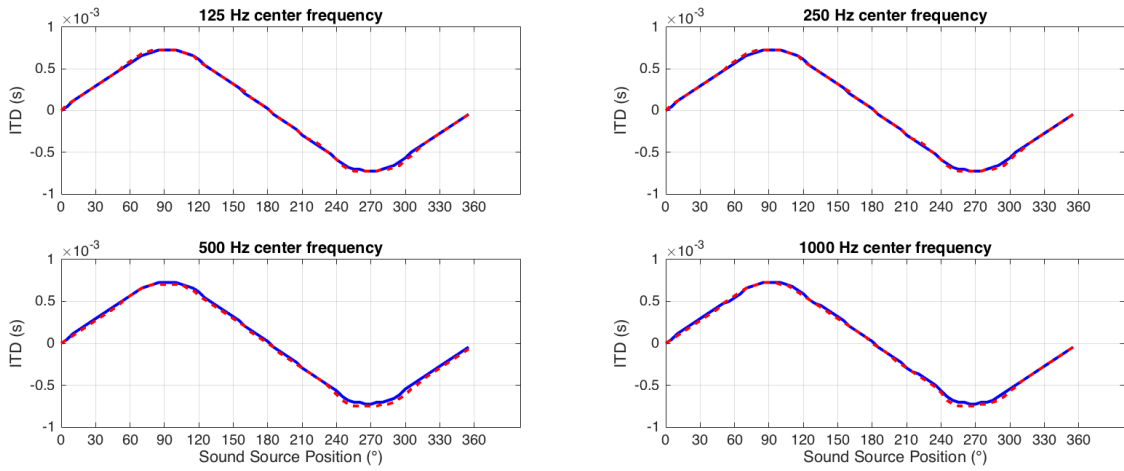


Fig. 4. Interaural time difference in octave bands measured with and without a pair of ARPs for each position of the sound source.

in other cases it occurs slightly before or after this positions. These observations are correlated with the results reported by Feddersen et al. [7].

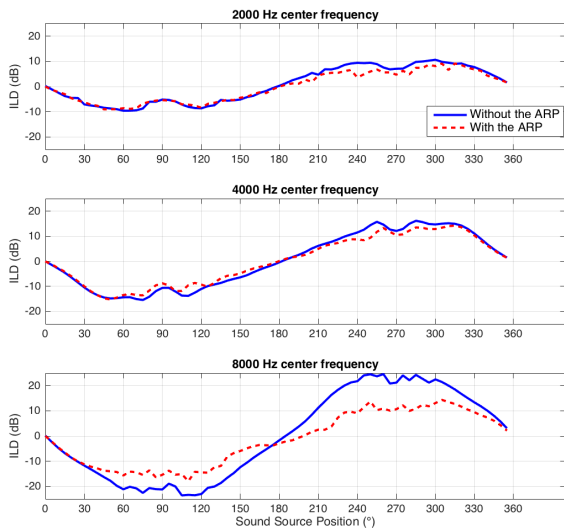


Fig. 5. Interaural level difference measured with and without the ARP in three octave bands: 2000, 4000 and 8000 Hz.

To see the effect of the ARP on ILD, the difference between the ILD calculated with and without the ARP is calculated and illustrated in Fig. 6. This figure shows that the difference between ILD with and without the ARP increases with the increase of the frequency, meaning that when wearing the ARP, the higher the sound frequency, the less accurate the localization is.

C. HRTF

Fig.7 illustrates an example of two HRTFs when the sound source is located at 0° azimuth and 0° elevation: one calculated with the ARP and the other one without the ARP. This figure

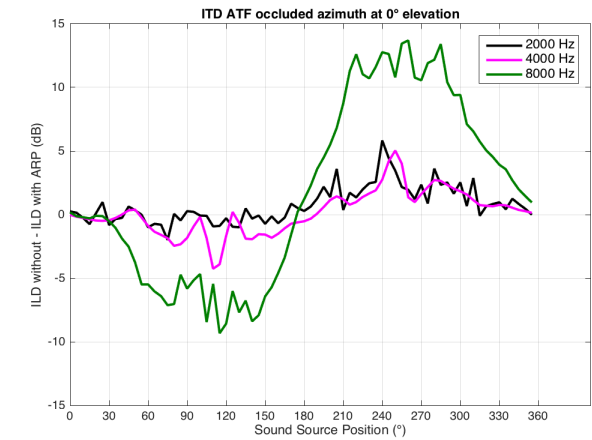


Fig. 6. Difference in octave bands between ILD of an open ear and ILD when wearing the ARP.

shows that the signal transmitted through the ARP is different from the signal transmitted through the acoustic path of the ear canal. The variation of these differences is mainly observed in the high frequencies.

To see how the ARP affects the frequency components of the incoming signal across the positions, the differences between HRTF with and without the ARP are calculated for each frequency band and for each sound source position. Fig.8 illustrates HRTF differences variations across horizontal positions of the sound source (elevation fixed to 0°) for each octave band. It shows that the higher the frequency band is, the higher the difference in HRTFs.

Fig.9 shows the variations across sagittal positions (azimuth fixed to 0°) of the differences between HRTF generated with open ears and HRTF with the ARP for each frequency band. This figure also shows that the variability increases with the increase of the frequency.

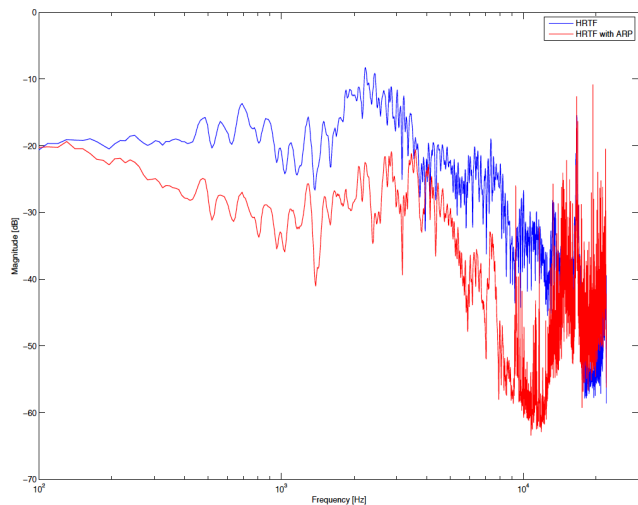


Fig. 7. HRTF of the left open ear and the HRTF of the left ear fitted with the ARP when the sound source is located at 0° azimuth and 0° elevation.

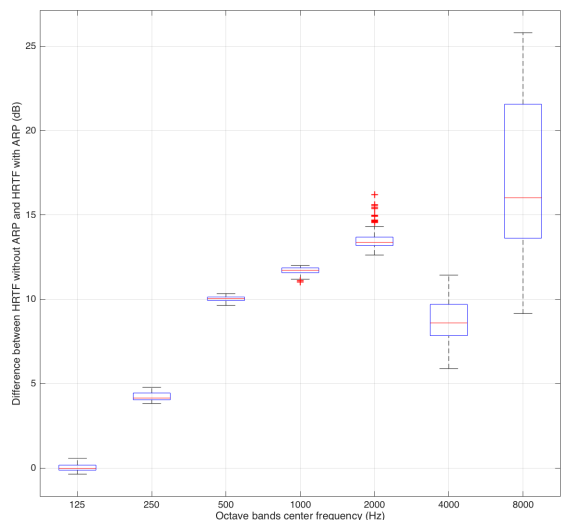


Fig. 8. Box and whisker plot of difference between HRTF with and without ARP across horizontal positions of the sound source and fixed elevation angle (0°).

VI. DISCUSSIONS AND CONCLUSIONS

This paper showed how an intra-aural AWD affects the localization of external sound sources in the horizontal plane in terms of ITD and the ILD and in the horizontal and sagittal plane in terms of HRTFs.

Analysis of ITD and ILD according to the "duplex theory" which states that the auditory system uses ITD cues to localize low frequency sounds and ILD cues to localize high frequency sounds showed that:

- **In the low frequencies** (frequency bands with center frequencies ≤ 1000 Hz): the interaural time difference between the right and left ears when wearing the ARP is

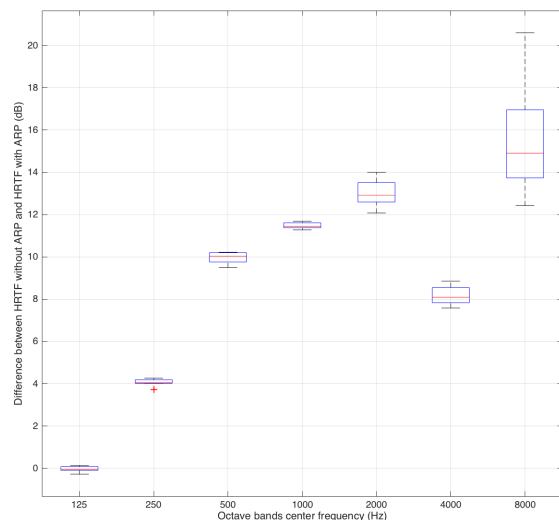


Fig. 9. Box and whisker plot of difference between HRTF with and without ARP across sagittal positions of the sound source and fixed azimuth angle (0°).

around $22 \mu\text{s}$ higher than the ITD between the left and right open ears, which corresponds to a localization error of $< 2^\circ$ in the horizontal plane.

- **In the high frequencies** (frequency bands with center frequencies ≥ 2000 Hz): changes of ILD across the azimuth when wearing the ARP are correlated with changes of ILD when having open ears. However, when using the ARP, localization of high frequency sounds depends on the azimuth of the sound source, and the frequency of the sound (8000 Hz introduces more errors than at 4000 Hz).

Analysis of HRTF is highly correlated with the ITD and ILD results showing that the ARP affects the high frequencies more than the low frequencies. Thus, these observations enable to conclude that the ARP barely affects the localization of low frequency sounds since it is an intra-aural device. However, the ARP introduces some localization errors in the high frequencies depending on the position of the sound source. Finally, for audio augmented reality applications, an equalization needs to be performed on the high frequencies to increase the acoustic "transparency" accuracy when using such intra-aural AWD.

ACKNOWLEDGMENT

The authors would like to thank Celine Lapotre for her involvement in the project, and EERS-ETS Industrial Research Chair in In-ear Technologies for its financial support.

REFERENCES

- [1] J. Wilson, B. N. Walker, J. Lindsay, C. Cambias, and F. Dellaert, "SWAN : System for Wearable Audio Navigation," in *11th IEEE International Symposium on Wearable Computers*, 2007, pp. 91 – 98.
- [2] J. Voix, "Did you say "Bionic Ear"?" *Canadian Acoustics*, vol. 42, no. 3, pp. 68–69, 2014.

- [3] M. Carbonneau, N. Lezzoum, J. Voix, and G. Gagnon, "Detection of alarms and warning signals on a digital in-ear device," *International Journal of Industrial Ergonomics*, vol. 43, no. 6, pp. 503–511, sep 2013.
- [4] N. Lezzoum, G. Gagnon, and J. Voix, "Voice Activity Detection System for Smart Earphones," *IEEE Transactions on Consumer Electronics*, vol. 60, no. 4, pp. 737–744, 2014.
- [5] C. Kyriakakis, "Fundamental and Technological Limitations of Immersive Audio Systems," *Proceedings of the IEEE*, vol. 86, no. 5, pp. 941–951, 1998.
- [6] L. Rayleigh, "On our perception of sound direction," *Philosophical Magazine*, vol. 13, no. 74, pp. 214–232, 1907.
- [7] Feddersen and Sandel, "Localization of High-Frequency Tones," *Journal of Acoustical Society of America*, vol. 29, no. 9, pp. 988–991, 1957.
- [8] E. A. Macpherson and J. C. Middlebrooks, "Listener weighting of cues for lateral angle : The duplex theory of sound localization revisited a)," *Journal of Acoustical Society of America*, vol. 111, no. 5, pp. 2219–2236, 2002.
- [9] H. Gamper, "Head-related transfer function interpolation in azimuth, elevation, and distance," *The Journal of the Acoustical Society of America*, vol. 134, no. 6, p. EL547, 2013.
- [10] J. C. Chen, K. Yao, and R. E. Hudson, "Source localization and beamforming," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 30 – 39, 2002.
- [11] J. Valin, F. Michaud, and J. Rouat, "Robust 3D localization and tracking of sound sources using beamforming and particle filtering," in *International Conference on Acoustics Speech and Signal Processing*, vol. 4, 2006.
- [12] J. Voix and F. Laville, "Expandable Earplug With Smart Custom Fitting Capabilities," in *Internoise*, 2002, p. 9.
- [13] J. Zwislocki and R. Feldman, "Just noticeable differences in dichotic phase," *Journal of Acoustical Society of America*, vol. 28, pp. 860–864, 1956.
- [14] R. Klumpp and H. Eady, "Some measurements of interaural time difference thresholds," *Journal of Acoustical Society of America*, vol. 28, pp. 859–860, 1957.
- [15] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. I: Stimulus synthesis," *The Journal of the Acoustical Society of America*, vol. 85, no. 2, pp. 868–878, 1989.
- [16] D. N. Zotkin, R. Duraiswami, and L. S. Davis, "Creation of virtual auditory spaces," in *International Conference on Acoustics Speech and Signal Processing*, vol. 2, 2002, pp. 2113–2116.
- [17] C. Faller and J. Merimaa, "Source localization in complex listening situations: selection of binaural cues based on interaural coherence." *The Journal of the Acoustical Society of America*, vol. 116, no. 5, pp. 3075–3089, 2004.