LOCALIZATION OF A SINGLE SOUND SOURCE FROM BINAURAL RECORDINGS

Gergana Vasileva, Ivo R. Draganov

Radiocommunications and Videotechnologies Dept.
Faculty of Telecommunications, Technical University of Sofia
8 Kliment Ohridski Blvd., 1756 Sofia, Bulgaria

e-mail: gergana_v95@abv.bg, idraganov@tu-sofia.bg

Abstract

In this paper is investigated the possibility of localizing a single sound source from binaural recordings using only low cost components and computationally light weighted applications. Interaural time difference and interaural intensity difference are in the basis of estimating the azimuth angle of the source equally distanced from the listener during testing. Low, mid and high frequencies are being generated by precise generator to analyze the accuracy of localization by the arranged environment. Positive results are being observed in comparison to more sophisticated methods demanding more resources.

1. INTRODUCTION

Binaural recordings have been used for a long time as a mean for producing more realistic experience in listeners close to the real environment where the sounds were recorded. Some recent studies suggest that they are also applicable for accurate sound source localization.

Minnaar et al. [1] discovered that here is a significant difference in audio recordings of this type when made from real human heads and artificial ones. Head-related transfer function in each case seems to be different. Considering both the interaural time differences and the interaural level differences Raspaud et al. [2] achieve better accuracy of the source localization. Further they propose an average parametric model based on personal estimates of these parameters. The model proved useful up to 6 kHz and its accuracy significantly decreases for azimuth angles approaching -90º and +90º respectively. Joint error averages vary between 3.35 and 11.34 by the type of sound.

More complex sound scenes, consisting of numerous sparse-spectrum sources, have been decomposed by Deleforge et al. [3] following by their localization with the use of acoustic space learning. Robot heads were used for the recordings and then dimensionality reduction via non-linear processing was applied. The mean angular error is changing between 3º and 12º. Deleforge and Horaud [4] also proposed simpler approach to the 2D localization problem by establishing a connection between sources' positions and interaural information registered in space with larger number of dimensions. Deviation of the angular error here rises up to 14º as or the azimuth detection with the increase of the training set density.

Hammershøi et al. [5] posed attention not only to the medium of sound propagation and elements for recording in relation to the personal hearing properties but also to the chain for playback. Inter-individual parameters were considered while making a correction prior to headphone reproduction. Deviation in the median-plane localization varied between 25% and 47% depending on the type of the head while recording.

Mandel et al. [6] use expectation-maximization on a model for present sources separation from the scene in order to localize them. Selective points from the spectrogram of the recorded signals after clustering indicate their presence. Separated signals have SNR 1.6 dB and and PESQ 0.27 greater than other proven techniques.

Coherence within binaural recordings and the sensitivity of human hearing with regard to the interaural time difference, especially at the presence of noise, seemed to play a role in the localization of sound sources as Rakerd and Hartmann [7] show. Their study suggests that correction should be made to the stored signals modeling the whole chain of elements leading to reproducing.

Joris and Yin [8] found that internal delays occurring due to the change in waveform of the sound from one and the same sources reaching both ears plays also a role in the localizing process. So interaural

correlation should also be employed in the model used for binaural recording.

In addition to all the above mentioned effects which need to be considered in the channel modelling of binaural recordings, Baumgartner et al. [9] pay attention to the localization of sources in sagittal-plane. Thus, it becomes possible to discriminate their position within front and back subspace. It is considered that similar approaches may have role in some medical applications [12].

In this paper, a simplified experimental setup is proposed for binaural recording followed by single source localization, cheap enough to be implemented on a wider scale for mass-user applications. In Section 2 description of the proposed methodology is given followed by experimental results in Section 3 with comparison to other implementation used in practice and related discussion. Then, a conclusion is made in Section 4.

2. METHODOLOGY DESCRIPTION

Influence over the process of a sound source localization (Fig. 1) have both the interaural time difference (ITD) and the interaural intensity difference – IID). The first factor is a sequence of the different length of the paths the sound wave is travelling to both ears. The second difference follows from the shadowing effect of the listener's head over one of the ears along the sound propagation direction.

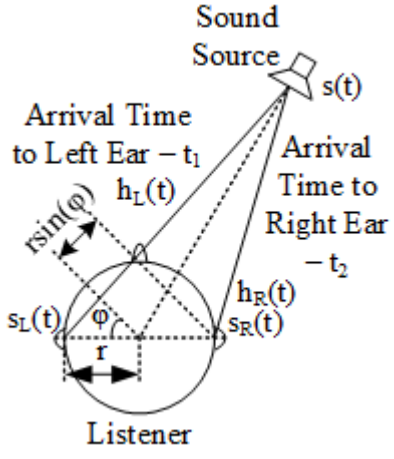

Fig. 1. Geometric model of the sound scene

On that basis, the interaural phase delay – IPD could be defined:

$$\Delta t = \left(\frac{r}{v}\right) 2\sin(\varphi), \qquad (1)$$

where v is the peed of the sound in the surrounding medium. It is obvious that $\Delta t = t_1 - t_2$. Decreasing of the frequency of the emitted signal leads to comparability between the wavelength and the distance from the source to the listener and from a given pint it could be considered that:

$$2\sin(\varphi) \approx \varphi + \sin(\varphi). \qquad (2)$$

The average diameter of the head $2r$ relates to the frequency below which the information for the spatial configuration interpreted by the user about the source is obtained mainly due to the ITD:

$$f_s = \frac{v}{2r} = \frac{340}{0.25} \approx 1.5\ kHz. \qquad (3)$$

Only the azimuth angle could be estimated taking into account (1)-(3) of the source in relation to the head – a result from the presence of ITD and IID.

In order to model the propagation channel and the transform by the hearing apparatus it is needed to introduce the Head-Related Impulse Response – HRIR. It is different for each ear. If for the left and right ear of a particular listener $h_L(t)$ and $h_R(t)$ denote the respective responses then $H_L(f)$ and $H_R(f)$ will be their representation in frequency domain. Sound wave pressure for the left and right ear in time domain could be represented by $s_L(t)$ and $s_R(t)$ at a source emission $s(t)$. Then:

$$s_{L,R} = h_{L,R}(t) * s(t) = \\ = \sum h_{L,R}(t - \Delta t)s(t)\Delta t \qquad (4)$$

Which in frequency domain is transformed to:

$$S_{L,R}(f) = \mathcal{F}\{h_{L,R}(t) * s(t)\} = \\ = H_{L,R}(f)S(f). \qquad (5)$$

HRTF can be found experimentally by changing the azimuth angle of the source with regard to the position of each ear using the following expression:

$$H_{L,R}(f,\varphi) = S_{L,R}(f,\varphi)/S(f,\varphi), \qquad (6)$$

where $S_{L,R}(f,\varphi)$ is calculated from the Fourier transform of the registered by microphone signal $s_{L,R}(t)$ in the position corresponding to the left and right ear separately. These microphones are located within the pinna of a dummy head which is a part of the experimental testing described in the next section. Calibrated generator connected with a loud-

speaker acts as a signal source with the option of changing the generated frequency *f* and level *s(f)*. The typical setup includes testing inside the far field (> 1m) of the source in which case *s(f)* decreases inversely with the distance to the microphone. The levels inside both ears then are expected to be closer. At distances less than a meter this difference considerably increases for the whole sound range.

Given a binaural recording, it becomes possible for a human to localize a sound source from a virtual sound space. The recording of the one channel $s_{L1}(t)$ is played through a headphone placed in the left year and the other – $s_{R1}(t)$ – by a headphone in the right year. The generated sound signals which affect the ear drums of both ears are $s_{l2}(t)$ and $s_{r2}(t)$ respectively. It is required that $s_{L2}(t) = s_{L1}(t)$ if HRTF has been properly estimated. In frequency domain the following expression holds:

$$\boldsymbol{S_{L1}(f) = S(f).H_M}(\mathrm{F})\,.\boldsymbol{H_L}(\mathrm{F})\,.\boldsymbol{H_S}(\mathrm{F}), \quad (7)$$

where $S_{L1}(f)$ is the spectrum of $s_{L1}(t)$, $H_M(f)$ – is the transfer function of the microphone from the left ear during recording, $H_S(f)$ – the transfer function of the loudspeaker playing the role of a sound source during the experimental testing.
ANALOGOUS TO THAT:

$$\boldsymbol{S_{L2}(f) = S_{L1}(f).H_{HPHL}(f).H_{CL}(f)}, \quad (8)$$

where $H_{HPHL}(f)$ is the transfer function of the headphone from the left ear and $H_{CL}(f)$ – that of a correcting filter applied during playback. Letting $s_{L1}(f) = s_{L2}(f)$ leads to:

$$\boldsymbol{H_{CL}(f) = 1/H_{HPHL}(f)}. \quad (9)$$

The transfer function of the correcting filter for the right ear $H_{CR}(f)$ could be found in the same way in order to apply the correction prior to the emission. In our study we employ the techniques described in [5] for finding all the transfer functions described above. Then, localizing the source from the recordings is done by using the relations presented in [10]. The general flowchart of all processing steps is given in Fig. 2.

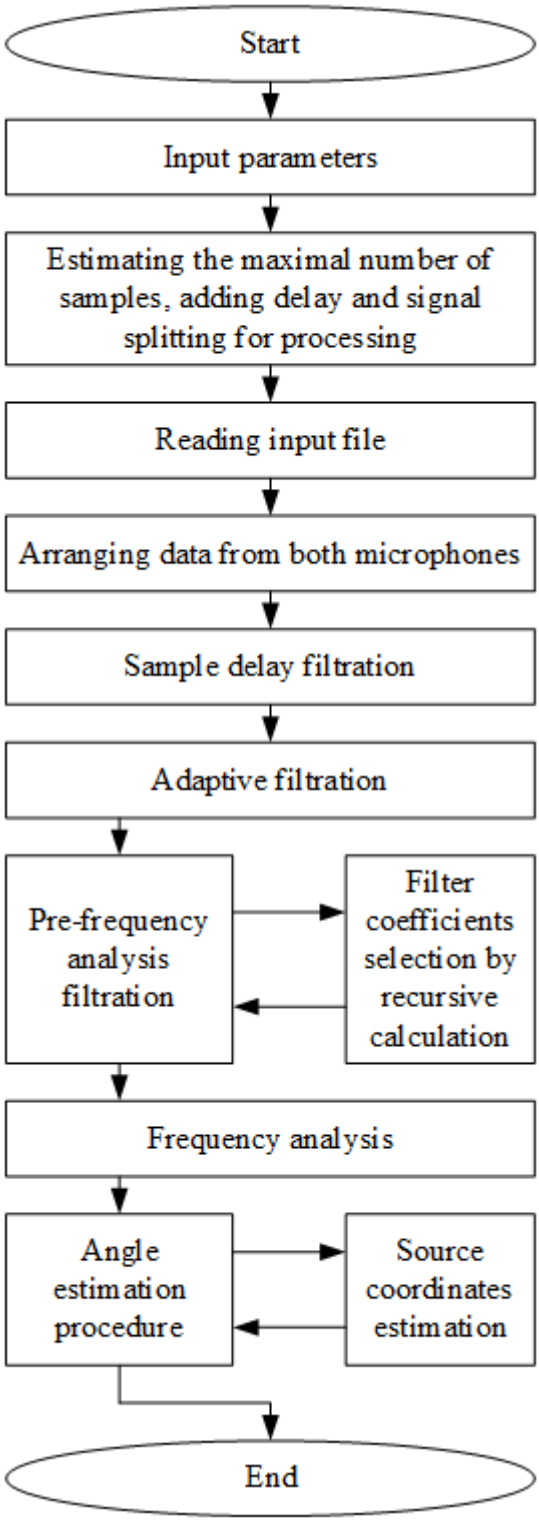

Fig. 2. Sound source localization algorithm

3. EXPERIMENTAL RESULTS

In order to accomplish the assigned task the model shown in Fig. 3 is used.

Instead of a human head, a mannequin head of approximately the same size is selected. The ear canal of the manikin is punctured and the two microphones are placed there. In Fig. 4, the head is

located at a distance of 50 cm from the speaker and rotated at an angle of 0 degrees to it.

Fig. 3. Head of a dummy imitating human head

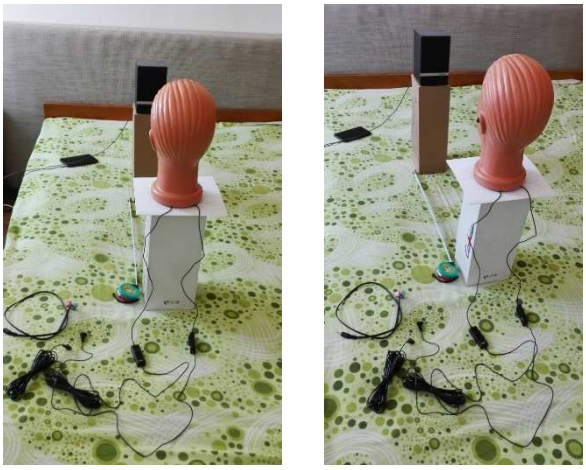

Fig. 4. Placement of the model

In order to obtain two channels (left and right) for the stereo signal, the microphones are connected by a circuit using three-way cable (Fig. 5).

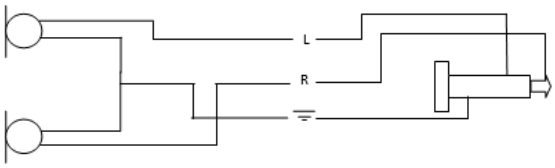

Fig. 5. Electrical circuit of a three-way cable

The three-way cable (Fig. 6) is connected to a computer for recording by the GoldWave program [11].

Recordings are made for signals with frequencies (400 Hz, 500 Hz, 700 Hz, 1 kHz, 2 kHz), and the head is located at (0, 5, 10, 15, 20, 25, 30) degrees to the speaker (the source of the sound signal). The speaker distance is 50 cm and 100 cm.

The sensitivity of the microphones is +/- 3 db (-65 db) with operational range from 50 Hz to 16 kHz and impedance of 900 Ω.

Fig. 6. Three-way cable

In the first experiment, when the sound source is located at 0.5 m from the dummy head the absolute estimation error of the azimuth angle changes for virtually all tested frequencies (Fig. 7). Most notably, the change occurs for 400 Hz with 18.48º on average. For higher frequencies the variation within testing range is smaller and closer one to the other – 8.87º.

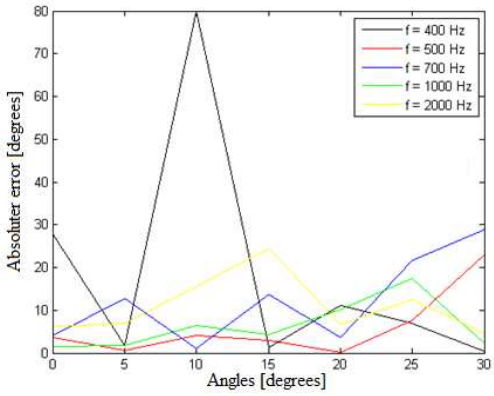

Fig. 7. Absolute estimation error of the orientation angle at 50 cm distance

The relative error for all the cases at 0.5 m is given in Fig. 8. The mid frequency of 1 kHz produces smallest error, equal to 6.32º.

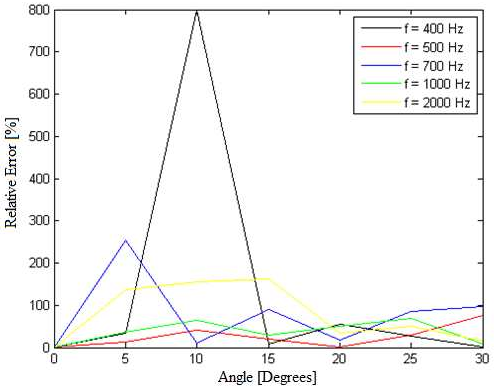

Fig. 8. Relative estimation error of the orientation angle at 50 cm distance

Testing at 1 m distance from the sound source gives smaller accuracy of localizing it (Fig. 9).

The relative estimation error in the latter case is presented in Fig. 10.

With the exception of the mid frequency of 1 kHz and the lower one of 400 Hz, all the others tend to increase the error – with 11.39º. The mean absolute error of the first two frequencies is 6.86º.

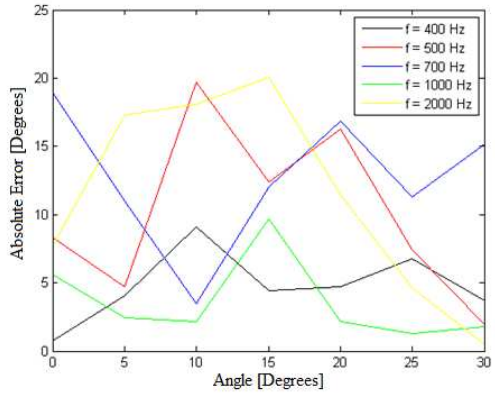

Fig. 9. Absolute estimation error of the orientation angle at 100 cm distance

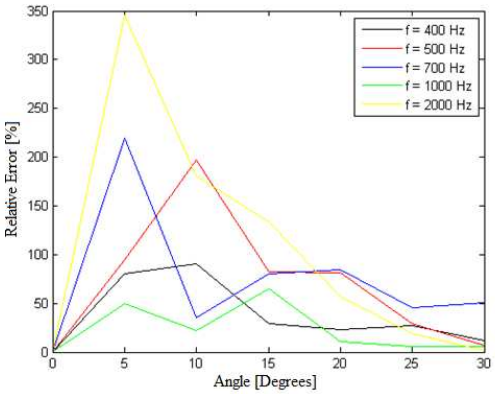

Fig. 10. Relative estimation error of the orientation angle at 100 cm distance

The probabilistic approach for localizing sound sources in [2] gives for mid frequencies error in azimuth angle from 3º to 4º when rotating the sound detector system from 0º to 30º. Given the more sophisticated approach suggested in that study and hardware used the average difference of 2º obtained within our approach seems negligible.

4. CONCLUSION

Binaural sound brings considerably more information than ordinary stereo recording. This gives a more realistic sense of the real environment in which the recording is made. The recordings made are also sufficient to locate a sound source after program processing, but with some frequency and directional limitations. For more accurate localization, it would take longer recording time for more precise analysis. The deviations in the results are due to the lateral noises, the presence of standing waves, and the relatively short length of the analyzing windows.

References

[1] *MINNAAR, Pauli, et al. Localization with binaural recordings from artificial and human heads. Journal of the Audio Engineering Society, 2001, 49.5: 323-336.*

[2] *RASPAUD, Martin; VISTE, Harald; EVANGELISTA, Gianpaolo. Binaural source localization by joint estimation of ILD and ITD. IEEE Transactions on Audio, Speech, and Language Processing, 2010, 18.1: 68-77.*

[3] *DELEFORGE, Antoine; FORBES, Florence; HORAUD, Radu. Acoustic space learning for sound-source separation and localization on binaural manifolds. International journal of neural systems, 2015, 25.01: 1440003.*

[4] *DELEFORGE, Antoine; HORAUD, Radu. 2D sound-source localization on the binaural manifold. In: Machine Learning for Signal Processing (MLSP), 2012 IEEE International Workshop on. IEEE, 2012. p. 1-6.*

[5] *HAMMERSHØI, Dorte; MØLLER, Henrik. Binaural technique—Basic methods for recording, synthesis, and reproduction. In: Communication Acoustics. Springer, Berlin, Heidelberg, 2005. p. 223-254.*

[6] *MANDEL, Michael I.; WEISS, Ron J.; ELLIS, Daniel PW. Model-based expectation-maximization source separation and localization. IEEE Transactions on Audio, Speech, and Language Processing, 2010, 18.2: 382-394.*

[7] *RAKERD, Brad; HARTMANN, William M. Localization of sound in rooms. V. Binaural coherence and human sensitivity to interaural time differences in noise. The Journal of the Acoustical Society of America, 2010, 128.5: 3052-3063.*

[8] *JORIS, Philip; YIN, Tom CT. A matter of time: internal delays in binaural processing. Trends in neurosciences, 2007, 30.2: 70-78.*

[9] *BAUMGARTNER, Robert; MAJDAK, Piotr; LABACK, Bernhard. Assessment of sagittal-plane sound localization performance in spatial-audio applications. In: The technology of binaural listening. Springer, Berlin, Heidelberg, 2013. p. 93-119.*

[10] *KWOK, Ngai M., et al. Sound source localization: microphone array design and evolutionary estimation. In: Industrial Technology, 2005. ICIT 2005. IEEE International Conference on. IEEE, 2005. p. 281-286.*

[11] *GoldWave - Audio & Video Editing Software and Fun Games, https://www.goldwave.com/, Last access Sept. 14th, 2018.*

[12] *IANTOVICS, Barna László. Agent-based medical diagnosis systems. Computing and Informatics, 2012, 27.4: 593-625.*