BINAURAL SPEAKER LOCALIZATION AND SEPARATION BASED ON A JOINT ITD/ILD MODEL AND HEAD MOVEMENT TRACKING

Mehdi Zohourian, Rainer Martin

Institute of Communication Acoustics
Ruhr-Universität Bochum, Germany
{mehdi.zohourian, rainer.martin}@rub.de

ABSTRACT

In this paper we present a novel algorithm to localize and separate simultaneous speakers using hearing aids when the head is subject to rotational movement. Most of the algorithms used in hearing aids are able to extract target signals that are in the look direction of the user and suffer from a reduced performance in localizing sounds received from other directions. Moreover, head-shadowing as well as variations like head movements may lead to significant distortions. The proposed binaural GSC beamformer includes an MMSE-based localization algorithm using an ITD/ILD model and is controlled by an inertial measurement unit. The localization algorithm can effectively localize multiple speakers in the presence of reverberation. The estimated source locations are used to adapt the GSC beamformer which extracts the desired speaker. Experimental results demonstrate the performance of the new system and especially the benefits of ILD information.

Index Terms— Binaural source localization, beamforming, source separation

1. INTRODUCTION

Speech enhancement for hearing aids has received significant attention in the past decade [1]. While it has been shown that single-channel methods improve the signal quality and reduce listener fatigue, multi-channel methods also enable the attenuation of fast fluctuating interferences such as competing speakers and thus bear the promise of improved intelligibility [2]. Moreover, with the advent of the wireless link in hearing devices binaural adaptive beamformers are of increasing interest [3]. However, a common assumption of most algorithms is that the target source is in front of the listener.

First and second-order adaptive differential microphone arrays [4], [5] are broadly used in current hearing aids, and they perform well for target sources located in the look direction. Other types of beamformer e.g. minimum variance distortionless response (MVDR) [6], multi-channel Wiener filter (MWF) [7], and the generalized sidelobe canceller [8] have also been employed successfully in hearing aids. A binaural MWF, for instance is proposed in [9] that also deals with the problem of binaural cue preservation. Furthermore, a superdirective beamformer is introduced in [10] that integrates binaural cues of a spherical head model into the MVDR beamformer and generates binaural signals.

In this paper, we aim at localization and separation of simultaneous speakers using behind-the-ears (BTE) hearing aids while we also account for rotational movements (yaw) of the head. Head movements are tracked by means of an inertial measurement unit (IMU) which provides the relative position of the head at each time step with respect to its initial position. We adapt our IMU-based beamforming approach [11] from the previously used linear array of microphones to binaural hearing aid microphones. We show that the reduction in the number of microphones from five microphones (in case of the linear array) to two microphones (new binaural configuration) can be compensated to a large extent if the head-shadowing effect is properly taken into account.

The binaural configuration causes a delay and an attenuation between the microphones which will no longer follow the free-field scenario [12]. In order to account for the head-shadowing effect we integrate binaural cues, i.e. the interaural time or phase difference (ITD / IPD) and the interaural level difference (ILD) into our system. Both cues are characterized in the form of a spherical head model [13] and merged in a novel minimum mean-square error (MMSE)-based localization approach that results in a simple addition of contributions from both cues. Our localization approach bears similarity with [14] where also ILD and ITD cues are combined in the Fourier domain, however, not in a simple addition but in a two-stage approach. Other systems, e.g. [15] and [16], integrate the binaural information into a statistical model that requires prior training. In [17] binaural room impulse responses estimated by means of blind channel identification are used instead of the microphones signals to compute binaural cues and to estimate the DOA of a single source. In work [17] the head model is employed to evaluate ITD cues while measured HRTFs are used for the evaluation of ILD cues. The proposed MMSE-based localization approach using the head model does not need a training step and provides a flexible integration of ITD and ILD cues in low and high frequency ranges. It effectively estimates the direction of arrival (DOA) of two or more speakers in each time frame. Thus, our approach enables the localization and the separation of target signals across a wide range of frequencies.

The remainder of this paper is organized as follows: In Section 2 we describe the binaural signal model as well as the HRTF model used in this paper. Section 3 will discuss the proposed system integrating the MMSE-based localization algorithm with the IMU-based GSC beamformer. Experimental results and conclusion will be presented in Sections 4 and 5, respectively.

2. BINAURAL SIGNAL AND HRTF MODEL

In the scenario depicted in Fig. 1 we consider binaural signals from two sources received by the front microphones of two BTE hearing aids. Using the convolution operator $*$ the received signal at each microphone m is written as

$$x_m(n) = \sum_{i=1}^{2} s_i(n) * h_{im}(n) + \nu_m(n) \tag{1}$$

This work has received funding from the People Programme (Marie Curie Actions) of the European Unions Seventh Framework Programme FP7/2007-2013/ under REA grant agreement n PITN-GA-2012-31752.

where $s_i(n)$ represents point source signal, $h_{im}(n)$ indicates a binaural room impulse response (BRIR) from the source i to the microphone m, $m \in \{L, R\}$, $\nu_m(n)$ is the noise at microphone m, and n is the sampling index. To analyze signals in the STFT domain, we take a $K-$point discrete Fourier transform (DFT) on overlapped and windowed signal frames. Using matrix notation we thus obtain

$$\begin{pmatrix} X_L(k,b) \\ X_R(k,b) \end{pmatrix} = \begin{pmatrix} H_{1L}(k,b) & H_{2L}(k,b) \\ H_{1R}(k,b) & H_{2R}(k,b) \end{pmatrix} \begin{pmatrix} S_1(k,b) \\ S_2(k,b) \end{pmatrix} + \begin{pmatrix} V_L(k,b) \\ V_R(k,b) \end{pmatrix}. \quad (2)$$

Here, $H_{im}(k,b)$ are the transfer functions of the left and right ears and (k,b) indicate frequency and frame index. The received signals are analyzed through the proposed binaural IMU-based GSC beamformer which is depicted in the lower part of Fig. 1 and will be discussed in Section 3.

2.1. Head-related transfer function (HRTF) model

In contrast to an HRTF database which depends on individual features e.g. the size of the head, pinnae, etc. we use a model that is more general and makes us free of measuring the HRTF in specific situations [10]. An HRTF model is proposed by Brown and Duda [13] that approximates the ITD and ILD using two filter blocks. A first-order recursive head-shadow filter cascaded by a delay element. Taking the coordinate system in Fig. 1 into account, the HRTF of the right ear is expressed as

$$H_R(\omega,\theta) = \frac{1 + j\frac{\omega}{2\omega_0}\gamma_R(\theta)}{1 + j\frac{\omega}{2\omega_0}} e^{-j\omega\tau_R(\theta)}. \quad (3)$$

In this equation we have $\omega_0 = c/a$, where c is the speed of sound, a is the radius of the head, and $\theta = \theta_{S_1}$ is the angle between the first source and the right ear. $\gamma_R(\theta)$ and $\tau_R(\theta)$ are two angle-dependent parameters that are defined as ($\theta_{min} = 150^\circ$, and $\beta_{min} = 0.1$)

$$\gamma_R(\theta) = \left(1 + \frac{\beta_{min}}{2}\right) + \left(1 - \frac{\beta_{min}}{2}\right)\cos\left(\frac{\theta}{\theta_{min}} 180^\circ\right) \quad (4)$$

$$\tau_R(\theta) = \begin{cases} -\frac{a}{c}\cos(\theta) & \text{if} \quad 0^\circ \le |\theta| \le 90^\circ \\ \frac{a}{c}\left(|\theta| - 90\right)\frac{\pi}{180} & \text{if } 90^\circ \le |\theta| \le 180^\circ. \end{cases} \quad (5)$$

Then, the HRTF of the left ear is given by $H_L(\omega,\theta) = H_R(\omega,\pi - \theta)$. As an example, Fig. 2 compares the HRTF model with one sample from an HRTF database [18]. It can be observed that the binaural cues from the model fit the measured HRTF well.

3. BINAURAL IMU-BASED SOURCE LOCALIZATION AND SEPARATION

Principally, the proposed algorithm is an extension of the IMU-based GSC beamformer [11] for the binaural configuration using hearing aids. In this work, however, we integrate the beamformer with a new localization system that is able to estimate the DOA of all speakers while taking the head-shadowing effect into account. The binaural IMU-based GSC beamformer is composed of two parts: first a GSC with a beamformer $\mathbf{W}_f(k,b)$ looking into the target direction, an adaptive blocking matrix $\mathbf{B}(k,b)$, and an adaptive noise canceler $\mathbf{W}_V(k,b)$. Secondly, a frequency-wise localization-tracking algorithm comprises an MMSE-based localization algorithm that estimates source angles $\hat{\theta}(k,b)$, a head tracking sensor (IMU) and an estimator of the posterior probability of speaker presence ($P_{\theta_{s_i}|\hat{\theta}}(k,b)$) which is updated via the expectation-maximization (EM) algorithm [19, 11]. All of these components jointly estimate and track DOAs while the head moves.

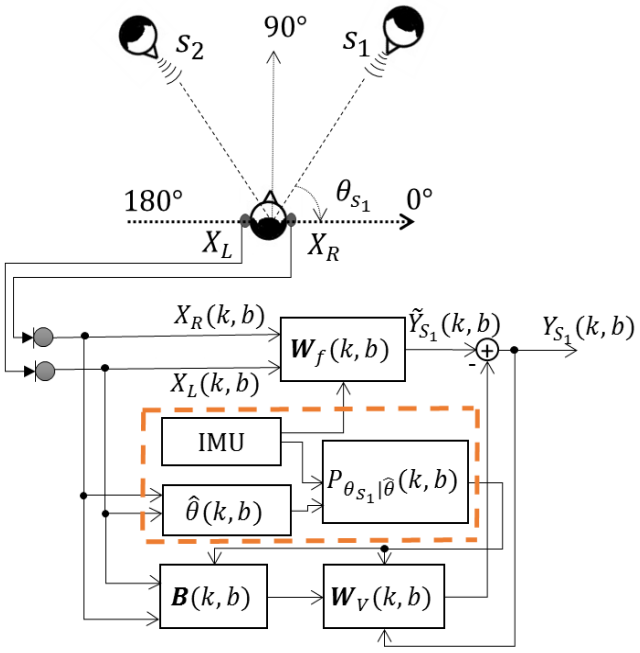


Fig. 1. Coordinate system and the proposed processing scheme.

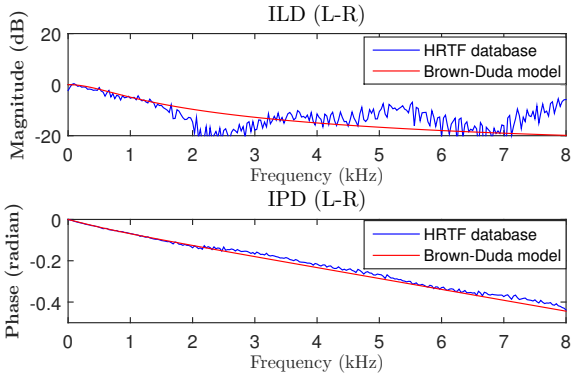


Fig. 2. The comparison between a measured HRTF [18] and the HRTF model [13] for $\theta = 30^\circ$.

3.1. MMSE-based Localization algorithm using ITD/ILD model

In this section we derive an MMSE estimator to localize multiple speakers using the binaural cues provided by the head model. It has been shown before that instead of evaluating the generalized cross-correlation (GCC) we may also evaluate the mean-squared error between the microphone signals when they are compensated with an ITD model [20]. Here, we extend the MMSE approach to the joint ITD/ILD model by means of the following objective function,

$$J(\Omega_k,\theta) = \left| \frac{X_L(\Omega_k)}{|H_L(\Omega_k,\theta)|} e^{-j\Omega_k\tau_L(\theta)} - \frac{X_R(\Omega_k)}{|H_R(\Omega_k,\theta)|} e^{-j\Omega_k\tau_R(\theta)} \right|^2 \quad (6)$$

where $|H_m|$ and τ_m ($m \in \{L, R\}$) are the magnitude and the time-delay of the HRTF for the angle θ, respectively. For simplicity, the time index b has been eliminated in the equation and $\Omega_k = 2\pi k f_S/M$ where f_s is the sampling rate. Expanding the objective function in (6) and exploiting the phase ϕ_m of the received signals we have

$$J(\Omega_k,\theta) = \left|\frac{X_L(\Omega_k)}{H_L(\Omega_k,\theta)}\right|^2 + \left|\frac{X_R(\Omega_k)}{H_R(\Omega_k,\theta)}\right|^2 - 2\Re\left\{\frac{|X_L(\Omega_k)|}{|H_L(\Omega_k,\theta)|}\frac{|X_R(\Omega_k)|}{|H_R(\Omega_k,\theta)|}e^{j\Delta\phi}\right\}. \quad (7)$$

where $\Delta\phi = \big(\phi_R(\Omega_k) - \phi_L(\Omega_k) - \Omega_k(\tau_R(\theta) - \tau_L(\theta)\big)$ and $\Re(.)$ denotes the real part. The objective function (7) can be factorized as

$$J(\Omega_k,\theta) = \frac{|X_L(\Omega_k)||X_R(\Omega_k)|}{|H_L(\Omega_k,\theta)||H_R(\Omega_k,\theta)|} \times \left(A(\Omega_k,\theta) + \frac{1}{A(\Omega_k,\theta)} - 2\Re\left\{e^{j\Delta\phi}\right\}\right) \quad (8)$$

with $A(\Omega_k,\theta) = \frac{|X_L(\Omega_k)||H_R(\Omega_k,\theta)|}{|X_R(\Omega_k)||H_L(\Omega_k,\theta)|}$. For the purpose of minimization we remove the first term and thus may simplify (8) to

$$\widetilde{J}(\Omega_k,\theta) = \left(A(\Omega_k,\theta) + \frac{1}{A(\Omega_k,\theta)}\right) - 2\cos(\Delta\phi) \quad (9)$$

which is now independent of the overall microphone and HRTF gains. For $A(\Omega_k,\theta) > 0$ the function $f(A) = A(\Omega_k,\theta) + \frac{1}{A(\Omega_k,\theta)}$ is always positive and attains its minimum value of $f(A) = 2$ for $A(\Omega_k,\theta) = 1$ and thus represents the effects of ILD deviations. Therefore, the objective function $\widetilde{J}(\Omega_k,\theta)$ attains its minimum value, i.e. $\min \widetilde{J}(\Omega_k,\theta) = 0$ when both the amplitudes and the phases match the head model.

In a more general formulation we add frequency-dependent weighting functions $0 \le \alpha(\Omega_k)$ and $0 \le \beta(\Omega_k)$ that control the contribution of the phase term and the amplitude term, respectively:

$$\widetilde{J}(\Omega_k,\theta) = \beta(\Omega_k)\big(A(\Omega_k,\theta) + \frac{1}{A(\Omega_k,\theta)}\big) - 2\alpha(\Omega_k)\cos(\Delta\phi) \quad (10)$$

Since the phase shows ambiguities for high frequencies the phase contribution can be reduced in this frequency range. Vice versa, at low frequencies the contribution of the ILD term can be reduced.

In Fig. 3 the performance of the MMSE-based localization approach using the ITD/ILD model is evaluated in each time-frequency bin for the estimation of two sources at 30° and 90° and compared to the *steered response power* approach [21] that only uses the ITD model. The estimation is performed over 5s of the speech data. The parameters of the MMSE solution were selected as $\alpha(\Omega_k) = 1, \beta(\Omega_k) = 0.1$ for $f \le$2 kHz and $\alpha = 0.1, \beta = 1$ for $f >$2 kHz. The value of 2 kHz is selected such that the phase difference has a uniform relation with the DOA considering the distance between the two microphones [6]. According to Fig. 3, a significant improvement in DOA estimation is attained specially in high frequencies.

Fig. 4 compares the performance of the two localization algorithms for one signal frame. According to this figure we find a much better concentration of estimated angles around the true source DOAs for the proposed method. Next, we integrate the localization method in the IMU-based GSC beamformer to extract the target speech signal.

3.2. IMU-based GSC beamformer

The IMU-based GSC beamformer is composed of a GSC beamformer whose parts are controlled through the localization-tracking algorithm. The GSC structure consists of a fixed beamformer, an adaptive blocking matrix and an adaptive noise canceler. We design

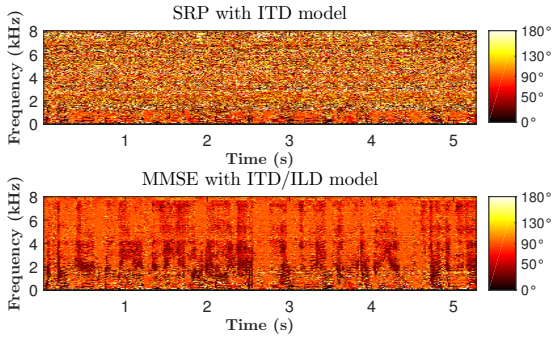


Fig. 3. $\hat{\theta}(k,b)$ as a function of time and frequency for the SRP method (top) and the proposed MMSE method (bottom) for two sources at 30° and 90° in a reverberated room.

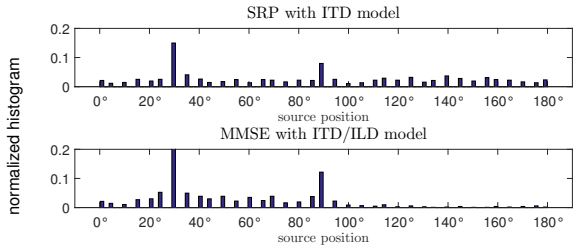


Fig. 4. The performance of the two localization approaches for the estimation of two sources at 30° and 90° for one signal frame.

the fixed beamformer based on the MVDR approach and the binaural cues attained from the head model. The general solution for the MVDR beamformer is [6]

$$\mathbf{H}_{MVDR} = \frac{\mathbf{\Phi}_{nn}^{-1}\mathbf{a}}{\mathbf{a}^H\mathbf{\Phi}_{nn}^{-1}\mathbf{a}} \quad (11)$$

where $\mathbf{a} = \begin{pmatrix} |H_L(\Omega_k,\theta)|e^{j\Omega_k\tau_L} \\ |H_R(\Omega_k,\theta)|e^{j\Omega_k\tau_R} \end{pmatrix}$ denotes the propagation vector and $\mathbf{\Phi}_{nn}$ is the noise covariance matrix. With the assumption of the uncorrelated noise at both microphones, we obtain the beamformer output for source S_1 as $\tilde{Y}_{S_1} = \mathbf{W}_f^H(k,b)\mathbf{X}(k,b)$ with

$$\mathbf{W}_f(k,b) = \frac{1}{E_H}\begin{pmatrix} |H_L(\Omega_k,\theta_{S_1})|e^{j\Omega_k\tau_L(\theta_{S_1})} \\ |H_R(\Omega_k,\theta_{S_1})|e^{j\Omega_k\tau_R(\theta_{S_1})} \end{pmatrix} \quad (12)$$

where $E_H = |H_L(\Omega_k,\theta_{S_1})|^2 + |H_R(\Omega_k,\theta_{S_1})|^2$. The beamformer is updated by the head tracker during the head rotation.

The blocking matrix provides a noise reference for the adaptive noise canceler and therefore should block the target signal. Once the posterior probability of target source presence $p_{\theta_{s_1}|\hat{\theta}}(k,b)$ is estimated at each time-frequency bin, the target signal subspace is recursively estimated as follows

$$\mathbf{P}(k,b) = (1 - P_{\theta_{S_1}|\hat{\theta}}(k,b))\mathbf{P}(k,b-1) + P_{\theta_{S_1}|\hat{\theta}}(k,b)\frac{\mathbf{X}(k,b)\mathbf{X}^T(k,b)}{\|\mathbf{X}^2(k,b)\|}. \quad (13)$$

Then, the blocking matrix $\mathbf{B}$ is computed by projection to the complementary subspace and selecting the first $(M-1)$ rows and M columns of the matrix argument (using operator $\kappa_{(M-1)M}(\cdot)$)

$$\mathbf{B}(k,b) = \kappa_{(M-1)M}\left(\mathbf{I}_{M\times M} - \mathbf{P}(k,b)\right), \quad (14)$$

where $\mathbf{I}_{M\times M}$ is an identity matrix.

The adaptive noise canceler uses a normalized least mean-square (NLMS) algorithm [22]

$$\mathbf{W}_V(k,b+1) = \mathbf{W}_V(k,b) + \alpha \frac{Y^*_{S_1}(k,b)\mathbf{B}(k,b)\mathbf{X}(k,b)}{||\mathbf{B}(k,b)\mathbf{X}(k,b)||^2} \quad (15)$$

with an adaptive step-size $\alpha = \left(1 - P_{\theta_{S_1}|\hat{\theta}}(k,b)\right)\alpha_f$, where α_f denotes a fixed stepsize factor. The above processing scheme is implemented twice to account for the two sources.

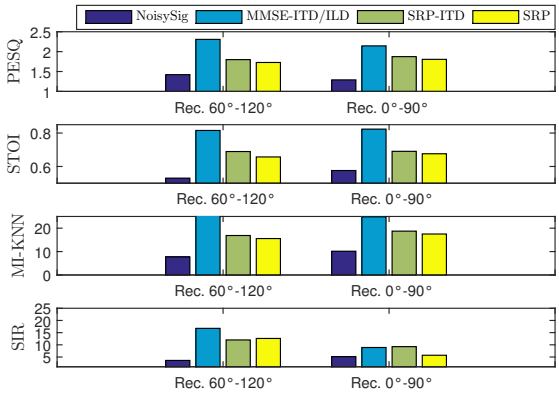
The posterior probability of each source presence is estimated in each frame b using a Gaussian mixture model (GMM) whose parameters are estimated through the *expectation-maximization* (EM) algorithm. The mean parameters of the GMM that represents the DOA of the signal are adapted by the head tracker sensor during the movement. The variance and the weighting factors of the GMM are re-estimated at each frame using the EM algorithm and then smoothed over previous frames using a first-order recursive system to enhance the posterior probability estimation [11].

4. EXPERIMENTAL RESULTS

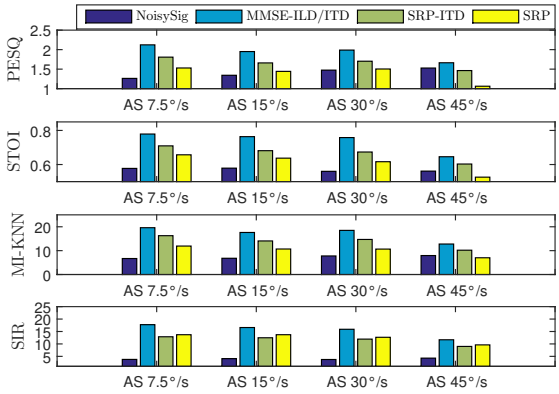
We conducted our experiments in an acoustically treated room with $T_{60} = 0.5$ s. A male and a female speaker were placed at a height of 1.2 m and a distance of 1.5 m from a *head acoustics* dummy head. The dummy head was located on a turntable to test four different head rotation speeds: $7.5^\circ/s$, $15^\circ/s$, $30^\circ/s$, and $45^\circ/s$. Each cycle of rotation starts with one speaker in the front direction of the dummy head and ends when the other speaker is in the front direction of the head. The audio were recorded by BTE hearing aid dummies. We attached a Sparkfun 9-axis IMU (SEN-10736) to the top of the dummy head to measure the relative azimuth position of the head with respect to the initial position every 0.02 s using open-source firmware [23]. Audio recordings were made at 48 kHz and later downsampled to 16 kHz. Speech material was taken from [24]. The total recording time was approximately 9 minutes.

The performance of the algorithm has been evaluated for two cases, that is, with and without the head movement. It is reported in terms of the perceptual evaluation of speech quality (PESQ) [25], intelligibility measurement using the short-time objective intelligibility (STOI) [26] and the mutual information using k-nearest neighbors (MI-KNN) [27], and a separation measurement using signal-to-interference ratio (SIR) [28]. For the static experiments we consider two speakers located at 60°, 120° and 30°, 90° w.r.t. to the coordinate system in Fig. 1. For the dynamic experiments we investigate four head rotation speeds. The results are shown in Fig. 5. We compare the performance of the proposed method (MMSE-ILD/ITD) to the input signal (NoisySig) and two other methods: The GSC beamformer controlled by the SRP algorithm using the ITD model (SRP-ITD) only, and the GSC beamformer controlled by the SRP algorithm using the free field model (SRP).

The results for the stationary recordings indicate that the adaptive beamformer which is controlled by the MMSE-based localization algorithm using the ITD/ILD model has better performance than the other two methods in terms of quality, intelligibility, and separation measurements. Therefore, this validates the idea of using the joint ITD/ILD model in the localization system. Furthermore, when there are significant head movements (especially with the speed of $15^\circ/s$ to $30^\circ/s$ that corresponds to realistic scenarios) the proposed localization-tracking framework can lock to the desired speaker and consequently is able to extract it while taking the head-shadowing effect and the head movements into account.



(a) Results for different stationary recordings (Sources are positioned either at 60° and 120° (Rec. 60°-120°) or at 0° and 90° (Rec. 0°-90°)).



(b) Results for the head movement at several angular speeds (AS).

Fig. 5. The comparison between different methods based on objective measurements.

5. CONCLUSION

In this paper we contribute a novel binaural IMU-based beamformer to localize, track, and separate simultaneous speakers. The approach requires an efficient binaural localization system that is able to estimate the azimuth DOA of speakers across a wide range of frequencies. A novel MMSE-based localization algorithm integrates joint ITD/ILD cues corresponding to their importance in different frequencies. Next, we utilize a head tracker sensor to measure the rotational movement of the head and adapt the estimated head position to the actual one. The information from the localization-tracking system is then gathered in a GMM-based posterior probability estimation of source presence in all frequency bins that is subsequently used to adapt the adaptive part of the GSC beamformer. We also employ an MVDR structure considering the binaural configuration to design the fixed beamformer. Informal audio tests as well as objective measurements over different recordings with and without head movement corroborate the efficiency of our system. Results averaged over the various recordings with and without head movement show the improvement of 0.35 PESQ, 0.9 STOI, and 3.5 dB SIR of the proposed algorithm using the ITD/ILD model with respect to the SRP algorithm using the ITD model only. The performance of the system is slightly lower than the performance of the beamformer using a linear array of microphones [11]. However, the use of ILD information leads to significant improvements at higher frequencies.

6. REFERENCES

[1] V. Hamacher, U. Kornagel, T. Lotter, and H. Puder, "Binaural signal processing in hearing aids: technologies and algorithms," *Advances in Digital Speech Transmission*, vol. 14, pp. 401–429, 2008.

[2] H. Luts, K. Eneman, J. Wouters, M. Schulte, M. Vormann, M. Büchler, N. Dillier, R. Houben, W. A. Dreschler, M. Froehlich, et al., "Multicenter evaluation of signal enhancement algorithms for hearing aids," *The Journal of the Acoustical Society of America*, vol. 127, no. 3, pp. 1491–1505, 2010.

[3] M. Aubreville and S. Petrausch, "Directionality assessment of adaptive binaural beamforming with noise suppression in hearing aids," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, April 2015, pp. 211–215.

[4] H. Teutsch and G. W. Elko, "First-and second-order adaptive differential microphone arrays," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*. Citeseer, 2001, pp. 35–38.

[5] G. W. Elko and A.-T Nguyen Pong, "A steerable and variable first-order differential microphone array," in *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*. IEEE, 1997, vol. 1, pp. 223–226.

[6] P. Vary and R. Martin, *Digital speech transmission: Enhancement, coding and error concealment*, John Wiley & Sons, 2006.

[7] K. U. Simmer, J. Bitzer, and C. Marro, "Post-filtering techniques," in *Microphone Arrays*, M. Brandstein and D. Ward, Eds., pp. 39–60. Springer, 2001.

[8] L.J. Griffiths and C.W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *Antennas and Propagation, IEEE Transactions on*, vol. 30, no. 1, pp. 27–34, Jan 1982.

[9] T.J. Klasen, T. Van Den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," *Signal Processing, IEEE Transactions on*, vol. 55, no. 4, pp. 1579–1585, April 2007.

[10] T. Lotter and P. Vary, "Dual-channel speech enhancement by superdirective beamforming," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 175–175, 2006.

[11] M. Zohourian, A. Archer-Boyd, and R. Martin, "Multi-channel speaker localization and separation using a model-based GSC and an inertial measurement unit," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, April 2015, pp. 5615–5619.

[12] J. Blauert, *Spatial hearing: The psychophysics of human sound localization*, The MIT Press, 1997.

[13] C.P. Brown and R.O. Duda, "A structural model for binaural sound synthesis," *Speech and Audio Processing, IEEE Transactions on*, vol. 6, no. 5, pp. 476–488, Sep 1998.

[14] M. Raspaud, H. Viste, and G. Evangelista, "Binaural source localization by joint estimation of ILD and ITD," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 1, pp. 68–77, Jan 2010.

[15] J. Woodruff and D. Wang, "Binaural localization of multiple sources in reverberant and noisy environments," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, no. 5, pp. 1503–1512, July 2012.

[16] T. May, S. van de Par, and A. Kohlrausch, "A probabilistic model for robust localization based on a binaural auditory front-end," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 1, pp. 1–13, Jan 2011.

[17] I. Merks, G. Enzner, and T. Zhang, "Sound source localization with binaural hearing aids using adaptive blind channel identification," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013, pp. 438–442.

[18] H. Kayser, S. D. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, pp. 6, 2009.

[19] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (methodological)*, pp. 1–38, 1977.

[20] N. Madhu and R. Martin, "Acoustic Source Localization with Microphone Arrays," in *Advances in Digital Speech Transmission*, R. Martin, U. Heute, and C. Antweiler, Eds. John Wiley, 2008.

[21] J.H. DiBiase, H.F. Silverman, and M.S. Brandstein, "Robust localization in reverberant rooms," in *Microphone Arrays*, M. Brandstein and D. Ward, Eds., pp. 157–180. Springer, 2001.

[22] N. Madhu and R. Martin, "A versatile framework for speaker separation using a model-based speaker localization approach," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 7, pp. 1900–1912, Sept 2011.

[23] P. Bartz, "Razor attitude and head rotation sensor," 2012, "https://github.com/ptrbrtz/razor-9dof-ahrs", accessed on 19.09.2014.

[24] P. Kabal, "TSP speech database," McGill University, Database Version 1.0, 2002.

[25] A.W Rix, J.G. Beerends, M.P. Hollier, and A.P. Hekstra, "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs," in *Acoustics, Speech, and Signal Processing, (ICASSP), 2001 IEEE International Conference on. IEEE*, 2001, vol. 2, pp. 749–752.

[26] C.H. Taal, R.C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 7, pp. 2125–2136, Sept 2011.

[27] J. Taghia and R. Martin, "Objective intelligibility measures based on mutual information for speech subjected to speech enhancement processing," *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, vol. 22, no. 1, pp. 6–16, Jan 2014.

[28] C. Févotte, R.I. Gribonval, E. Vincent, et al., "BSS_EVAL toolbox user guide–revision 2.0," 2005.