Proceedings of the 2015
IEEE Conference on Robotics and Biomimetics
Zhuhai, China, December 6-9, 2015

Binaural Cues Estimates Based on Interaural Matching Filter for Sound Source Localization

Ling Chen, Jie Zhang, Guodong Chen, Meng Zhang and Hong Liu*

Abstract— **Robot audition is a natural and convenient way for robot to interact with outside. Binaural sound source localization for robot audition plays an important role in human-robot interaction, speech capturing, enhancement and communication, etc. For the issue, interaural time difference (ITD) and interaural level difference (ILD) are two essential binaural (interaural) cues, yet they are still difficult to extract in complex environments. In this context, this paper proposes a new scenario to simultaneously estimate the ITD and ILD based on Interaural Matching Filter (IMF). The IMF is an optimal filter in view of eliminating the disparity of binaural signals so that it implies some information of binaural cues. Firstly, the IMF is decomposed into a minimum phase component and an all-pass component by homomorphic filtering. Subsequently, ITD is evaluated from the phase response of all-pass component, and ILD is yielded by the amplitude response of minimum phase component. Experiments verify the effectiveness of our method to preserve binaural cues, and it is robust for sound localization.**

I. INTRODUCTION

As an important part of artificial intelligence and human-robot interaction, robot audition has reached an undisputed level thanks to varieties of studies dealing with sound source localization, speech recognition and so on in the last decade. Sound source localization for robot audition is to point the sound source accurately using the received signals on microphone arrays of robot in many applications such as speech capturing, speech enhancement, communication and human-robot interaction [1]. To estimate the location of sound source, it is often stated that more than two microphones are needed [2]. One of the primary abilities of human auditory system is to localize sources by two ears. Thus as an important trend of sound localization, binaural or dual-channel sound source localization has become quite attractive and interesting since it uses only two sensors to capture signals to act as human-like auditive system [3], [4].

A large amount of binaural localization algorithms have been developed in various experimental environments like [5], [6] since "Duplex Theory" [7] was proposed, and most of them often used binaural cues to determine the relative position of a sound source. Similar to the fact that we cognise sound position by loudness, tone and orientation, there are two significant binaural (interaural) cues based on differences in time and level of the sound arriving at two ears called interaural time difference (ITD) and interaural level difference (ILD) [8], [9]. The ITD, which is caused by the different distances from the sound source to sensors, is commonly used in the time difference of arrival (TDOA)-based approaches [10], and ILD is often brought about by the distinct attenuation ratios of two ears. Exact estimation of binaural cues is the key to accurately localizing the sound source but ITD and ILD estimates are still very challenging and significative. Among the ITD approaches, the most popular is the generalized cross-correlation (GCC) method, and ILD is usually defined by the logarithmic energy ratio of two ears, which namely means that two free-running progresses are required to reckon binaural cues.

The paper proposes to model the difference between left and right ear signals through Interaural Matching Filter (IMF), which was proposed in [11]. The IMF is an optimal filter that takes the signal of left (right) ear as the input of a Wiener filter and the other one as the expectation signal. The principle of this design is to eliminate the disparities between binaural signals using the famous Minimum Means Square Error (MMSE) criterion. As the disparities are mainly reflected in delay and multiplier units, it is confirmed that the coefficients of IMF imply the ITD and ILD. Therefore, we try to estimate the ITD and ILD based on the IMF so that an integrated method is presented to binaural cues estimates simultaneously instead of the GCC or logarithmic energy ratio. Specifically, the IMF is decomposed into a minimum phase component and an all-pass component using homomorphic filtering [12]. Then, the ITD is obtained from the phase response of all-pass component, who has the unit amplitude response. The ILD can be derived form the amplitude response of minimum phase component as well, all of whose poles and zeros are located in the unit circle as it does not influence the phase of IMF. For the localization procedure, this paper locates sound source by a new joint estimation of ITD and ILD, since joint estimation of ITD and ILD has less time consumption in real-time [8], [13]–[15]. Accordingly, the novelty of our method lies in fore-

This work is supported by National Natural Science Foundation of China (NSFC, No. 60875050, 60675025, 61340046), National High Technology Research and Development Program of China (863 Program, No. 2006AA04Z247), Science and Technology Innovation Commission of Shenzhen Municipality (No. 201005280682A, No. JCYJ20120614152234873), Specialized Research Fund for the Doctoral Program of Higher Education (No. 20130001110011).

Ling Chen is with the Engineering Lab on Intelligent Perception for Internet of Things (ELIP), Shenzhen Graduate School, Peking University, Shenzhen, China chenling@sz.pku.edu.cn

Jie Zhang is with the Engineering Lab on Intelligent Perception for Internet of Things (ELIP), Shenzhen Graduate School, Peking University, Shenzhen, China zhangjie827@sz.pku.edu.cn

Guodong Chen is with the Engineering Lab on Intelligent Perception for Internet of Things (ELIP), Shenzhen Graduate School, Peking University, Shenzhen, China guodongxyz@gmail.com

Meng Zhang is with Shenzhen Hanwuji Intelligence Technology Co., Ltd

*Hong Liu is with the Engineering Lab on Intelligent Perception for Internet of Things (ELIP), Shenzhen Graduate School, Peking University, Shenzhen, China hongliu@pku.edu.cn

978-1-4673-9675-2/15/$31.00 © 2015 IEEE

most coming up with an original scheme for binaural cues simultaneous estimates to reduce the realization complexity of sound localization. The experiments has demonstrated the robustness of our method in both noisy and reverberant environments.

The rest of this paper is organized as follows: The IMF is discussed in Sec. II. The binaural cues estimates and localization are presented in Sec. III and Sec. IV, respectively. Experiments and discussions are shown in Sec. V. At last, the conclusions are drawn in Sec. VI.

II. INTERAURAL MATCHING FILTER

To obtain binaural cues means collecting the interaural differences between binaural signals. Here we do not intend to extract the interaural differences directly, but compose an Interaural Matching Filter (IMF) to eliminate the disparities between binaural signal first. With the MMSE criterion the impulse response of IMF which includes delay and attenuation is acquired and the details of the IMF will be introduced in this section.

If we denote the received signals on the two ears as $x_i(n), i \in \{l,r\}$ when there is a sound source signal $s(n)$, then the acoustic propagation model can be formulated as

$$x_i(n) = h_i(\theta,\varphi,n) * s(n), i \in \{l,r\}, \tag{1}$$

where θ and φ represent the azimuth and elevation of the sound source, and the interaural differences are characterized by $h_i(\theta,\varphi,n)$. In order to eliminate the disparity between binaural signal $x_i(n), i \in \{l,r\}$, we compose an optimal filter derived from the Wiener filter as Fig. 1 shows. Here the left ear signal $x_l(n)$ is taken as the input of IMF to predict the right ear signal $x_r(n)$, and our goal is to make the output $y(n)$ be the best prediction.

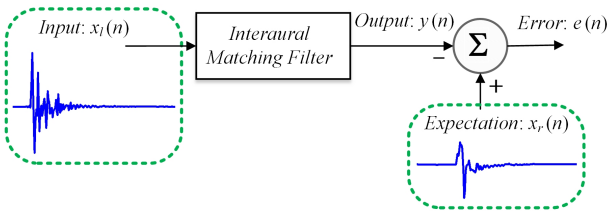

Fig. 1. Linear discrete time Interaural Matching Filter. Taking $x_l(n)$ as the input of IMF and $x_r(n)$ as the expectation is equivalent to the contrast situation in theory.

Let $\boldsymbol{w} = [w_0, w_1, \ldots, w_{M-1}]$ be the impulse response of IMF in the time domain such that M is the frame length of $x_i(n), i \in \{l,r\}$. Hereby the output of IMF is easily expressed as

$$y(n) = \sum_{i=0}^{M-1} w_i^* x_l(n-i), \quad n = 0,1,\ldots,M, \tag{2}$$

where $*$ denotes the conjugate operator. By the way, we define the error function of IMF as

$$e(n) = x_r(n) - y(n). \tag{3}$$

If the recorded binaural signals are pre-normalized, the error function defined above often acts as Gaussian noise with unit mean. In the context, we can calculate the cost function of IMF using its variance as

$$J(n) = E\{|e(n)|^2\} = E\{e(n)e^*(n)\}, \tag{4}$$

where E is the expectation operator. The famous Wiener-Hopf equation is solved by the Maximum Likelihood estimate as

$$\sum_{i=0}^{\infty} w_i R_{x_l,x_l}(i-k) = R_{x_l,x_r}(-k), \quad k = 0,1,\ldots,M-1, \tag{5}$$

where R_{x_l,x_l} is the autocorrelation matrix of x_l and R_{x_l,x_r} is the cross-correlation vector between the binaural signals. If the left ear signal is set as

$$\boldsymbol{x}_l(n) = [x_l(n), x_l(n-1), \ldots, x_l(n-M+1)]^T, \tag{6}$$

then the autocorrelation matrix R_{x_l,x_l} of $x_l(n)$ is given by

$$\begin{aligned}\boldsymbol{R}_{x_l,x_l} &= E\{x_l(n)x_l^H(n)\} \\ &= \begin{bmatrix} R_{x_l,x_l}(0) & R_{x_l,x_l}(1) & \ldots & R_{x_l,x_l}(M-1) \\ R^*_{x_l,x_l}(1) & R_{x_l,x_l}(0) & \ldots & R_{x_l,x_l}(M-2) \\ \vdots & \vdots & \vdots & \vdots \\ R^*_{x_l,x_l}(M-1) & R^*_{x_l,x_l}(M-2) & \ldots & R_{x_l,x_l}(0) \end{bmatrix}.\end{aligned} \tag{7}$$

Similarly, the cross-correlation vector is calculated as

$$\begin{aligned}\boldsymbol{r}_{x_l,x_r} &= E\{x_l(n)x_r^*(n)\} \\ &= [R_{x_l,x_r}(0), R_{x_l,x_r}(-1), \ldots, R_{x_l,x_r}(-M+1)]^T.\end{aligned} \tag{8}$$

Therefore, the coefficients of IMF can be solved by the Wiener-Holf equation in the time domain as

$$\boldsymbol{w} = \boldsymbol{R}^{-1}_{x_l,x_l} \boldsymbol{r}_{x_l,x_r}. \tag{9}$$

So far, we have accomplished the design of IMF and computed its impulse response, i.e., obtained disparity information between binaural signals. In the following, the emphasis should be concentrated on how to resolve the binaural cues from it.

III. BINAURAL CUES ESTIMATION

As the interaural differences are generally characterized by the ITD and ILD, the function of IMF can be thought as the combination of delayers and multiplicators as well. Thereout, the IMF is an linear time invariant system such that it can be decomposed into a minimum phase component (MPC) and an all-pass component (APC) [12]. Intuitively, the MPC has zero-phase, which does not influence the time-delay of interaural signals but the intensity difference. At the same time, the APC has unit amplitude response, which does not affect the intensity difference but the time-delay only. Thus, we can estimate the ITD and ILD by evaluating the phase response of the APC and the amplitude of the MPC, respectively.

Based on the theories in signal processing, the impulse response $w(n)$ of IMF can be decomposed in the time domain as

$$w(n) = w_{min}(n) * w_{all}(n), \ n = 0,1,\ldots,M-1. \tag{10}$$

where $*$ represents the convolution, $w_{min}(n)$ is the minimum phase component (MPC) of $w(n)$ and $w_{all}(n)$ is the all-pass component (APC). In the frequency domain, Eq. (10) can be rewritten as

$$W(f) = W_{min}(f)W_{all}(f), \tag{11}$$

where $W(f)$, $W_{min}(f)$ and $W_{all}(f)$ are the fast Fourier transforms (FFTs) of $w(n)$, $w_{min}(n)$ and $w_{all}(n)$, respectively. On one hand, the MPC contains all poles and zeros within the unit circle, thus the amplitude response of IMF is merely caused by the MPC, i.e.

$$|W(f)| = |W_{min}(f)|. \tag{12}$$

On the other hand, the APC has the image symmetrical zero-pole pairs and the unit amplitude response, thus the phase response of IMF is merely induced by the APC, i.e.

$$arg\big(W(f)\big) = arg\big(W_{all}(f)\big). \tag{13}$$

Using the homomorphic filtering [12] to decompose $w(n)$ into its MPC and APC, and the detailed processes are depicted in Fig. 2. Firstly we compute the FFT of $w(n)$ to get $W(f)$, then calculate its complex logarithm which is expressed by $W_R(f)$. And calculating the inverse FFT of $W_R(f)$, we can produce the cepstrum sequence $c(n)$. These above three steps are shown in the red box of Fig. 2. The complex cepstrum of MPC $c_{min}(n)$ is obtained by multiplying $c(n)$ with $2u(n) - \delta(n)$, where $u(n)$ and $\delta(n)$ are the unit step and Dirac delta functions, respectively. Then, we begin the following three steps as shown in the blue box which is inverse operation of the red box. Conducting the FFT of $c_{min}(n)$ and then exponentiating, $W_{min}(f)$ is settled, that is the MPC $w_{min}(n)$ in the time domain can be deduced by the inverse FFT of $W_{min}(f)$. Finally, the APC in the frequency domain $W_{all}(f)$ is divided $W(f)$ by $W_{min}(f)$, and the corresponding form $w_{all}(n)$ can be also easily obtained. In the next, the ITD and ILD will be evaluated from $w_{all}(n)$ and $w_{min}(n)$, respectively.

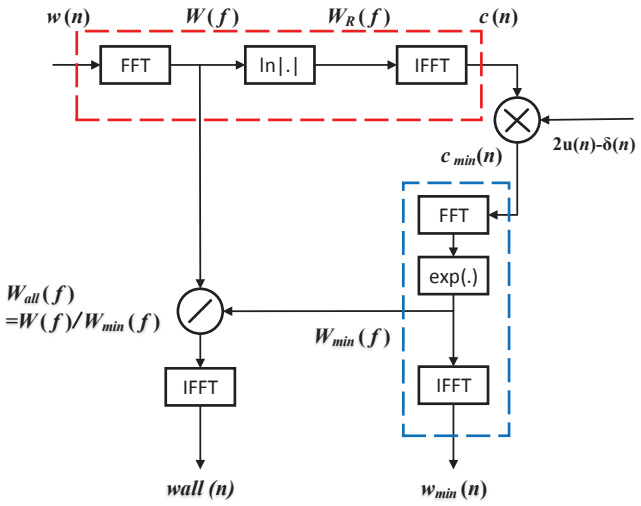

Fig. 2. Decomposition of minimum phase component and all-pass component using homomorphic filtering.

A. ITD Estimation

According to the aforehand analysis, the interaural phase difference (IPD) can be calculated first using the phase response of the APC as

$$IPD = arg\big(W_{all}(f)\big). \tag{14}$$

Then since the ITD is the slope coefficient of unwrapped phase difference, the time-delay estimate (TDE) can be realized by unwrapping IPD, i.e.

$$ITD = \frac{1}{2\pi} f^{+} IPD, \tag{15}$$

where f is the frequency of binaural signals and $(.)^{+}$ denotes the Moore-Penrose pseudo inverse. Actually, the ITDs are obtained by a least square operator performed on the IPD.

A comparison of binaural estimates of CIPIC Head-Related Transfer Functions (HRTFs) [16] for the office environment between the proposed method and typical research is shown in Fig. 3. The CIPIC database includes 1250 directions (25 azimuths $\times$ 50 elevations). From Fig. 3 it can be seen that the ITDs obtained by our method vary systematically and have less fluctuation such that they are more robust and adaptive to the challenging environments.

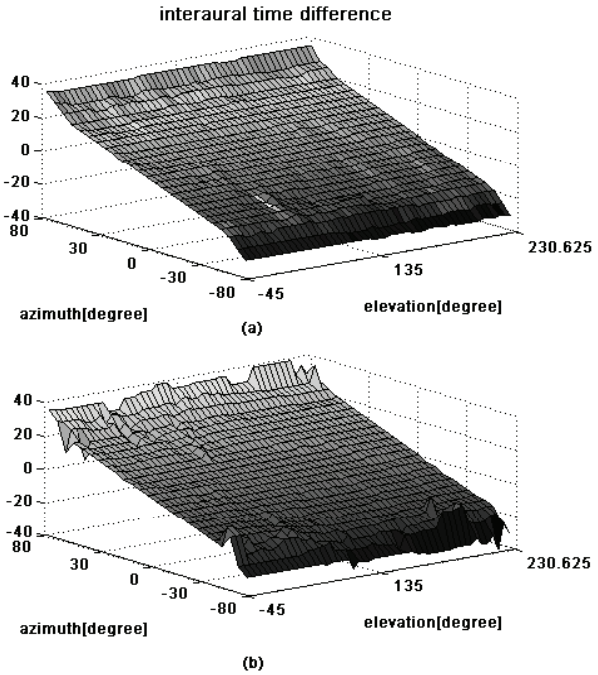

Fig. 3. Comparisons for ITD. Up: obtained by our proposed method. Down: attained by GCC-PHAT.

B. ILD Estimation

The ILD is directionally independent, and it does not have a salient geometrical distribution with the azimuth or elevation. Yet it is a frequency dependent cue that reflects the intensity difference of the signals reaching the two ears. We can extract the ILD from the amplitude response of the MPC of IMF using the Eq. (12) as

$$ILD = 20\log_{10}|W_{min}(f)|, \tag{16}$$

where $W_{min}(f)$ has already been estimated. And from Fig. 4 it can be concluded that the ILDs solved by the two methods have the approximate envelopes, that is, the ILDs decomposed from the IMF are able to represent the realistic ILDs (i.e. denoted by the logarithmic energy ratio).

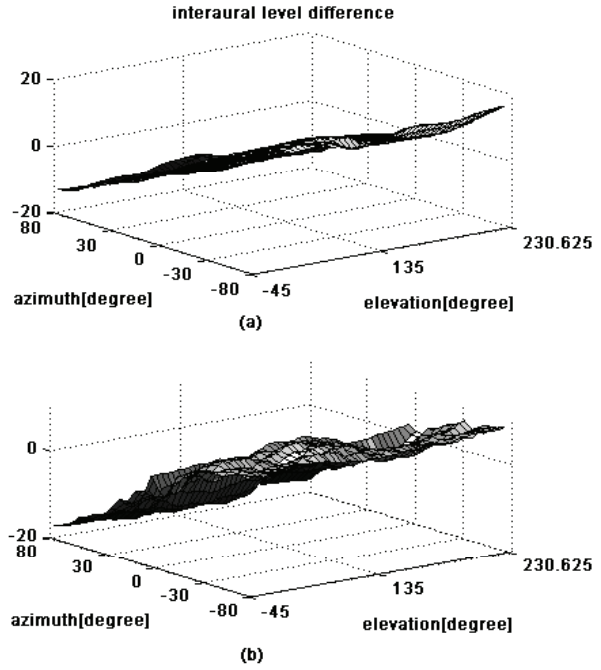

Fig. 4. Comparisons for ILD. Up: obtained by our proposed method. Down: attained by the logarithmic energy ratio.

IV. JOINT AZIMUTH LOCALIZATION

In most practical applications, the azimuth localization is much more significant than elevation localization. For instance, we control a robot to rotate horizontally so as to interact with human beings according to the localization results. And the azimuth of a desired speech is needed to assured to be enhanced or separated from other sources. Besides, the elevation localization is much more difficult than the other one, because it is only related with the ILDs which has little robustness to the environment elements. Fortunately, the azimuth localization has already satisfied the realistic requirements, thus we mainly concentrate on the localization of sounds situated on the azimuth plane.

If the ITD is only used for orientation, we can train the ITDs offline for all directions using the HRTFs or recorded binaural audio. And then we store them as templates. Furthermore, it is obvious that the similarity between the ITDs obtained from the binaural signals and the ITD templates can be utilized to describe the spatial distance between the sound source and an arbitrary direction. Thus, it means that if the ITD of the received binaural signals resembles a certain one, the sound source is localized. Here a simple but effective cosine-based similarity is adopted as

$$\alpha = \frac{< ITD_1, ITD_2 >}{\|ITD_1\| \|ITD_2\|}, \tag{17}$$

where $<,>$ denotes the inner product of vectors and $\|\cdot\|$ denotes 2nd order norm. If ITD_1 is from the templates and ITD_2 is computed from the received sound, α will be the probabilistic distribution of the ITDs. Thereby, the ITDs can be only involved in localization . Similarly, we can also consider merely using the ILD for localization, and the cosine-based similarity of ILD is given by

$$\beta = \frac{< ILD_1, ILD_2 >}{\|ILD_1\| \|ILD_2\|}, \tag{18}$$

where the ILD_1 and ILD_2 should be the ILD templates and ILD of binaural signals, respectively.

Once ITD and ILD templates are stored, the azimuth localization is simplified as matching the received binaural cues with templates. Since the ITD and ILD are mostly based on the STFT spectra of the input signals, they should be estimated for each spectral coefficient. On one hand, the ILD-based acoustic localization has a relatively large standard deviation, especially at low frequencies. On the other hand, the ITD-based acoustic localization has smaller standard deviation, but is ambiguous due to phase wrapping in the Fourier transform. Since both the ILD and the ITD are related to the azimuth, they can also be related to each other. Therefore, for practice jointly evaluating of these quantities would be more effective in order to provide good source estimations. The ILDs are used to resolve the ITD ambiguities, and the ITDs are taken to overcome the invisible ILD distribution. This joint evaluation for azimuth localization can be formulated by

$$\theta = arg \max_{\theta} \alpha\beta. \tag{19}$$

Then the detailed sound source localization process is drawn in Algorithm 1.

Algorithm 1: *Sound Source Localization based on IMF*

Input: left ear signal $x_l(n)$, right ear signal $x_r(n)$
Output: azimuth θ

Templates: $ITDs, ILDs$;
Design Interaural Matching Filter ;
$w_{min}(n)$, $w_{all}(n) \leftarrow$ decompose $w(n)$ into MPC and APC;
$IPD \leftarrow arg\big(W_{all}(f)\big), ITD \leftarrow \frac{1}{2\pi} f^{+} IPD$;
$ILD \leftarrow 20\log_{10}|W_{min}(f)|$;
while θ_i *exists* **do**
  $\alpha_i = \frac{<ITD_i, ITD>}{\|ITD_i\| \|ITD\|}$;
  $\beta_i = \frac{<ILD_i, ILD>}{\|ILD_i\| \|ILD\|}$;
end
$\theta = arg \max_{\theta_i} \alpha_i \beta_i$;
return θ

V. EXPERIMENTS AND DISCUSSIONS

The CIPIC database [16] is used in experiments to verify the performance of our method. The database is measured by

the U. C. Davis CIPIC Interface Laboratory, which includes head-related impulse responses (HRIRs) for 45 different subjects (including 27 males, 16 females, and 2 KENARs with large and small pinna). The HRIRs are tested at 1m with 25 azimuths and 50 elevations, i.e. totally 1250 directions for each subject. In experiments, the sound source is sampled at 44.1kHz and enframed to 256 sample points for each frame, because the fact that binaural signals with long frame length will make designing IMF more difficult as well as the unavoidable increasing computational complexity. The detailed parameters used here are shown in TABLE I.

TABLE I

PARAMETERS USED IN EXPERIMENTS

Parameter	Value
Sampling frequency	44.1kHz
Frame length (STFT length)	256 points
Frame shift	128 points
Block length (observation time)	2 s
Processor type	i5-2320 @ 3.00GHz

A. *Efficiency of Joint Method*

Firstly, we should analyze the influence of different binaural cues on the localization correct rate. As mentioned in Sec. IV, the ITD-based azimuth estimates are ambiguous and there is a larger standard deviation for ILD-based localization. Thus we select three methods, i.e. ITD, ILD and $ITD+ILD$, for azimuth estimate. Fig. 5 compares the localization accuracy between ILD, ITD and the joint of ILD and ITD. The noise used in our experiments is white Gaussian noise and the signal-to-noise ratio (SNR) ranges from 0dB to 40dB. We can see that all of the correct rates increase along with the increases of SNR. On one hand, the ITD-based method achieves more preferable effects than the ILD-based one, and in general the joint method acts as the most precise solutions. Specifically, the joint method can even achieve nearly 100% correct rate with 5° tolerance in the quiet enclose. On the other hand, the random noise would corrupt the available binaural signals and lead to an incoherent design of the IMF, which makes the estimation of ITD and ILD more difficult. Yet we can conclude that the joint one effectively raises the localization accuracy, especially with the low SNR it reaching 68.83%. That means both ITD and ILD are useful for azimuth localization.

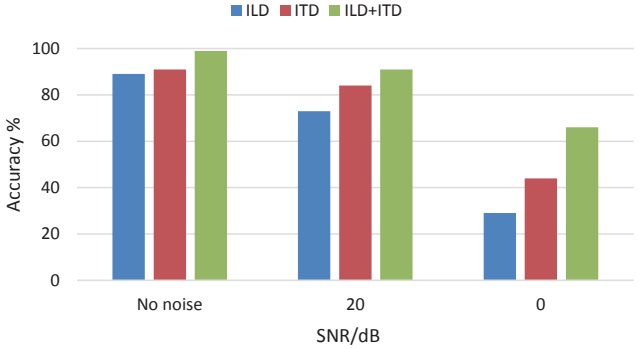

Fig. 5. Comparing the azimuth localization accuracy of ILD, ITD and joint of ILD and ITD in different SNRs with 5° tolerance.

B. *Comparison with Other Methods*

Secondly, we compare our method with several state-of-the-art algorithms including TDC [7], Hierarchical System (HS) [17] and Probability Model (PM) [19]. The experimental sound sources are speech utterances captured in an office environment with different SNRs. The detailed comparison results are illustrated in Fig. 6 and Fig. 7. It can be concluded that in most cases our method has achieved the best results. In detail, when Tolerance=0° (i.e., the localization resolution is within 1°), our method displays a tremendous superiority among these four algorithms as shown in Fig. 6. We have achieved 98.79% localization performance when with no noise, that promote 6% better than HS approximately, because our method greatly benefits from the effective simultaneous binaural cues extractions from the IMF and ability of joint estimate. Indeed, the IMF is an optimal filter, which can alleviate the additional noises to some extend. Yet the ITD and ILD are evaluated by the GCC and logarithmic energy ratio. Therefore, we can provide more precise ITD and ILD by decomposing the IMF in the noisy environments than the others, and the join of ITD and ILD can improve the localization issue. Besides, when Tolerance=10° and $SNR \geq 20$dB, all of the performances of these methods exceed 99% as shown in Fig. 7. This is satisfactory to the practical applications already.

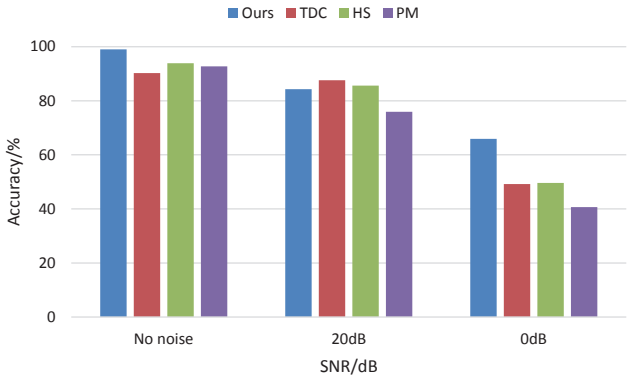

Fig. 6. Comparing localization correct rate using the proposed method and several popular methods when the tolerance is 0°.

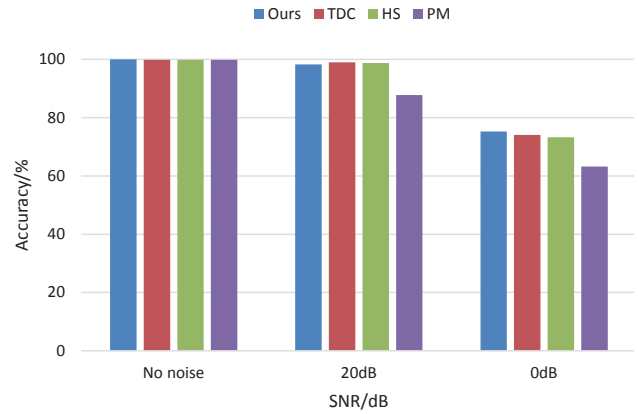

Fig. 7. Comparing localization correct rate using the proposed method and several popular methods when the tolerance is 10°.

C. Reverberant Localization

Finally, we also test our approach in a reverberant enclose, which is simulated as a room of size $(10 \times 6 \times 3)$ m with the image method [17]. Different reverberation times are considered range from 0 to 500 ms. The head is put at the point of $(2 \times 3 \times 1)$ m. The Parisi's method [18] is chosen as a reference, because it also utilizes the joint scheme but with different binaural cues estimates in the reverberant environments. When the sound sources is positioned at $\theta = -15°$, Fig. 8 illustrates the comparing results. It is clear that the localization performances of ours are better than that of [18], especially $T_R = 0ms$ our method get 47.35% accuracy which is higher than Parisi's. That means the IMF can still work effectively in the reverberant environments. Accordingly, our method is adaptive to both the noisy and reverberant surroundings, and it can be applied to the practical scenarios.

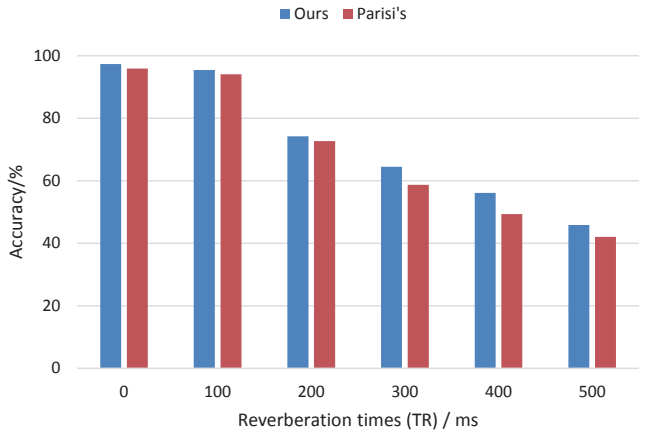

Fig. 8. Comparing localization results at azimuth $\theta = -15°$ at different reverberation times (from $T_R = 0$ ms to $T_R = 500$ ms) by the proposed method and Parisi's method.

VI. CONCLUSIONS

This paper proposes a new binaural cues estimates method based on IMF for the sound source localization. The IMF begins with the differences between binaural signals such that it implies the information of ITD and ILD. Thus we decompose it into a minimum phase component and an all-pass component to deduce ILD and ITD, respectively. From the experiments, it is observed that both of the binaural cues are useful for localization and ITD is more effective. The joint estimate of ITD and ILD is a better selection to improve the robustness of localization method. Theoretically, the IMF is an optimal Wiener filter, which can eliminate the influence of noise or reverberation to some degree, so that means our method could extract more robust binaural cues in the complex environments. In the future, we will consider the other design of IMF and try to use the IMF for localization directly. In addition, for the complex environments adding a noise or reverberation suppression unit for our localization algorithm may be more effective.

REFERENCES

[1] J. M. Valin, F. Michaud, J. Rouat, and et al., "Robust sound source localization using a microphone array on a mobile robot," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, vol. 2, pp. 1228–1233, 2003.

[2] A. Ovcharenko, S.J. Cho, and U.P. Chong, "Front-back confusion resolution in three-dimensional sound localization using databases built with a dummy head," *The Journal of the Acoustical Society of America*, vol. 122, no. 1, pp. 489–495, 2007.

[3] H. Sun, P. Yang, L.N Zu, and Q.Q Xu, "An auditory system of robot for sound source localization based on microphone array," *IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 629–632, 2010.

[4] B.G. Shinn-Cunningham, S. Santarelli, and N. Kopco, "Tori of confusion: Binaural localization cues for sources within reach of a listener," *The Journal of the Acoustical Society of America*, vol. 107, no. 3, pp. 1627–1636, 2000.

[5] T. Rodemann, G. Ince, F. Joublin, and et al., "Using binaural and spectral cues for azimuth and elevation localization," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2185–2190, 2008.

[6] H. Liu, Z. Fu, and X.F. Li, "A two-layer probabilistic model based on time-delay compensation for binaural sound localization," *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2705–2712, 2013.

[7] L.A. Jeffress, "A place theory of sound localization," *Journal of comparative and physiological psychology*, vol. 41, no. 1, pp. 35, 1948.

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

[9] R.F. Lyon and C. Mead, "An analog electronic cochlea," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, no. 7, pp. 1119–1134, 1988.

[10] C.H. Knapp and G.C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.

[11] H. Liu, J. Zhang, and Z. Fu, "A new hierarchical binaural sound source localization method based on interaural matching filter," *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1598–1605, 2014.

[12] A.V. Oppenheim and R.W. Schafer, *Digital signal processing*, Prentice Hall, Englewood Cliffs, NJ, 1975.

[13] D. Li and S.E. Levinson, "A bayes-rule based hierarchical system for binaural sound source localization," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 521–524, 2003.

[14] N. Roman and D. Wang, "Binaural tracking of multiple moving sources," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 4, pp. 728–739, 2008.

[15] V. Willert, J. Eggert, J. Adamy, and et al., "A probabilistic model for binaural sound localization," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 36, no. 5, pp. 982–994, 2006.

[16] V.R. Algazi, R.O. Duda, D.M. Thompson, and C. Avendano, "The CIPIC HRTF database," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASSAP)*, pp. 99–102, 2001.

[17] J.B. Allen and D.A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[18] F. Parisi, F. Camoes, M. Scarpiniti, and A. Uncini, "Cepstrum prefiltering for binaural source localization in reverberant environments," *IEEE Signal Processing Letters*, vol. 19, no. 2, pp. 99–102, 2012.