IMPROVE EQUALIZATION-CANCELLATION-BASED SOUND LOCALIZATION IN NOISY REVERBERANT ENVIRONMENTS USING DIRECT-TO-REVERBERANT ENERGY RATIO

Duc Thanh Chau, Masato Akagi

School of Information Science
Japan Advanced Institute of Science and Technology

Junfeng Li

Institute of Acoustics
Chinese Academy of Sciences

ABSTRACT

We previously proposed an algorithm for binaural sound source localization based on the equalization-cancellation (EC) binaural hearing model. Though this sound source localization approach exhibits relatively good results in noisy conditions, its performance in the presence of reverberation dramatically degrades. To deal with this problem, in this paper, the EC procedures on which the sound localization approach was designed are firstly analyzed. Subsequently, we propose to further improve the previous sound source localization algorithm in reverberant conditions by adapting the parameters of the EC model to the present conditions based on the direct-to-reverberant energy ratio (DRR). Experimental results demonstrate that the improved sound source localization algorithm outperforms our previously-proposed and other traditional sound source localization algorithms under noisy reverberant environments.

Index Terms— Binaural sound localization, Equalization-Cancellation model, EC-BEAM, reverberation, DRR.

1. INTRODUCTION

Sound source localization (SSL) has been extensively researched and applied in many fields of signal processing in which one of the important applications is humanoid robot [1, 2]. For human-robot interaction, the robot is required to have some basic human-like behaviours, e.g. facing the speaker during communication. In such kinds of system, sounds received at sensors on the robot are normally affected by the robot's shape, such as head-related transfer function (HRTF). Although a large number of SSL algorithms have been proposed as in the review of Dibiase *et al.* [3], very few of them are able to well adapt to these effects. Methods relying on beamforming approaches (e.g. delay and sum) and spectral analysis approach (e.g. MUSIC) achieve relatively accurate localization but they require a large microphone-array and/or high computational complexity [3] which are not in the manner of binaural system. The well-known general cross-correlation (GCC) method [4] does not consider the HRTF-like effects. There has been some effort to make GCC-based methods adapting to HRTF in which the robot head is strictly assumed to be spherical [1]. Others attempted adaptive HRTF localization system by inverse-HRTF [5]. Such HRTF-dependent methods with strict assumptions seem to fit to few systems since there are a variety of robot shapes.

Motivated by binaural hearing studies, particularly the equalization-cancellation (EC) model [6], we have proposed a binaural SSL algorithm, namely EC-BEAM [7]. In order to adapt to the physical effect from device, the EC-BEAM was designed with a prior training process to pre-calibrate the interaural differences compensator, namely equalizer, between two microphones. Although experimental results of EC-BEAM in noisy conditions are relatively promising [7], its performance in the presence of reverberation is still limited due to the effect of reverberation on the interaural differences [8].

Psychoacoustic researches have revealed evidences that human hearing system adapts to reverberation level when being in room for a relatively short time [9]. These evidences suggested that a simulating binaural system should be able to adjust itself correspondingly to the environment. One of the most important factors characterizing for reverberant level is direct-to-reverberant energy ratio (DRR). So far, studies on DRR estimation have been investigated and successfully applied in distance estimation [10, 11]. These results provide a motivation to improve signal processing applications in practical conditions by taking advantage of such information. In this study, the DRR information is exploited to modify the equalizer appropriately to improve the robustness of the EC-BEAM algorithm against reverberation. Our approach differs from previous approaches by understanding the effect of reverberation and adapting to it rather than processing at onset segments to avoid high reverberation or increasing the number of microphones to achieve higher spatial information [3]. This investigation is one step to make the EC-BEAM algorithm toward a more practical localization method, which is normally required (1) adaptivity to physical effects from the device, and (2) robustness under noisy reverberant conditions.

This research is partially supported through the A3 Foresight Program by the Japan Society for the Promotion of Science (JSPS), the National Natural Science Foundation of China (NSFC), and the National Research Foundation of Korea (NRF).

2. EC-BEAM ALGORITHM

The mechanism for localization based on EC model was firstly mentioned as "steering the null" in the research of Durlach [6]. Based on this mechanism, the EC-BEAM localization algorithm was designed in frequency domain [7] including two processes: training and estimating.

The training process is carried out in the condition that only target signal is present to learn the effect that the device may cause on the observed signals. For each direction θ_i, an equalizer, $W_0(\omega, \theta_i)$, is constructed so that the left and right signals are equalized,

$$\begin{aligned} S_L(\omega, \theta_i) - W_0(\omega, \theta_i) S_R(\omega, \theta_i) &\approx 0 \\ \text{or} \quad W_0(\omega, \theta_i) &= \frac{S_L(\omega, \theta_i)}{S_R(\omega, \theta_i)} \end{aligned} \tag{1}$$

where ω denotes the frequency band. In this manner the equalizer $W_0(\omega, \theta_i)$ plays a role as a compensator for interaural phase difference (IPD) and level difference (ILD).

In estimation, the observed signal is supposed to consist of the source at direction ϕ and noise,

$$X_p(\omega) = S_p(\omega, \phi) + N_p(\omega), \quad p = L, R \tag{2}$$

The EC-BEAM steers the null to each direction θ_i by firstly applying the equalizer, then subtracting one signal from the other. Suppose that the equalizer is well-constructed in the training process as in Eq. (1), if the equalizer matches to direction of the source, i.e. $\theta_i = \phi$, the source's energy should be approximately eliminated, remaining only energy of noise from other directions.

$$\begin{aligned} \underset{\theta_i = \phi}{Z(\omega, \theta_i)} &= X_L(\omega) - W_0(\omega, \theta_i) X_R(\omega) \\ &\approx N_L(\omega) - W_0(\omega, \theta_i) N_R(\omega) \end{aligned} \tag{3}$$

Finally, the *direction-of-arrival* (DOA) is realized as the angle at which the residual energy of the steered null is minimal.

$$\widehat{\phi} = \underset{\theta_i}{\operatorname{argmin}} \left[E(\theta_i) \right] \quad \text{with} \quad E(\theta_i) = \int_{-\infty}^{\infty} |Z(\omega, \theta_i)|^2 d\omega \tag{4}$$

3. EC-BEAM UNDER REVERBERATION

In reverberant conditions, the observed signal consists of direct signal, reverberation and possibly diffuse noise.

$$Y_p(\omega) = S_p(\omega, \phi) + R_p(\omega, \phi) + N_p(\omega), \quad p = L, R \tag{5}$$

In the matching case where considering angle is equal to that of the sound source, the direct signal component is approximately eliminated. The cancellation output consists of residual reverberation and residual noise.

$$\begin{aligned} Z(\omega, \theta_i) \approx\ & R_L(\omega, \phi) - W_0(\omega, \theta_i) R_R(\omega, \phi) \\ & + N_L(\omega) - W_0(\omega, \theta_i) N_R(\omega) \end{aligned} \tag{6}$$

Because the reverberation component was not taken into consideration when constructing the equalizer, the target signal cannot be completely eliminated in Eq. (6). This leads to the fact that the steered null with minimal residual energy may not be the null to the true sound source.

4. PROPOSED MODIFICATION OF EQUALIZER

In order to cancel all the energy from the source at direction θ_i (hereafter, θ_i is omitted for simplicity), including direct and reverberant components, the ideal equalizer should be

$$W_m(\omega) = \frac{S_L(\omega) + R_L(\omega)}{S_R(\omega) + R_R(\omega)} \tag{7}$$

Let $\Delta(\omega) = S_L(\omega) - S_R(\omega), \quad \delta(\omega) = R_L(\omega) - R_R(\omega)$ Mathematically, $W_m(\omega)$ can be rewritten as

$$\begin{aligned} W_m(\omega) &= \frac{S_L(\omega) + Q(\omega)}{S_R(\omega) + Q(\omega)} \\ \text{with} \quad Q(\omega) &= R_R(\omega) - \frac{\delta(\omega)[S_L(\omega) + S_R(\omega)]}{\Delta(\omega) + \delta(\omega)} \end{aligned} \tag{8}$$

In Eq. (8), the component $Q(\omega)$ characterizes for the effect of reverberation on the equalizer. Let $\kappa(\omega)$ be the coefficient describes the size of $Q(\omega)$ relative to the size of reverberation at the right receiver,

$$Q(\omega) = \kappa(\omega) R_R(\omega) \tag{9}$$

The reverberant component can be expressed by the sum of multiple copies of direct signal in which each copy is delayed by τ and decayed by a coefficient $\alpha(\tau)$, that is

$$\begin{aligned} R_R(\omega) &= \int_{0^+}^{\infty} \left[\alpha(\tau) e^{-j\omega\tau} \right] S_R(\omega) d\tau \\ &= \lambda(\omega) S_R(\omega) \\ \text{where} \quad \lambda(\omega) &= \int_{0^+}^{\infty} \alpha(\tau) e^{-j\omega\tau} d\tau \end{aligned} \tag{10}$$

The ideal equalizer in Eq. (8) is rewritten as follows

$$\begin{aligned} W_m(\omega) &= \frac{S_R(\omega) + \Delta(\omega) + [\kappa(\omega)\lambda(\omega)] S_R(\omega)}{S_R(\omega) + [\kappa(\omega)\lambda(\omega)] S_R(\omega)} \\ &= 1 + \frac{1}{1 + \kappa(\omega)\lambda(\omega)} \frac{\Delta(\omega)}{S_R(\omega)} \end{aligned} \tag{11}$$

In a binaural localization system, the distance between two sensors is normally much smaller than that between the system and the source, hence $|\Delta(\omega)| \ll |S_R(\omega)|$. By applying Taylor expansion, following equation can be obtained

$$[W_0(\omega)]^\beta = \left[1 + \frac{\Delta(\omega)}{S_R(\omega)} \right]^\beta \approx 1 + \beta \frac{\Delta(\omega)}{S_R(\omega)} \tag{12}$$

From Eq. (11) and Eq. (12), the relation between the equalizer in anechoic condition $W_0(\omega)$ and that in reverberant condition $W_m(\omega)$ can be described as

$$W_m(\omega) = [W_0(\omega)]^{\beta} \quad \text{with} \quad \beta = \frac{1}{1 + \kappa(\omega)\lambda(\omega)} \tag{13}$$

It is difficult to specify β in practice since $\kappa(\omega)$ and $\lambda(\omega)$ are unknown. Therefore, we look for an approximated value $\widehat{\beta}$ which is expected to closely play a role like β as follows

$$\widehat{\beta} = \frac{1}{1 + |\kappa(\omega)|.|\lambda(\omega)|} \tag{14}$$

From the Eq. (10),

$$|\lambda(\omega)| = \frac{|R_R(\omega)|}{|S_R(\omega)|} = 10^{-\frac{DRR}{20}} \tag{15}$$

where DRR is the direct-to-reverberant energy ratio (in dB) defined as

$$DRR = 10 \log \left(\frac{|S_R(\omega)|^2}{|R_R(\omega)|^2} \right) \tag{16}$$

Consider the condition where the reverberation component (especially late reflection) is large enough relative to the direct signal component, the size of $Q(\omega)$ will approach closely to the reverberant component and $|\kappa(\omega)| \to 1$. In this case

$$\widehat{\beta} \approx \frac{1}{1 + 10^{-\frac{DRR}{20}}} \tag{17}$$

As a result, given the equalizer $W_0(\omega, \theta_i)$ in anechoic condition, the correspondent one in reverberant conditions, $W_m(\omega, \theta_i)$, can be approximately obtained as follows

$$W_m(\omega, \theta_i) = [W_0(\omega, \theta_i)]^{\widehat{\beta}} \text{ with } \quad \widehat{\beta} = \frac{1}{1 + 10^{-\frac{DRR}{20}}} \tag{18}$$

Intuitively, the Eq. (18) is consistent with the variation of equalizer corresponding to the DRR condition. When $DRR \to +\infty$ (anechoic condition), $\widehat{\beta} \to 1$, which means $W_m(\omega, \theta_i) = W_0(\omega, \theta_i)$. Similarly, $DRR \to -\infty$ indicates that the condition is extremely high reverberant, when $\widehat{\beta} \to 0$ and $W_m(\omega, \theta_i) = 1$. This is reasonable because in such condition the binaural differences between left and right signals are completely destroyed due to the overlap masking caused by the late reverberation component.

5. EXPERIMENTS AND RESULTS

The experiments are divided into two parts. The first part is to evaluate the effectiveness and the feasibility of the modified method with real recorded data in reverberant conditions. The second one is to exam the applicability of the modified EC-BEAM under noisy reverberant environment using binaural impulse response measured by dummy head, as well as to compare it with the traditional GCC-PHAT method.

5.1. Experiments part I

5.1.1. Configuration

Directional signals were recorded using two microphones spacing at 0.34m in anechoic room and reverberant room with reverberation time $T_{60} = 0.4s$. Sound source varied from 0^o (the front) to 90^o (the right side), 10^o of increment.Utterances were five 10-second speech sentences selected from ATR database [12] in which one was used for training and the others were for test. The training stage of EC-BEAM was carried out with anechoic signals using *normalized least mean square* (NLMS) method as in [7], while the estimating stage was performed with reverberant signals. The DRR was computed manually using distance perception model proposed by Bronkhorst and Houtgast [13]. A window length of 500ms was used for each estimation.

5.1.2. Results

The *average estimation errors* (AEEs) regarding to DOA of the original and modified EC-BEAM are shown in Fig. 1. It can be observed that the estimates of the original EC-BEAM in the frontal area (from 0^o to 50^o) are relatively accurate while those of the side area (from 60^o to 90^o) are quite poor. This results are consistent with the results of binaural hearing researches [14, 15, 16] since the azimuth change at the side area results in a smaller change in binaural cues, e.g. IPD and ILD, than that at the front. Therefore, the effect of reverberation at the side area is more emphasized. By adjusting the equalizer appropriately, considerable improvement was achieved in the modified EC-BEAM. It is noticed that in the frontal region AEEs of the modified EC-BEAM are slightly increased comparing with those of original EC-BEAM. This may because in experiment the approximated exponent $\widehat{\beta}$ in the Eq. (18) was used instead of the ideal exponent β in the Eq. (13). However, in overall the modified EC-BEAM provides more reliable estimation and moderates the estimation error at the acceptable levels for all directions.

The effect of reverberation depends upon its energy in the whole signal, which is characterized by the DRR factor. Consequently, the analysis of this effect based on DRR is reasonable. However, since DRR is a kind of distance-dependent factor [13], if the improvement of EC-BEAM is strictly relied on DRR, its applicability would be relatively limited. Therefore, the feasibility of the modification is further investigated. Revealed from the Eq. (18), the value of $\widehat{\beta}$ decreases very slowly when the distance is large enough, as shown in Fig. 2a. This leads to a possibility of using only one appropriate value of $\widehat{\beta}$ for each room in the scope of far-field localization. Fig. 2b shows the overall AEEs of the modified EC-BEAM in comparison with the original one along various distance, using the DRR information at fixed 2m. From Fig. 2b, the best improvement of the modified EC-BEAM is remarked at the matched DRR value (2m). Its performance at

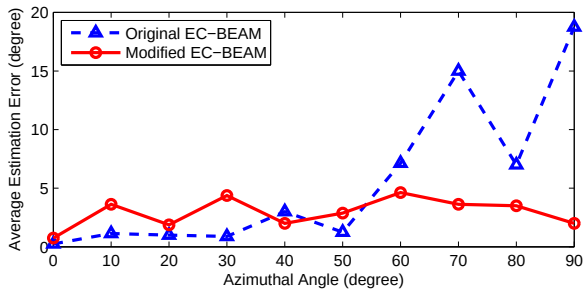

Fig. 1. AEE along the azimuths of the original and modified EC-BEAM with sound recorded at 3m.

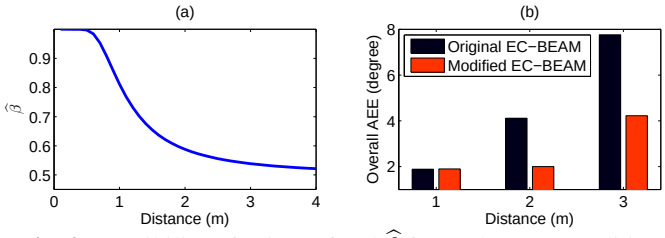

Fig. 2. Feasibility of using a fixed $\widehat{\beta}$ for each room condition: (a) Value of $\widehat{\beta}$ along distances; (b) Overall AEE, modification using DRR information at fixed 2m.

1m (over-estimated) and 3m (under-estimated) is not as significant as that at 2m, however, in general the modified EC-BEAM outperformed the original one.

5.2. Experiments part II

5.2.1. Configuration

Head-related impulse responses (HRIRs) measured in the room conditions *'Anechoic'* and *'Office I'* from the database of University of Oldenburg [17] were employed. Directional signals were generated by convoluting the HRIRs captured by the a of microphones located at the rears of the dummy ears with the ATR speech utterances used in the experiments part I. The training stage of EC-BEAM was carried out with *anechoic* data while the estimating stage was performed with reverberant data specified by *'Office I'*. Babble noise was added into reverberant signals at a SNR of 10dB to simulate noisy reverberant data. DRR for modification was calculated in the same way as in part I. Since the observed signals were affected by HRTF, to fairly compare with GCC-PHAT method, a "training process" was designed to generate a mapping from time delays to azimuthal angles using *anechoic* data. In estimation stage of GCC-PHAT, the time delays firstly were computed, then mapped to azimuthal angles. Three algorithm used the same 200ms window length for each estimation.

5.2.2. Results

Fig. 3 shows the *error rate* of original EC-BEAM, GCC-PHAT and modified EC-BEAM along error thresholds. Because EC-BEAM relies on the compensation of equalizer regarding to ILD and IPD, its performance degrades when mismatching occurs. This is the common limitation of methods based on training approach. The GCC-PHAT can well estimates the time delay and seems less suffered from noise and reverberation. However, under the effect of the dummy head, especially when the microphones are placed at the rears of ears (not on the diameter of the head), the ITD seems an ambiguous cue for DOA estimation due to the fact that time delays of several directions in the side area are almost identical. In this case, ILD is a necessarily additional cue for DOAs distinguishing. In the modified EC-BEAM, since the equalizer is adjusted to well compensate for ILD and IPD, its localization ability is improved comparing with the original EC-BEAM and outperforms the GCC-PHAT method. These results suggested that the modified EC-BEAM is potential to satisfy the adaptivity and robustness requirements for practical sound localization.

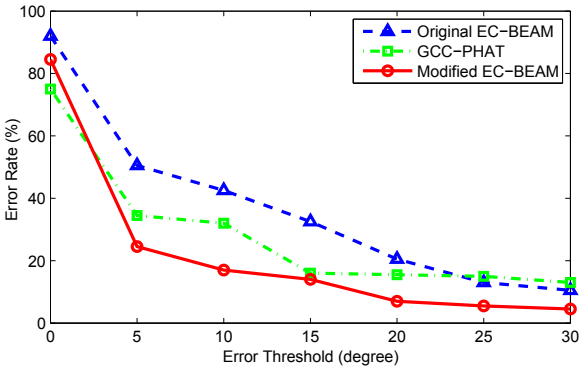

Fig. 3. Error rate along the thresholds of estimates provided by original EC-BEAM, GCC-PHAT and modified EC-BEAM in noisy *'Office I'* condition.

6. CONCLUSION

In this study, the problem of application of EC model in EC-BEAM under reverberant conditions was analyzed. A modification method to improve the adaptability of EC-BEAM to hearing environment was proposed. Specifically, the DRR information was exploited to adjust the EC-equalizer to reduce mismatching between training and testing conditions. Experimental results using measured impulse responses and real recorded data showed that the proposed modified EC-BEAM provided higher accuracy in estimation and remained estimation error at an acceptable level. These results are promising for a practical localization method in diverse environments.

In the modification, the complicated parameter was approximated by a simple one. Although the experimental results supported this simplification, this way of approximation may not always well account for all acoustic factors in rooms. There are still remaining rooms related to this issue to be further investigated for a more effective localization method.

7. REFERENCES

[1] U.H. Kim, T. Mizumoto, T. Ogata, and H.G. Okuno, "Improvement of speaker localization by considering multipath interference of sound wave for binaural robot audition," in *International Conference on Intelligent Robots and Systems (IROS)*, 2011, pp. 2910 – 15.

[2] K. Nakamura, K. Nakadai, and G. Ince, "Real-time super-resolution sound source localization for robots," in *International Conference on Intelligent Robots and Systems (IROS)*, 2012, pp. 694–699.

[3] J.H. DiBiase, H.F. Silverman, and M.S. Brandstein, *Microphone Arrays*, chapter Robust Localization in Reverberant Rooms, Springer, 2001.

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

[5] F. Keyrouz, Y. Naous, and K. Diepold, "A new method for binaural 3d localization based on hrtfs," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2006.

[6] N.I. Durlach, *Foundations of Modern Auditory Theory*, vol. 2, chapter Binaural signal detection: equalization and cancellation theory, New York: Academic Press, 1972.

[7] D.T. Chau, J. Li, and M. Akagi, "A DOA estimation algorithm based on equalization cancellation theory," in *Interspeech*, 2010, pp. 2770–2773.

[8] H.S. Colburnand A. Kulkarni, *Sound Source Localization*, vol. 25, chapter Model of Sound Localization, pp. 276–316, Springer, 2005.

[9] B.G Shinn-Cunningham, "Learning reverberation: considerations for spatial auditory displays," in *International Conference on Auditory Displays*, 2000, pp. 126–134.

[10] Y. Hioka, K. Niwa, S. Sakauchi, K. Furuya, and Y. Haneda, "Estimating direct-to-reverberant energy ratio using d/r spatial correlation matrix model," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 2374 - 2384, no. 8, 2012.

[11] Y.C. Lu and M Cooke, "Binaural estimation of sound source distance via the direct-to-reverberant energy ratio for static and moving sources," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1793 – 1805, 2010.

[12] A. Kurematsu, K. Takeda, H. Kuwabara, K. Shikanoand, Y. Sagisaka, and S. Katagiri, "Atr japanese speech database as atool of speech recognition and synthesis," *Speech Communication*, vol. 9, no. 4, pp. 357–363, 1990.

[13] A.W. Bronkhorst and T. Houtgast, "Auditory distance perception in rooms," *Nature*, vol. 397, no. 6719, pp. 517–520, 1999.

[14] A.W. Mills, "On the minimum audible angle," *Journal of the Acoustical Society of America*, vol. 30, pp. 237–246, 1958.

[15] J.D. Harris, "A florilegium of experiments on directional hearing," *Acta Otolaryngologia*, vol. Suppl. 298, pp. 1–26, 1972.

[16] D.W. Chandler and D.W. Grantham, "Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, source azimuth, and velocity," *Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1624–36, 1992.

[17] H. Kayser, S.D. Ewert, J. Anemller, 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*, , no. 6, 2009.