Journal of Applied Sciences

ISSN 1812-5654

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 13 (21): 4897-4901, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.4897.4901

Improved Sound Source Localization Using Classifier in Reverberant Noisy Environment

Xinwang Wan and Juan Liang
College of Telecommunication and Information Engineering,
Nanjing University of Posts and Telecommunications, Nanjing, 210003, China

Abstract: Sound source localization is very important in many microphone arrays application, ranging from speech enhancement to human-computer interface. The Steered Response Power (SRP) using the phase transform (SRP-PHAT) method has been proved robust but its performance degrades in highly reverberant noisy environment. The Naive-Bayes and Euclidean localization algorithms based on classification of cross-correlation functions outperform the SRP-PHAT in highly reverberant noisy environment. This study proposes the improved Naive-Bayes and Euclidean localization algorithms using principal eigenvector. Simulation results have demonstrated that the improved Naive-Bayes and Euclidean algorithms provide higher localization accuracy than the Naive-Bayes and Euclidean algorithms in reverberant noisy environment.

Key words: Microphone array, Naive-Bayes classifier, Euclidean distance classifier, cross-correlation function, principal eigenvector, sound source localization

INTRODUCTION

Sound source localization is useful for most microphone array applications such as speech enhancement, video-conferencing, hands-free speech recognition and human-computer interface. Many approaches for sound source localization using microphone arrays have appeared in the literature (Dmochowski and Benesty, 2010). For example, method based on energy measurements, distributed microphone network, blind multiple-input multiple-output filtering and algorithms based on particle filtering are used to locate and track acoustic source (Lombard *et al*., 2006; Valenzise *et al*., 2008). The time-difference-of-arrival estimation methods (Knapp and Carter 1976; Chen *et al*., 2006; Brutti *et al*., 2008) are the most popular methods in practice. The method based on Steered Response Power (SRP) is more robust than that based on time-difference-of-arrival estimation. The steered response power using the phase transform (SRP-PHAT), also known as global coherence field (De Mori and Angelini, 1998; DiBiase *et al*., 2001), is one of the most popular modern localization algorithms. Another three SRP-based acoustic source localizers are more robust than SRP-PHAT (Mungamuru and Aarabi, 2004; Zhang *et al*., 2008; Wan and Wu, 2010). However, the above mentioned methods may fail to locate the sound source in adverse noise and reverberation conditions.

Recently, to improve the localization performance in adverse acoustic conditions, many classification-based approaches (Strobel and Rabenstein 1999; Brutti *et al*., 2007; Takiguchi *et al*., 2009; Wan and Wu, 2013) are proposed. We proposed the Naive-Bayes and Euclidean localization algorithms based on classification of cross-correlation functions in (Wan and Wu, 2013). This type of approach involves two phases: training and localization.

In this study, the improved Naive-Bayes and Euclidean localization algorithms using principal eigenvector are proposed. The principal eigenvector is used to estimate the cross-correlation function which forms the feature vector and then the source location is estimated by the Naive-Bayes classifier or the Euclidean distance classifier based on the feature vector.

This study is organized as follows. In section 2, we describe the signal model. The Naive-Bayes and Euclidean localization algorithms are formulated in section 3. The proposed localization algorithms using principal eigenvector are presented in section 4. The results of localization experiments exhibit the performance of the proposed algorithms in section 5. Finally, Section 6 gives the conclusions of the study.

SIGNAL MODEL

The signal received at the nth microphone in an array of two microphones can be modeled as:

$$x_n(k) = h_n(r_s, k) \times s(k) + w_n(k), \quad n = 1, 2 \tag{1}$$

Corresponding Author: Xinwang Wan, College of Telecommunication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, 210003, China

where, $h_n(r_s, k)$ represents the acoustic room impulse response between the source and the nth microphone, "*" denotes linear convolution, $s(k)$ is the source signal located at r_s and $w_n(k)$ is uncorrelated additive noise component received at the nth microphone.

The received signal can also be expressed in the frequency-domain. Transforming (1) to the frequency-domain, gives:

$$X_n(\omega) = H_n(r_s, \omega)S(\omega) + W_n(\omega), \quad n = 1, 2 \tag{2}$$

where, $X_n(\omega)$, $H_n(r_s, \omega)$, $S(\omega)$ and $W_n(\omega)$ are the Fourier transform of $x_n(k)$, $h_n(r_s, k)$, $s(k)$ and $w_n(k)$, respectively.

Let the stacked vector of microphone signals be denoted as:

$$X(\omega) = H(\omega)S(\omega)+W(\omega) \tag{3}$$

Where:

$$\begin{aligned} X(\omega) &= [X_1(\omega), X_2(\omega)]^T \\ H(\omega) &= [H_1(r_s, \omega), H_2(r_s, \omega)]^T \\ W(\omega) &= [W_1(\omega), W_2(\omega)]^T \end{aligned} \tag{4}$$

the superscript "T" denotes transpose.

SOURCE LOCALIZATION USING CLASSIFIER

Estimation of the cross-correlation function: In the frequency-domain, the Generalized Cross-Correlation (GCC) function between $x_1(k)$ and $x_2(k)$ can be calculated:

$$R_{1,2}(\tau) = \int_{-\infty}^{\infty} \Psi_{1,2}(\omega)X_1(\omega)X_2^*(\omega)e^{j\omega\tau}d\omega \tag{5}$$

where, $\psi_{u,\ v}(\omega)$ is the weighting function and the superscript "*" denotes complex conjugation.

The Generalized Cross-Correlation (GCC) function can be made more immune to reverberation using the phase transform (PHAT). The PHAT weighting is:

$$\Psi_{1,2}(\omega) = \frac{1}{\left|X_1(\omega)X_2^*(\omega)\right|} \tag{6}$$

Inserting Eq. 6 into 5, we get:

$$R_{1,2}(\tau) = \int_{-\infty}^{\infty} \frac{X_1(\omega)X_2^*(\omega)}{\left|X_1(\omega)X_2^*(\omega)\right|}e^{j\omega\tau}d\omega \tag{7}$$

Naive-bayes classifier for source localization: The generalized cross-correlation function $R_{1,2}(\tau)$ forms the feature vector:

$$\begin{aligned} y &\triangleq [R_{1,2}(-\tau_{max}), R_{1,2}(-\tau_{max}+1), \ldots, \\ &\quad R_{1,2}(\tau_{max}-1), R_{1,2}(\tau_{max})]^T \\ &\triangleq [y_1, y_2, \ldots, y_j, \ldots, y_{2\tau_{max}}, y_{2\tau_{max}+1}]^T \end{aligned} \tag{8}$$

and:

$$\tau_{max} = \text{round}\,(\alpha Df_s/c) \tag{9}$$

where, round(-) is rounding function, the scale factor α is set to 1.67 in the next experiments, f_s is the sampling frequency, D is the distance between the two microphones and c is the acoustical velocity.

The Gaussian probability density function of individual feature y_j is:

$$p(y_j) = \frac{1}{\sqrt{2\pi}\sigma_j}\exp\left(-\frac{(y_j-\mu_j)^2}{2\sigma_j^2}\right), \quad j = 1, 2, \ldots, 2\tau_{max}+1 \tag{10}$$

where, μ_j is the mean value and σ^2_j is the variance of the feature y_j.

For each position r_i, the mean value $\mu_j(r_i)$ and the variance $\sigma^2_j(r_i)$ are estimated using M frames data in the training phase:

$$\mu_j(r_i) = \frac{1}{M}\sum_{m=1}^{M} y_j^m, \ i = 1, 2, \ldots, K \tag{11}$$

$$\sigma_j^2(r_i) = \frac{1}{M}\sum_{m=1}^{M}(y_j^m - \mu_j(r_i))^2, \ i = 1, 2, \ldots, K \tag{12}$$

Assuming that the individual features y_j, j = 1, 2,..., $2\tau_{max}+1$ are statistically independent, the probability density function of the feature vector y is:

$$p_{r_i}(y) = \prod_{j=1}^{2\tau_{max}+1} p_{r_i}(y_j) = \prod_{j=1}^{2\tau_{max}+1} \frac{1}{\sqrt{2\pi}\sigma_j(r_i)}\exp\left(-\frac{(y_j-\mu_j(r_i))^2}{2\sigma_j^2(r_i)}\right) \tag{13}$$

The location that maximizes the probability density function $p_{r_i}(y)$ will be a good source's location estimate:

$$\hat{r}_s = \underset{r_i}{\text{argmax}}\, p_{r_i}(y) \tag{14}$$

Euclidean distance classifier for source localization: For each position r_i, the mean vector μ_{r_i} is:

$$\boldsymbol{\mu}_{r_i} = [\mu_1(r_i), \mu_2(r_i), \ldots, \mu_{2\tau_{max}+1}(r_i)]^T \tag{15}$$

The Euclidean distance between the mean vector μ_{r_i} and the feature vector y is:

$$d_{r_i}(y) = \sqrt{(y-\mu_{r_i})^T(y-\mu_{r_i})} \tag{16}$$

The location that minimizes the Euclidean distance $d_{r_i}(y)$ will be a good source's location estimate:

$$\hat{r}_s = \underset{r_i}{\arg\min}\, d_{r_i}(y) \tag{17}$$

PROPOSED ALGORITHM

Estimation of the cross-correlation function using principal eigenvector: The frequency-domain correlation matrix $R_{xx}(\omega)$ is given by the expectation:

$$R_{xx}(\omega) = E\{X(\omega)X^H(\omega)\} \tag{18}$$

In practice, the frequency-domain correlation matrix estimate at the mth update $R_{xx}(\omega, m)$ is recursively obtained:

$$R_{xx}(\omega,m) = \alpha R_{xx}(\omega,m-1) + (1-\alpha)X(\omega,m)X^H(\omega,m) \tag{19}$$

The recursion is initialized as $R_{xx}(\omega, m) = X(\omega, 1)$ $X_H(\omega, 1)$ and the smoothing factor α is set to 0.4 in the next experiments.

The eigen-decomposition of the correlation matrix $R_{xx}(\omega, m)$ is given by:

$$R_{xx}(\omega,m) = \sum_{i=1}^{2} \lambda_i(\omega,m) q_i(\omega,m) q_i^H(\omega,m) \tag{20}$$

where, $\lambda_i(\omega, m)$ is eigenvalue and $q_i(\omega, m)$ is corresponding eigenvector with $\lambda_1(\omega, m) \geq \lambda_2(\omega, m)$. By the principal component method, an approximate reverberant signal $\hat{Y}(\omega,m)$ is equal to the first principal component (Wolfgang and Leopold, 2007), i.e.:

$$\hat{Y}(\omega,m) = \sqrt{\lambda_1(\omega,m)}\, q_1(\omega,m) \tag{21}$$

where, $q_1(\omega, m) = [q_1(\omega, m), q_2(\omega, m)]^T$ is the principal eigenvector. The approximate reverberant signal $\hat{Y}_n(\omega,m)$ of the nth microphone is:

$$\hat{Y}_n(\omega,m) = \sqrt{\lambda_1(\omega,m)}\, q_n(\omega,m), \quad n = 1, 2 \tag{22}$$

Similarly to Eq. 7, using approximate reverberant signal $\hat{Y}_n(\omega,m)$, the cross-correlation function can be obtained:

$$R_{PE1,2}(\tau, m) = \int_{-\infty}^{\infty} \frac{q_1(\omega,m)q_2^*(\omega,m)}{\left|q_1(\omega,m)q_2^*(\omega,m)\right|} e^{j\omega\tau} d\omega \tag{23}$$

Improved naive-bayes classifier for source localization: The cross-correlation function $R_{PE1,2}(\tau, m)$ forms the feature vector:

$$\begin{aligned} z &\triangleq [R_{PE1,2}(-\tau_{max}), R_{PE1,2}(-\tau_{max}+1), \ldots, \\ &\quad R_{PE1,2}(\tau_{max}-1), R_{PE1,2}(\tau_{max})]^T \\ &\triangleq [z_1, z_2, \ldots, z_j, \ldots, z_{2\tau_{max}}, z_{2\tau_{max}+1}]^T \end{aligned} \tag{24}$$

The Gaussian probability density function of individual feature z_j is:

$$p(z_j) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(z_j - \mu_j)^2}{2\sigma_j^2}\right), \quad j = 1, 2, \ldots, 2\tau_{max}+1 \tag{25}$$

The individual features z_j, $j = 1, 2, \ldots, 2\tau_{max}+1$ are assumed to be statistically independent. For each position r_i, the probability density function of the feature vector z is:

$$p_{r_i}(z) = \prod_{j=1}^{2\tau_{max}+1} p_{r_i}(z_j) = \prod_{j=1}^{2\tau_{max}+1} \frac{1}{\sqrt{2\pi}\sigma_j(r_i)} \exp\left(-\frac{(z_j - \mu_j(r_i))^2}{2\sigma_j^2(r_i)}\right) \tag{26}$$

The location that maximizes the probability density function $p_{r_i}(z)$ will be a good source's location estimate:

$$\hat{r}_s = \underset{r_i}{\arg\max}\, p_{r_i}(z) \tag{27}$$

Improved euclidean distance classifier for source localization: The Euclidean distance between the mean vector μ_{r_i} and the feature vector z is:

$$d_{r_i}(z) = \sqrt{(z - \mu_{r_i})^T (z - \mu_{r_i})} \tag{28}$$

The location that minimizes the Euclidean distance $d_{r_i}(z)$ will be a good source's location estimate:

$$\hat{r}_s = \underset{r_i}{\arg\min}\, d_{r_i}(z) \tag{29}$$

SIMULATION RESULTS

Here, the performances of the proposed source location algorithms are evaluated by simulation. The dimensions of the simulated rectangular room in meters are 7×6×3 m. There is 0.3 m between the two microphones which are located, respectively at (3.85, 2.5, 1.2) and (4.15, 2.5, 1.2). There is 2 m from the speaker to the midpoint between the two microphones. The image method (Allen and Berkley, 1979) is used to generate the room impulse

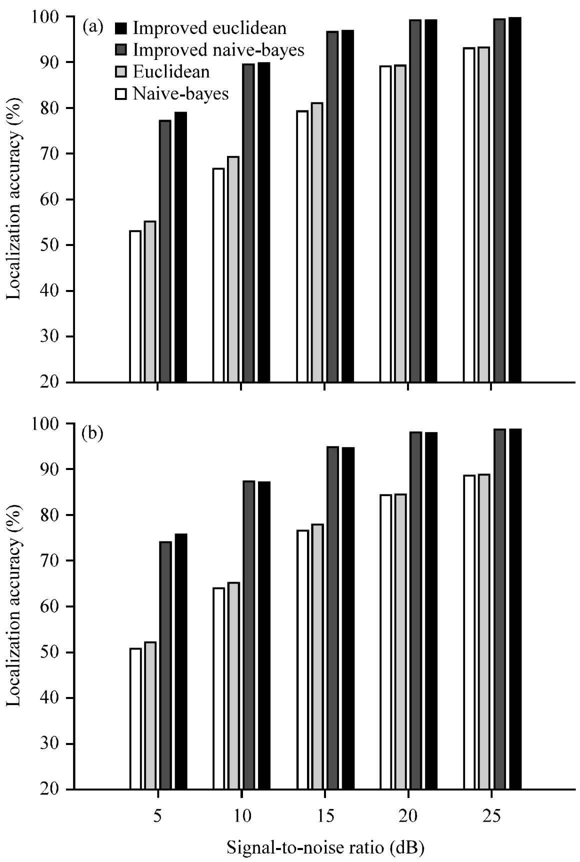

Fig. 1(a-b): Comparison of the performance for the naive-bayes, euclidean and the proposed algorithms with 9-position (a) T 60 = 0.3 sec and (b) T 60 = 0.6 sec

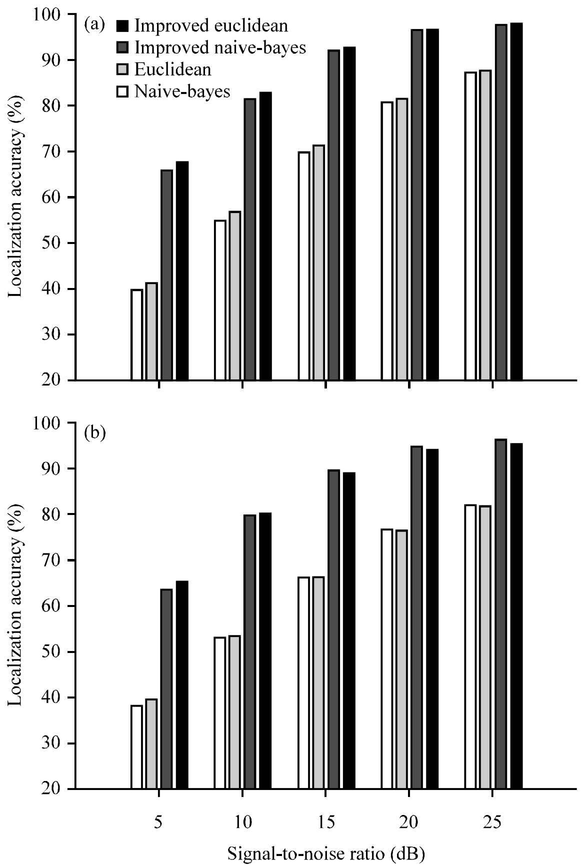

Fig. 2(a-b): Comparison of the performance for the naive-bayes, euclidean and the proposed algorithms with 17-position (a) T 60 = 0.3 sec and (b) T 60 = 0.6 sec

responses. By adjusting the frequency-independent reflection coefficient, two levels of room reverberation time (T_{60}): 0.3 and 0.6 sec are achieved. A 16 kHz sampled "clean" speech is convolved with the room impulse responses to generate reverberant signals. The additive white Gaussian noises of the two microphones are uncorrelated with each other and the noises are uncorrelated with the desired signal. The zero mean noises are then added to the reverberant signals and the average signal-to-noise ratios of the two microphones vary from 5-25 dB. Each frame is windowed by a Hanning window and the frame size is 512 samples (32 msec). The speaker's position for training and testing consists of nine positions (10, 30, 50,..., 150 and 170 degrees) and seventeen positions (10, 20, 30,..., 160 and 170 degrees).

Figure 1 and 2 depict the localization accuracy as a function of signal-to-noise ratio for the Naive-Bayes, Euclidean and the proposed algorithms, where the number of training data is 100 frames. As expected, each of the algorithms performs well at high signal-to-noise ratio levels. As reverberation time increases, the performances for all algorithms become bad. It can be observed that the proposed algorithms outperform Naive-Bayes and Euclidean algorithms for each signal-to-noise ratio condition.

CONCLUSION

This study has presented an improved sound source localization method based on principal eigenvector. The principal eigenvector is used to estimate the cross-correlation function and then the cross-correlation function forms the feature vector. Based on the feature vector, the source location is estimated by the Naive-Bayes classifier or the Euclidean distance classifier. Simulation results have demonstrated that the principal eigenvector based algorithm offers improved sound source localization accuracy over the Naive-Bayes and Euclidean algorithms in reverberant noisy environment.

ACKNOWLEDGMENT

This study is supported by Nanjing University of Posts and Telecommunications (NUPT) Scientific Research Foundation for the Introduced Talents under Grant No. NY211040.

REFERENCES

Allen, J.B. and D.A. Berkley, 1979. Image method for efficiently simulating small-room acoustics. J. Acoust. Soc. Am., 65: 943-950.

Brutti, A. and M. Omologo and P. Svaizer, 2008. Comparison between different sound source localization techniques based on a real data collection. Proceedings of the Conference on Hands-Free Speech Communication and Microphone Arrays, MAy 6-8, 2008, Trento, pp: 69-72.

Brutti, A. and M. Omologo, P. Svaizer and C. Zieger, 2007. Classification of acoustic maps to determine speaker position and orientation from a distributed microphone network. Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Volume 4, April 15-20, 2007, Honolulu, HI., pp: 493-496.

Chen, J., J. Benesty and Y. Huang, 2006. Time delay estimation in room coustic environments: An overview. EURASIP J. Applied Signal Process., 10.1155/ASP/2006/26503

De Mori, R. and B. Angelini, 1998. Spoken Dialogue with Computers. Academic Press, London, UK., ISBN: 9780122090554, Pages: 702.

DiBiase, J.H. and H. F. Silverman and M.S. Brandstein, 2001. Robust Localization in Reverberant Rooms. In: Microphone Arrays: Signal Processing Techniques and Applications, Brandstein, M. and D. Ward (Eds.). Springer, Berlin, ISBN: 978-3-642-07547-6, pp: 157-180.

Dmochowski, J.P. and J. Benesty, 2010. Steered Beamforming Approaches for Acoustic Source Localization. In: Speech Processing in Modern Communication, Cohen, I., J. Benesty and S. Gannot (Eds.). Vol. 3, Springer, Berlin, ISBN: 978-3-642-11129-7, pp: 307-337.

Knapp, C. and G.C. Carter, 1976. The generalized correlation method for estimation of time delay. IEEE Trans. Acoust. Speech Signal Process., 24: 320-327.

Lombard, A. and H. Buchner and W. Kellermann, 2006. Multidimensional localization of multiple sound sources using blind adaptive MIMO system identification. Proceedings of the International Conference on Multisensor Fusion and Integration for Intelligent Systems, September 3-6, 2006, Heidelberg, pp: 7-12.

Mungamuru, B. and P. Aarabi, 2004. Enhanced sound localization. IEEE Trans. Syst. Man Cybernetics, Part B., 34: 1526-1540.

Strobel, N. and R. Rabenstein, 1999. Classification of time delay estimates for robust speaker localization. Proceedings of the International Conference on Acoustics Speech and Signal Processing, Volume 6, March 15-19, 1999, Phoenix, AZ., pp: 3081-3084.

Takiguchi, T., Y. Sumida, R.Takashima and Y. Ariki, 2009. Single-channel talker localization based on discrimination of acoustic transfer functions. Eurasip J. Adv. Signal Proces., Vol. 2009. 10.1155/2009/918404

Valenzise, G. and G. Prandi, M. Tagliasacchi and A. Sarti, 2008. Resource constrained efficient acoustic source localization and tracking using a distributed network of microphones. Proceedings of the International Conference on Acoustics, Speech and Signal Processing, March 31-April 4, 2008, Las Vegas, NV., pp: 2581-2584.

Wan, X. and Z. Wu, 2010. Improved steered response power method for sound source localization based on principal eigenvector. Applied Acoust., 71: 1126-1131.

Wan, X. and Z. Wu, 2013. Sound source localization based on discrimination of cross-correlation functions. Applied Acoustics, 74: 28-37.

Wolfgang, H. and S. Leopold, 2007. Factor Analysis. In: Applied Multivariate Statistical Analysis, Hardle, W. and L. Simar (Eds.). 2nd Edn., Springer, Berlin, ISBN: 978-3-540-72243-4, Pages: 251-270.

Zhang, C. and D. Florencio, D.E. Ba and Z. Zhang, 2008. Maximum likelihood sound source localization and beamforming for directional microphone arrays in distributed meetings. IEEE Trans. Multimedia, 10: 538-548.