Efficient Implementation of GMDA-based DOA Technique Using Pre-training Phase Unwrapping for Source Localization

Sang-Ick Kang, Seongbin Kim, Sangmin Lee

Department of Electronic Engineering, Inha University, Republic of Korea
rkdtkddlr@gmail.com, zelabean@naver.com, sanglee@inha.ac.kr

Abstract

In this paper, a novel technique that improves the performance of generalized mixture decomposition algorithm (GMDA) based on pre-training phase unwrapping. From the investigation of the GMDA scheme, it was discovered that the conventional GMDA technique cannot fully consider phase unwrapping, because the estimated inter-channel phase difference (IPD) slope is initialized randomly. To avoid this phenomenon, the proposed GMDA approach initialized the IPD slope from the data of low-frequency bins. Experimental results show that comparing to the conventional GMDA technique, the proposed GMDA technique based on pre-training phase unwrapping obtains a lower estimation error. When integrated into a source localization system, the result of source localization is improved.

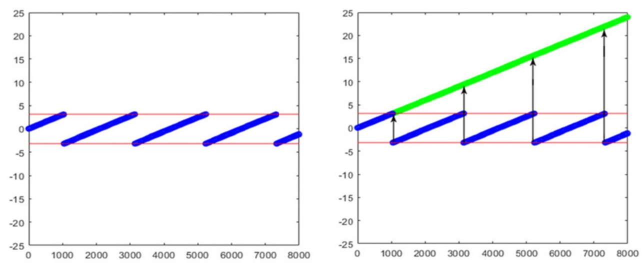
Keywords: Source localization, Inter-channel phase difference, Generalized mixture decomposition algorithm, Pre-training

1 Introduction

Source localization is an important tool used in many multichannel signal processing systems. It may include other functions such as source tracking, signal separation, speech enhancement, and noise suppression for artificial intelligence speakers, sound-tracking CCTVs, hearing aids, etc. A number of source localization algorithms have been proposed, e.g., adaptive eigenvalue decomposition algorithm associated with blind channel identification [1-2], least mean square (LMS)-type adaptive time delay estimation (TDE) algorithm [3-4], the generalized cross-correlation (GCC) method [5-9], steered response power-phase transform (SRP-PHAT) method [10] and etc [11].

Recently, one of the successful source localization techniques used is the generalized mixed decomposition algorithm (GMDA), which obtains an estimate of the direction of arrival (DOA) of the sources by utilizing inter-channel phase difference (IPD) between dual channels in sinusoidal tracking [12-14]. The GMDA-based source localization technique is refined using the sinusoidal track method in the existing IPD distribution, and it shows robust performance in white noise environment.

For the GMDA-based algorithm, the IPD must be calculated, but the calculated IPD is in the range of [–π, π]. Therefore, to conduct a precise phase analysis, a discontinuous signal phase need to be added with a value obtained by multiplying 2π by integer. Then, a continuous signal phase becomes a continous signal, as illustrated in Figure 1. This process is known as phase unwrapping. The result of the DOA estimation is sensitive to the accuracy of the phase unwrapping. However, because the existing method randomly initialize the IPD slope, the convergence elapse time to find proper IPD slope is long and slope is not accurate when GMDA is performed in the whole frequency range.



(a) Wrapped signal (b) Unwrapped signal

Figure 1. Phase unwrapping

In this paper, we propose a GMDA-based DOA estimation by incorporating pre-training phase unwrapping for source localization. In order to efficiently determine the phase unwrapping of the GMDA technique, pre-training technique is performed in low-frequency bins where phase unwrapping does not occur. The proposed technique is substantially adopted in the source localization technique and evaluated under various conditions. The following structure of this papers is as follows. In Section 2, review GMDA for estimate DOA. The proposed idea by us will be explained in Section 3. In Section 4, We will explain the experimental method for estimate

*Corresponding Author: Sangmin Lee; E-mail: sanglee@inha.ac.kr
DOI: 10.3966/160792642020052103021

angle and the conclusion will be explained the last section of this paper.

2 Review of Generalized Mixture Decomposition Algorithm (GMDA)

Let $x_1[\lambda]$ received at one microphone denote a noisy signal, which is the sum of a desired source signal $s[\lambda]$ and an uncorrelated additive noise signal $d_1[\lambda]$; $x_1[\lambda]=s[\lambda]+d[\lambda]$. Another signal $x_2[\lambda]$ is the sum of a delayed version of $s[\lambda]$ and noise signal $d_2[\lambda]$, where λ is frame index. Applying a short-time discrete Fourier transform (DFT), we have in the time-frequency domain

$$x_1(\omega)=S(\omega)+D_1(\omega) \quad \textbf{(1)}$$

$$x_2(\omega)=S(\omega)e^{-j\omega\tau}+D_2(\omega) \quad \textbf{(2)}$$

where ω represents angular frequency and τ is the time delay of the desired source.

The IPD $\psi_x(\omega)$ can be used for estimating the DOA between the two channel signals and is calculated as

$$\psi_X(\omega)=\angle X_1(\omega)-\angle X_2(\omega) \quad \textbf{(3)}$$

where $\angle X_1(\omega)$ and $\angle X_2(\omega)$ are the phases of $X_1(\omega)$ and $X_2(\omega)$, respectively. Considering the noise components, equation (3) is expressed as follows:

$$\psi_X(\omega)=\omega\tau+2\pi n+v(\omega) \quad (n=\ldots,-1,0,1,\ldots) \quad \textbf{(4)}$$

where $2\pi n(n=\ldots,-1,0,1,\ldots)$ represents possible phase unwrapping and $v(\omega)$ is the IPD error. The DOA of the desired source can be derived from τ using the following equation:

$$\tau=d\ \sin\theta/c \quad \textbf{(5)}$$

where d is the inter-microphone distance, θ denotes the DOA of the desired source, and c is the sound speed.

In contrast to the white signal, a speech signal has sparsity in the time-frequency domain. The speech signals include many short pauses and silence segments in the time domain [15]. In the frequency domain, the power of the signal is concentrated on the harmonics of the pitch frequency in the voiced speech [16, 17]. The white noise signal can be effectively removed by applying the sinusoidal track method considering the sparsity of the speech signal [18].

According to equation (4), the points on the IPD versus frequency plot are spread over several lines based on the DOA information of the sources. The GMDA is adopted for clustering and estimation of the directions of multiple sources.

The parameters of the mixture model are to be trained from the data. By employing the maximum-likelihood approach and using the expectation-maximization (EM) algorithm [19-21], the conditional expectation of complete data log-likelihood given the observed data under the previous parameter value is

$$Q(\Theta;\Theta(T))=\sum_{i=1}^{N}\sum_{j=1}^{m}P(C_j\mid \boldsymbol{y}_i;\Theta(t))\ln(p(\boldsymbol{y}_i\mid C_j;\boldsymbol{\theta})P_j) \quad \textbf{(6)}$$

where $\Theta=[\boldsymbol{\theta}^T,\boldsymbol{P}^T]^T$ are the parameters for the whole mixture model, $\boldsymbol{\theta}$ is the parameter vectors for all clusters, and $\boldsymbol{P}$ is *a priori* probability vectors. $P(C_j\mid \boldsymbol{y}_i;\Theta(t))$ is the posterior probability for class C_j given the previous parameter value $\Theta(t)$ and data point $\boldsymbol{y}_i$. t is the iteration number.

The GMDA-based DOA technique was proposed to select the phase unwrapping that yields the highest probability for the all observed data points. To determine a proper phase unwrapping factor for the IPD, this probability density function is reversed as

$$P(\boldsymbol{y}_i\mid C_j;\boldsymbol{\theta}_j)=\max_n\frac{1}{\sqrt{2\pi\sigma_j}}\times\exp\left\{-\frac{(\psi_{x,i}(\omega_i)+2\pi n_i-a_j(\omega_i))^2}{2\sigma_j^2}\right\} \quad \textbf{(7)}$$

where α_j is the slope of the line, σ_j^2 is the variance of the model, and

$$n_i^j=\arg\max_{ni}\frac{1}{\sqrt{2\pi\sigma_j}}\times\exp\left\{-\frac{(\psi_{x,i}(\omega_i)+2\pi n_i-a_j(\omega_i))^2}{2\sigma_j^2}\right\} \quad \textbf{(8)}$$

Let

$$J=\arg\max_j p(\boldsymbol{y}_i\mid C_j;\boldsymbol{\theta}_j) \quad \textbf{(9)}$$

Then $\boldsymbol{y}_i$ is chosen as

$$\boldsymbol{y}_i=[\omega_i,\psi_{x,i}(\omega_i)+2\pi n_i^j]^T \quad \textbf{(10)}$$

The GMDA using the unwrapped data $\boldsymbol{y}_i$ can be written as follows.

Generalized Mixture Decomposition Algorithm (GMDA):

1. $t=0$. Choose initial estimates for the model parameters, $\boldsymbol{\theta}=\boldsymbol{\theta}(0)$, as $\boldsymbol{P}=\boldsymbol{P}(0)$.

2. Repeat until convergence is achieved.

—Compute

$$P(C_j\mid \boldsymbol{y}_i;\Theta(t))=\frac{p(\boldsymbol{y}_i\mid C_k;\boldsymbol{\theta}_j(t))(P_j(t))}{\sum_{k=1}^{m}p(\boldsymbol{y}_i\mid C_k;\boldsymbol{\theta}_k(t))(P_k(t))'} \quad \textbf{(11)}$$

$$i=1,\ldots,N,\ j=1,\ldots,m.$$

—Set $\boldsymbol{\theta}_j(t+1)$ equal to the solution of the equation

$$\sum_{i=1}^{N} P(C_j \mid \boldsymbol{y}_i; \Theta(t)) \frac{\partial}{\partial \boldsymbol{\theta}_j} \ln(p(\boldsymbol{y}_i \mid C_j; \boldsymbol{\theta}) P_j) = 0 \quad \textbf{(12)}$$

with respect to $\boldsymbol{\theta}_j$, for $j = 1, \ldots, m$.

— Set

$$P_j(t+1) = \frac{1}{N} \sum_{i=1}^{N} P(C_j \mid \boldsymbol{y}_i; \Theta(t)) \quad \textbf{(13)}$$
$$j = 1, \ldots, m.$$

— $t = t+1$.

3. Convergence criteria: $\| \Theta(t+1) - \Theta(t) \| < \varepsilon$, where ε is a small threshold.

3 Enhanced GMDA Based on Pre-training Phase Unwrapping

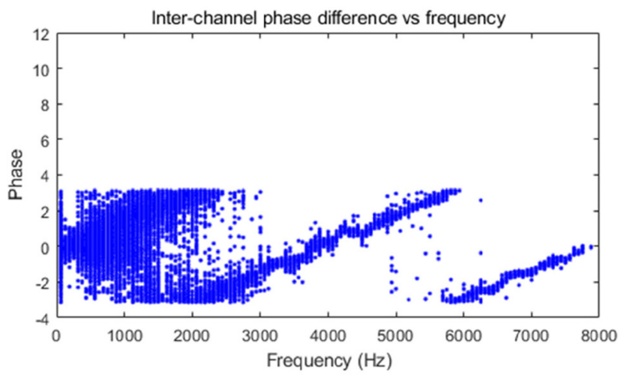
In the previous section, we note that the important parameter in the GMDA approach is α_j, which denotes the slope of the line iteratively calculated based on the Gaussian model from the data point $\boldsymbol{y}_i$. The conventional GMDA approach proposed three methods to initialize α_j such as an IPD histogram method, a random initialization method, and uniformly distributed initialization method to determine the initial α_j. However, when the slope is randomly initialized, it takes a long time to converge when selected in the other direction, and it is not accurate when the distribution is symmetric. When the histogram method is used, a large error is expected if the noise concentrated at a certain frequency is added.

Figure 2 shows the IPD versus frequency for the plot original distribution and after the phase unwrapping using the pre-training approach. Because it is known that a possible phase unwrapping parameter $2\pi n$ is zero in low frequency, we first consider the data point $\boldsymbol{y}_i'$ where the distribution in low frequency is as follows:

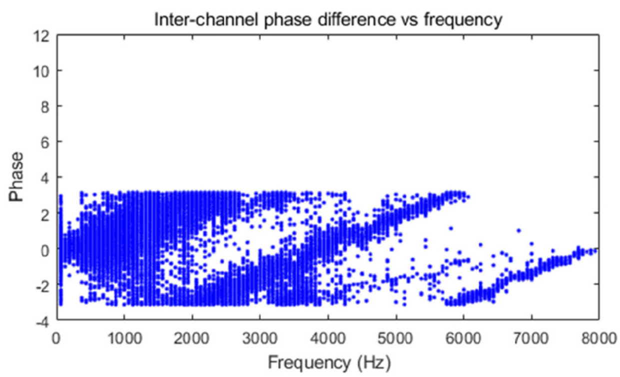
$$\boldsymbol{y}_i' = [\omega_i, \psi_{X,i}(\omega_i)]^T, \omega_i < \eta \quad \textbf{(14)}$$

where η is the threshold of the angular frequency. The initial slope $\alpha_j(0)$ is given by

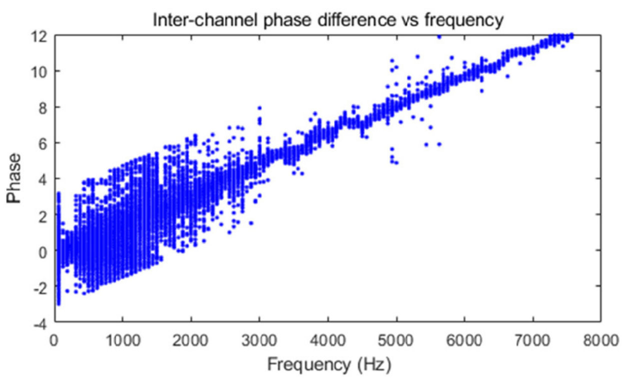
$$\alpha_j(0) = \arg\max_{\alpha_j} \frac{1}{\sqrt{2\pi\sigma_j}} \times \exp\left\{ -\frac{(\psi_{x,i}(\omega_i) - a_j(\omega_i))^2}{2\sigma_j^2} \right\} \quad \textbf{(15)}$$



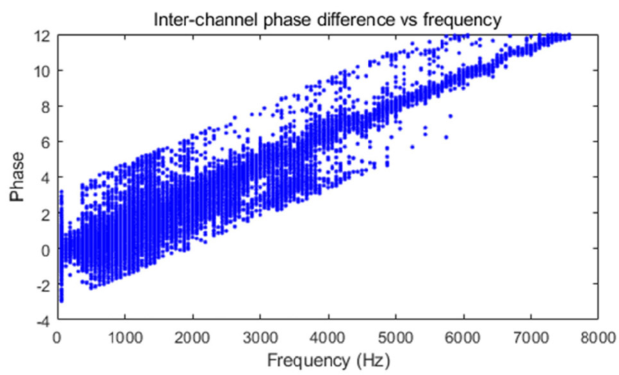
(a) are wrapped IPDs in clean and noisy environment, respectively



(b) are wrapped IPDs in clean and noisy environment, respectively



(c) are adjusting phase-unwrapping by using the pre-training approach



(d) are adjusting phase-unwrapping by using the pre-training approach

Figure 2. Phase-unwrapping on real data

The slope $\alpha_j(0)$ derived from the low-frequency range is obtained from equation (8), and the n values in all the frequency bins are found and the data is unwrapped through equation (10). This method results in faster convergence than the conventional algorithm, which randomly sets the initial slope $\alpha_j(0)$.

The block diagram of the proposed algorithm is illustrated in Figure 3.

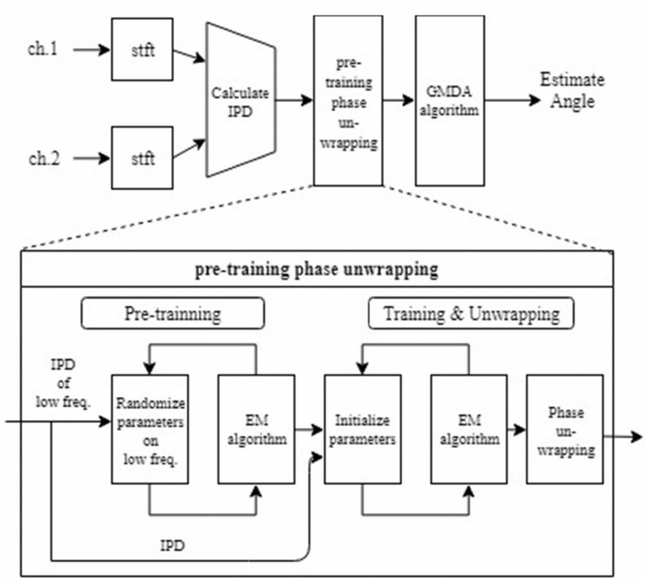


Figure 3. Block diagram of the proposed algorithm

4 Experiments

4.1 Simulation Experiment

Before the start of the actual experiment, we created toy-data to test the algorithm and conducted a simulation experiment. The toy data was created using a noisy speech recorded with a sampling rate of 48kHz from CSTR's VCTK Corpus (Centre for Speech Technology Voice Cloning Toolkit) [22].

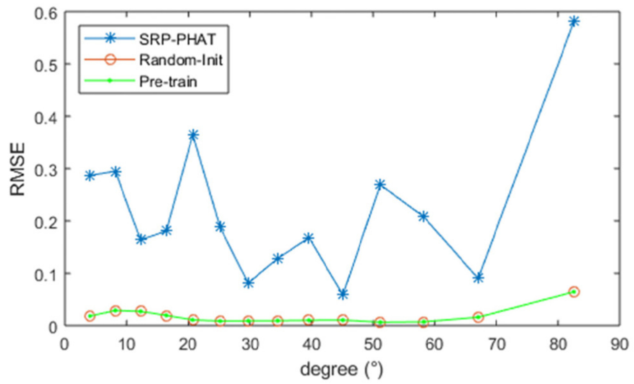
To generate binaural toy data, put the audio signal into channel 1, put the audio signal shifted by 1 to 14 samples into channel 2, and resample the signal to 16 kHz. Then we obtained 14 toy data ranging from 4.0619°to 82.5980°in accordance with the following equations:

$$\text{sample delay} = \tau \times F_s \qquad \textbf{(16)}$$

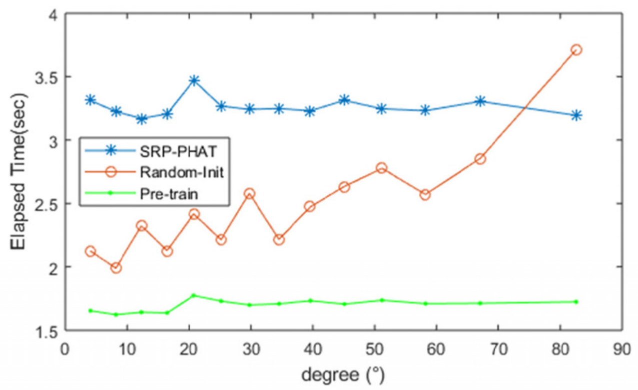
$$\text{sample delay} = d \times \frac{\sin\theta \times F_s}{c} \qquad \textbf{(17)}$$

$$\theta = \sin^{-1}\left(\frac{\text{sample delay} \times c}{d \times F_s}\right) \qquad \textbf{(18)}$$

where F_s is sampling rate. The time delay for each sample is determined from equation (16) and τ is determined from equation (5). Substituting τ into equation (16), we obtain equation (17). Finally, θ is determined from equation (18). Figure 4 illustrates the performance measured by the SRP-PHAT algorithm, conventional GMDA and proposed algorithm with the generated data.



(a) DOA RMSE



(b) DOA elapsed time

Figure 4. Benchmark of toy data compared to various algorithms

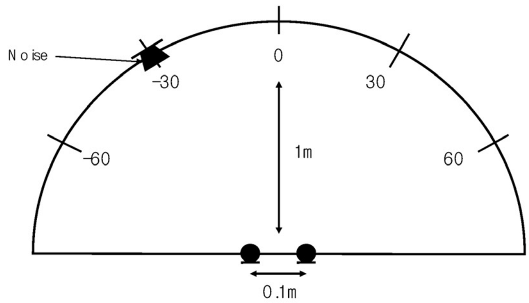
To measure the performance, the toy data of 15 s length was estimated using SRP-PHAT algorithm, conventional GMDA and the proposed algorithm at 1 s interval, and the average of 10 times results was calculated using RMSE. For the estimation using the SRP-PHAT, the resolution of the angle was set to 0.1 units, and for STFT the window size of conventional GMDA and the proposed algorithm are set to 25 ms, and the frames were shifted by 10 ms. The distance between the microphones was assumed to be 0.1 m and the sound velocity was assumed to be 340 m/s. The SRP-PHAT algorithm demonstrated the lowest estimation accuracy, whereas the conventional GMDA and the proposed algorithm demonstrated similar performance on RMSE. The proposed algorithm was the fastest in measuring time. In the proposed algorithm, the estimation accuracy was slightly degraded at the elevation angle, but the estimation time was stable and good.

4.2 Actual Experiment

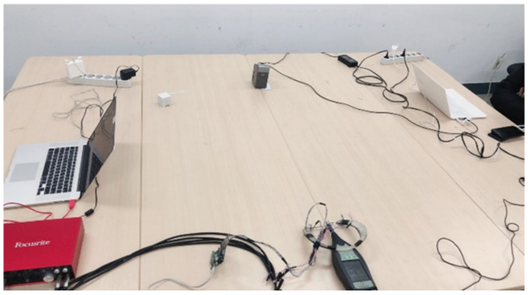
Several experiments were performed to evaluate the performance of the SRP-PHAT algorithm, conventional GMDA and proposed algorithm under various conditions. All experiments use two microphone channels and the source signal as the speech sound that constitutes the TIMIT database [23] with a sampling rate of 16 kHz. The average sound pressure level (SPL) of the speech is 65 dB SPL and the average SPL of noise (white, babble, factory1, volvo, leopard) is 55 dB SPL (SNR ≒ 10 dB). The experimental environment of the DOA estimation system and actual experiment setting are provided in Figure 5. The direction of the speech signal was tested at intervals of 30° from -60° to 60° from 1 m. The noise location was fixed at -30°. Experiments were conducted in the echo room to evaluate the performance depending on the presence or absence of echoes. The size of the echo room was $6m \times 6m \times 2m$. The settings for actual experiment were identical to those of the simulation experiment. The RMSE in the DOA estimation was used to evaluate the performances of the various algorithms.

Table 1 presenting the average DOA RMSE and elapsed time performance shows that the proposed pre-training-based GMDA approach outperformed the original GMDA-based approach, which randomly chose the initial slope, and SRP-PHAT algorithm. All three algorithms used the average of 10 times and 10 people experimental results to compensate for the random variation in performance.

The experimental results showed that the proposed pre-training GMDA approach improved for low angle of the echoic environment in comparison with the SRP-PHAT algorithm and conventional GMDA, respectively.



(a) DOA estimation system



(b) picture of experiment

Figure 5. Experimental environment of DOA estimation system based on two microphones

Table 1. Relative RMSE and mean of elapsed time for DOA estimation obtained from randomly initialized GMDA, SRP-PHAT and proposed method

Method	noise	-60		-30		0		30		60	
		RMSE	Elapsed	RMSE	Elapsed	RMSE	Elapsed	RMSE	Elapsed	RMSE	Elapsed
SRP-PHAT	clean	6.094	9.048	3.514	9.102	1.624	8.993	8.307	8.988	14.865	8.889
	white	25.445	8.881	0.956	9.131	28.559	9.167	57.024	8.951	43.784	8.908
	babble	10.416	9.029	1.647	9.112	13.539	8.980	26.983	8.974	43.784	8.908
	factory	20.124	9.054	1.227	9.127	22.124	9.014	45.782	9.103	73.982	8.850
	leopard	12.030	9.036	1.511	9.147	13.258	9.010	26.225	8.976	41.792	9.015
	volvo	9.324	9.071	1.666	9.128	11.751	9.011	25.584	8.956	39.578	8.976
mean		**13.906**	**9.020**	**1.754**	**9.125**	**15.143**	**9.029**	**31.651**	**8.991**	**42.964**	**8.924**
Random	clean	11.466	2.274	6.662	2.231	5.032	2.198	4.735	2.208	8.760	2.208
	white	10.575	2.217	6.580	2.255	4.087	2.256	5.156	2.203	9.437	2.262
	babble	12.223	2.276	7.131	2.248	4.882	2.203	7.076	2.207	14.284	2.458
	factory	15.831	2.264	7.311	2.233	4.896	2.217	6.014	2.228	11.635	2.202
	leopard	11.053	2.266	7.613	2.264	5.290	2.204	6.331	2.203	12.229	2.235
	volvo	10.929	2.272	7.778	2.258	4.403	2.194	6.367	2.194	10.681	2.225
mean		**12.013**	**2.262**	**7.179**	**2.248**	**4.765**	**2.212**	**5.947**	**2.207**	**11.171**	**2.265**
Pretrain	clean	9.714	2.256	3.870	2.222	1.501	2.195	2.294	2.195	7.107	2.185
	white	10.673	2.210	4.173	2.231	1.767	2.245	2.644	2.189	7.399	2.236
	babble	12.843	2.276	3.802	2.244	1.225	2.193	5.568	2.203	13.886	2.448
	factory	14.791	2.249	3.771	2.232	1.619	2.197	4.625	2.193	10.454	2.173
	leopard	11.479	2.271	4.466	2.258	1.280	2.214	5.529	2.200	12.380	2.222
	volvo	10.580	2.261	4.362	2.238	1.302	2.197	6.429	2.202	10.642	2.207
mean		**11.680**	**2.254**	**4.074**	**2.238**	**1.449**	**2.207**	**4.515**	**2.197**	**10.311**	**2.245**

The SRP-PHAT algorithm significantly degraded performance in noisy environments. This is because the noise direction is incorrectly estimated as the direction of the signal when the DOA is estimated in a noisy environment [24]. The average estimation time of the proposed algorithm was 6.79 s faster than SRP but similar to conventional GMDA. The proposed algorithm showed better performance than the existing algorithms in terms of accuracy and showed particularly good performance at low angles.

5 Conclusion

We proposed a robust approach to incorporate pre-training into the conventional GMDA-based phase unwrapping estimation scheme for source localization. The initial slope for phase unwrapping was estimated by using the Gaussian distribution in low-frequency bins. Compared to the conventional GMDA technique, it was demonstrated that the proposed technique provides better phase unwrapping estimates in the source localization systems.

Furthermore, if more number of microphones are used, or if we apply the deep neural network model [25] to replace the GMDA, or if we analyze the time domain instead of the frequency domain a more precise angle estimation will be possible.

Acknowledgements

This work was supported by INHA UNIVERSITY Research Grant and by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2020R1A2C2004624).

References

[1] Y. Huang, J. Benesty, G. W. Elko, Adaptive Eigenvalue Decomposition Algorithm for Real Time Acoustic Source Localization System, *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings*, Phoenix, AZ, USA, 1999, pp. 937-940.

[2] J. Benesty, Adaptive Eigenvalue Decomposition Algorithm for Passive Acoustic Source Localization, *The Journal of the Acoustical Society of America*, Vol. 107, No. 1, pp. 384-391, January, 2000.

[3] F. Reed, P. Feintuch, N. Bershad, Time Delay Estimation Using the LMS Adaptive Filter—Static Behavior, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 29, No. 3, pp. 561-571, June, 1981.

[4] C. Weiwei, C. Zhigang, W. Jianqiang, Time Delay Estimation Techniques in Source Location, *Journal of Data Acquisition & Process*, Vol. 1, pp. 90-99, 2007.

[5] C. Knapp, G. 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, August, 1976.

[6] Y. Chan, R. Hattin, J. Plant, The Least Squares Estimation of Time Delay and Its Use in Signal Detection, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 26, No. 3, pp. 217-222, June. 1978.

[7] D. Hertz, Time Delay Estimation by Combining Efficient Algorithms and Generalized Cross-correlation Methods, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 34, No. 1, pp. 1-7, February, 1986.

[8] G. Jacovitti, G. Scarano, Discrete Time Techniques for Time Delay Estimation, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 41, No. 2, pp. 525-533, February, 1993.

[9] Kenneth W. K. Lui, Frankie K. W. Chan, H. C. So, Accurate Time Delay Estimation Based Passive Localization, *Signal Processing*, Vol. 89, No. 9, pp. 1835-1838, September 2009.

[10] J. P. Dmochowski, J. Benesty, S. Affes, A Generalized Steered Response Power Method for Computationally Viable Source Localization, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 15, No. 8, pp. 2510-2526, November. 2007.

[11] Y. Zou, H. Xie, L. Wan, G. Han, High Accuracy Frequency and 2D-DOAs Estimation of Conformal Array Based on PARAFAC, *Journal of Internet Technology*, Vol. 16, No. 1, pp. 107-119, January, 2015.

[12] W. Zhang, B. D. Rao, A Two Microphone-Based Approach for Source Localization of Multiple Speech Sources, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 18, No. 8, pp. 1913-1928, October, 2010.

[13] M Ren, Y. X. Zou, A Novel Multiple Sparse Source Localization Using Triangular Pyramid Microphone Array, *IEEE Signal Processing Letters*, Vol. 19, No. 2, February, 2012.

[14] Y. X. Zou, W. Shi, B. Li, C. H. Ritz, M. Shujau, Jiangtao Xi, Multisource DOA Estimation Based on Time-frequency Sparsity and Joint Inter-sensor Data Ratio with Single Acoustic Vector Sensor, *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, Canada, 2013, pp. 4011-4015.

[15] J. L. Flanagan, *Speech Analysis Synthesis and Perception*, Springer Science & Business Media, 1972.

[16] S. T. Roweis, One Microphone Source Separation, *Advances in Neural Information Processing Systems*, Denver, CO, USA, 2000, pp. 793-799.

[17] A. Reddy, B. Raj, Soft Mask Methods for Single-channel Speaker Separation, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 15, No. 6, pp. 1766-1776, August, 2007.

[18] R. McAulay, T. Quatieri, Speech Analysis/Synthesis Based on a Sinusoidal Representation, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 34, No. 4, pp. 744-754, August, 1986.

[19] N. M. Laird, A. P. Dempster, D. B. Rubin, Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society. Series B*

(Methodological), Vol. 39, No. 1, pp. 1-38, December 1977.

[20] C. F. J. Wu, On the Convergence Properties of the EM Algorithm, *The Annals of Statistics*, Vol. 11, No. 1, pp. 95-103, 1983.

[21] Richard A. Redner and Homer F. Walker, Mixture Densities, Maximum Likelihood and the EM Algorithm, *SIAM Review*, Vol. 26, No. 2, pp.195-239, 1984.

[22] C. Veaux, J. Yamagishi, K. MacDonald et al., Superseded-cstr Vctk Corpus: English Multi-speaker Corpus for Cstr Voice Cloning Toolkit, 2016.

[23] DARPA-ISTO, *The DARAPA TIMIT Acoustic-Phonetic Continuous Speech Corpus (TIMIT) Speech Disc Cd1-1.1 edition*, 1990.

[24] C. Zhang, D. Florencio, Z. Zhang, Why Does PHAT Work Well in Lownoise, Reverberative Environments?, *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, NV, USA, 2008, pp. 2565-2568.

[25] S. Chakrabarty, E. A. P. Habets, Broadband doa Estimation Using Convolutional Neural Networks Trained with Noise Signals, *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, USA, 2017, pp. 136-140.

Biographies

Sang-Ick Kang He received the B.S., M.S. and Ph.D. degree in Electronic Engineering from Inha University, Incheon, Korea in 2007, 2009 and 2019, respectively. He is researcher in Korea Telecom, Korea. His main research interests are speech signal processing, machine learning, hearing aids.

Seongbin Kim He received the B.S. degree in 2019 and has been received M.S. course from 2019 in Electonic Engineering from Inha University, Incheon, Korea. His main research interests are speech signal processing, machine learning, hearing aids.

Sangmin Lee He received the B.S., M.S., Ph.D. degree in Electronic Engineering from Inha University, Incheon, Korea in 1987, 1989 and 2000, respectively. He is an a professor in Electronic Engineering of Inha University. His main research interests are speech signal processing, healthcare system design, psyco-acoustics, and brain-machine interface.