9

Source Localization for Dual Speech Enhancement Technology

Seungil Kim[1], Hyejeong Jeon[2], and Lag-Young Kim[2]
[1]*Speech Innovation Consulting Group*
[2]*Information Technology Lab. LG Electronics*
Korea

1. Introduction

Many researchers have investigated multi-channel speech enhancement techniques which can be used for the pre-processing of the speech recognition system. Numerous microphones can give high performance, but they require additional hardware costs and generate the design problem about microphone position. Therefore speech enhancement technique using two microphones is preferred in mobile phone such as LG KM900, iPhone 4 and Nexus One. For enhancing the speech with two or more microphones, the spatial information from the input signal's incident angle should be used. Therefore, various sound source localization(SSL) methods have been used to estimate the talker's direction-of-arrival(DOA). There are two main approaches to localization (Brandstein, 1995), (Dibase, 2000): the steered-beamformer approach, which includes various kinds of beamformers; and time-difference of arrival (TDOA) approach, which includes a generalized cross-correlation (GCC). The steered-beamformer approach has the capability of enhancing a desired signal that originates from a particular direction. The beamformer can steer its response at a particular angle; it can then find the spatial information required to maximize the beamformer output by scanning over a predefined spatial region. For this purpose, we can use a simple conventional delay-and-sum beamformer or many optimum beamfomers (Naguib, 1996). The TDOA approach uses classical time delay estimation techniques, such as cross-correlation, GCC, adaptive time delay estimation, and the adaptive eigenvalue decomposition algorithm (Chen et al., 2006). The most common time delay estimation method is the GCC, which consists of various types such as the unfiltered type, the maximum likelihood (ML) type, and the phase transform (PHAT) type. The GCC-PHAT is a widely used for TDOA estimation method because it works well in a realistic environment.

The resolution of the DOA estimator is deeply related to the aperture size of the array and the number of microphone. A large aperture size and microphones make an accurate estimation result. Therefore, SSL method using two microphones cannot give the accurate direction-of-arrival (DOA) estimation result. Moreover, the implementation of a TDOA estimator requires a voice activity detector (Araki et al., 2007) or a speech/non speech detector (Lathoud, 2006). However, the TDOA estimation often shows a failed result in spite of these kinds of additional processing. Hence, reliable SSL algorithm is needed for dual channel speech enhancement system.

In this chapter, we will define the reliability measure based on waterbed effect of DOA estimator and then show a method of increasing the accuracy of DOA estimation by using reliability measure(Jeon et al., 2007).

2. Dual Speech Enhancement technology

Dual Speech Enhancement (DSE™) is a trademark of advanced two-channel speech enhancement technology developed by LG Electronics. It has been shown that DSE would be competitive to the other state-of-art speech enhancement technologies. DSE technology can be divided into two sub-technologies according to its function and aim. One is the Dual Speech Enhancement for Talk (DSE.T™) and the other is the Dual Speech Enhancement for Recording (DSE.R™).

DSE.T is a solution for speech communication system. Comfortable call is unfortunately impossible in noisy environments. DSE.T can be a new solution to enhance speech quality in noisy place. DSE.T technology was introduced at CES 2009 in Las Vegas via Woo-hyun Baek (LG CTO) as one of the representative technologies prepared by LG Electronics.

DSE.R makes clear video recording with directionality. In DSE.R, two omni-directional microphones are processed and virtually make them as one directional microphone. One more useful thing in DSE.R is the function of electrical steering. Therefore, the user can select the direction of sound focusing. If user wants to record the voice of person who is pictured, user only needs to select “Producer Mode”. If user wants to record the landscape or something else, user will select “Narrator Mode”.

DSE technology was applied to commercial LG mobile phone, KM900 Arena as shown in Fig. 1. Four related video clips are available in following link and they will be also presented in the multimedia appendix of this book.

LG DSE.T technology - http://goo.gl/TUEo

LG’s dual mic noise reduction demo at CES 2009 - http://goo.gl/QlFx

LG DSE.R technology - http://goo.gl/dbz3

DSE.R Test : LG KM900 Arena Video Recording - http://goo.gl/kwzJ

Fig. 1. LG KM900 Arena Phone

3. Sound source localization for Dual Speech Enhancement technology

The direction of talker can be used for DSE technology. However, two microphones are not enough to get high angular resolution at the DOA estimator. Therefore it is needed to reject some unreliable results and select the good one. For the reliable DOA estimation, we adopt the new scheme “Reliability Measure” which arises from waterbed effect. If the obtained reliability measure is lower than a predefined threshold, the result will be rejected. By using the reliable results only, we can decrease the detection failure of the signal.

3.1 Waterbed effect in DOA estimation

The waterbed effect means that if somewhere the amplification characteristic is pushed down, it goes up somewhere else. This term is usually used in filter response. Stoica and Ninness showed the waterbed effect would appear in spectral estimation (Stoica & Ninness, 2004) (Ninness, 2003). He proved that the power spectral density estimated by a periodogram has a constant average relative variance. The searching method for the DOA estimation is very similar to a spectral estimation such as a periodogram. Therefore the waterbed effect in DOA estimation can be obtained by similar process to spectral estimation (Jeon et al. 2007). Let $\Phi(\omega)$ be the power spectral density of a Gaussian white noise process and $\hat{\Phi}(\omega)$ be the periodogram estimate of $\Phi(\omega)$. We can then show that the variance of $\hat{\Phi}(\omega)$ is proportional to the square of the power spectral density (Hayes, 1996). Thus,

$$\text{var}\left\{\hat{\Phi}(\omega)\right\} = \Phi^2(\omega). \tag{1}$$

The average relative variance of $\hat{\Phi}(\omega)$ has the following form:

$$\begin{aligned}&\text{average relative variance}\left\{\hat{\Phi}(\omega)\right\}\\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\text{var}\left\{\hat{\Phi}(\omega)\right\}}{\Phi^2(\omega)}d\omega\\ &= 1.\end{aligned} \tag{2}$$

This phenomenon has been called the waterbed effect.
The waterbed effect in DOA estimation can be reduced by a similar process.
If $R(\theta)$ is the cross-correlation value of GCC-PHAT, then

$$R(\theta) = \sum_{k=0}^{K-1}\frac{X_1[k]X_2^*[k]}{\left|X_1[k]X_2^*[k]\right|}e^{-j\frac{2\pi k}{K}\frac{d\sin\theta}{c}}, \quad \theta \in [-\pi,\pi]. \tag{3}$$

In addition, a power pattern estimate of $R(\theta)$, $\hat{P}(\theta)$, is expressed by

$$\hat{P}(\theta) = \frac{1}{K}\left|R(\theta)\right|^2 = \frac{1}{K}\sum_{k=0}^{K-1}\sum_{l=0}^{K-1}\frac{X_1[k]X_2^*[k]}{\left|X_1[k]X_2^*[k]\right|}\frac{X_1^*[l]X_2[l]}{\left|X_1^*[l]X_2[l]\right|}e^{-j\frac{2\pi(k-l)}{K}\frac{d\sin\theta}{c}}. \tag{4}$$

Furthermore, the expected value of the power pattern is

$$E\left\{\hat{P}(\theta)\right\} = \frac{1}{K}\sum_{k=0}^{K-1}\sum_{l=0}^{K-1} E\left\{ e^{j\frac{2\pi(k-l)}{K}\frac{d\sin\theta_0}{c}} \right\} e^{-j\frac{2\pi(k-l)}{K}\frac{d\sin\theta}{c}}. \tag{5}$$

Let the input signal be a spatially white noise process, and note that the signal is assumed to be Gaussian white noise in the spectral estimation. For spatially white noise, the expected value of $e^{j\frac{2\pi(k-l)}{K}\frac{d\sin\theta_0}{c}}$ is equal to unity when $k = l$ only; otherwise it is equal to zero. Thus, the expected value of the power pattern in (5) is equal to unity. That is,

$$E\left\{\hat{P}(\theta)\right\} = 1. \tag{6}$$

The second-order moment of the power pattern estimate is

$$E\left\{\hat{P}^2(\theta)\right\} = \frac{1}{K^2}\sum_{k=0}^{K-1}\sum_{l=0}^{K-1}\sum_{m=0}^{K-1}\sum_{n=0}^{K-1} E\left\{ e^{j\frac{2\pi(k-l+m-n)}{K}\frac{d\sin\theta_0}{c}} \right\} e^{-j\frac{2\pi(k-l+m-n)}{K}\frac{d\sin\theta}{c}}, \tag{7}$$

and separated by sum of two parts as follows:

$$\begin{aligned} E\left\{\hat{P}^2(\theta)\right\} &= \frac{1}{K^2}\sum\sum_{k+m=l+n}\sum\sum E\left\{ e^{j\frac{2\pi(k-l+m-n)}{K}\frac{d\sin\theta_0}{c}} \right\} e^{-j\frac{2\pi(k-l+m-n)}{K}\frac{d\sin\theta}{c}} \\ &+ \frac{1}{K^2}\sum\sum_{o.w}\sum\sum E\left\{ e^{j\frac{2\pi(k-l+m-n)}{K}\frac{d\sin\theta_0}{c}} \right\} e^{-j\frac{2\pi(k-l+m-n)}{K}\frac{d\sin\theta}{c}}. \end{aligned} \tag{8}$$

The number that can be satisfy $k+m=l+n$ is $\frac{(K+1)(K+2)(2K+3)}{6} - 3K$. Hence, the equation (8) can be simplified to

$$E\left\{\hat{P}^2(\theta)\right\} = \frac{1}{K^2}\left[\frac{(K+1)(K+2)(2K+3)}{6} - 3K\right]. \tag{9}$$

The variance of $\hat{P}(\theta)$ is

$$\begin{aligned} Var\left\{\hat{P}(\theta)\right\} &= E\left\{\hat{P}^2(\theta)\right\} - E\left\{\hat{P}(\theta)\right\}^2 \\ &= \frac{2K^3 + 3K^2 - 5K + 6}{6K^2} \approx \frac{K}{3}. \end{aligned} \tag{10}$$

By using (6) and (10), we can calculate the average relative variance of $\hat{P}(\theta)$ as follows:

$$\begin{aligned} &\text{average relative variance}\left\{\hat{P}(\theta)\right\} \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\text{var}\left\{\hat{P}(\theta)\right\}}{P^2(\theta)}d\theta \\ &= \frac{2K^3 + 3K^2 - 5K + 6}{6K^2} \approx \frac{K}{3}. \end{aligned} \tag{11}$$

This equation is the waterbed effect in the DOA estimation.
Figure 2 shows the result of the DOA estimation. The input signal which had the source in the angle of 30° location was used. The result showed that the direction was correctly estimated and the waterbed effect appeared in the angle of -30°. Even though there was no other signals, the result showed that there is the negative value in the angle of -30°.

3.2 Reliability measure

The concept of reliability measure was presented in (Jeon et al., 2007) and (Jeon, 2008). Figure 3 shows the cross-correlation value of the GCC-PHAT when the speech source is present at a direction of 0° and when the speech source is absent.

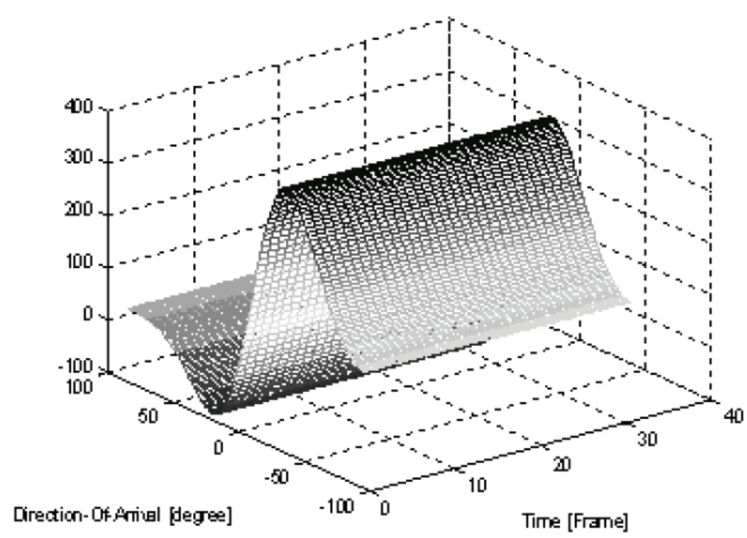


Fig. 2. Waterbed Effect in the DOA Estimation

To test the waterbed effect, we seated the talker in front of the dual microphone receiver. When a dominant source exists, the waterbed effect should cause the mainlobe to be prominent. If there is no directional source, $R(\theta)$ has a flat pattern for all directions. The reliability measure (z), which indicates the prominence of the lobe of $R(\theta)$, is defined as

$$z = f\left(R_{\max} - R_{\min}\right), \tag{12}$$

where $f(x)$ is any monotone-increasing function, $R_{\max}$ is the maximum value of R and $R_{\min}$ is the minimum. We used the formula $f(x) = \left|\frac{x}{K}\right|^2$.

In Fig. 3, the reliability (z) is 0.0177 when speech is absent and the reliability (z) is 0.9878 when speech is present. Because the reliability measure refers to the directivity of the sound source, we only selected the DOA estimation results that had a high reliability value and we clustered those results.

If we assume that a reliable DOA estimation result can be obtained when a dominant directional input exists, we can consider the following two hypotheses of reliability decision problem:

Assuming that reliable DOA estimation result can be obtained when dominant directional input exists, two hypotheses of reliability decision problem are as follow:

H_0 : unreliable DOA estimation result
H_1 : reliable DOA estimation result

And the hypothesis test equation can be defined as

$$z = \frac{1}{K^2}\left(R_{\max} - R_{\min}\right)^2 \underset{d_0}{\overset{d_1}{\gtrless}} \eta, \tag{13}$$

where η is the threshold for the selection of reliable results.

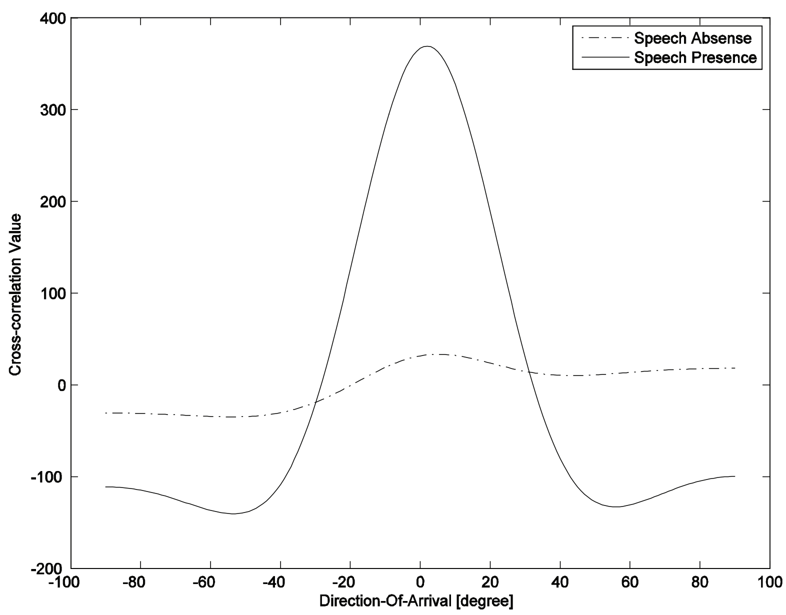


Fig. 3. The cross-correlation value when the speech source is present and when speech source is absent.

3.3 Determination of the threshold

To determine whether the estimate is reliable or not, we need to find the optimum threshold for detection. In (Kim et al., 2008), the optimum threshold was calculated based on maximum likelihood criteria. If we assume that the structure of z is known, reliable source detection can be considered as a simple binary decision problem. To determine which probabilistic model is fit to z, we made observations of z. The recorded data used to calculate the value of z was measured in a quiet conference room. The microphones were 8 cm apart and a single talker was located in front of the microphones. We visually determined that z could be modeled with a Rayleigh pdf as follows:

$$p(z \mid H_0) = \frac{z}{\sigma_0^2}\exp\left(-\frac{z^2}{2\sigma_0^2}\right) \tag{14}$$

$$p(z \mid H_1) = \frac{z}{\sigma_1^2} \exp\left(-\frac{z^2}{2\sigma_1^2} \right). \tag{15}$$

The ML estimation for the unknown parameter $\left(\sigma_0^2, \sigma_1^2\right)$ is given by the maximum value of the log-likelihood function (Schmidt et al., 1996). If we have N_0 items of observation data for z, which is in a decision region Z_0, then

$$\sigma_0^2 = \frac{1}{2N_0} \sum_{i=1}^{N_0} z_i^2 \quad , z_i \in Z_0. \tag{16}$$

Similarly, σ_1^2 can be easily obtained as follows:

$$\sigma_1^2 = \frac{1}{2N_1} \sum_{j=1}^{N_1} z_i^2 \quad , z_j \in Z_1. \tag{17}$$

Figure 4 depicts the observation data distributions fitted with a Rayleigh model. In the quiet conference room, the estimated variances σ_0 and σ_1 are 0.0183 and 0.1997, respectively.
If we make use of the likelihood ratio

$$\Lambda(z) = \frac{p(z \mid H_1)}{p(z \mid H_0)}, \tag{18}$$

the decision rule can be represented by

$$\Lambda(z) = \frac{\sigma_0^2}{\sigma_1^2} \exp\left(\frac{\sigma_1^2 - \sigma_0^2}{2\sigma_0^2 \sigma_1^2} \cdot z^2 \right) \underset{d_0}{\overset{d_1}{\gtrless}} \lambda. \tag{19}$$

If we take the natural logarithm of both sides of (19), then

$$\ln\left(\frac{\sigma_0^2}{\sigma_1^2} \right) - \left(\frac{\sigma_1^2 - \sigma_0^2}{2\sigma_0^2 \sigma_1^2} \cdot z^2 \right) \underset{d_0}{\overset{d_1}{\gtrless}} \ln \lambda. \tag{20}$$

Because the reliability measure, z, always has a positive value in (13),

$$z \underset{d_0}{\overset{d_1}{\gtrless}} \sqrt{\frac{2\sigma_0^2 \sigma_1^2}{\sigma_1^2 - \sigma_0^2} \cdot \left\{ \ln \lambda + \ln\left(\frac{\sigma_1^2}{\sigma_0^2} \right) \right\}} = \eta. \tag{21}$$

When $\ln \lambda$ is equal to zero, the threshold of the ML decision rule (Melsa & Cohn, 1978) can be determined by

$$\eta_{ML} = \sqrt{\frac{2\sigma_0^2 \sigma_1^2}{\sigma_1^2 - \sigma_0^2} \cdot \ln\left(\frac{\sigma_1^2}{\sigma_0^2} \right)}. \tag{22}$$

If we use $\left(\sigma_0^2, \sigma_1^2\right) = (0.0183, 0.1997)$, which is previously calculated, η_{ML} becomes 0.0567 for Fig. 4.

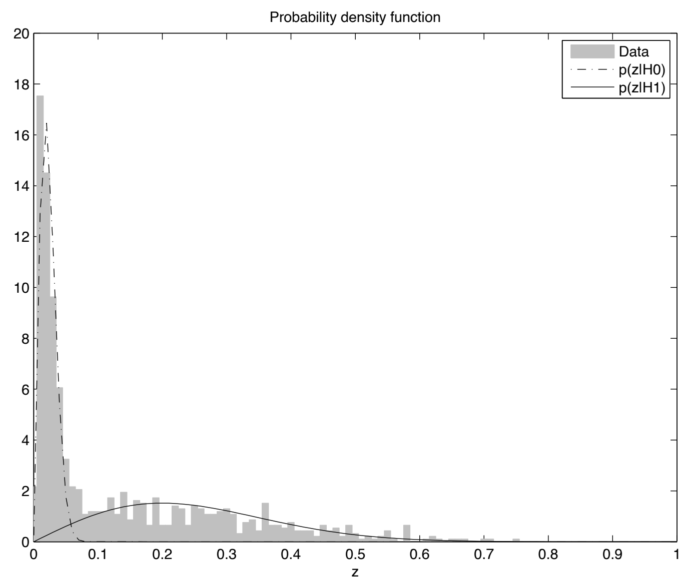


Fig. 4. The cross-correlation value when the speech source is present and when speech source is absent.

4. Performance evaluations

4.1 Simulations

The simulation was performed with the male talker's speech signal. The input speech came from the 30° and the spatially white random noise was mixed to make the SNR of 5dB, 10 dB, 15 dB, and 20 dB. The distance between two microphones was assumed to be 8cm.

The comparison of the estimated DOA is shown in Fig. 5. When the reliability measure and the threshold selection were applied, the average value of the estimated DOA was close to the speech direction. Also, the standard deviation and the RMS error was drastically reduced.

4.2 Experiments

To evaluate the performance of the proposed method, we applied it to the speech data recorded in a quiet conference room. The size of room was 8.5m x 5.5m x2.5m. This conference room, which was suitable for a conference with the several people, generated a normal reverberation effect. The impulse response of the conference room is shown in Fig. 6. The room had various kinds of office furniture such as tables, chairs, a white board standing on the floor, and a projector fixed to the ceiling. The two microphones were placed on the table in the center of the room, and the distance between the microphones was set to 8 cm. Figure 7 shows the experimental setup. The sampling rate of the recorded signal was 8 kHz, and the sample resolution of the signal was 16 bits.

Because the proposed method worked efficiently for the probabilistic model of reliability, we found it useful to eliminate the perturbed results of the estimated DOA in the speech recorded in this room. We compared the results with the normal GCC-PHAT method.

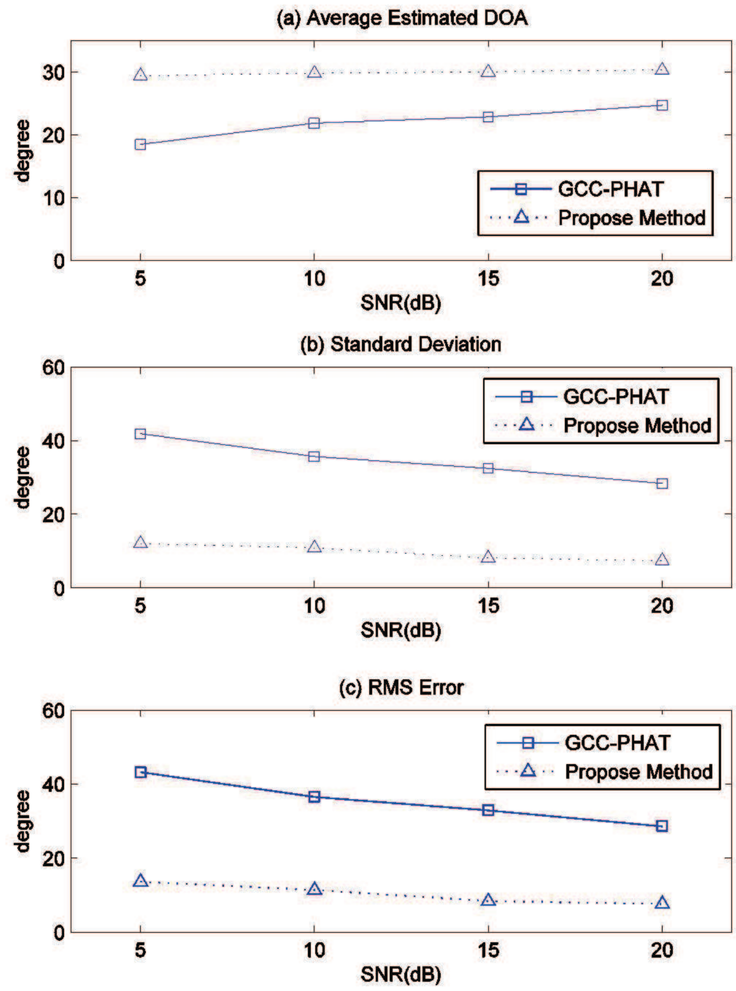


Fig. 5. (a) The average estimated DOA (b) The standard deviation (c) The RMS error when the SNR was 5 dB, 10 dB, and 20 dB

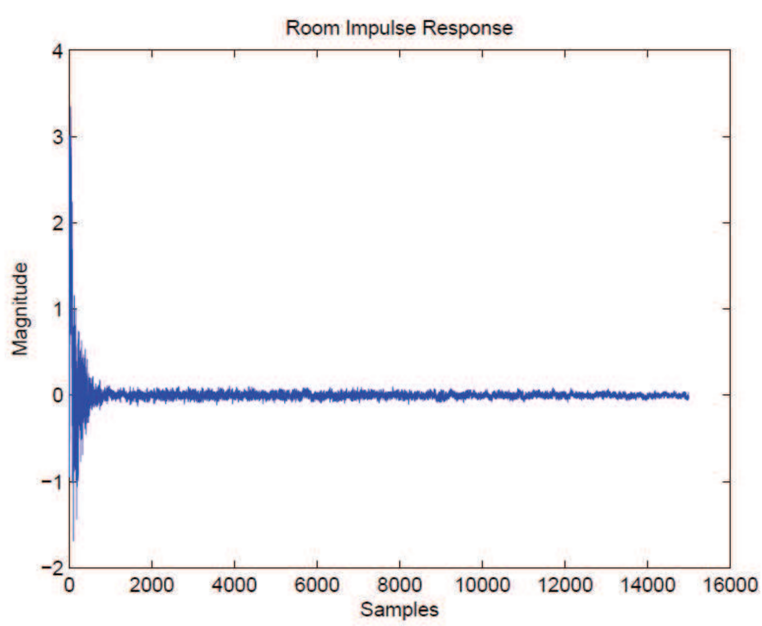


Fig. 6. Impulse response of the conference room for the experiments

4.2.1 Reliability

As shown in Fig. 7 and Fig. 8, we performed the experiment of the DOA estimator for a talker's speech from a direction of 60˚. White noise and tone noise resulted from the fan of the projector.

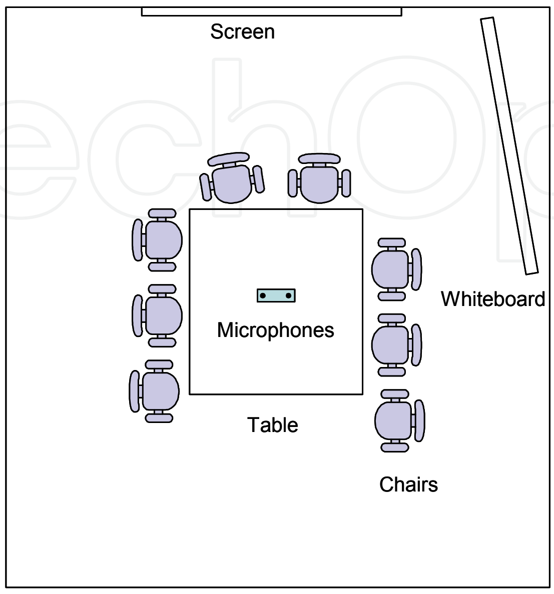


Fig. 7. The Experimental Setup

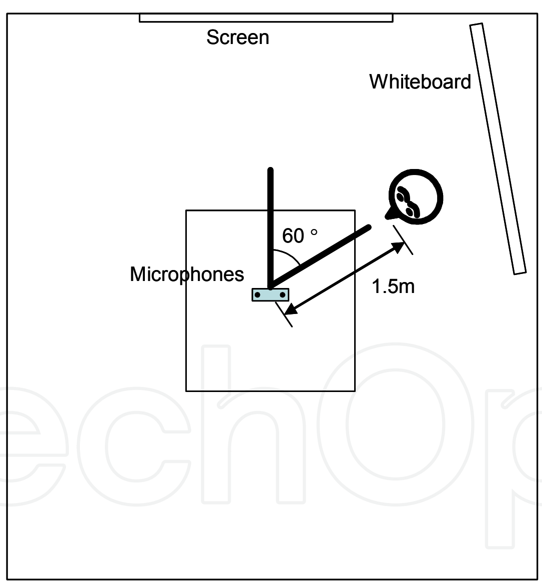


Fig. 8. The Recording Setup for Fixed Talker's Location

Figure 9(a) shows the waveform of the talker's speech. We calculated the direction of the talker's speech on the basis of the GCC-PHAT, and the result is shown in Fig. 9(b). The small circles in the figure indicate the results of the estimated DOA. There are many incorrect results for the estimated DOA, especially in periods when the talker didn't talk. Because of the estimated DOA results for when the talker didn't talk, there was a drastic drop in the performance of the estimated DOA. We calculated the reliability values of the given speech and applied the results to the estimated DOA.

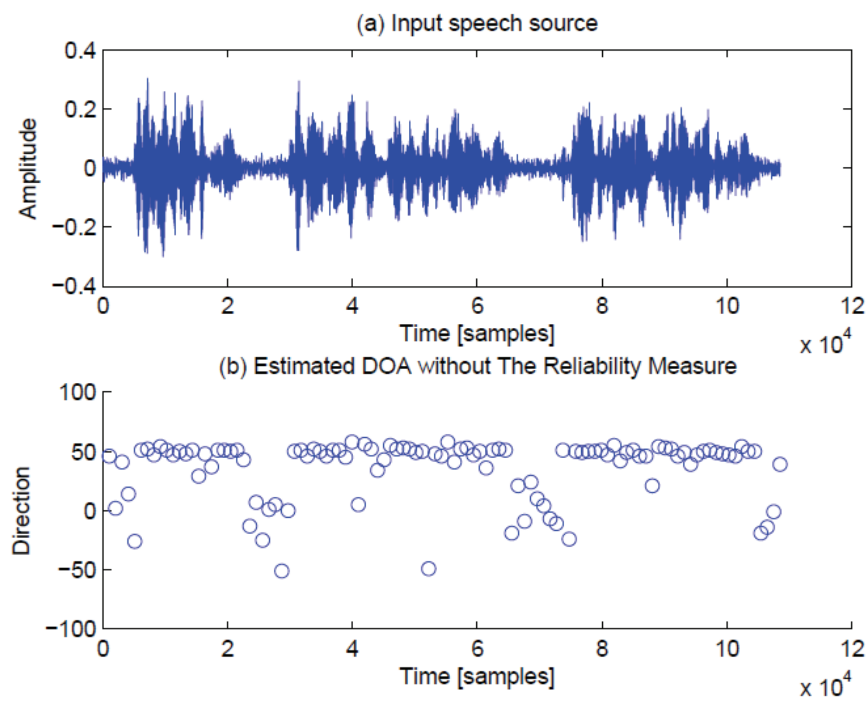


Fig. 9. (a) A waveform of the talker's speech (b) DOA estimation results of GCC-PHAT. It doesn't use the reliability measure.

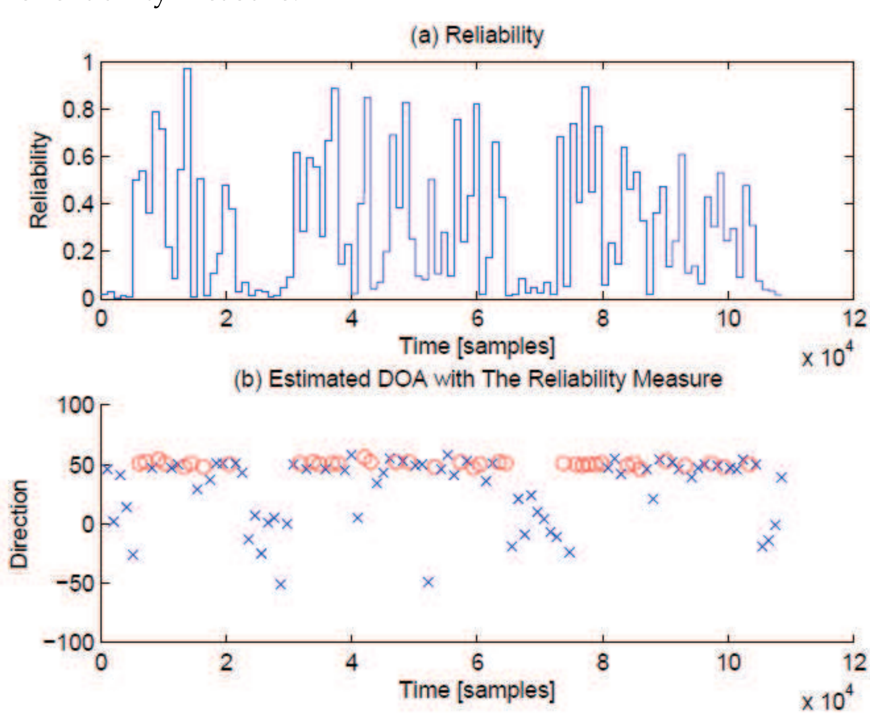


Fig. 10. (a) The calculated reliability for Fig. 9(a). (b) DOA estimation results of GCC-PHAT. It uses the reliability measure and eliminates unreliable estimates.

Figure 10(a) shows the reliability measures of the given speech, and Fig. 10(b) shows the estimated DOA after the removal of any unreliable results. We set the threshold, η, to 0.15. The x-marks indicate the eliminated values; these values were eliminated because the reliability measure revealed that those results were perturbed.

We can trace the talker's direction by using this method. In the experiment, the talker spoke some sentences while walking around the table, and the distance from the talker to the microphones was about 1.5 m. Figure 11 shows the talker's path in the room.

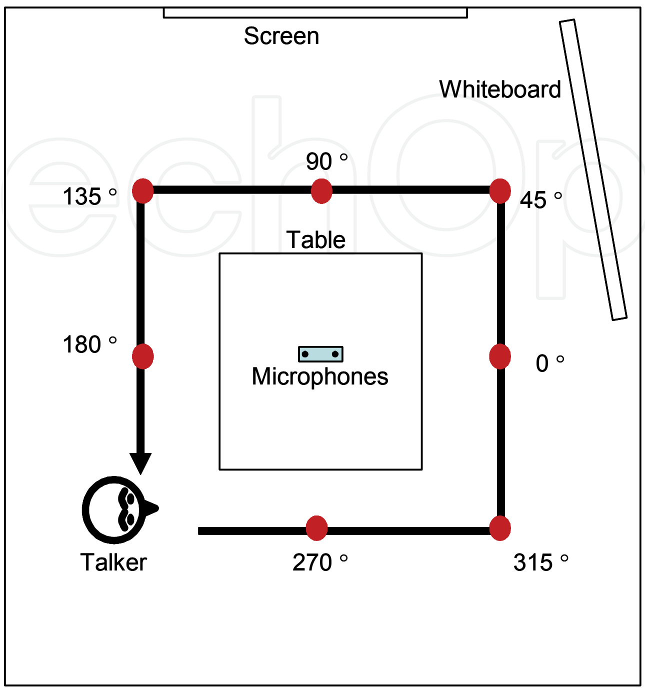


Fig. 11. The Recording Setup for Moving Talker

Figure 12(a) and Fig. 12(b) show the waveform and the estimated DOA based on the GCC-PHAT. The results of the estimated DOA are very disturbed because of the perturbed results. Figure 13(a) shows the calculated reliability values for the speech. By applying the reliability measure, as shown in Fig. 13(b), we can eliminate the perturbed values and produce better results for the estimated DOA. The x-marks represent the eliminated results. By eliminating the perturbed results, we can ensure that the estimated DOA is more accurate and has a smaller variance.

There is a degree of difference between the source direction and the average estimated DOA value. The difference occurs with respect to the height of the talker's mouth. Basically, we calculated the direction of the source from the phase difference of the two input signals. When we set the source direction, we thought the source was located on the same horizontal plane as the microphones. Thus, when the height of the source is not the same as the table, the phase difference cannot be the intended value as shown in Fig. 14. Even though we set the source direction at 90°, the actual source direction was 90°- θ_h , where θ_h is

$$\theta_h = \tan^{-1}\left(\frac{h}{d}\right) \tag{23}$$

Because we used the source signal incident from the direction of 60° in Fig. 8, the actual source direction would be 48.5507° by using (23). The same phenomenon also occured in the next experiment; hence, the estimated DOA range was reduced to (-90°+ θ_h , 90°- θ_h), not (-90°, 90°).

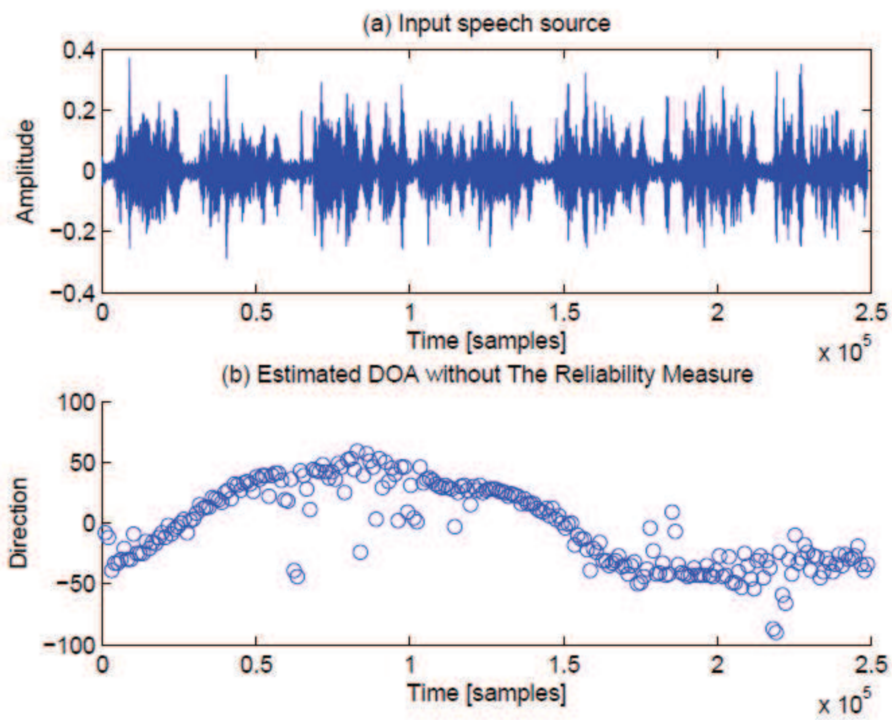


Fig. 12. A waveform of the talker's speech (b) DOA estimation results of GCC-PHAT. It doesn't use the reliability measure.

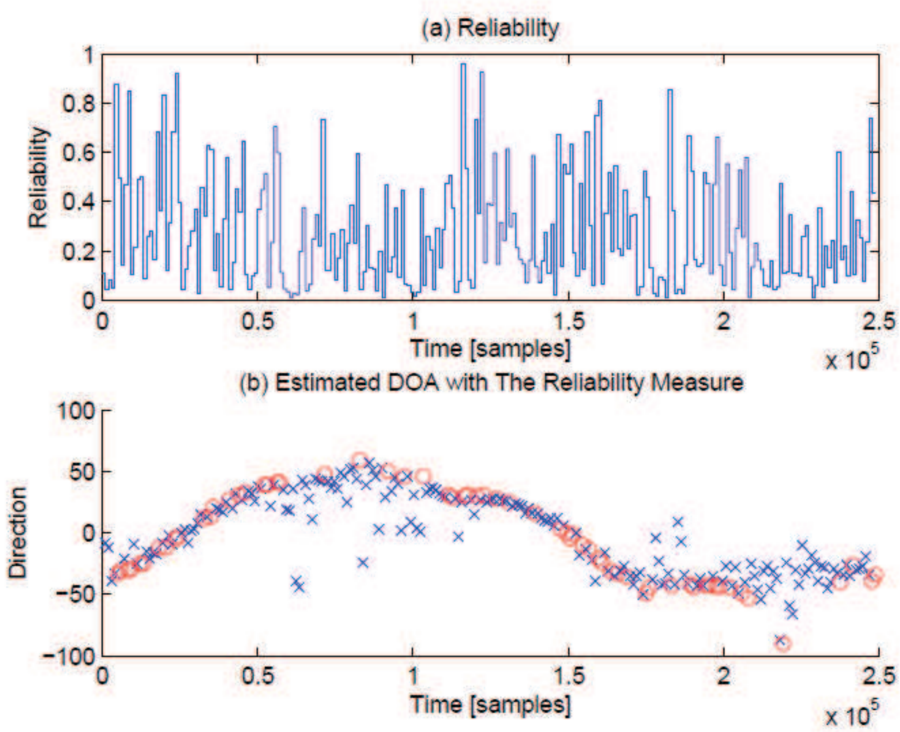


Fig. 13. (a) The calculated reliability for Fig. 11(a). (b) DOA estimation results of GCC-PHAT. It uses the reliability measure and eliminates unreliable estimates.

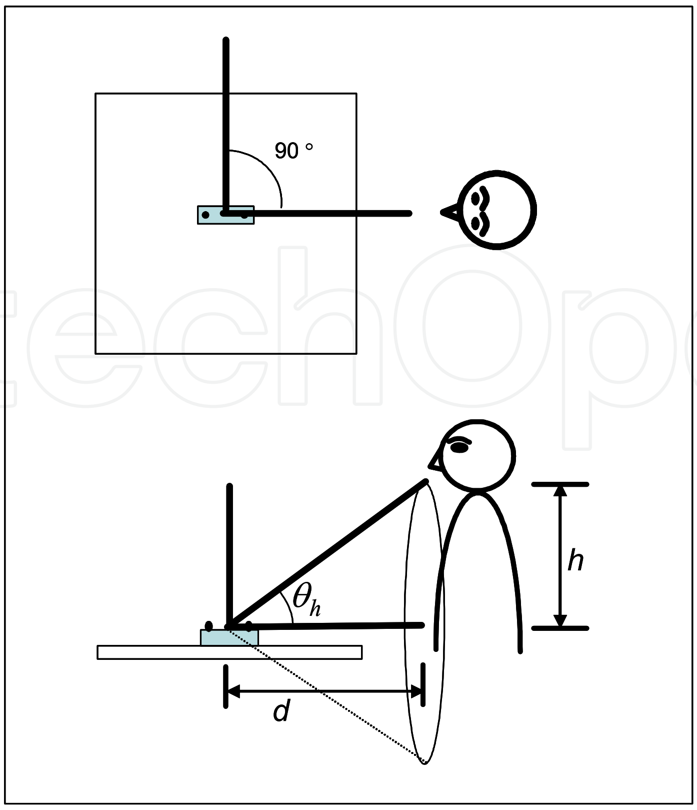


Fig. 14. The Recording Setup for Moving Talker

4.2.2 Speech recognition with DSE technology

The source localization has played an important role in the speech enhancement system. We applied the proposed localization method to the speech recognition system and evaluate its performance in a real car environment (Jeon, 2008).

The measurements were made in a mid-sized car. The input microphones were mounted on a sun visor for speech signal to impinge toward the input device (at the direction of 0˚) as shown in Fig. 15. And a single condenser microphone was mounted between the two microphones. It was installed for the comparison with DSE output. The reference microphone was set in front of speaker. We controlled the background noise with the driving speed. In the high and low noise condition, the speed of car was 80-100km/h and 40-60km/h, respectively.

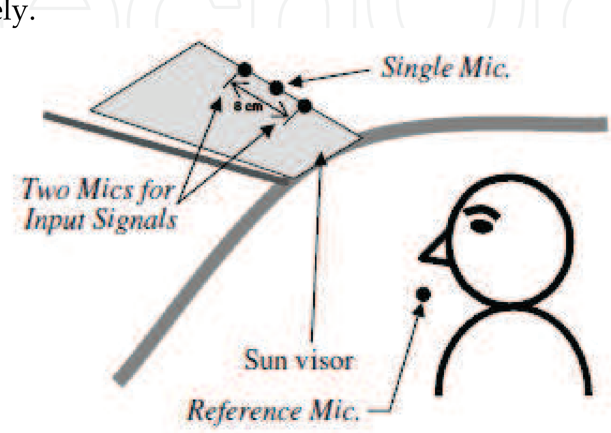


Fig. 15. The experiment setup in a car

For speech recognition test, we used the Hidden Markov Model Toolkit (HTK) 3.4 version as speech recognizer. HTK is a portable toolkit for building and manipulating hidden Markov models. HTK is primarily used for speech recognition research (http://htk.eng.cam.ac.uk/). We used 30 Korean phonemes word set for the experiments. The 30 words were composed of commands which were indispensable to use the telematics system. The speech recognition result is shown in Table 1. The speech recognition rate was decreased according as the background noise was increased.

Noise Type	Speech Recognition Rate
Low (low speed)	73.33
High (high speed)	58.83

Table 1. The speech recognition rate results : No pre-processing

We tested the DSE technology and source localization method using reliability measure. For evaluation, signal-to-noise ratio (SNR) and speech recognition rate were used. The SNR results are shown in table 2. The SNR for the low noise environment was increased from 9.5 to 18.5 and for the high noise from 1.8 to 14.9.
The increased performance of the DSE technology affected to the speech recognition rate. The speech recognition rate is shown in table 3 when the DSE technology was adopted. Without reliability measure, the speech recognition system for the high noise environment didn't give a good result as table 1. However the speech recognition rate was increased from 58.83 to 65.81 for the high noise environment when DSE technology was used.

Method	Low Noise	High Noise
Single Microphone	9.5	1.8
DSE w/o reliability measure	5.2	2.7
DSE with reliability measure	18.5	14.9

Table 2. SNR comparison results

Noise Type	Speech Recognition Rate
Low (low speed)	77.42
High (high speed)	65.81

Table 3. Speech recognition rate results : DSE pre-processing with reliability measure

5. Conclusions

We introduced a method of detecting a reliable DOA estimation result. The reliability measure indicates the prominence of the lobe of the cross-correlation value, which is used to find the DOA. We derived the waterbed effect in the DOA estimation and used this effect to calculate the reliability measure. To detect reliable results, we then used the maximum likelihood decision rule. By using the assumption of the Rayleigh distribution of reliability, we calculated the appropriate threshold and then eliminated the perturbed results of the

DOA estimates. We evaluated the performance of the proposed reliability measure in a fixed talker environment and a moving talker environment. Finally we also verified that DSE technology using this reliable DOA estimator would be useful to speech recognition system in a car environment.

6. References

S. Araki, H. Sawada, and S. Makino (2007). "Blind speech separation in a meeting situation with maximum SNR beamformers," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. I, p. 41-44.

M. Brandstein (1995). A Framework for Speech Source Localization Using Sensor Arrays, *Ph. D Thesis*, Brown University.

J. Chen, J. Benesty, and Y. Huang (2006). "Time delay estimation in room acoustic environments: An overview," *EURASIP Journal on Applied Signal Processing*, Vol. 2006, pp. 1-19.

J. Dibase (2000). A High-Accuracy, Low-Latency Technique for Talker Localization in reverberant Environments Using Microphone Arrays, *Ph. D Thesis,* Brown University.

M. Hayes (1996). Statistical Digital Signal Processing and Modeling, *John Wiley & Sons*.

H. Jeon, S. Kim, L. Kim, H. Yeon, and H. Youn (2007). "Reliability Measure for Sound Source Localization," IEICE Electronics Express, Vol.5, No.6, pp.192-197.

H. Jeon (2008). Two-Channel Sound Source Localization Method for Speech Enhancement System, *Ph. D Thesis*, Korea Advanced Institute of Science and Technology.

G. Lathoud (2006). Spatio-Temporal Analysis of Spontaneous Speech with Microphone Arrays, *Ph. D Thesis*, Ecole Polytechnique Fédérale de Lausanne.

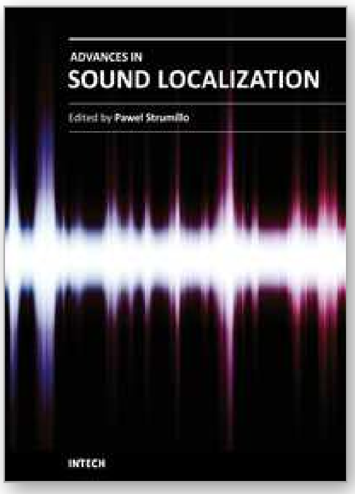
J. Melsa, and D. Cohn (1978). Decision and Estimation Theory, *McGraw-Hill*.

A. Naguib (1996). Adaptive Antennas for CDMA Wireless Networks, *Ph. D Thesis,* Stanford University.

B. Ninness (2003). "The asymptotic CRLB for the spectrum of ARMA processes," IEEE Transactions on Signal Processing, Vol. 51, No. 6, pp. 1520-1531.

F. Schmitt, M. Mignotte, C. Collet, and P. Thourel (1996). "Estimation of noise parameters on SONAR images", *in SPIE International Society for Optical Engineering - Technical Conference on Application of Digital Image Processing XIX - SPIE'96* , Vol. 2823, pp. 1-12, Denver, USA.

P. Stoica, J. Li, and B. Ninness (2004). "The Waterbed Effect in Spectral Estimation," *IEEE Signal Processing Magazine*, Vol. 21, pp. 88-100.



Advances in Sound Localization

Edited by Dr. Pawel Strumillo

ISBN 978-953-307-224-1
Hard cover, 590 pages
Publisher InTech
Published online 11, April, 2011
Published in print edition April, 2011

Sound source localization is an important research field that has attracted researchers' efforts from many technical and biomedical sciences. Sound source localization (SSL) is defined as the determination of the direction from a receiver, but also includes the distance from it. Because of the wave nature of sound propagation, phenomena such as refraction, diffraction, diffusion, reflection, reverberation and interference occur. The wide spectrum of sound frequencies that range from infrasounds through acoustic sounds to ultrasounds, also introduces difficulties, as different spectrum components have different penetration properties through the medium. Consequently, SSL is a complex computation problem and development of robust sound localization techniques calls for different approaches, including multisensor schemes, null-steering beamforming and time-difference arrival techniques. The book offers a rich source of valuable material on advances on SSL techniques and their applications that should appeal to researches representing diverse engineering and scientific disciplines.