THE JOURNAL OF THE ACOUSTICAL SOCIETY OF KOREA Vol.20, No.4E 2001. 12. pp. 52~60

A DSP Implementation of Subband Sound Localization System

Kyusik Park*

*Dept. of Computer Science and Statistics, Dankook University

(Received 9 October 2001; accepted 8 January 2002)

Abstract

This paper describes real time implementation of subband sound localization system on a floating-point DSP TI TMS320C31. The system determines two dimensional location of an active speaker in a closed room environment with real noise presents. The system consists of an two microphone array connected to TI DSP hosted by PC. The implemented sound localization algorithm is Subband CPSP which is an improved version of traditional CPSP (Cross-Power Spectrum Phase) method. The algorithm first split the input speech signal into arbitrary number of subband using subband filter banks and calculate the CPSP in each subband. It then averages out the CPSP results on each subband and compute a source location estimate. The proposed algorithm has an advantage over CPSP such that it minimize the overall estimation error in source location by limiting the specific band dominant noise to that subband. As a result, it makes possible to set up a robust real time sound localization system. For real time simulation, the input speech is captured using two microphone and digitized by the DSP at sampling rate 8192 hz, 16 bit/sample. The source location is then estimated at once per second to satisfy real-time computational constraints. The performance of the proposed system is confirmed by several real time simulation of the speech at a distance of 1m, 2m, 3m with various speech source locations and it shows over 5% accuracy improvement for the source location estimation.

Keywords: *Subband CPSP, Sound localization, Subband, CPSP*

I. Introduction

The sound localization technology is a fundamental step for the advancement of teleconferencing, acoustic surveillance systems, and hands-free speech recognition, etc. These applications require capabilities such as automatic speaker location which can be performed by the acoustic signal processing from the microphone array technology.

Corresponding author: Kyusik Park (kspark@dankook.ac.kr)
Dankook University, Seoul 140-714, Korea

The way of locating sound source in a closed room environment is consists of two major steps as shown in Figure 1. The objective of the system is a detection of acoustic events that can occur in a given environment as well as localization of the acoustic source that generated them.

The first step is to estimate the TDOA (Time Delay of Arrival) between the received active speech signal from the two or more microphone arrays. The next and final step is to calculate the three dimensional location

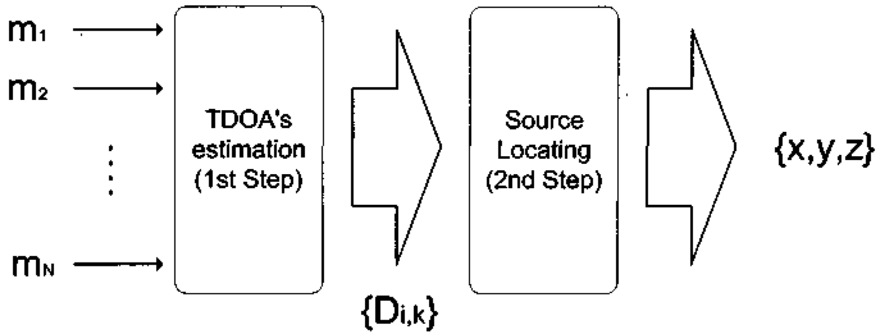


Figure 1. Sound source localization.

(x, y, z) of an active speaker or sound generator using the TDOA estimation from step 1 and the geometrical coordination of microphone array. The accuracy of the second step is dependent on the error incurred in obtaining a set of time delay estimates.

In real environment, the performance of sound localization degrades dramatically because of two main reasons. One is the environmental noise and the other is acoustic reverberation. The environmental noise occurred from such as ventilation, computer air-conditioning, and refrigerator, are normally dominated in low frequency below 500 Hz. These kind of environmental noise enter the microphone array with the interested speaker's speech and they make it difficult to estimate accurate time delay of arrival and in turn affect the accuracy of the location estimates. On the other hand, the acoustical reverberation results from the reflection of sound to the wall or other objectives in the room. These two degradation factors occur in the speech propagation process from the interested speaker's speech (acoustic events) to the microphone sensors and it gives more serious effects as a distance between them increased.

In the last decade, some research effort has been devoted to microphone array processing technique mainly focused on the topics of minimization of those degradation effects[1-6]. Among them, a popular method for sound localization is Cross-Power Spectrum Phase (CPSP) of Omologo[5] and various modified version of this algorithm has been proposed[7,8].

This paper presents real time implementation of subband sound localization system on a floating-point DSP TI TMS320C31. The system determines two dimensional location, that is to say source direction in bearing angle from the center of microphone, of an active speaker in a closed room environment with real noise presents. The implemented sound localization algorithm is Subband CPSP (Cross Power Spectrum Phase) which is a development of CPSP[] and it takes advantage of subband filter banks to minimize the effects of the environmental noise.

The remainder of this paper is organized as follows : Section 2 introduces a fundamental principle of sound localization algorithm and discuss several time delay estimation algorithm. Section 3 proposes the Subband CPSP for the improved sound localization algorithm and the corresponding real time DSP implementation has been explained. Section 4 shows the localization experiments conducted in a real environment and the system performance is evaluated at various different source location conditions. Finally, section 5 gives some conclusion of this paper.

II. Fundamentals of Sound Localization Technique

2.1. Time Delay Estimation and Sound Source Location in Two Dimension

For a given sound source signal $s(t)$, propagate in a generic noisy and reverberation environment as shown in figure 2, the signals received in two microphone $x_1(t)$,

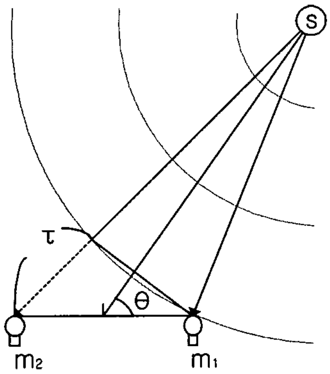


Figure 2. Time delay estimation.

$x_2(t)$ can be expressed as

$$x_1(t) = s(t) + n_1(t)$$
$$x_2(t) = s(t+\tau) + n_2(t) \quad (1)$$

where τ is relative time delay of arrival between two microphones and $n_1(t), n_2(t)$ is noise signal present at each microphone. The goal of the first step of sound localization is to determine time delay τ and if it is known, then the second step, sound source location, can be estimated using time delay τ from step 1 and the geometrical coordination of microphone array m_i, m_j as follows[9].

$$\theta \approx \cos^{-1}\left(\frac{v \cdot \tau}{|\mathrm{m}_i - \mathrm{m}_j|}\right) \quad (2)$$

From equation (2), v is 340m/sec sound speed and θ is an bearing angle of sound source from the center of two microphones. As seen on equation (2), the accuracy of the source location estimation is heavily dependent on the estimation of time delay τ of step 1. Therefore the most efforts for sound localization technique has mainly been focused on the estimation of time delay between the microphone arrays.

One most common method of determining the time delay τ and hence, the arrival angle relative to the microphone sensor axis is to compute the following cross correlation function[2]

$$R_{x_1x_2}(\tau) = E[x_1(t)\, x_2(t-\tau)]$$
$$= \int_{-\infty}^{\infty} S_{x_1x_2}(f)\, e^{j2\pi f\tau} d\tau \quad (3)$$

where $S_{x_1x_2}(f)$ is input Cross Power Spectrum. And the argument τ that maximize equation (3) provides an estimate of time delay. However, the cross-correlation method is known to have a serious problems on the environmental noise and acoustic reverberation.

2.2. Cross Power Spectrum Phase (CPSP) Method

Among many time delay estimation algorithm, Cross-Power Spectrum Phase (CPSP) has been popular because of its simple and reliable time delay estimation result with environmental noise. The CPSP method is to estimate time delay τ that maximize the Generalized Cross-Correlation (GCC) given in equation (4)[5]

$$R_{x_1x_2}^{(g)}(\tau) = \int_{-\infty}^{\infty} \phi_g(f) S_{x_1x_2}(f) e^{j2\pi f\tau} df \quad (4)$$

where $\phi_g(f)$ is a weighting function defined as an inverse of an input signal power spectrum as follows

$$\phi_C(f) = \frac{1}{|S_{x_1x_2}(f)|} \quad (5)$$

With the assumption of an uncorrelatedness between the source signal and the white noise, the equation (4) reduces to

$$R_{x_1x_2}^{(C)}(\tau) = \delta(t-\hat{\tau}) \quad (6)$$

and again the time delay estimation τ is chosen so that maximize equation (6).

III. Real Time DSP Implementation of Subband Sound Localization System

3.1. CPSP and Subband CPSP Algorithm

The weighting function of CPSP method described in section 2.2 is based on the assumption that the statistical behavior of both signal and noise is uniform across the entire spectrum. However, in real environment, there is commonly large amount of acoustic noise below 500 Hz and sometimes a noise concentrated on a specific frequency bands. These kind of noise seriously degrades the overall performance of CPSP. It is therefore desirable to provide a way of alleviating the effects of the specific noise band and it can be attained with subband filter banks.

In this paper, a modified version of CPSP method, Subband CPSP algorithm is proposed and it takes over the advantage of subband filter banks. Subband CPSP first split the input speech signal into arbitrary N subband and apply CPSP method onto each subband. Then the time delay estimation is calculated base on the average of each estimation results on each subband. The algorithm efficiently limit the effect of the specific band noise to itself and make it possible to set up robust sound localization system.

3.2. Subband CPSP Structure

Figure 3 shows the basic structure of proposed Subband CPSP method.

In figure 3, $x_1(t)$, $x_2(t)$ are the speech signals received in two microphones, $H_1(z), H_2(z), \ldots, H_N(z)$ is N subband bandpass filter bank and $Y_1(z), Y_2(z), \ldots, Y_N(z), Z_1(z), Z_2(z), \ldots, Z_N(z)$ represents the signal after the bandpass filter. As seen on the figure, the final result of the algorithm is calculated by taking average of each Subband CPSP as follows

$$S^{(SC)}_{x_1x_2}(f) = \frac{1}{N} \sum_{i=0}^{N-1} S^{(C)}_{y_iz_i}(f) \tag{7}$$

where $S_{y_iz_i}$ represents subband Cross-Power Spectrum on ith subband.

By taking inverse fourier transform of equation (7), the generalized cross-correlation of subband CPSP can be found as

$$R^{(SC)}_{x_1x_2}(\tau) = \frac{1}{N} \sum_{i=0}^{N-1} [R^{(C)}_{y_iz_i}(\tau)] \tag{8}$$

and the desired time delay estimation is a τ that maximize a generalized cross correlation in equation (8).

3.3. Real Time DSP Implementation of Subband CPSP

The proposed Subband CPSP algorithm is implemented in real time DSP using TI TMS320C31. The received speech signal through two microphones is first amplified and digitized by the DSP at 8192hz, 16 bit/sample. Then the Subband CPSP algorithm is applied to get the bearing angles of sound source relative to microphone sensor in real time and the result is transmitted and displayed to PC UI. Figure 4 shows the implemented DSP system block diagram and figure 5 shows the actual hardware system and user interface built in for the subband sound localization.

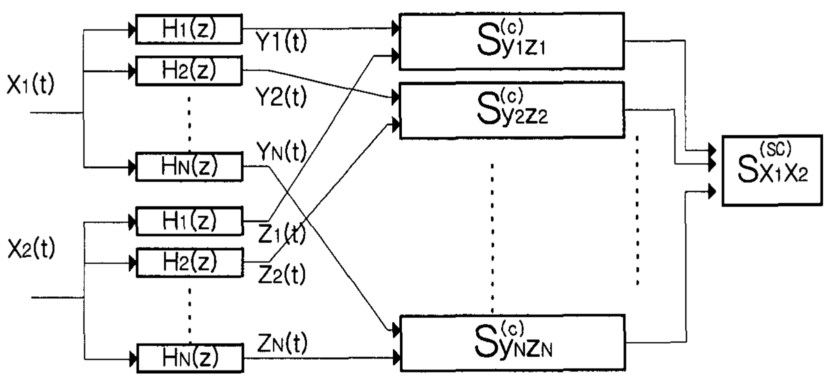


Figure 3. Basic structure of Subband CPSP.

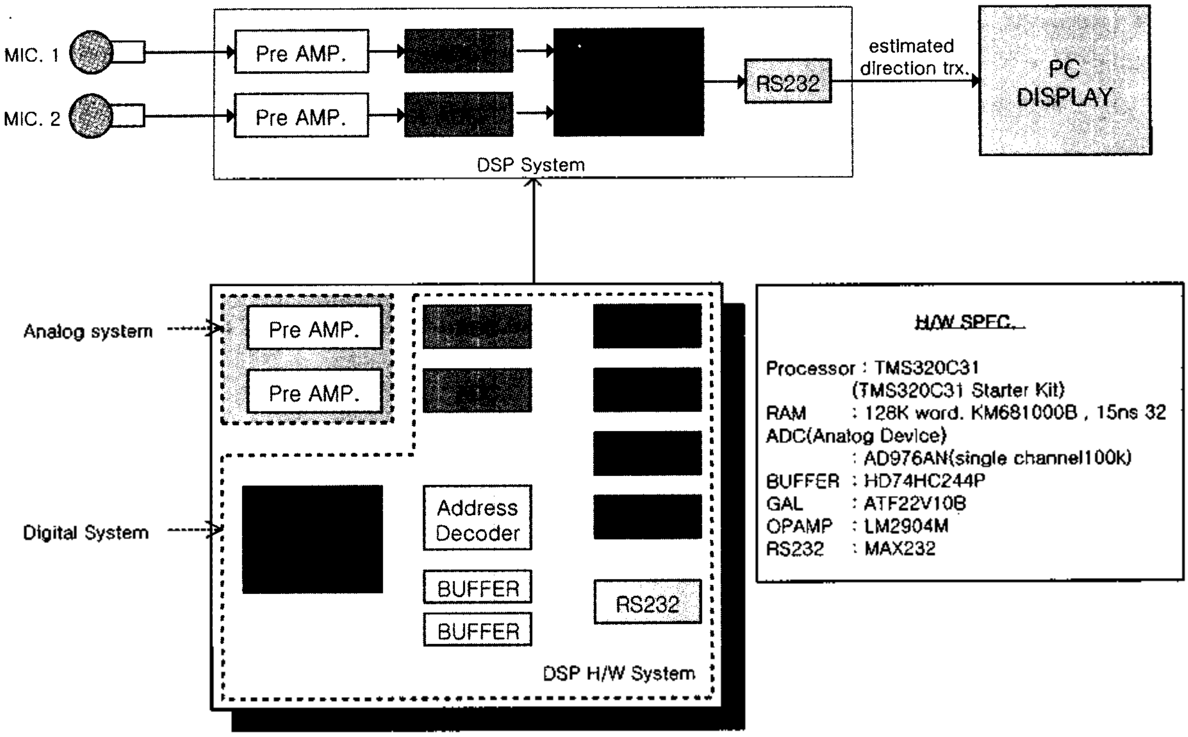


Figure 4. DSP system block diagram.

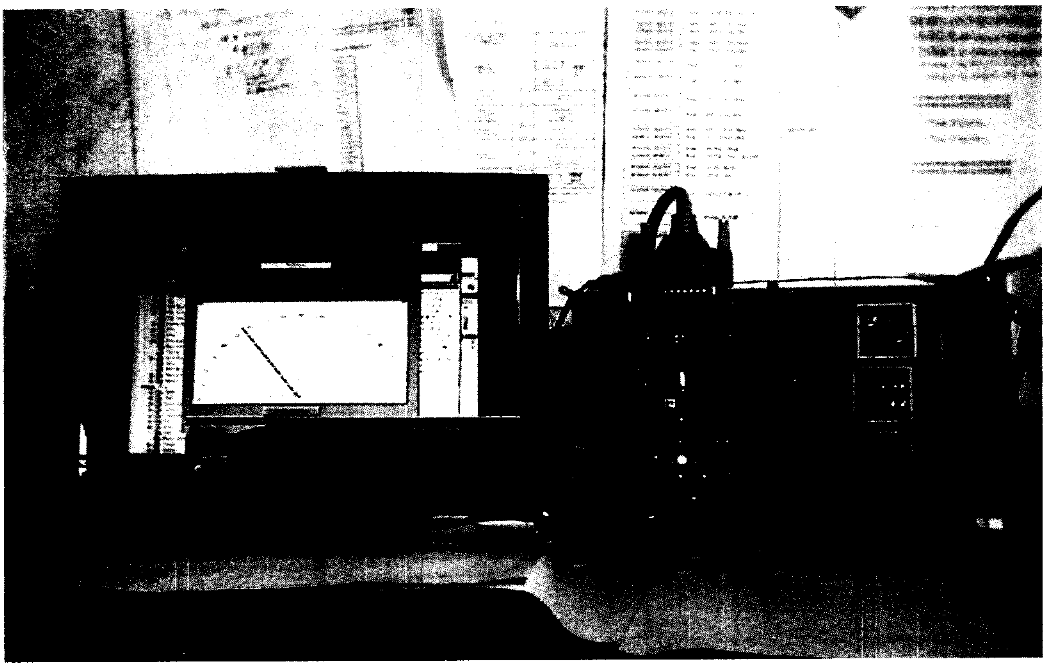


Figure 5. Implemented H/W and S/W UI for real time Subband CPSP System.

IV. Experiments and Performance Evaluation

4.1. Experimental Setup

Figure 6 shows a experimental setup for the real time subband sound localization system. Two microphone is placed apart 0.3m to capture the active speech. The receiving microphone signal is connected to analog input part of the system and amplified using LM2904 operational amplifier. The actual estimation of the sound source

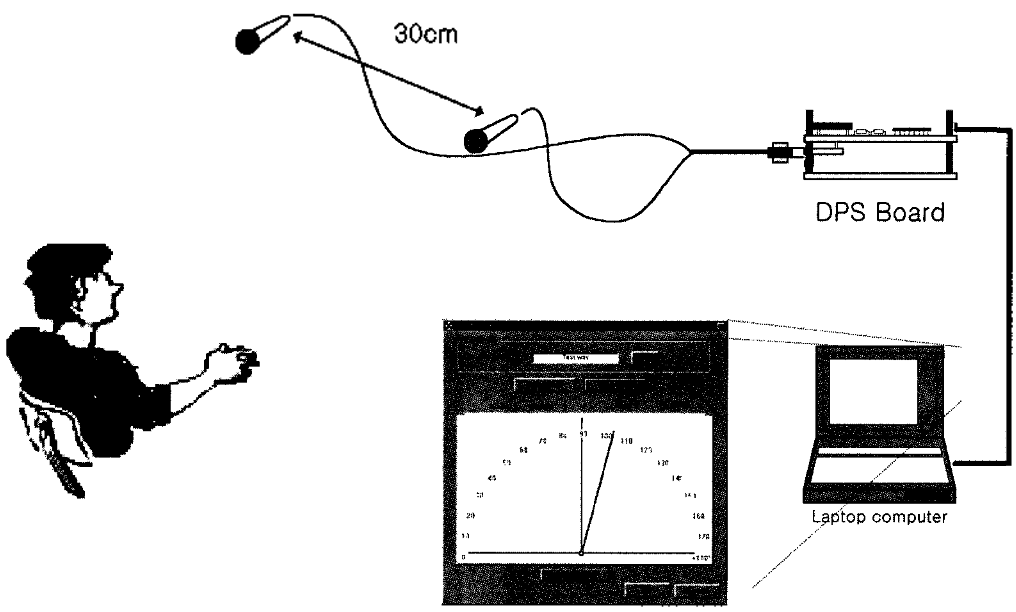


Figure 6. Subband CPSP Sound Localization System.

direction is performed in TI TMS320C31 DSP. The estimation result is transferred to PC through RS-232C serial port, and displayed and stored.

Figure 7 shows the experimental room environment setup for the real time simulation. To evaluate the performance of an algorithm, the system is set up in room about 5X6 meter size with two PCs on the office desks and refrigerator in the left side of the room. The reason for the PC and refrigerator is to simulate noise source in general office environment.

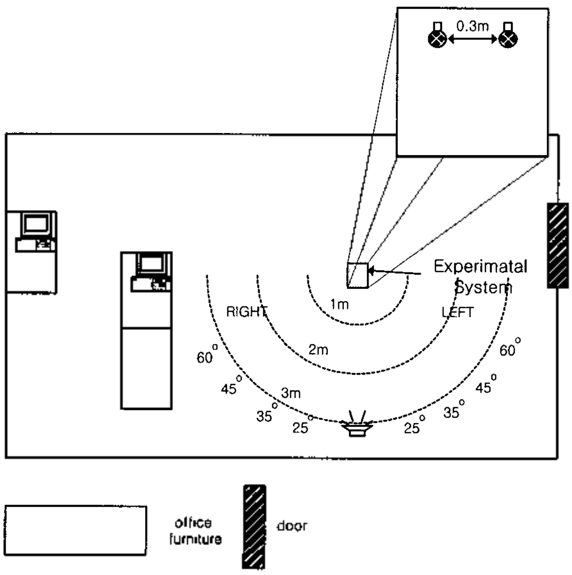


Figure 7. Experimental room setup to test sound localization system.

For the simulation, an 30 minutes long english conversation tape including man and woman's voice interactively was used as an test speech and played repeatedly at the predetermined locations. The input speech signal is first sampled at 8192 Hz and 2048-sample (0.25 sec) frame of input speech data is gathered and the Subband CPSP is performed in 0.75 sec. Therefore the system produces an output sound source direction in every 1 sec. For the experiment, the sound source is located with 8 different angles of -60°, -45°, -35°, -25°, 25°, 35°, 45°, 60° with three different distance of 1m, 2m, 3m from the center of microphone pair. The experiment was performed 300 times at each source direction setup and in total 2400 times of experiments was performed at each distance. Note that this setup with 8192hz sampling rate, integer value of time delay estimation is possible with each estimated source direction such that 3 sample delay with estimated source direction of $\pm 25°$ and 5 sample delay with estimated source direction of $\pm 45°$. For the simulation, Subband CPSP algorithm was implemented with $N=3$ subband filter bank which split the input speech signal into Low, Middle, High band. Figure 8 shows flowchart of procedure to estimate the sound source direction with Subband CPSP.

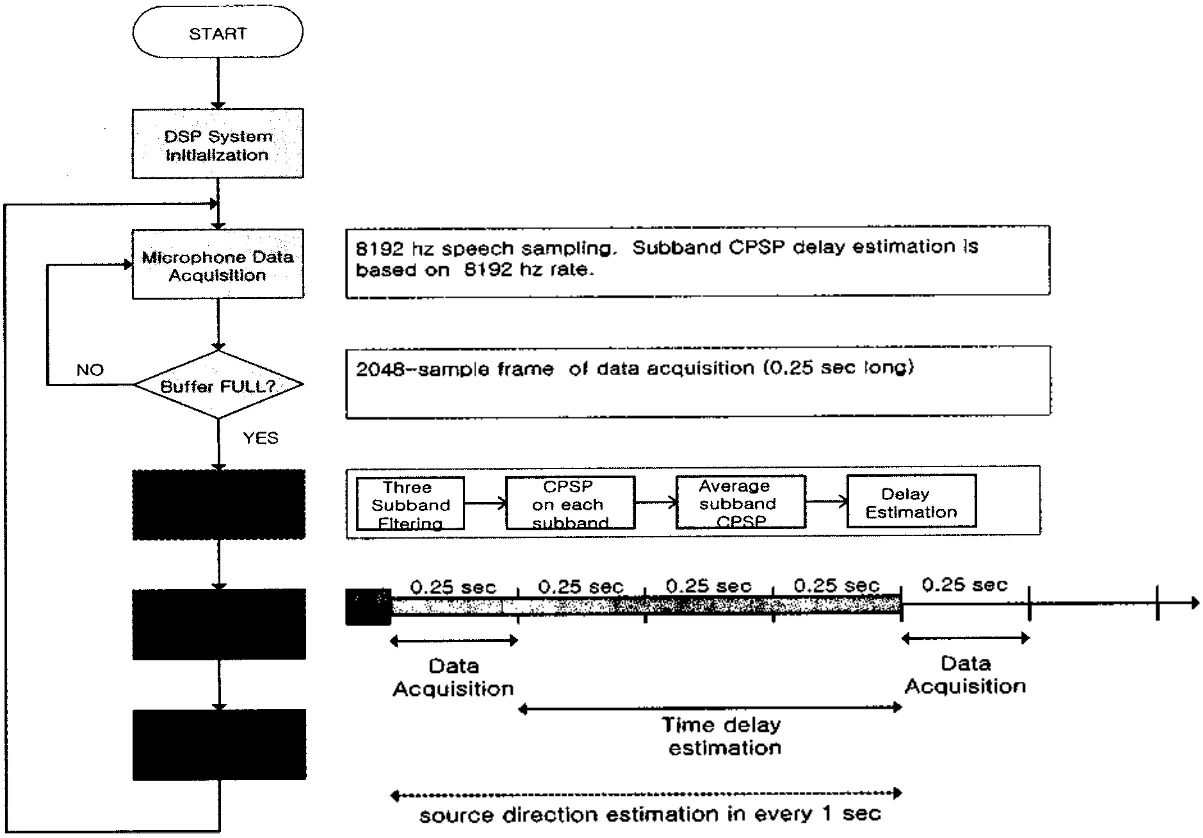


Figure 8. Flowchart of procedure to estimate the sound source direction with Subband CPSP.

4.2. Performance Evaluation

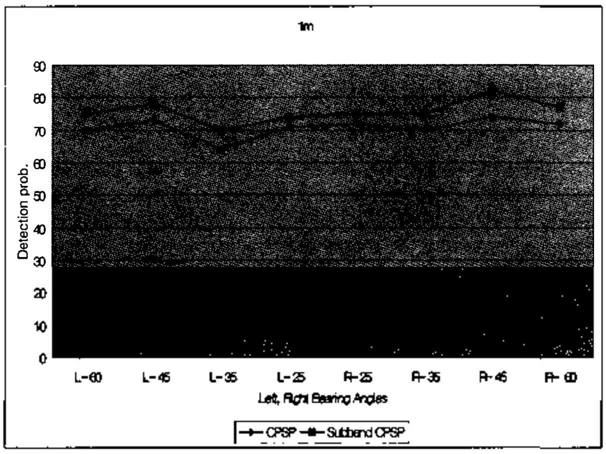
Table 1 summarize the entire experimental results of real time sound localization system. In the table, L-xx represents accurate source detection probability in percentage when the source is assumed to locate in xx degree in left side. On the other hand, R-yy represents those for the case of source assumed to be in right side yy angle. Table 1 also compares the estimated results of Subband CPSP with those of CPSP at a distance of 1m, 2m, 3m. As seen on the table, Subband CPSP provides around 9% improvement at maximum over traditional CPSP and in overall, Subband CPSP guarantee 5% superiority. From the table, as a distance gets longer, the performance of the system is getting worse and below 2m, the system shows reliable performance. By taking account of silence periods of testing speech which we have not considered in this paper and usually takes 20 to 30% of whole speech, the proposed system performs fairly good enough below 2m distance.

Table 1. Source detection prob. with bearing angles.

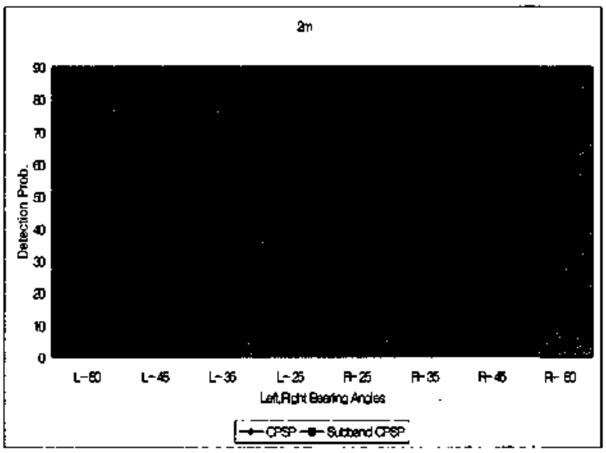
	Bearing Angles	L-60	L-45	L-35	L-25	R-25	R-35	R-45	R-60
1m	CPSP	70	73	64	71	72	70	74	72
	Subband CPSP	75	78	70	74	75	75	82	77
		L-60	L-45	L-35	L-25	R-25	R-35	R-45	R-60
2m	CPSP	55	63	61	60	61	57	62	60
	Subband CPSP	68	73	68	69	67	54	72	65
		L-60	L-45	L-35	L-25	R-25	R-35	R-45	R-60
3m	CPSP	30	32	30	30	28	29	32	29
	Subband CPSP	34	35	31	32	31	32	34	30

Figure 9 displays the experimental results of Subband CPSP and CPSP for the case of sound source location at 1m, 2m, 3m respectively.

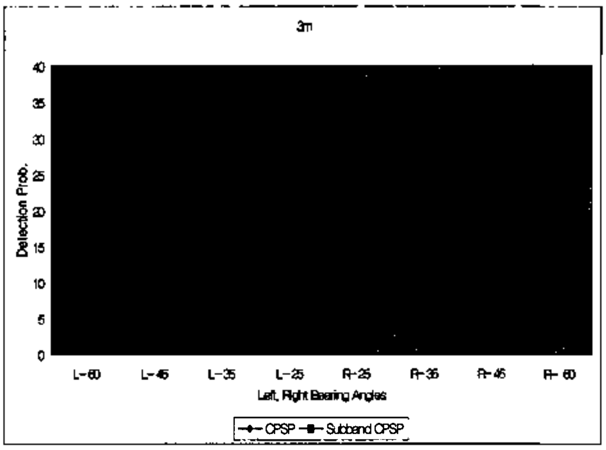
From figure 9(a), at 1m source distance from the microphone, both Subband CPSP and CPSP shows over 60% detection probability and Subband CPSP shows 2~3% better performance over CPSP. Figure 9(b) shows the experimental results at a source distance of 2m and again the Subband CPSP system shows better performance than CPSP. An experimental results at a distance of 3m source distance is shown in figure 9(c). It shows somewhat degraded performance than that of 1m, 2m. The main reason for this degradation is due to the relative increasement of environmental noise power as a source distance gets longer. Due to the low SNR and so the dominance of the noise in each received microphone signal, it yields false 0° source direction for almost 50% of experiment.



(a) Detection prob. at 1m distance



(b) Detection prob. at 2m distance



(c) Detection prob. at 3m distance

Figure 9. Source detection prob.

V. Conclusion

In this paper, a low-cost real time sound localization system using two microphone array was presented with a floating-point DSP TI TMS320C31. A new Subband CPSP algorithm was proposed and compared with traditional CPSP method. A proposed Subband CPSP method takes advantage of subband filter banks to provide a way of alleviating the effects of the specific band dominant noise and make it possible to set up robust sound localization system. The real time system has been tested successfully in a typical 5X 6 meter size room with typical environment noise and it shows 5%~9% accuracy improvement for the source location estimation. It is believed that the performance of the system can be further improved by implementing the speech detector that takes the silence period of speech into consideration. Another way of improving the system might be selective processing of specific subband of CPSP that has highest speech signal energy. All these works are under investigation.

References

1. A. papoulis, "Probability, Random Variables and Stochastic Process," New York, McGraw-Hill, 1965.
2. C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans., Acoust.,*

Speech, Signal Processing, vol. **ASSP-24**, pp. 320-327, Aug., 1976.
3. P. R. Roth, "Effective measurements using digital signal analysis," *IEEE Spectrum*, vol. **8**, pp. 62-70, Apr. 1971.
4. G. C. Carter, A. H. Nuttal and P. G. Cable, "The smoothed coherence Transform," *Proc. IEEE*, vol. **6**, pp. 1497-1498, Oct. 1973.
5. M. Omologo and P. Svaizer, "Acoustic event localization using a cross-power spectrum phase based technique," *Proceedings of ICASSP'94, Adelaide, Australia*, vol. 2, pp. 273-276, 1994.
6. C. Eckhart, "Optimal rectifier systems for detection of steady signals," Scripps. Inst Oceanography, Marine Physical Lab., Univ. California, Rep. SIO 12692, Ref. 52-11, 1952.
7. D. V. Rabinkin, Richard J. Renomeron, and J. L. Flanagan, "Estimation of wavefront arrival delay using the cross-power spectrum phase technique," 132nd ASA, Honolulu, USA, Dec. 4, 1996.
8. M.S. Brandstein, J. E. Adcock, and H. F. Silverman, "A practical time delay estimator for localizing speech sources with a microphone array," *Computer, Speech*, and Language, 9(2):153-169, April 1995.
9. M. Brandstein, *A framework for speech source localization using sensor arrays.*, Doctoral Dissertation, Brown University, May 1955.

[Profile]

● **Kyusik Park**

Kyusik Park received B.S, M.S and Ph.D degrees in 1986, 1988, and 1994, respectively, all from the department of Electrical Engineering of Polytechnic University, Brooklyn, NY USA. In 1994, he joined the semiconductor division of Samsung Electronics as a staff engineer. He is currently an assistant professor with the department of computer science in Dankook University, Seoul Korea. His research interests are digital signal, speech, and audio processing, and digital communication.