Electronic Journal «Technical Acoustics»
http://www.ejta.org

2006, 11

Doh-Hyoung Kim[1], Youngjin Park[2]

Korea Advanced Institute of Science and Technology,
ME3076, 373-1, Guseong-Dong, Yuseong-Gu, Daejon, 305-701, Republic of Korea

Development of moving sound source localization system

Received 13.04.2005, published 12.05.2006

This paper proposes a novel approach of moving sound source localization using adaptive time delay estimation (TDE) algorithm and active-positioning of microphone arrays. Using the adaptive TDE that continuously estimates the time differences between the captured signals in the microphones sensors, the active-positioning controller keeps track of the source direction by rotating arrays mechanically. Theoretical analysis and computer simulations of the convergence characteristics of the proposed TDE algorithm are presented. The active-positioning array guarantees the highest delay-position sensitivity with smaller number of microphones than the fixed arrays. The overall performance is shown by using an experimental prototype system.

1. INTRODUCTION

Sound source localization is to estimate the location of sound sources using the measurements of the acoustic signals by microphone arrays [1]. Moving sound source localization [2–5] has several additional difficulties because the signals captured are non-stationary. The proposed method employs an explicit adaptive time delay estimation (EATDE) algorithm [6–10] and active array positioning. EATDE methods explicitly parameterize an adaptive delay estimate to minimize some kind of delay-error functions. It has the ability of tracking time-varying delay parameter very fast without a prior knowledge of the statistical characteristics of signals [8]. However, it may converge to incorrect delay estimation when the delay-search range or the signal bandwidth is wide [6]. In this paper, a modified EATDE using wavelet transform is proposed to avoid such local convergence phenomenon. The algorithm uses Haar wavelet [11] transform of cross-correlations of captured signals instead of its simple gradient. We also present an active positioning method of microphone array. One of the disadvantages of a source localization using delay information is that the sensitivity of position estimation depends on the relative positions of microphones and a sound source since the relationship between the delay and positions is nonlinear [1]. An active positioning controller can track the optimal array direction according to the source movement using a feedback loop of delay estimator and motor-driven microphone array. The characteristics of the proposed system are investigated through theoretical analysis, computer simulations and the experiment results using a prototype experimental system.

[1] corresponding author, dh_kim@kaist.ac.kr
[2] yjpark@kaist.ac.kr

2. ADAPTIVE TIME DELAY ESTIMATION

Time delay estimation (TDE) between signals received at two spatially separated sensors can be mathematically modeled as

$$\begin{aligned} x(t) &= s(t) + w_1(t), \\ y(t) &= s(t - d_0) + w_2(t), \end{aligned} \tag{1}$$

where $s(t)$ is the source signal, $w_1(t)$ and $w_2(t)$ are the corrupting white noises, d_0 is the time difference between the received signals [12].

It is assumed that $s(t)$, $w_1(t)$, $w_2(t)$ are mutually uncorrelated, zero-mean, stationary processes. The task is to estimate and track the time delay d_0. The existing adaptive TDE methods are divided into two general categories: the implicitly-adaptive methods [13–15] and the explicitly-adaptive methods [6–10]. The implicitly-adaptive method uses adaptive filters for modeling the cross-correlations or delays between the received signals. The delay is estimated as the location of the maximum. Alternatively, in the explicitly-adaptive time delay estimation (EATDE) method, the delay d is explicitly parameterized and adapted to minimize a delay-error function $g(d[n])$ as

$$d[n+1] = d[n] + \mu g(d[n]), \tag{2}$$

where μ is an adaptation size. The adaptation iterates until d converges to true delay d_0. n stands for an integer time index, a positive step-size. The delay-error function based on gradient of the cross-correlations of signals and a steepest-descent method is generally used. EATDE is simple, computationally efficient, and suitable for time-varying delay applications like moving platforms. The conventional EATDE algorithms assume that the true delay d_0 and delay estimate d are limited to ensure that the correlations would have only one maximum at $d = d_0$.

This assumption is not valid when the actual delay search range is wider than the *unimodal* region, which is a part of main lobe of the cross-correlation where the slope of left part of a maximum is positive, and the slope of right part is negative, hence, a gradient search method can converge to a global maximum. The search range of delay estimation is determined by the distance between two sensors while the unimodal region is determined by the signal bandwidth. If two sensors are placed too wide with respect to the signal bandwidth, the correlations would have multiple local maxima in the search range. Therefore, a gradient search based optimization method cannot converge to these local maxima. A sample cross-correlation function is shown in Fig. 1. The global maximum or true delay is at $d_0 = 0$, but conventional EATDEs using simple gradient search method cannot converge to zero if the initial delay estimate is outside the convergence range (– 0.5, 0.5).

We propose a new EATDE algorithm which converges to a global maximum even in this case. This algorithm employs the wavelet transform of the correlations instead of its simple gradient:

$$d[n+1] = d[n] + \mu R(d[n]). \tag{3}$$

Continuous wavelet transform $R(d)$ of cross-correlation r_{xy} is defined as

$$R(d) = \int_{-\infty}^{\infty} r_{xy}(\tau) w(\tau - d) d\tau \,, \tag{4}$$

where $w(t)$ is a wavelet function.

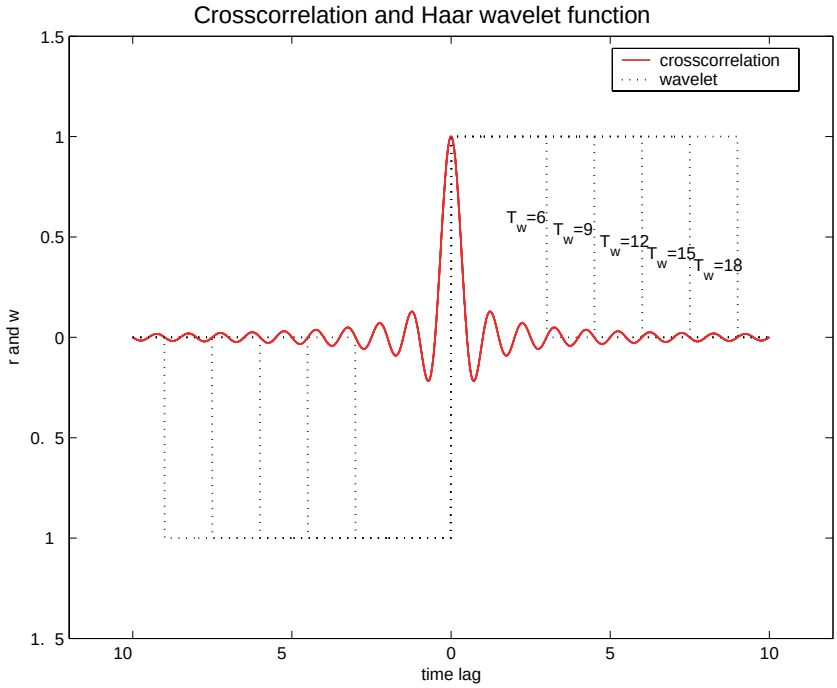

Figure 1. Haar wavelet and cross-correlation of baseband signal

We also propose a prefiltering implementation of wavelet transform in order to reduce the computational load of transform. This method was originally devised for Hilbert transform in [16], but can be applied to general linear integral transforms as well. If *x* and *y* are stationary, transform $R(d)$ of r_{xy} is the same as a cross-correlation of signal *y* and signal x_f filtered by the reversed wavelet $w(d-t)$:

$$R(d) = r_{x_f y}(d) \,. \tag{5}$$

The cross-correlation $r_{x_f y}(d)$ is estimated with a one-point sample mean

$$\hat{r}_{x_f y}(d) = x_f[n] y[n - d/T_s] \,, \tag{6}$$

where T_s is a sampling time, $x_f[n]$ is calculated from a convolution of a sampled signal $x[n]$ and a wavelet prefilter.

The delayed value of y is computed by using an FIR fractional delay filter. The additional computational burden of the proposed method to the conventional EATDE for pre-filtering is very small, with no multiplication/division, but only addition/subtraction. In this research, we chose a sampled Haar wavelet [11]:

$$w[n]=\begin{cases}1 & \text{if } 0<n<\mathrm{L}\\ -1 & \text{if } -\mathrm{L}<n<0\\ 0 & \text{otherwise (including } n=0)\end{cases}, \tag{7}$$

where $L=T_w/T_s$ and T_w is a support of Haar wavelet.

Consider an ideal baseband signal $s(t)$ with a cutoff frequency f_c. Then the autocorrelation $r_s(\tau)$ is a sinc function and the cross-correlation $r_{xy}(\tau)$ is $r_s(\tau-d_0)$. Without loss of generality, the delay d_0 can be assumed zero.

In this case, the sign of $R(d)$ is

$$R(d)\begin{cases}<0 \text{ if } 0<\mathrm{d}<\max(T_w,T_c)\\ =0 \text{ if } d=0\\ >0 \text{ if } -\max(T_w,T_c)<d<0\end{cases}, \tag{8}$$

where $T_c=1/f_c$ [17].

For instance, a cross-correlation function with $T_c=1$ and Haar wavelet with $T_w=6,9,12,15,18$ and their corresponding wavelet transforms $R(d)$ are presented in Fig. 1. and Fig. 2. This numerical result illustrates above inequalities, eq. (8).

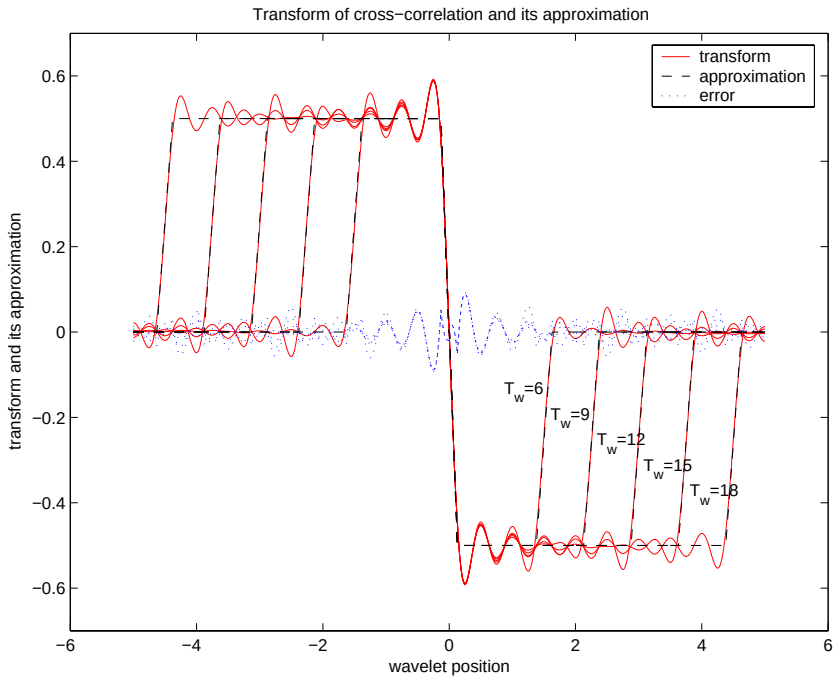

Figure 2. Wavelet transforms and approximations

These inequalities show that $R(d)$ has the similar sign-characteristics with a gradient of the cross-correlation, and hence it can be used as a maximum searching information like gradients. But the region of convergence when wavelet is applied is wider than when the gradient is applied, as shown in eq. (8) and Fig. 2. Moreover, it can be controlled by the user parameter T_w, width of wavelet. This means that the proposed algorithm statistically converges to a true delay if we set the width T_w wider than the delay search range. Simulation tests were carried out to verify this convergence property.

We compared the proposed algorithm with a conventional gradient search method. Ten experiments with different initial points $d[0]$ from 1 to 10 are described in Fig. 3. The source signal was a baseband signal with a normalized cutoff frequency $f_{cn} = f_c/f_s = 0.5$ and the sampling frequency $f_s = 10\ kHz$, so the unimodal region of correlation was about $[-3T_s, 3T_s]$. Corrupting white noises with $SNR = 20\,dB$ were added. The proposed EATDE with $T_w = 10T_s$ and the conventional EATDE with gradient search were compared. As shown in the simulation results in Fig. 3, the proposed method converged to true delay $d_0 = 0$ for all the initial point $d[0] < T_w = 1\ ms = 10$ samples, while the conventional method converged only for $d[0] = 0,1,2,3$, or $d[0] \le 3T_s$. These findings led us to conclude that the wavelet-based EATDE provides wider convergence range.

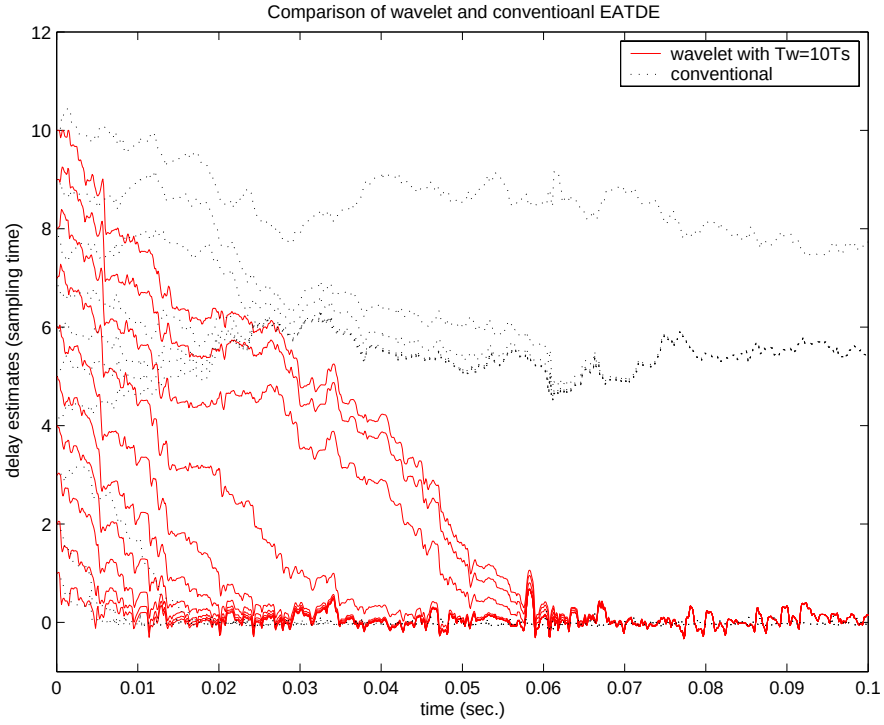

Figure 3. Comparison of the proposed method with conventional EATDE
$f_s = 10\ kHz$, $f_{cn} = 0.5$, $T_w = 1\ ms$, $\mu = 0.05$, SNR = 10 dB, $d[0] = 1 \ldots 10$

3. ACTIVE MICROPHONE ARRAY CONTROL

A conventional time delay based localization algorithm computes the source position from a relation of delay and the relative position of source and sensors. The relation is nonlinear and implicit and the solution is not unique in general. For example, in two dimensional case with two sensors as shown in figure 4, the relation is

$$\tau = \frac{1}{c}\sqrt{2(l^2 + r^2) - 2\sqrt{(l^2 + r^2)^2 - 4l^2 r^2 \sin^2\theta}}\,, \tag{9}$$

where τ, $2l$, r are delay, distance between the sensors, and distance of source from the center of two sensors.

If $r >> 2l$, the equation can be approximated to

$$\tau \cong \frac{2l}{c}\sin\theta \tag{10}$$

for $-\pi < \theta < \pi$. A sensitivity of the source with respect to r goes to zero and the delay becomes a function of source angle θ only. In general 3D case, a nonlinear optimization technique for minimizing the estimation errors is used and this is a heavy computational burden for a real-time source tracking system.

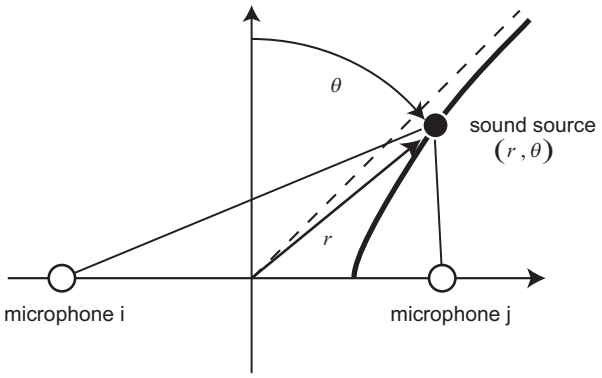

Figure 4.

A hyperbola determined by two microphones and the TODA measured in a two-dimensional plane

In this section, an active positioning method of microphone arrays is described to reduce the computational burden for 3D real-time source tracking applications. Active positioning means to change the physical positions of sensors using rotation mechanism, etc. It also improves the estimation performance with a smaller number of sensors.

The active positioning of microphones makes it possible to integrate TDE and GPE more systemically and consequently makes the localization process more efficient. In this research, an active positioning of microphones is implemented using a mechanically steerable structure. The microphone sensors can be rotated by motorized rotational base.

The basic concept is simple. Assume that two microphones can rotate about the center axis of two microphones. A source direction angle θ defined as the angle between the source direction and the line perpendicular to the line segment between two sensors at midpoint is a control variable. The control objective is to drive the microphone angle into the direction of the sound source.

We use the estimated delay $\hat{d}[n]$ to regulate the source direction angle θ:

$$\theta[n+1] = \theta[n] - \mu'\hat{d}[n], \tag{11}$$

where μ' is an adaptation step size for θ.

If the delay estimate $\hat{d}$ is the same with the actual delay d from the eq. (10),

$$\theta[n+1] \cong \theta[n] - \mu'\frac{2l}{c}\sin\theta[n]. \tag{12}$$

This algorithm converges to the correct source direction or the system is stable with the equilibrium point $\theta = 0$ and $\dot{\theta} = 0$.

The above discrete-time difference equation is approximated to the differential equation

$$\dot{\theta} \cong -k\sin\theta, \tag{13}$$

where k is a positive constant.

The convergence of this algorithm can be proved using Lyapunov stability theorem [18]. Considering the Lyapunov function as the square of the angle

$$V = \theta^2, \tag{14}$$

its derivative

$$\dot{V} = 2\theta\dot{\theta} = -2k\theta\sin\theta \tag{15}$$

is negative except the singular point $\theta = \pi$. Therefore, this first-order nonlinear system is globally asymptotically stable.

This active-position method has the advantages that less number of microphones are necessary because it has virtually infinite number of sensors. Note that three sensors are required in the above 2D case for the conventional localization method. Additionally, the maximum estimation sensitivity is guaranteed with regardless to the source position. From eq. (2), delay sensitivity is largest when the source is just in front of the microphone ($\theta = 0$).

4. EXPERIMENTS

This section explains the experiment system and the experiment result. The experiment system consists of three components: (1) microphone array which is steered by stepping motors, (2) microphone sensor amplifiers and motor drivers, (3) digital signal processing board for time delay estimation and source position calculation. Fig. 5 shows the block diagram of the experiment system. The microphone array has four microphones sensors each of which is placed at each vertex of the square of 20×20 cm dimension. The array is actuated by two stepping motors.

The experiment was performed in a conventional office room with small reverberation. The source sound is a baseband signal with cutoff frequency $f_c = 5$ kHz. Fig. 6 shows the experiment result plots. The loudspeaker is place at the 75° angle from the line normal at the center of two microphones. The dotted line denotes the angle of microphone array and moves to the source angle with a time constant $\tau = 0.95$ *s*. The solid line is the estimated time delay.

It first becomes to the delay corresponding to the angle 75° once with a time constant $\tau = 0.02\ s$ and converges to zero as the microphone array rotates to the source direction. The narrow solid line denotes the delay estimation in the case of no rotation and shown as the reference. As a demonstration, 3-dimensional space moving sound source directing test was carried out. Fig. 7 shows the results of this test. A portable radio speaker was used as a sound source. A Korean traditional mask (Hahoe-tal) was attached to the microphone array for emphasizing the direction of microphone array. This mask may cause a sound diffraction effect near the microphone sensors but this is a practical case for the applications such as robot, CCTV cameras. As shown in the pictures, the source localization using EATDE and active positioning works well in 3-dimensional space test. The direction of the array converges to the source direction within about 2 *s*.

5. CONCLUSIONS

This paper is a study to provide new explicit adaptive time delay estimation (EATDE) algorithm and an active positioning method for sound source localization. EATDE method is suitable for fast tracking of a time-varying delay and hence for the mobile platform. EATDE using Haar wavelet transform is proposed to avoid this convergence failure. The theoretical analysis and numerical simulations show that the proposed algorithm converges globally to an unbiased delay estimate. The on-line wavelet scale adaptation is also proposed to combine both the fast convergence and small estimation error at the same time and to avoid the bias error due to secondary sound sources.

This algorithm is developed for 1-dimensional audio signal, but it may also be applied to 2-dimensional image processing. For instance, two images captured from a moving camera are the translated versions of each other and the cross-correlation of two images is a 2-dimensional peak function. The peak finding or optimization technique proposed in this paper may be extended for such cases.

By active positioning of microphone array, calculation of position become easy and the precision of position estimation improved because the sensitivity of time delay with respect to the relative position is maximized. The convergence of this method is proven theoretically and the computer simulations are provided also.

Finally, the experiment system with a steerable microphone array which consists of 4 microphones and 2 stepping motors are developed and the performance are tested in real environment.

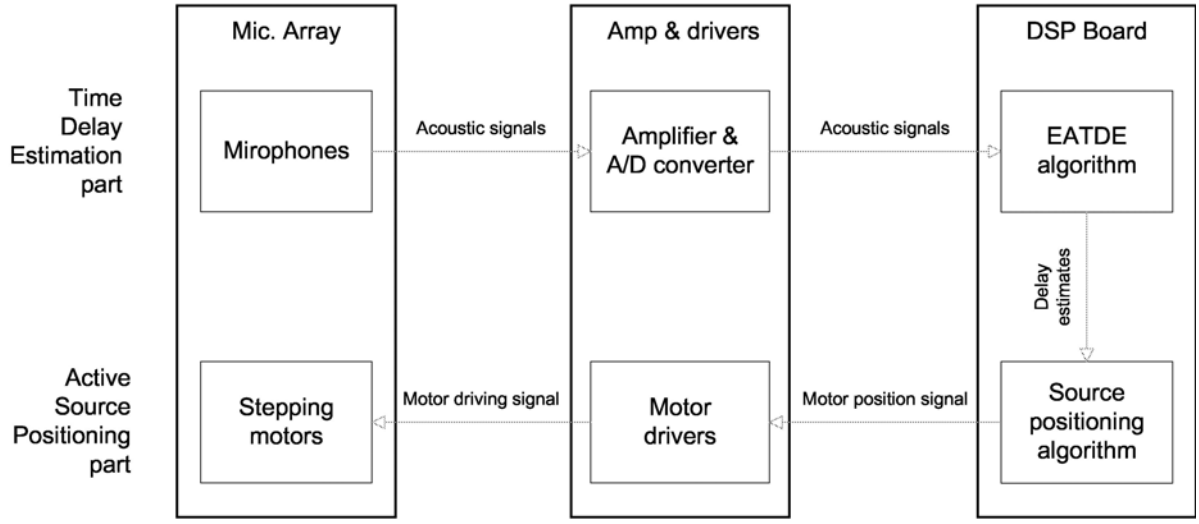

Figure 5. The overall structure of experiment system

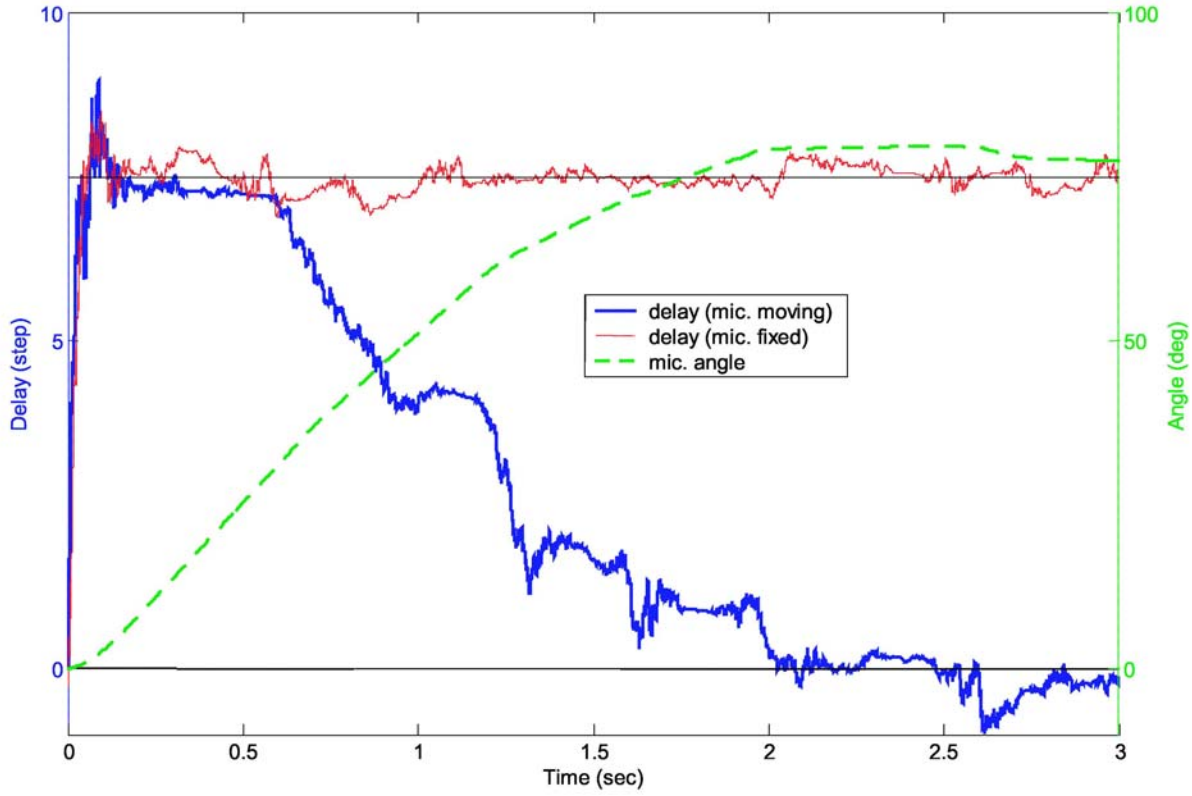

Figure 6. Time delay estimation and active source positioning experiment result

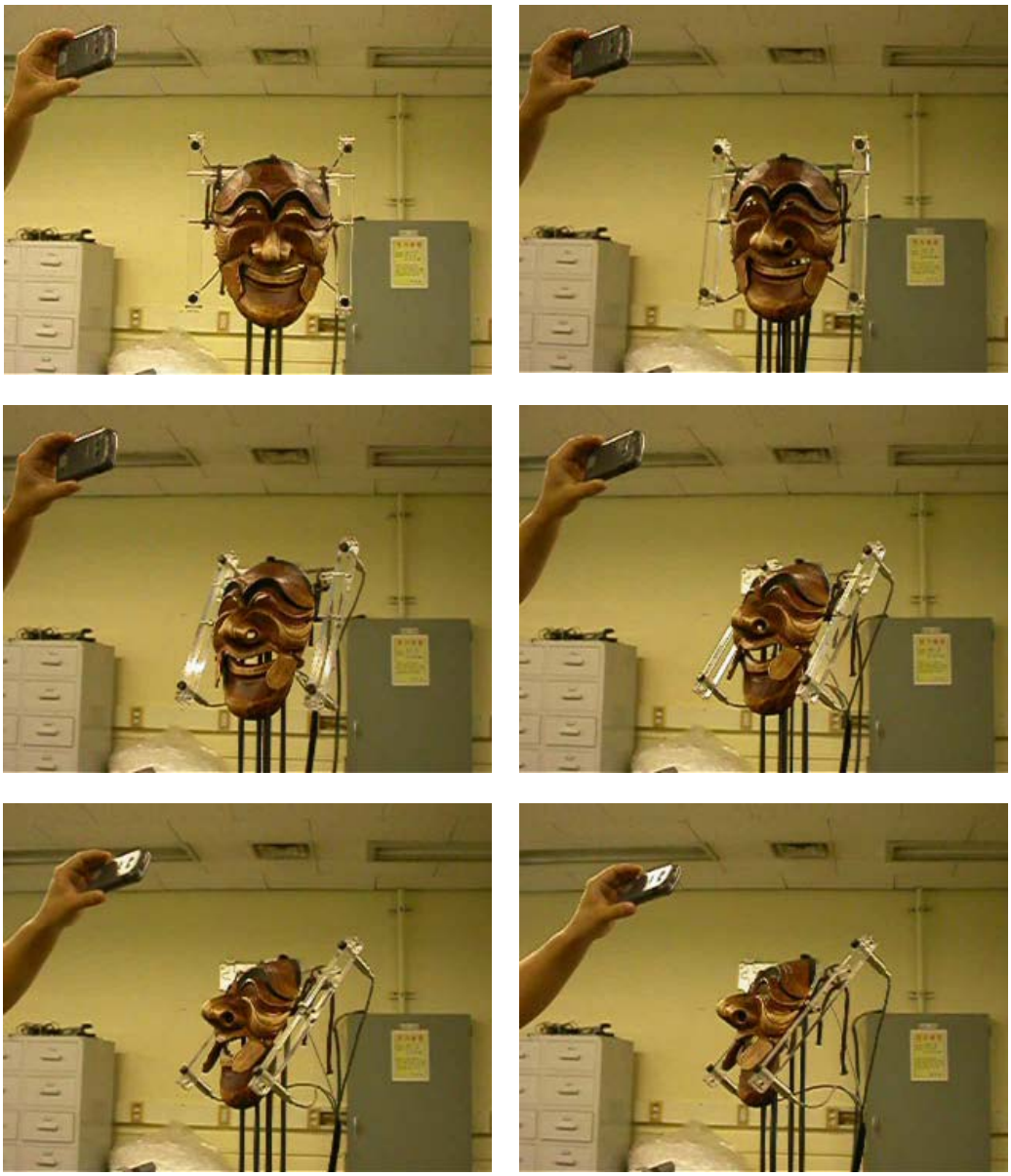

Figure 7. Example of sound source tracking

REFERENCES

[1] Michael S. Brandstein, Harvey F. Silverman. A practical methodology for speech source localization with microphone arrays. *Computer Speech and Language*, May 1997.
[2] R. D. Short. Sting ray – a sound-seeking missile. *IEE Review*, 35(11), 419–423, December 1989.
[3] J. Borenstein, Y. Koren. Obstacle avoidance with ultrasonic sensors. *IEEE Journal of Robotics and Automation*, 4(2), 213–218, 1988.
[4] H. G. Okuno, K. Nakadai, K. I. Hidai, H. Mizoguchi, H. Kitano. Human-robot interaction through real-time auditory and visual multiple-talker tracking. *Proceedings of 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol.3, 1402–1409, 2001.
[5] Yiteng Huang, J. Benesty, G. W. Elko. Passive acoustic source localization for video camera steering. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages II909–II912, 2000.
[6] H. Meyr. Delay-lock tracking of stochastic signals. *IEEE Transactions on Communications*, 24(3), 331–339, March 1976.
[7] D. Etter, S. Stearns. Adaptive estimation of time delays in sampled data systems. *IEEE Transactions on Acoustic Speech and Signal Processing*, 29(3), 582–587, June 1981.
[8] H. Messer, Y. Bar-Ness. Closed-loop least mean square time-delay estimator. *IEEE Transactions on Acoustic Speech and Signal Processing*, 35(4), 413–424, April 1987.
[9] H. Messer. A unified approach to closed-loop time delay estimation systems. *IEEE Transactions on Acoustic Speech and Signal Processing*, 36(6), 854–861, June 1988.
[10] H. C. So, P. C. Ching, Y. T. Chan. A new algorithm for explicit adaptation of time delay. *IEEE Transactions on Signal Processing*, 42(7), 1816–1820, July 1994.
[11] M. Vetterli, J. Kovacević. *Wavelets and subband coding*. Prentice Hall, 1995.
[12] C. Knapp, G. Carter. The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustic Speech and Signal Processing*, 24(4), 320–327, August 1976.
[13] F. A. Reed, P. L. Feintuch, N. J. Bershad. Time delay estimation using the LMS adaptive filter – static behavior. *IEEE Transactions on Acoustic Speech and Signal Processing*, 29(3), 561–571, June 1981.
[14] P. L. Feintuch, N. J. Bershad, F. A. Reed. Time delay estimation using the LMS adaptive filter – dynamic behavior. *IEEE Transactions on Acoustic Speech and Signal Processing*, 29(3), 571–576, June 1981.
[15] D. H. Youn, Nasir Ahmed, G. Cliford Carter. On using the LMS algorithm for time delay estimation. *IEEE Transactions on Acoustic Speech and Signal Processing*, 30(5), 1982.
[16] Richard C. Cabot. A note on the application of the Hilbert transform to time delay estimation. *IEEE Transactions on Acoustic Speech and Signal Processing*, 29(3), 1981.
[17] Doh-Hyoung Kim, *Sound Source Direction Estimation for Mobile Systems*. Ph.D. Thesis, Korea Advanced Institute of Science and Technology, 2005.
[18] Hassan K Khalil, *Nonlinear Systems*, 3rd ed., Prentice Hall, 2001.

TABLE OF SYMBOLS

t *or* (t)	continuous time index
n or $[n]$	time step or discrete time index
$s(t)$	continuous source signal
$w_1(t)$, $w_2(t)$	corrupting noise signals
τ or d_0	real time difference between the received signals
$d[n]$	delay estimation at step n
$g(d[n])$	delay-error function for $d[n]$
μ , μ'	positive step-size for adaptation
$w(t)$	Haar wavelet function
$R(d)$	wavelet transform of a signal with the delay wavelet function $w(t–d)$
r_{xy}	cross-correlation between signal x and y
x_f	signal filtered by the reversed wavelet $w(t–d)$
T_s	sampling time
T_w	support of Haar wavelet
$L = T_w/T_s$	normalized Haar wavelet support
f_c	cutoff frequency of baseband signal
$f_{cn} = f_c/f_s$	normalized cutoff frequency
l	distance of sensor from the center of two sensors
r	distance of source from the center of two sensors
c	speed of sound
θ	source direction angle
V	Lyapunov function