Proceedings of 20th International Congress on Acoustics, ICA 2010

23–27 August 2010, Sydney, Australia

A Sample-wise Acoustic Positioning Method Using Natural Gradient Adaptation

Toshiharu Horiuchi and Tsuneo Kato

User Interface Laboratory, KDDI R&D Laboratories Inc.
2-1-15 Ohara, Fujimino, Saitama 356-8502, Japan
to-horiuchi@kddilabs.jp

PACS: 43.60.Fg, 43.60.Jn, 43.60.Mn

ABSTRACT

This paper presents a sample-wise acoustic positioning method using natural gradient adaptation to track fast source location change. We are studying a sound imaging system with binaural reproduction setup for virtual and augmented reality applications such as a navigation system for pedestrians on mobile devices. The system uses stereo earphones, binaural microphones, and a hand-held mobile phone that emits a measurement signal for head tracking and source positioning. In this study, we developed a sample-wise acoustic positioning method using multiple receivers based on natural gradient adaptation. This method directly estimates the three-dimensional source position on the spherical coordinate system defined by the receivers' positions by minimizing the cost function. We derived this method under the constraint that the relative positions of the receivers are spatially-fixed on the spherical coordinate system. The cost function is defined as a residual sum of squares between the actual and estimated source signals. The estimated source signals are namely those for which the time delay and the amplitude are compensated depending on the estimated source position. The proposed method executed sample-wise processing for 48 kHz sampling in real-time. It reduces 90% of the azimuth error compared to the conventional correlation-based method at a movement speed of over 30 degrees per second, which corresponds to natural head turning. The azimuth error was within 1 degree, which means this method produces sufficient accuracy for the sound imaging system.

INTRODUCTION

With the recent rapid growth of signal processing technologies, three-dimensional sound technology has been highlighted in order to realize high performance audio-visual systems. The binaural reproduction technique is indispensable for truly immersive three-dimensional sound systems. This technique creates a sound image anywhere around the listener by using stereo headphones or earphones. The applications of this technique are diverse and in great demand such as virtual and augmented reality, tele- and video-conferencing, and other entertainments. We are studying a sound imaging system with binaural reproduction setup for virtual and augmented reality applications such as a navigation system for pedestrians on mobile devices.

An important property required in typical applications of the binaural reproduction technique is to position one's head precisely in order to produce a spatially-fixed sound image [1]. For that reason, many acoustic head tracking and positioning systems have been developed. Most of these systems need the listener to wear equipment that consists of acoustic transmitters or receivers, and base stations are also acoustic receivers or transmitters to be placed at fixed positions. The systems are technically based on estimation of the time delay of arrival and/or the time difference of arrival. Generally, the three-dimensional position is determined as the intersection of multiple spherical or hyperbolic planes given by estimates of the time delay of arrival or the time difference of arrival [2, 3, 4, 5, 6].

Conventional estimators assume that the movements of sources and/or receivers are slow and approximately constant within a short-time. This assumption works effectively in many practical situations [4, 7]. The generalized cross-correlation (GCC) method proposed by Knapp and Carter [7] is the most popular technique. The GCC method finds the maximum of the normalized cross-correlation between two signals within the short-time. However, this assumption fails when the receivers and/or the source move quickly [8] as in our application.

On the other hand, adaptive algorithms have also been applied to estimation of the time delay in many previous works [8, 9, 10, 11, 12, 13, 14, 15, 16, 17]. Chan *et al*. introduced a parameter estimation approach to time delay estimation by modeling the delay as a finite impulse response filter whose coefficients are samples of a sinc function [9]. Using this condition in adaptation, So *et al*. proposed an adaptive algorithm for explicit time delay estimation, where the filter coefficients are given by a function of the estimated delay only [14]. For estimating the time-varying delay, adaptive algorithms are regarded as the most effective techniques as seen in previous studies. Thus, we previously reported an adaptive algorithm for explicit two-dimensional direction estimation using the steepest descent method [18, 19]. However, the stability and the convergence time of this method depend on the step-size parameter. It was difficult to choose the step-size parameter in such a way as to provide fast convergence.

In this paper, we develop a sample-wise acoustic positioning method based on an adaptive filtering technique to provide fast convergence using natural gradient adaptation [20]. This method directly estimates the three-dimensional source position to the positions of the receivers on the spherical coordinate system by minimizing the cost function of the natural gradient algorithm. First, we describe the proposed sample-wise

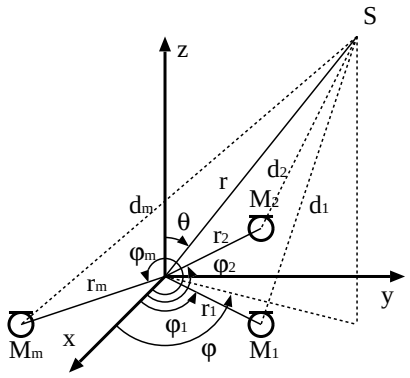

Figure 1: Geometric relationship between the source S existing at the position $\boldsymbol{p} = (r, \theta, \varphi)^T$ and the receiver M_i $(i = 1, 2, \cdots, m)$ located at the position $(r_i, \pi/2, \varphi_i)^T$ on the spherical coordinate system.

acoustic positioning method. Next, we perform experiments on the accuracies of the proposed method and the conventional method. In the experiment, we use binaural microphones for receivers and a mobile phone that emits the sound for the source. Finally, we summarize the paper.

SAMPLE-WISE POSITIONING METHOD

In this section, we develop the proposed sample-wise acoustic positioning method. First, we define the geometric relationship between the source and receivers, and these signals. Next, we describe the details of the proposed method.

Definition

As shown in Fig. 1, let us assume that the source S and the signal $s(k)$ exist in free space. We can obtain the signal $x_i(k)$ $(i = 1, 2, \cdots, m)$ observed by the receiver M_i as

$$x_i(k) = \frac{1}{d_i} s\left(k - \frac{d_i}{c}\right) + n_i(k), \tag{1}$$

where k is the discrete time index, d_i $(i = 1, 2, \cdots, m)$ represents the distance between the source S and the receiver M_i, c is the sound velocity, and $n_i(k)$ $(i = 1, 2, \cdots, m)$ is the interference signal observed by the receiver M_i. Here, the time delay of arrival τ_i and the distance d_i can be expressed as a function of the position $\boldsymbol{p} = (r, \theta, \varphi)^T$ of the source S on the spherical coordinate system

$$\tau_i(\boldsymbol{p}) = \frac{d_i(\boldsymbol{p})}{c} = \frac{\sqrt{r^2 + r_i^2 - 2 r r_i \sin\theta \cos(\varphi - \varphi_i)}}{c}, \tag{2}$$

where $(r_i, \pi/2, \varphi_i)^T$ is the position of the receiver M_i.

Our task is to find the position $\boldsymbol{p}$ of the source S from the source signal $s(k)$ and the observed signal $x(k)$.

Algorithm

By using an adaptive filtering technique, the estimated observed signal $\hat{x}_i(k)$ is modeled as a sinc function that is expressed as the estimated position $\boldsymbol{p}$ of the source S:

$$\hat{x}_i(k) = \sum_{n=-\infty}^{\infty} \mathrm{sinc}\left(n - \frac{d_i(\boldsymbol{p})}{cT}\right) \frac{1}{d_i(\boldsymbol{p})} s(k - n), \tag{3}$$

where all adaptive filter coefficients are expressed in terms of the estimated position $\boldsymbol{p}$ only, $\mathrm{sinc}(x)$ is defined as $\mathrm{sinc}(x) = \sin(\pi x)/\pi x$, and T is the sampling period. Here, we define a cost function $J(\boldsymbol{p})$ as

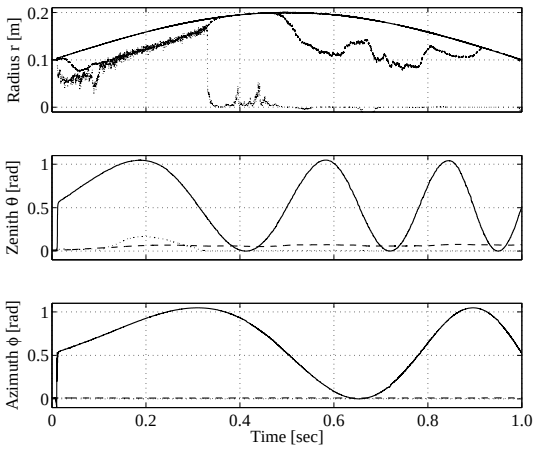

Figure 2: Estimated position trajectories for the proposed method based on the natural gradient adaptation (solid line: $\mu(k)$=1e-4) and the standard gradient adaptation (dotted line: $\mu(k)$=1e-4 and dashed line: $\mu(k)$=1e-5) for comparison. Top, middle, and bottom figures show the radius r, the zenith θ, and the azimuth φ of the source S.

$$J(\boldsymbol{p}) = \sum_i (x_i(k) - \hat{x}_i(k))^2. \tag{4}$$

This cost function $J(\boldsymbol{p})$ is minimized when the estimated position $\boldsymbol{p}$ matches the actual source position.

Using the cost function $J(\boldsymbol{p})$, the estimated position $\boldsymbol{p}$ updates for the natural gradient adaptation [20] are given by

$$\boldsymbol{p}(k+1) = \boldsymbol{p}(k) - \mu(k) \boldsymbol{G}^{-1}(\boldsymbol{p}(k)) \frac{\partial J(\boldsymbol{p}(k))}{\partial \boldsymbol{p}}, \tag{5}$$

where $\mu(k)$ is the step-size parameter, and $\boldsymbol{G}(\boldsymbol{p})$ is the Riemannian metric tensor for the position $\boldsymbol{p}$ on the spherical coordinate system given by

$$\boldsymbol{G}(\boldsymbol{p}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2\theta \end{bmatrix}. \tag{6}$$

Moreover, the partial differentiation $\partial \mathrm{sinc}(x)/\partial x$ in Eq.(5) is given by

$$\mathrm{sinc}'(x) = \frac{\cos(\pi x) - \mathrm{sinc}(x)}{x}. \tag{7}$$

This method is able to estimate the three-dimensional position of the source to the positions of the receivers on the spherical coordinate system by minimizing the cost function of the natural gradient algorithm.

EVALUATIONS

Computer simulation

First, we applied the proposed method to a three-channel planar array for 3-D position estimation. Three receivers are located on the vertex of an equilateral triangle as shown in Fig. 1. The receiver positions are $(\rho, \pi/2, 0)^T$, $(\rho, \pi/2, 2\pi/3)^T$, and $(\rho, \pi/2, 4\pi/3)^T$, where ρ is 0.08 m. For the source signal $s(k)$, we used a white Gaussian noise. The sampling condition is

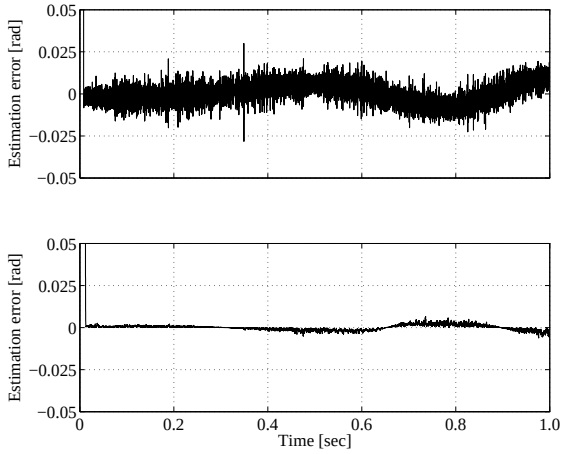

Figure 3: Estimation error for the azimuth φ of the source *S*. The upper figure shows the GCC method and lower figure shows the proposed method.

48 kHz/16 bit. The observed signals $x_i(k)$ $(i = 1,2,3)$ of receiver M_i are generated to give a time delay in the computer. In this simulation, the source moves as follows:

$$\boldsymbol{p}(t) = \begin{pmatrix} r(t) \\ \theta(t) \\ \varphi(t) \end{pmatrix} = \begin{pmatrix} 0.1(1+\sin(2\pi(0.5t))) \\ \pi(1+\sin(2\pi(t+2t^2)))/6 \\ \pi(1+\sin(2\pi(0.5t+t^2)))/6 \end{pmatrix}, \tag{8}$$

where t is the time index; i.e., $t = kT$. The other conditions are as follows: The initial position $\boldsymbol{p}(0)$ of the source S is $(0.1, 0.01, 0.01)^T$. The adaptive filter length is 128 samples.

Figure 2 shows the estimated position trajectories for the proposed method based on the natural gradient adaptation (solid line: step-size parameter $\mu(k)$=1e-4) and the standard gradient adaptation (dotted line: $\mu(k)$=1e-4 and dashed line: $\mu(k)$=1e-5). In the standard gradient adaptation, the Riemannian metric tensor is $\boldsymbol{G}(\boldsymbol{p}) = \boldsymbol{I}$. The stability and the convergence time depend on the step-size parameter. This figure indicates that the proposed method can estimate the 3-D position of the rapid moving source.

Next, we compared the accuracy of the proposed method with the conventional method. For comparison, we applied the proposed method to a two-channel array for 2-D position estimation. The receiver positions are $(\rho, \pi/2, 0)^T$ and $(\rho, \pi/2, \pi)^T$, where ρ is 0.08 m. The interval between neighboring microphones is 0.16 m. For the conventional method, we use the generalized cross-correlation (GCC) method described in the literature [7]. The GCC method finds the maximum of the normalized cross-correlation between the two observed signals. Additionally, to improve the precision of the time delay estimation and to make a fair comparison, we performed interpolation of the normalized cross-correlation before finding the maximum by using a windowed sinc function filter. The analysis window used in our experiments has a duration of 128 samples with an overlap of 127 samples. The tap length of the sinc function filter for the proposed method and the GCC method is also 128 samples. In this simulation, the source moves as follows:

$$\boldsymbol{p}(t) = \begin{pmatrix} r(t) \\ \theta(t) \\ \varphi(t) \end{pmatrix} = \begin{pmatrix} 0.1 \\ \pi/2 \\ \pi(3+\sin(2\pi(0.5t+t^2)))/6 \end{pmatrix}. \tag{9}$$

The other conditions were the same as those in the first simulation.

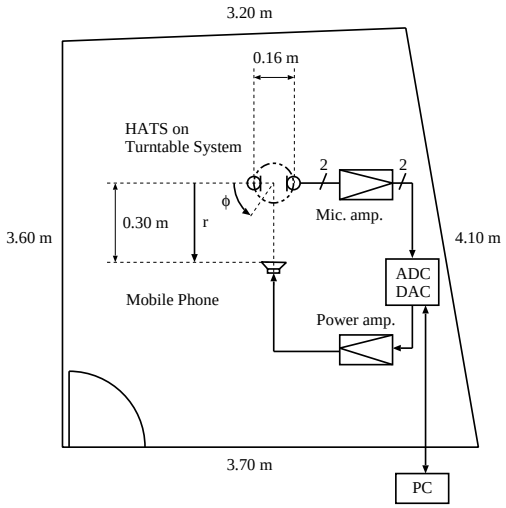

Figure 4: The experimental setup. The HATS stands on the turntable system and wears the binaural microphones for receivers, and the mobile phone emits the quasi-ultrasonic sound for the source.

Table 1: The equipment for the experiment.

Equipment	Manufacture	Type
HATS	B&K	4128D
Turntable system	B&K	9640
Binaural microphones	Adphox	BME-200
Microphone amplifier	audio-technica	AT-MA2
Power amplifier	audio-technica	AT-HA20
AD/DA converter	PreSonus	FIRESTDIO

In Fig. 3 the upper panel shows the estimation error for the azimuth φ of the GCC method, and the lower panel shows the estimation error of the proposed method. We define the estimation error as the difference between the actual source position and the estimated position. Here, we consider the difference between the GCC method and the proposed method. The GCC method is based on the assumption that the time delay does not change or remains approximately constant within the analysis window. Consequently, movement causes a large estimation error. The proposed method reduces 90% of the azimuth error compared to the GCC method at a movement speed of over 30 degrees per second.

Experiment

We also investigated the availability of the proposed method in our target application. We applied the proposed method to a two-channel array for 2-D head tracking with source position estimation. Figure 4 shows the experimental setup. In a soundproof room, with reverberation time of 0.1 s, the HATS stands on the turntable system and wears binaural microphones for receivers, and the mobile phone emits the sound for the source.

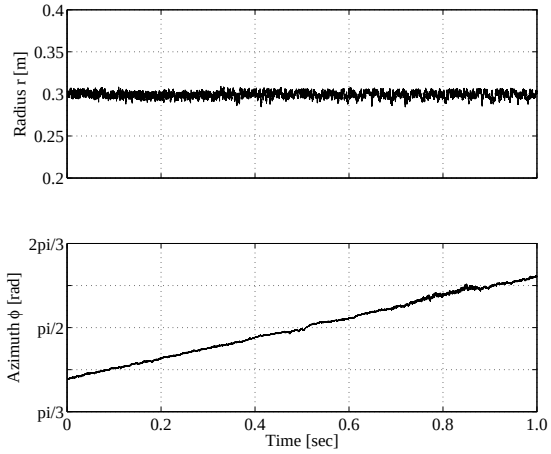

Figure 5: Estimated position trajectories. The HATS rotates $\pi/5$ rad/s. Upper and lower figures show the radius r and the azimuth φ.

The receiver positions are $(\rho, \pi/2, 0)^T$ and $(\rho, \pi/2, \pi)^T$, where ρ is 0.08 m. The interval between receivers is 0.16 m. For the source signal $s(k)$, we used a band-limited Gaussian noise (18–20 kHz). The HATS rotates $\pi/5$ rad/s about the azimuth φ. Table 1 shows the equipment for the experiment. The other conditions were the same as those in the simulation.

Figure 5 shows the estimated position trajectories for the proposed method in the experiment. This figure indicates that the proposed method can track the position of moving receivers at a movement speed of over 30 degrees per second, which corresponds to natural head turning. The azimuth error was within 1 degree, which means this method produces sufficient accuracy for the sound imaging system.

CONCLUSIONS

This paper presents a sample-wise acoustic positioning method based on an adaptive filtering technique to provide fast convergence using natural gradient adaptation. This method directly estimates the three-dimensional source position to the positions of the receivers on a spherical coordinate system by minimizing the cost function. The simulation result indicates that the proposed method can track the three-dimensional position of a rapid moving source. The estimation error of the proposed method is smaller than that of the conventional method. In the experiments, we use binaural microphones for receivers and a mobile phone that emits the sound for the source. The experimental result supports the effectiveness of the method for two-dimensional head tracking and source position estimation.

REFERENCES

[1] M. Karjalainen, M. Tikander, and A. Härmä, "Head-tracking and subject positioning using binaural headset microphones and common modulation anchor sources," Proc. IEEE ICASSP, vol. 4, pp. 101–104, 2004.

[2] A. Ward, A. Jones, and A. Hopper, "A new location technique for the active office," IEEE Pers. Commun., vol. 4, no. 5, pp. 42–47, 1997.

[3] N. B. Priyantha, A. K. Miu, H. Balakrishnan, and S. Teller, "The cricket compass for context-aware mobile applications," Proc. ACM MobiCom, pp. 1–14, 2001.

[4] M. Omologo and P. Svaizer, "Use of the Crosspower-Spectrum Phase in Acoustic Event Localization," IEEE Trans. Speech & Audio Process., vol. 5, no. 3, pp. 288–292, May 1997.

[5] Y. T. Chan and K. C. Ho, "A simple and efficient estimator for hyperbolic location," IEEE Trans. Signal Process., vol. 42, no. 8, pp. 1905–1915, Aug. 1994.

[6] M. S. Brandstein, J. E. Adcock, and H. F. Silverman, "A closed-form location estimator for use with room environment microphone arrays," IEEE Trans. Speech & Audio Process., vol. 5, no. 1, pp. 45–50, Jan. 1997.

[7] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," IEEE Trans. Acoust., Speech & Signal Process., vol. 24, no. 4, pp. 320–327, Aug. 1976.

[8] H. Kagiwada, H. Ohmori, and A. Sano, "A recursive algorithm for tracking DOA's of multiple moving targets by using linear approximations," IEICE Trans. Fundamentals, vol. E81–A, no. 4, pp. 639–648, Apr. 1998.

[9] Y. T. Chan, J. Riley, and J. B. Plant, "A parameter estimation approach to time delay estimation and signal detection," IEEE Trans. Acoust., Speech & Signal Process., vol. 28, no. 2, pp. 8–16, Feb. 1980.

[10] Y. T. Chan, J. Riley, and J. B. Plant, "Modeling of time delay and its application to estimation of non-stationary delays," IEEE Trans. Acoust., Speech & Signal Process., vol. 29, no. 3, pp. 577–581, Jun. 1981.

[11] P. L. Feintuch, N. J. Bershad, and F. A. Reed, "Time delay estimation using the LMS adaptive filter –Dynamic behavior," IEEE Trans. Acoust., Speech & Signal Process., vol. 29, no. 3, pp. 571–576, Jun. 1981.

[12] D. M. Etter and S. D. Stearns, "Adaptive estimation of time delays in sampled data systems," IEEE Trans. Acoust., Speech & Signal Process., vol. 29, no. 3, pp. 582–587, Jun. 1981.

[13] P. C. Ching and Y. T. Chan, "Adaptive time delay estimation with constraints," IEEE Trans. Acoust., Speech & Signal Process., vol. 36, no. 4, pp. 599-602, Apr. 1988.

[14] H. C. So, P. C. Ching, and Y. T. Chan, "A new algorithm for explicit adaptation of time delay," IEEE Trans. Signal Process., vol. 42, no. 7, pp. 1816-1820, July 1994.

[15] S. Affes, S. Gazor, and Y. Grenier, "Robust adaptive beamforming via LMS-like target tracking," Proc. IEEE ICASSP, vol. 4, pp. 269–272, 1994.

[16] M. Zhang and M. H. Er, "An alternative algorithm for estimating and tracking talker location by microphone arrays," J. Audio Eng. Soc., vol. 44, no. 9, pp. 729–736, Sep. 1996.

[17] Y. T. Chan and K. C. Ho, "TDOA-SDOA estimation with moving source and receivers," Proc. IEEE ICASSP, vol. 5, pp. 153–156, 2003.

[18] T. Horiuchi, M. Mizumachi, and S. Nakamura, "Iterative Compensation of Microphone Array and Sound Source Movements Based on Minimization of Arrival Time Differences," Proc. IEEE SAM, pp. 566-570, 2004.

[19] T. Horiuchi, M. Mizumachi, and S. Nakamura, "Iterative Estimation and Compensation of Signal Direction for Moving Sound Source by Mobile Microphone Array," IEICE Trans. Fundamentals, vol. E87-A, no. 11, pp. 2950-2956, 2004.

[20] S. Amari and S. C. Douglas, "Why natural gradient?," Proc. IEEE ICASSP, vol. 2, pp. 1213–1216, 1998.