

로봇 플랫폼에서 마이크로폰 위치를 고려한 음원의 방향 검지 방법

Considering Microphone Positions in Sound Source Localization Methods: in Robot Application

권병호† · 김경호* · 박영진**

Byoungho Kwon, Gyeongho Kim and Youngjin Park

Key Words : Sound Source Localization(음원 방향 검지), TDOA(Time Difference of Arrival, 도달 시간지연), Opened Microphone Position(개방된 위치의 마이크로폰), Closed Microphone Position(폐쇄된 위치의 마이크로폰).

ABSTRACT

Many different methods for sound source localization have been developed. Most of them mainly depend on time delay of arrival (TDOA) or on empirical or analytic head related transfer functions (HRTFs). In real implementation, since the direct path between a source and a sensor is interrupted by obstacles as like a head or body of robot, it has to be considered the number of sensors as well as their positions. Therefore, in this paper, we present the methods, which are included sensor position problem, to localize the sound source with 4 microphones to cover the 3D space. Those are modified two-step TDOA methods. Our conclusion is that the different method has to be applied in case to be different microphone position on real robot platform.

1. 서 론

3 차원 공간상에 존재하는 음원 방향 검지 방법은 여러 가지 음향학적 특성들을 이용하여 연구되고 있다. 이런 연구에는 다수의 마이크로폰을 이용하여 각 마이크로폰 사이의 도달 시간지연(Time Difference of Arrival, TDOA)을 이용하는 방법이 있다(one-step TDOA and two-step TDOA)[1]. 또 마이크로폰 사이의 도달 시간지연뿐만 아니라 특정 형상에 대한 신호의 세기 정보까지 이용하는 방법도 있다[2]. 뿐만 아니라, 특정 형상에 대한 머리전달함수(Head Related Transfer Function, HRTF) 데이터베이스를 이용하여 3 차원 공간상의 음원의 방향을 추정하는 방법도 널리 연구되고 있다[3], [4].

또한 이런 방법들이 그 성능을 확인하기 위해서 실제 로봇에 적용된 사례들도 많이 있다. 로봇에 청각 시스템을 가장 먼저 적용한 휴머노이드 로봇 SIG[5]가 대표적이다. SIG 에는 머리전달함수 데이터베이스 방법이 적용되었으며, 사람과 같이 두 개의 마이크로폰만을 사용하여 3 차원 공간상의 음원방향검지가 가능하다는 점에서 상당히 성공적인 사례로 평가 받고 있다. 또 LABOIOUS 연구실

에서는 8 개의 마이크로폰을 사용하여 움직이는 로봇에서 음원의 방향을 추정할 수 있는 시스템을 개발하여 그 성능을 실제 플랫폼에서 검증하였다 [6]. 하지만 위에서 소개한 사례들은 머리전달함수를 측정하고 데이터베이스화하고 있어야 하고, 너무 많은 마이크로폰이 사용된다는 점에서 다양한 로봇에 적용하는데 제약이 될 수 있다. 이런 점들을 개선하기 위해서 3 차원 공간상에 존재하는 음원의 방향을 추정하기 위한 최소한의 마이크로폰으로 다양한 로봇들에 효율적으로 적용될 수 있는 TDOA 방법을 적용한 연구가 필요하다. 뿐만 아니라, 실제 로봇 플랫폼에 적용될 때 설치되는 마이크로폰의 위치가 음원 방향 검지 방법의 성능에 영향을 미칠 수 있기 때문에 마이크로폰의 위치도 고려되어 연구될 필요가 있다.

따라서 본 연구에서는 음원에서 마이크로폰까지의 직접 도달 경로의 유무에 따라 두 가지 경우로 구분하고, 각 경우에 대해 적용될 방법에 대해 제시하고자 한다.

2. 3 차원 공간에서 음원 방향 검지

3차원 공간상의 음원의 방향을 추정하기 위해서는 각 마이크로폰 사이의 TDOA를 이용해 수평각과 고도각으로 구분하여 구한다. 먼저, 각 마이크로폰 사이의 TDOA는 신호들 사이의 상호상관계수(cross-correlation)를 이용하여 구한다. 다음 식으로 표현되는 상호상관계수의 값이 최대가 되는 지연시간이 두 마이크로폰 사이의

† 권병호; 한국과학기술원
E-mail : bhkown@kaist.ac.kr
Tel : (042) 869-3056, Fax : (042) 869-8220

* 한국원자력연구소

** 한국과학기술원

TDOA 가 된다.

$$R_{ij}(\tau) = \sum_{n=0}^{N-1} x_i[n]x_j[n-\tau] \quad (1)$$

x_i : signal received by microphone i
 τ : correlation lag in samples

일반적으로 식(1)은 계산 량이 N^2 에 비례하므로 계산 량이 적은 상호 스펙트럼 밀도함수(cross spectral density function)의 푸리에 역변환 (inverse Fourier transform)을 이용하여 근사화한다. 상관계수의 근사화는 다음과 같이 표현된다.

$$R_{ij}(\tau) \approx \sum_{n=0}^{N-1} X_i(k)X_j(k)^* e^{i2\pi k\tau/N} \quad (2)$$

$X_i(k)$: discrete Fourier transform of $x_i[n]$
 $X_i(k)X_j(k)^*$: cross-spectrum of $x_i[n]$ and $x_j[n]$

위의 방법은 음원 신호의 통계학적 특성에 상당히 의존적인 경향이 있다. 이 연구에서 목표로 하고 있는 음성신호는 대부분 5kHz 미만의 신호이므로 상관계수의 최대값이 넓게 분포할 수 있다. 이런 문제점을 보완하기 위해 서 가중치함수(PHAT)를 적용할 수 있다[7]. PHAT은 일반적으로 잔향과 반사 영향이 포함되어 있는 신호에서 상호상관계수를 구할 때 널리 이용되는 가중치 함수이다.

$$R_{ij}(\tau) \approx \sum_{n=0}^{N-1} \psi_p(k)X_i(k)X_j(k)^* e^{i2\pi k\tau/N} \quad (3)$$

where, $\psi_p(k) = 1/|X_i(k)X_j(k)|$

위에 제시된 방법으로 구해진 두 마이크로폰 사이의 상관계수 값으로부터 음원의 방향을 추정한다. 수평면상에서 TDOA 방법을 이용해 음원의 위치를 앞 뒤 혼동현상(front back confusion) 없이 추정하기 위해서는 최소한 세 개의 마이크로폰이 필요하며, 고도각까지 추정하기 위해서는 세 개의 마이크로폰이 이루는 평면과 다른 평면에 위치하는 또 다른 하나의 마이크로폰이 필요하게 된다. 따라서 3차원 상에서 음원의 위치를 추정하기 위해서는 최소한 4개의 마이크로폰이 필요하다. 먼저 수평각은 한 평면 상에 존재하는 세 개의 마이크로폰을 이용하여 구한다. 이 때 각 마이크로폰 쌍의 도달 시간지연 조합을 이용하여 구하게 된다. 그런 다음 고도각을 추정하는데 계산의 효율성을 증대시키기 위해서 가상 마이크로폰 방법을 이용한다. 가상 마이크로폰 방법은 Fig. 1과 같이 앞서서 구한 수평각과 설치된 마이크로폰들의 형상을 이용하여 가상 마이크로폰의 위치를 결정하고, 그

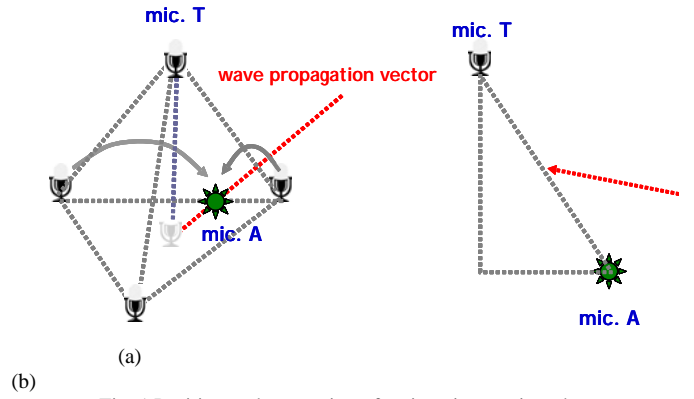


Fig. 1 Position and generation of an imaginary microphone

신호는 주위의 두 마이크로폰 신호를 보간 법을 이용해 만들어 낸다. 가상 마이크로폰을 이용하여 고도각을 구함에 있어 수평각의 측정에러가 포함될 수 있지만, 그 영향은 마이크로폰 사이의 거리가 상당히 크지 않는 이상 크지 않다. TDOA 방법을 이용하여 음원의 방향을 추정함에 있어서 가장 큰 가정은 자유음장조건(free field condition)이다. 이 가정을 이용해 실제 마이크로폰에서 측정된 신호를 음파 전달 벡터(source wave propagation vector)와 같은 방향에 위치하는 가상 마이크로폰의 신호를 추정할 수 있다. 여기에서 오차를 줄이기 위해 인접한 실제 마이크로폰 두 개로부터 각각 추정하여 중간 값을 사용한다. (Fig. 1.(a)) 시간지연 값을 구하기 위해 상관계수를 계산할 때는 신호 전체에 대해서 하지 않고 일정 길이로 나뉘서 중복적으로 계산되게 함으로써 좀 더 강인하게 시간 지연을 구할 수 있게 하였다. 이 때 계산 상의 효율성을 고려해 계산되는 한 frame의 샘플수는 $2^N(2048)$ 이 되게 하였다. 그런 다음 Fig. 1. (a) 에 보이는 같은 평면에 존재하지 않는 mic.T를 이용하여 Fig. 1. (b)와 같이 단순한 두 마이크로폰 사이의 시간 지연으로부터 고도각을 추정할 수 있다.

앞에서 설명한 방법을 이용하여 3차원 공간상의 음원의 위치를 추정하게 된다. 하지만 여기에서 고려되어야 하는 것은 실제 로봇에 적용되는 경우에 마이크로폰이 설치되는 위치이다. 마이크로폰을 특정 플랫폼에 설치하면서 자유음장이라는 가정이 깨지기 때문에 음원의 방향을 추정하는데 상당한 오차가 발생하게 된다. 수평각을 추정할 때 발생한 오차는 고도각을 추정할 때에도 영향을 줄 수 있으므로 마이크로폰의 설치 위치에 따른 각기 다른 방법을 적용할 필요가 있다.

3. 마이크로폰 위치에 따른 음원 방향 검지 방법

3.1 개방된 위치에 마이크로폰이 존재하는 경우

음원에서 마이크로폰까지 직접 전달 경로 (direct path)가 존재하는 경우에는 ‘개방된 위치에 마이크로폰이 존재한다’고 하고 다음과 같은 방법으로 음원의 위치를 추정한다. 먼저 한 평면상에 존재하는 세 개의 마이크로폰으로부터 수평각을 추정한다. 그런 다음 Fig. 2에서 보는 것과 같이 각 마이크로폰 사이의 상관계수 값을 마이크로폰 배열의 형상을 고려하여 각도로 환산하고, 각 상관계수 값을 합하면, 실제 음원의 위치 방향으로 peak이 생기게 된다.

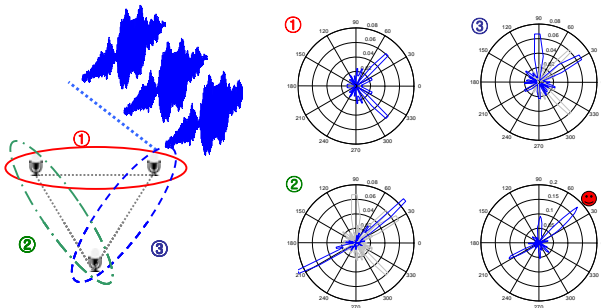


Fig. 2 Method to estimate azimuth angle using cross correlations (The case of opened microphone position)

오른쪽 그림에서 ①, ②, ③의 계산된 상관계수 값을 보면 두 개의 peak을 가지는 것을 볼 수 있다. 이것은 두 개의 마이크로폰을 이용하여 음원의 위치를 추정할 때 생기는 앞 뒤 혼돈현상 때문이다. 이 방법은 하나의 마이크로폰 쌍의 상관계수 값에 어느 정도의 오차가 발생해도 나머지 마이크로폰 쌍들이 이를 보상해주기 때문에 실제 잡음환경에서 강인한 특성을 가진다.

3.2 폐쇄된 위치에 마이크로폰이 존재하는 경우

다음은 마이크로폰이 로봇의 머리카나 몸통에 의해 음원의 직접 전달 경로가 존재하지 않는 경우(Fig. 3. (a))에는 ‘폐쇄된 위치에 마이크로폰이 존재한다’고 하고 다음과 같은 방법으로 음원의 위치를 추정한다. 각 마이크로폰 쌍의 시간 지연 값을 이용하여 음원이 위치하는 영역을 Fig. 3.의 (b)와 같이 구분하고[8], 그 영역에 포함되는 마이크로폰 쌍만을 이용하여 음원의 위치를 추정하게 된다.

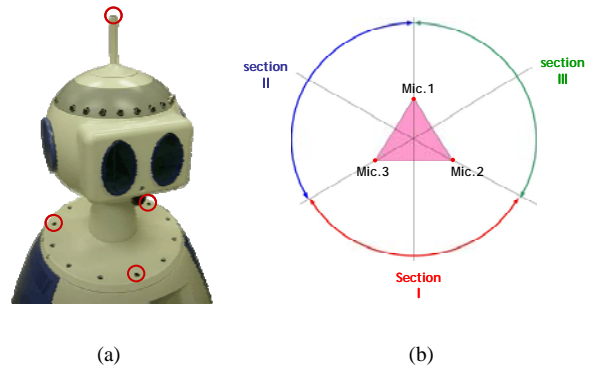


Fig. 3 (a) A case of closed microphone position; adapted form [8] (b) Region separation: * Information Robot Platform Version. 1, IRP-2004

위의 그림 (b)와 같이 마이크로폰을 정삼각형 모양으로 설치하면 하나의 마이크로폰 쌍이 커버해야 하는 부분은 120°가 된다. 일반적으로 TDOA 방법으로 음원의 방향을 추정할 경우 음원이 한 쪽 마이크로폰으로 치우치면 one sample delay가 유발하는 오차가 커진다. 이를 고려하여 마이크로폰을 설치하는 것도 중요하다. 앞서서도 언급했듯이 음원이 존재하는 영역이 정해지면 그 영역에 포함되어 있는 한 쌍의 마이크로폰만을 이용하여 음원의 위치를 추정해야 하므로 신호를 여러 frame으로 나눠서 TDOA를 구한다.

4. 무향실 실험을 통한 검증

실험은 위에 제시한 두 경우의 성능 비교를 위해 Fig. 3. (a)의 로봇 platform 을 이용하여 무향실에서 1.5m 거리만큼 떨어져 있는 음원에 대해 수평각 0°~ 180°, 고도각 0°~ 30°사이에서 10°씩 변화시키면서 측정된 데이터를 이용하였다(Fig. 4.). 이 데이터는 로봇의 화자 위치 추정을 위한 HRTF 측정에 사용된 것으로 한국과학기술원 구조동역학 및 응용제어 연구실에 보관되어 있다 [9]. 로봇 플랫폼은 아래에 장착되어 있는 turn table 에 의해 정확하게 회전하며, 음원의 위치는 구조물을 이용



Fig. 4 Experimental setup in anechoic chamber; adapted form [8]

해 고도각에 변화를 주었다. 신호는 random white noise 를 사용하였으며 data acquisition board 의 sampling frequency 는 44.1kHz 이다. 계산 시간을 고려해서 6000samples 을 가지고 음원 인식을 하였으며, 20 개의 data set 을 가지고 분석하였다.

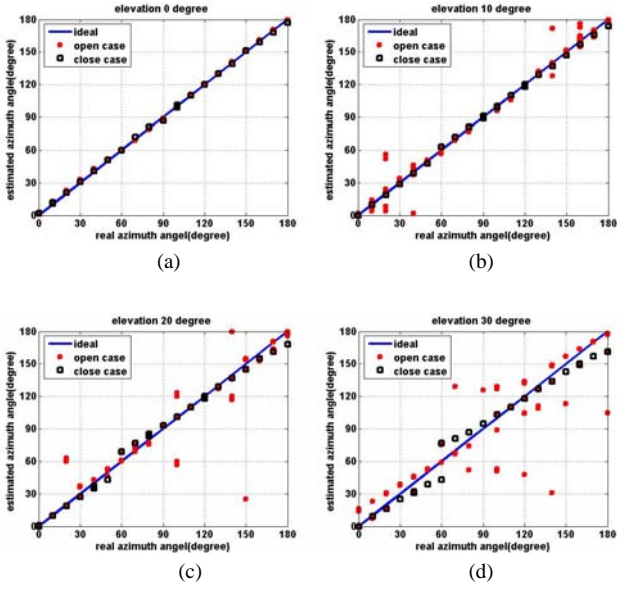


Fig. 5 Experimental results for azimuth angle at 1.5m: (a) elevation 0degree, (b) elevation 10degree, (c) elevation 20degree, (d) elevation 30degree

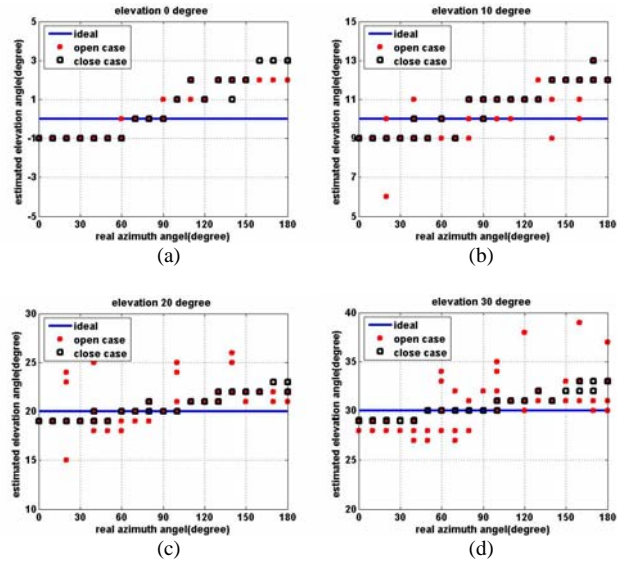


Fig. 6 Experimental results for elevation angle at 1.5m: (a) elevation 0degree, (b) elevation 10degree, (c) elevation 20degree, (d) elevation 30degree

Fig. 5 는 수평각을 변화시켜 가며 수평각을 추정 한 결과이고, Fig. 6 수평각을 변화시켜 가며 고도각을 추정한 결과이다. 또한 특정한 고도각에 대한 20 개의 결과를 한 그래프에 나타내고 있다. 고도각이 높아 질수록 음원은 로봇의 머리 부분에

의해 더 많이 방해 를 받게 되는데, 이런 영향이 결과에 그대로 반영되어 있다. 로봇 머리에 의해 방해 받는 부분이 많아 질수록 폐쇄된 위치에 마이크가 존재하는 경우의 방법이 더 정확하게 추정함을 알 수 있다. 일반적으로 음원의 방향을 추정함에 있어서 실험 공간에 존재하는 환경적인 요소 (반사와 잔향)를 반드시 고려해야 하지만, 실제 로봇에 적용함에 있어서 로봇 플랫폼이 가지는 제약 조건만을 고려하기 위해 무향실 데이터가 적용되었다. 이를 이 실험결과가 잘 보여주고 있다.

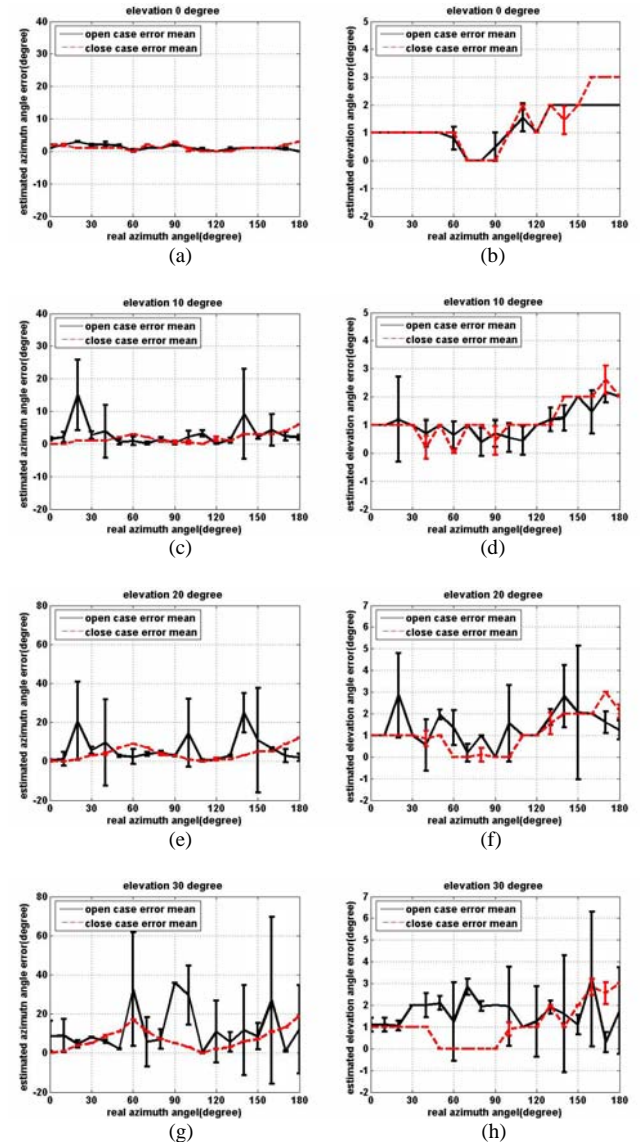


Fig. 7 Mean and Standard deviation of azimuth estimation error as each elevation angle.

실험결과를 Fig. 7 에서 다시 정리하였다. 각 수평각에서 20 개 data set 오차의 평균과 분산을 고도각에 따라 나타내었다. 고도각 0° 에서의 결과 (Fig.7 (a), (b))에서 보듯이 로봇의 머리에 의한 방해가 없는 부분에서는 두 가지 방법에 대한 결과

가 상당히 비슷함을 알 수 있다. 그러나 고도각 $10^\circ \sim 30^\circ$ 에서의 결과(Fig. 7 (c) ~ (h))는 로봇 머리에 의해 방해 받는 부분이 많아 질수록 개방된 위치에 마이크론이 존재하는 경우의 방법이 더 많은 오차가 발생함을 알 수 있다. 이는 로봇 머리에 의해 가려지게 되는 마이크론의 신호가 상당히 왜곡되어 나타나게 되는 결과이다. 고도각이 30° 인 경우에 폐쇄된 위치에 마이크론이 존재하는 경우의 방법도 상당한 오차를 포함하게 되는데, 이는 이 각도에서 자유음장이라는 가정이 깨지기 때문이다. 다시 말해서 음원이 존재하는 부분의 두 개의 마이크론도 음원의 직접 전달 경로가 완벽하게 존재하지 않기 때문이다. 위 결과에서 보듯이 고도각에 대한 추정 오차는 수평각에 대한 오차보다 상당히 적은 것을 볼 수 있다. 이것은 로봇 플랫폼에 장착되어 있는 마이크론 사이의 거리가 30cm 이하여서 수평각의 추정 오차가 고도각의 추정에 크게 영향을 미치지 않았고, 마이크론이 고도각이 변함에 따라 직접 전달 경로에 영향을 받지 않게 장착되어 있기 때문이다.

5. 결 론

본 연구에서는 TDOA 방법을 이용하여 3 차원 공간상에 존재하는 음원의 방향을 추정하는 두 가지 경우에 대한 방법을 제시하였고, 실제 로봇 플랫폼(IRP-2004)에 장착된 마이크론을 가지고 무향실에서 실험한 데이터를 이용해 그 특성들을 비교하였다. 로봇의 머리카 몸통에 의해 간섭 받지 않는 자유음장에서는 두 가지 방법이 거의 비슷한 결과를 보였지만 실제 로봇에 의한 간섭이 발생하는 경우에는 폐쇄된 위치에 마이크론이 존재하는 경우의 방법이 더 좋은 결과를 보여주었다. 이런 결과를 바탕으로 앞으로 로봇에 적용될 음원의 방향을 추정하는 연구는 실제 플랫폼의 영향을 고려하여 진행되어야 할 것이다. 그리고 이번 실험은 무향실에서 이루어 졌지만, 실제 사용 환경을 고려하여 환경적 요소가 포함되어 있는 공간에서 음성신호를 이용한 실험을 통해 두 가지 경우에 대한 비교실험도 필요하리라 생각된다.

후 기

이 논문은 2007 년도 정부(과학기술부)의 재원으로 한국과학재단의 국가지정연구실사업(M1050000112-05J0000-1121), 두뇌 한국 21 프로젝트와 전자부품연구원(KETI) 위탁 과제로 공동 수행된 연구임

참고문헌

- (1) Y. Rui and D. Florencio, "New direct approaches to robust sound source localization," in *proceeding of IEEE ICME*, 2003, pp. 730-740
- (2) A. A. Handzel, "High acuity sound source localization by means of a triangular spherical array," in *proceedings of IEEE ICASSP*, 2005, pp. 1057-1060.
- (3) F. Keyrouz and K. Diepold, "An enhanced binaural 3D sound localization algorithm," in *proceedings of IEEE International Symposium on Signal Processing and Information Technology*, 2006, pp 662-665.
- (4) 황성목 등, 2005, "머리전달함수를 이용한 로봇의 화자 위치 추정" 추계학술대회논문집, 한국소음진동공학회, pp.637-640
- (5) K. Nakadai, H. G. Okuno, and H. Kitano. "Real-time sound source localization and separation for robot audition," in *proceeding of IEEE International Conference on Spoken Language Processing*, 2002, pp. 193-196
- (6) J.M. Valin, F. Michaud, J. Rouat, D. Letourneau, "Robust sound source localization using a microphone array on a mobile robot," in *proceedings of the IEEE/IRJ Intl. Conference on Intelligent Robots and Systems*, 2003, pp. 1228-1233.
- (7) C. H. Knnapp, G. C. Cater, "The generalized correlation method for estimation of time delay," in *proceeding of IEEE Trnas. on acoustics, speech and signal processing*, Vol. Assp-24, No.4 1976.
- (8) J. Yli-Hietanen, K. Koppinen, and J. Astola, "Time-delay selection for robust angle of arrival estimation," in *proceedings of the IASTED International Conference Signal and Image Processing(SIP'99)* Oct. 18-21, 1999 - Nassau, Bahamas
- (9) Sungmok Hwang, "Sound Source Localization Using HRTF Database," Master's Thesis, Korea Advanced Institute of Science and Technology, 2006