

논문 2009-46SP-2-13

영역 분할을 이용한 효율적인 음원 위치 추정 시스템

(Efficient Sound Source Localization System Using Angle Division)

김 용 은*, 조 수 현*, 정 진 균**

(Yong-Eun Kim, Su-Hyun Cho, and Jin-Gyun Chung)

요 약

서비스 로봇에서 사용되는 음원인식 시스템은 사람이 로봇을 향해 말할 때 화자의 위치를 추정한다. 로봇용 음원인식 알고리즘들 중에서 복수개의 마이크로폰에 소리가 도착하는 시간지연 정보를 이용하여 음원을 추정하는 방법이 널리 이용된다. 마이크로폰에 도달하는 지연시간 정보를 구하기 위해서 상관관계 함수가 사용된다. 상관관계 함수에서 최대 값의 위치를 각도로 변환할 때 코사인 역함수가 사용되며 코사인 역함수의 비선형적인 특성 때문에 음원의 위치에 따라 마이크로폰 쌍에서 추정하는 각도의 정확성이 다르게 된다. 본 논문에서는 음원의 위치에 따라 각도를 가장 잘 추정할 수 있는 각도 영역을 구분하여 정확하게 음원의 위치를 인식하는 방법을 제안한다. 또한 제안한 방법을 이용하면 각 마이크 쌍에서 60도 범위에 대해서만 각도를 계산하므로 기존의 180도에 대한 방법에 비하여 연산이 1/2로 줄어든다. 테스트 환경을 구축하여 제안한 영역 분할 위치 추정 알고리즘과 기존 알고리즘의 성능을 비교한 결과 제안한 알고리즘의 기존 알고리즘에 의한 평균오차의 31%에 불과함을 보인다.

Abstract

Sound source localization systems in service robot applications estimate the direction of a human voice. Time delay information obtained from a few separate microphones is widely used for the estimation of the sound direction. Correlation is computed in order to calculate the time delay between two signals. Inverse cosine is used when the position of the maximum correlation value is converted to an angle. Because of nonlinear characteristic of inverse cosine, the accuracy of the computed angle is varied depending on the position of the specific sound source. In this paper, we propose an efficient sound source localization system using angle division. By the proposed approach, the region from 0° to 180° is divided into three regions and we consider only one of the three regions. Thus considerable amount of computation time is saved. Also, the accuracy of the computed angle is improved since the selected region corresponds to the linear part of the inverse cosine function. By simulations, it is shown that the error of the proposed algorithm is only 31% of that of the conventional approach.

Keywords: 음원 위치 추정, 영역 분할, 도달시간지연, 각도 매핑

I. 서 론

최근 사람과 로봇 사이의 원활한 상호작용을 위하여 비전 센서(vision sensor)를 이용한 시각 정보 뿐만 아니라 청각 정보를 활용하는 연구도 활발히 진행되고 있

다^[1]. 청각 정보 활용의 중요한 응용 중의 하나는 음원 위치 추정 기법을 이용하여 화자의 음성이 발생한 위치를 판단하여 로봇의 움직임에 활용하는 것이다^[1~3].

기존의 마이크로폰 사이의 도달시간지연(TDOA)이나 빔포밍을 이용한 음원 위치 추정 방법은 다수의 마이크로폰 어레이를 이용하여 화상회의에서 화자의 위치로 카메라가 움직이게 하기 위해 사용되었다^[4]. 음원 추정 알고리즘 중 음성 도달시간지연을 이용하여 마이크로폰 사이의 시간지연과 마이크로폰이 놓여 있는 기하학적 형상만을 이용하여 간단하게 음원의 위치를 추정할 수 있는 방법이 널리 사용된다^[5~6].

* 학생회원, ** 정회원, 전북대학교 전자정보공학부 (Div. of Electronic & Information Engineering Chonbuk University)

* This work was supported by MKE & IITA. (08-GIBAN-13, JEONBUK IT research center project)

접수일자: 2008년8월4일, 수정완료일: 2009년2월27일

기하학적 형상을 이용하여 음원 도달 지연시간을 측정함으로써 음원을 추정하는 경우 front-back 혼동현상을 방지하기 위해서 최소 3개 이상의 마이크로폰을 이용한다. 3개의 마이크로폰을 이용하는 경우 각각의 쌍에서 상관관계 값을 계산한 후 계산된 최대 값의 위치를 각도로 매핑하고 더하여 가장 큰 값의 위치를 음원 위치로 매핑한다.

본 논문에서는 매핑 유도식에서 사용되는 코사인 역함수의 선형적인 구간만을 사용하여 음원의 위치를 추정할 수 있는 각도 영역 분할 방법에 대해서 제안한다. 제안한 방법을 이용하였을 때 원 신호 위치와의 오차, 하드웨어, 상관 관계 연산횟수가 감소함을 보인다.

II장에서는 음성 도달 지연 시간을 이용한 기존의 음원 추정 방법, III장에서는 제안된 영역 분할 음원 추정 방법에 대해 설명한다. IV장에서는 시뮬레이션 결과를 제시하고 V장에서는 결론을 맺는다.

II. 음성 지연 시간을 이용한 음원 추정 방법

음원의 위치를 추정하기 위해서는 도달시간지연을 이용한 방법이 널리 사용된다. 음원에서 나온 음파는 두 마이크로폰에서 측정된 신호 $x_1(t)$ 와 $x_2(t)$, 음원에서 첫번째 마이크로폰에 도달한 신호 $s_1(t)$, 잡음 $n_i(t)$ 를 사용하여 식 (1)과 같이 나타낼 수 있다.

$$\begin{aligned} x_1(t) &= s_1(t) + n_1(t), \\ x_2(t) &= \alpha s_1(t - D) + n_2(t) \end{aligned} \quad (1)$$

식 (1)에서 α 는 감쇠계수이고, D 는 추정하고자 하는 두 신호사이의 시간지연이다. 두 신호 $x_1(t), x_2(t)$ 의 상호 상관계수는 식 (2)와 같다.

$$R_{x_1x_2}(\tau) = E[x_1(t)x_2(t - \tau)] \quad (2)$$

그림 1에서와 같이 두 마이크로폰사이의 거리를 l_{mic} 음원 위치 각도를 ϕ 라 하면 지연거리 d 는 다음과 같이 표현된다.

$$d = l_{mic} \cos \phi \quad (3)$$

음속을 $c(c = 343m/s)$ 라고 하면 지연거리 d 에 해당하는 지연시간 τ 는 다음과 같다.

$$\tau = \frac{d}{c} = \frac{l_{mic} \cos \phi}{c} \quad (4)$$

따라서 구하고자 하는 음원위치각도 ϕ 는 다음과 같이 표현된다.

$$\phi = \cos^{-1} \frac{\tau c}{l_{mic}} \quad (5)$$

음원입력을 샘플링주파수 f_s 로 샘플링하고 음원신호가 n_d 샘플만큼 지연되었을 때 지연거리 d 는 다음과 같다.

$$d = \frac{cn_d}{f_s} \quad (6)$$

따라서 두 신호사이의 최대 지연 샘플 구간내에서 음원 추정 연산을 통해 음원위치를 추정 할 수 있다. 그림 1에서 d 는 직각삼각형의 한 변이므로 직각삼각형의 빗변인 l_{mic} 와는 $d < l_{mic}$ 인 관계가 있다. 따라서 최대지연

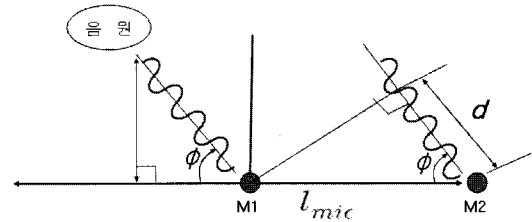


그림 1. 마이크로폰 2개를 이용한 음원 위치 추정
Fig. 1. Sound source localization system using two microphone.

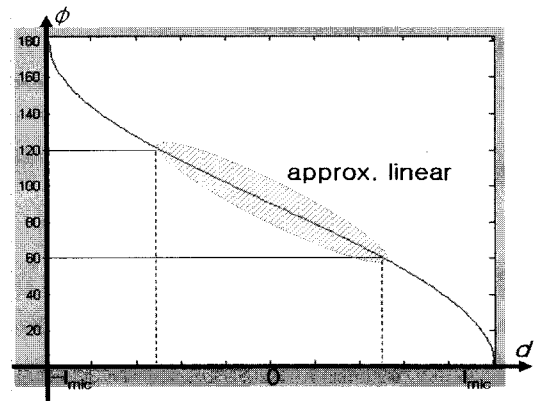


그림 2. 코사인 역함수 그래프
Fig. 2. Inverse cosine graph as a function of d.

샘플 수는 $d = l_{mic}$ 일 때 다음 식과 같다.

$$n_{d,max} = \frac{f_s l_{mic}}{c} \quad (7)$$

식 (5)에서 최종 각도 매핑은 코사인 역함수에 의해 수행 된다. 코사인 역함수의 입력을 -1에서 1까지 변화

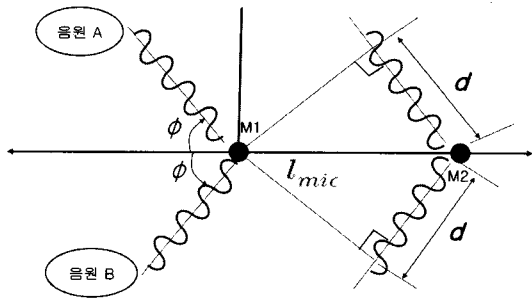


그림 3. 앞 뒤 혼동 현상
Fig. 3. Frond back confusing phenomenon.

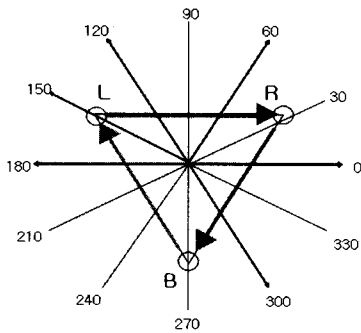


그림 4. 마이크로폰 3개를 이용한 음원 추정
Fig. 4. Sound source localization system using tree microphone.

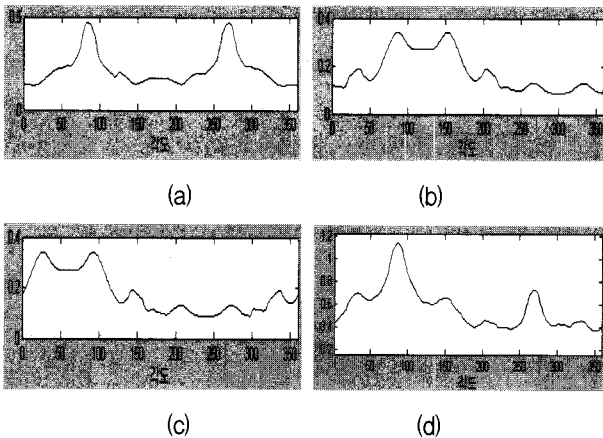


그림 5. 음원이 90° 일 때 각 마이크 쌍에서 구해진 각도: (a)LR, (b)BL (c)RB, (d)LR+BL+RB
Fig. 5. Angle estimation results obtained When sound source is 90°: (a)LR, (b)BL (c)RB, (d)LR+BL+RB.

시킴으로써 해당하는 각을 구하면 그림 2와 같이 비선형적으로 변화함을 알 수 있다. 하지만 그림 2의 60°~120°구간에서는 거의 선형적인 특징을 보인다.

그림 1과 같이 마이크 2개만을 이용하여 음원의 위치를 추정하면 그림 3과 같이 전방에서 ϕ 의 각으로 입사된 신호(음원 A)와 후방에서 ϕ 의 각으로 입사된 신호(음원 B)의 모양이 동일하게 되어 전방과 후방을 구분

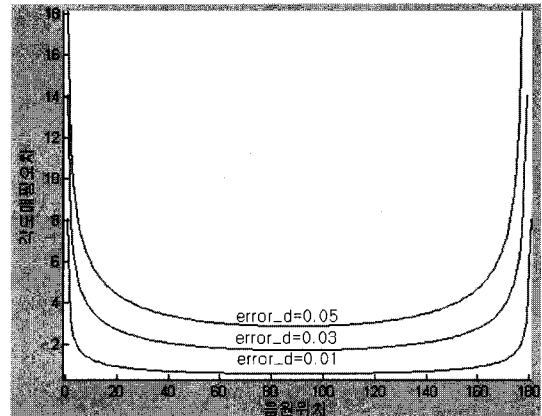


그림 6. d에 따른 변화할 때 각도 위치에 따른 음원
Fig. 6. Estimation error of sound source locations as d.

하지 못하는 앞뒤 혼동 현상이 발생한다.

음원 앞뒤 혼동 현상을 피하기 위하여 보통 그림 4 같이 마이크로 폰 3개(L, R, B)를 이용한다. 그림 4의 각 마이크로폰 페어의 상관관계 값은 각 페어에 표시된 화살표 꼬리의 수신 신호를 $x_1(t)$, 화살표 머리의 수신 신호를 $x_2(t)$ 라고 할 때 식(2)에 의해 계산된다.

그림 5는 음원이 90°에 위치해 있을 때 세 개의 마이크로폰을 사용한 실험 결과를 보여준다. 각각의 마이크로폰 페어에서 계산된 각도를 더한 값인 그림 5-(d)로부터 음원이 90°에 위치해 있음을 알 수 있다. 좀더 세밀할 위치를 추정하기 위해서는 가중함수를 사용할 수 있다^[8].

III 제안한 방법을 이용한 음원 추정 시스템

1. 각도 분할 방법을 이용한 음원인식

주위 환경 방해 요소나 마이크로폰이 장착되는 모형으로 인해 지연거리 d 의 계산과정에 오차가 발생한 경우가 존재한다^[8]. 이 때 지연거리 오차가 포함된 데이터의 상관관계 연산 값을 코사인 역함 수를 이용하여 각도로 매핑할 때 선형으로 근접화 되는 구간(그림 2의 approx. linear 영역)에서 각도 매핑 오차가 작다. 예를 들어 그림 2의 approx. linear 영역에 포함된 각도인 90°에 음원이 존재할 때 지연거리 d 에 0.01의 오차가 발생하였다면 각도로 변화하였을 때 89.427°로 실제 음원과 0.5730° 오차가 발생한다. 반면 각도 0°에 음원이 존재하고 지연거리 d 의 오차가 0.01일 때는 음원의 실제 각도와 8.11°만큼 오차가 발생한다.

그림 6은 d 의 오차가 0.01, 0.02, 0.03일 때 음원 위치

에 대한 각도오차를 나타낸다. 60°~120°사이에는 각도 매핑 오차가 상대적으로 적지만 음원 위치가 0°~60°와 120°~180°에 위치하면 각도 매핑 오차가 크게 증가함을 알 수 있다. 따라서 음원이 60°~120°사이에 위치해 있을 때 음원 위치 추정의 정확도가 증가하게 된다. 마이크로폰 3개를 이용하여 음원을 인식하는 경우 음원을 추정하는 3개의 마이크로폰 페어로부터 연산결과가 출력되고 각각 지연거리 오차가 발생한 경우 보다 정확한 음원 인식이 가능한 마이크로폰 페어가 존재한다. 그림 7은 각 마이크로폰 페어에 대해 정확한 음원인식이 가능한 구간을 보여준다. 그림 7을 정리하면 표 1과 같다.

그림 7에서 60°~120° 또는 240°~300°사이에 음원이 위치하고 지연시간 오차가 발생하였다면 마이크로폰 R-L의 음원 인식 결과가 가장 정확하다. 반면 0°~60° 또는 180°~240° 사이에 음원이 위치하고 지연시간 오차가 발생하였다면 마이크로폰 B-R의 음원 인식 결과가 가장 정확하다.

이렇게 음원 위치에 따라 마이크로폰 페어에서 연산된 정확한 결과를 선택하기 위해서는 음원이 발생한 위치가 어느 구간에 있는지 찾아야 한다. 음원이 발생한 위치 구간은 상관관계의 최대 값이 발생한 위치를 비교하여 쉽게 구분할 수 있다.

예를 들면 그림 7의 90°에 음원이 위치하게 되면 마이크로폰 R-L 페어를 고려할 때 R과 L 마이크로폰에 지연차가 거의 없는 신호가 입력된다. 따라서 식(2)를 이용하여 상관계수를 구할 때 $\tau=0$ 근처에서(즉, 신호들을 상대적으로 서로 시프트 시키지 않았을 때) 최대 값이 발생한다($\tau = n/f_s$, n 은 정수). 반면, 마이크로폰 R-B 페어를 고려하면 때 B 마이크로폰에 R에 비해 상대적으로 지연된 신호가 도착하게 된다. B의 수신신호

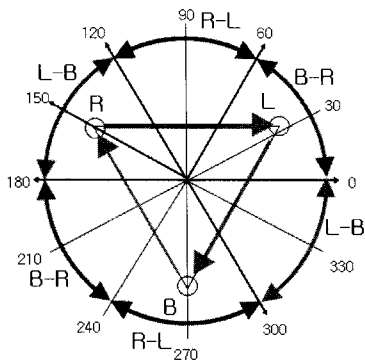


그림 7. 음원 추정 분할 영역
Fig. 7. The division part of sound source localization.

를 $x_1(t)$, R의 수신신호를 $x_2(t)$ 라하고 식 (2)를 이용하여 상관계수를 구할 때 $x_2(t)$ 를 오른쪽으로 시프트 하여야($\tau > 0$) 두 신호가 유사한 모양이 되어 상관계수 값이 최대가 된다. 결과적으로 그림 8와 같은 상관관계 결과 그래프가 출력된다. 그림 8에서 식 (7)의 최대 지연 샘플을 고려하면 64개의 연산 값에 해당함을 알 수 있다.

표 2는 음원영역에 따른 마이크로폰 RL, LB, BR의 상관관계 최대값 위치 ($n(RL_{max})$, $n(LB_{max})$, $n(BR_{max})$) 비교표이다. 표 2에서 상관관계 최대값의 위치에 따라

표 1. 음원 위치에 따라 정확히 음원 위치를 인식할 수 있는 마이크로폰 쌍
Table 1. Selection of proper microphone pair for six different source locations.

음원 위치(각도)	음원 위치 판단에 적합한 마이크로폰 페어
60~120 또는 240~300	R-L
120~180 또는 300~360	L-B
180~240 또는 0~60	B-R

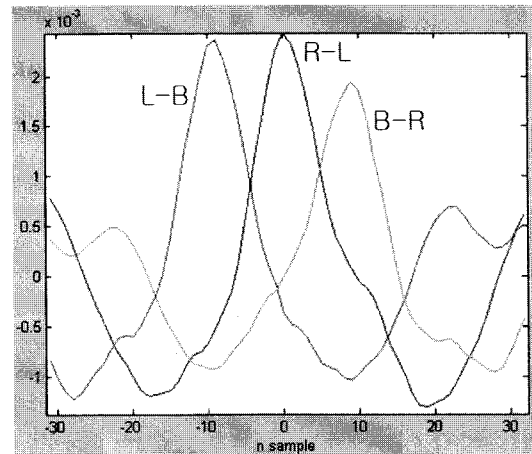


그림 8. 90°에서 음원이 입력될 때 세 쌍의 마이크로폰 사이의 상관관계 출력 비교
Fig. 8. Comparison of the correlation values obtained from three microphone pairs for the localization.

표 2. 마이크로 폰 영역 선택
Table 2. Selection of proper microphone pair.

상관관계 최대값 위치	영역	앞, 뒤
$n(LB_{max}) \leq n(RL_{max}) \leq n(BR_{max})$	R-L	앞
$n(BR_{max}) \leq n(LB_{max}) \leq n(RL_{max})$	L-B	앞
$n(LB_{max}) \leq n(BR_{max}) \leq n(RL_{max})$	B-R	앞
$n(BR_{max}) \leq n(RL_{max}) \leq n(LB_{max})$	R-L	뒤
$n(RL_{max}) \leq n(LB_{max}) \leq n(BR_{max})$	L-B	뒤
$n(RL_{max}) \leq n(BR_{max}) \leq n(LB_{max})$	B-R	뒤

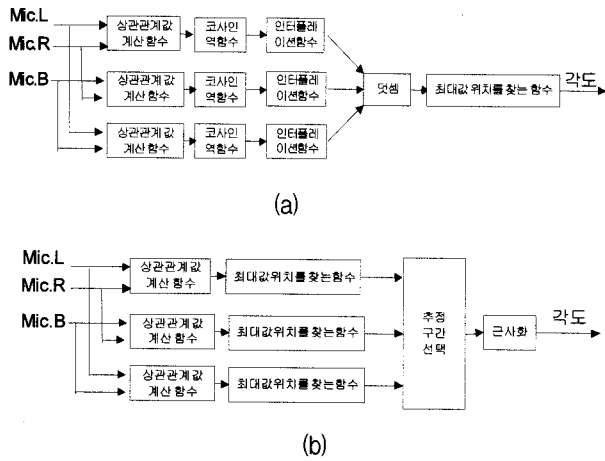


그림 9. 기존의 음원 인식 블록도와 제안한 음원 추정 시스템 블록도: (a) 기존의 음원인식 블록도, (b) 제안한 음원인식 블록도

Fig. 9. Block diagrams of conventional and proposed methods: (a) conventional method, (b) proposed method.

서 영역을 구분할 수 있음을 알 수 있다. 영역이 구분되면 마이크로폰 페어의 계산 결과 값 중에 그 영역에서 각도 매핑오차가 가장 적은 값을 선택할 수 있으므로 지연거리 오차에 대한 각도 매핑 오차를 줄일 수 있다.

2. 선형 식을 이용한 각도 매핑 함수 간소화

기존의 각도 매핑회로는 코사인 역함수가 필요하고 코사인 역함수의 출력을 1° 간격으로 매핑하기 위해 보간함수가 필요하다. 하지만 제안한 각도 분할 음원인식 방법은 코사인 역함수의 선형 부분(그림 2)만 이용하므로 선형 직선식을 이용하면 코사인 역함수 및 보간 함수를 사용할 필요가 없게된다. 다음 식은 코사인 역함수 대신 이용될 직선식을 나타낸다.

$$y = ax + b \tag{8}$$

식 (8)에서 각도가 120°일 때와 각도가 60°일 때 즉 y 가 120°, 60°일 때 지연 위치의 값 즉 x의 값이 존재하므로 두 점을 지나는 직선식을 이용하면 계수 a, b를 식 (9)와 같이 구할 수 있다.

$$a = \frac{-60}{(\cos(\frac{1}{3}\pi) - \cos(\frac{2}{3}\pi)) \times l_{mic}} \tag{9}$$

$$b = 120 - \frac{-60 \times \cos(\frac{2}{3}\pi) \times l_{mic}}{(\cos(\frac{1}{3}\pi) - \cos(\frac{2}{3}\pi)) \times l_{mic}}$$

이와 같이 식 (9)를 이용하면 코사인 역함수에 비해 간단한 하드웨어로 구현 가능하며 보간 함수가 제거되어 간단한 구현이 가능하다. 또한 기존의 음원 인식 시스템은 각 마이크로폰 쌍에서 0°~180° 사이의 상관 관계값을 모두 구해야 했지만 제안한 각도 분할 방법에서는 60°~120°사이 즉 60° 구간의 상관관계 값만 계산하고 선택하여 사용하므로 상관관계 연산 횟수가 2배 감소한다. 그림 9은 기존의 음원인식 블록도와 제안한 음원인식 블록도를 나타내고 있다. 제안한 음원 인식 블록도는 기존의 음원인식 블록도 그림 9-(a)의 코사인 역함수와 보간 블록이 제거됨을 알 수 있다.

IV. 실험 결과

그림 10은 음원추정을 위한 테스트 환경이다. 정삼각형 모양으로 배치된 3개의 마이크에 소리를 내면 이 소

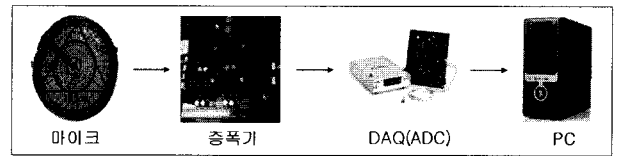


그림 10. 실험 환경

Fig. 10. Test environment.

표 3. 기존의 방식과 제안한 방식의 시뮬레이션 결과: (a) 기존 방식을 이용한 음원 추정 결과, (b) 제안한 방식을 이용한 음원 추정 결과

Table 3. The Simulation result of conventional method and proposed method: (a)conventional method, (b) proposed method.

(a)

음원위치 거리	0°	30°	60°	90°
1m	0°	27°	56°	88°
2m	0°	27°	59°	85°
3m	0°	27°	59°	88°
최대 절대 오차	0°	3°	4°	5°
평균 오차	0°	3°	2°	3°

(b)

음원위치 거리	0°	30°	60°	90°
1m	0°	32.7°	60°	87.2°
2m	0°	32.7°	60°	87.2°
3m	0°	32.7°	60°	87.2°
최대 절대 오차	0°	2.7°	0°	2.8°
평균 오차	0°	2.7°	0°	2.8°

리가 증폭기를 통해 증폭된 후, DAQ(ADC)로 보내진다. DAQ에서는 음성신호를 16 KHz로 샘플링 한 후 PC로 보낸다. PC에서는 음성 신호를 텍스트 형태로 받아 음원 추정 프로그램으로 보내게 된다. 따라서 신호를 더해서 일정 값 이상이 되면 이를 사람 음성으로 판단하고 음원의 위치를 추정한다. 그림 10과 같이 3개의 마이크로폰을 이용해 음원의 위치를 추정하기 위해서는 마이크 페어 R과 L, 마이크 페어 L과 B, 마이크 페어 B와 L로 부터 얻은 시간지연정보를 각각 각도로 변환한 후 더하여 최대 값을 찾는다. 다양한 음원의 위치와 거리에 대해 256개 샘플의 음성 데이터를 이용해 기존의 방법과 제안한 방법으로 실험한 결과를 표 3에 보였다. 실험결과 제안한 방법에 의한 평균오차는 기존 방법의 31%에 불과함을 알 수 있다.

V. 결 론

제안한 분할 영역 음원 위치 추정 방식을 이용하면 간단한 함수만을 이용하여 하드웨어를 제작할 수 있고 전력 소비를 줄일 수 있다. 또한 각도 오차를 줄여 음원 인식 시스템을 로봇 및 응용 분야에 널리 사용할 수 있다. 추후 제안한 음원 인식시스템을 칩으로 제작하여 제안한 알고리즘을 실용화할 계획이다.

참 고 문 헌

[1] K. Nakadai, et al, "Active audition for humanoid," *Proceedings of the 17th National Conference on Artificial Intelligence and 12th Conference on Innovative Applications of Artificial Intelligence*, pp. 832-839, 2000.

[2] 황성목, "머리전달함수를 이용한 로봇의 화자 위치 추정", *한국소음진동공학회 추계학술대회논문집*, pp. 637-640, 2005.

[3] Y. Sasaki, S. Kagami and H. Mizoguchi, "Multiple sound source mapping for a mobile robot by selfmotion triangulation," *Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Beijing, China, 2006.

[4] 권병호, 김경호, 박영진, "로봇 플랫폼에서 마이크로폰 위치를 고려한 음원의 방향 검지 방법" *한국소음진동공학회 2007년 추계학술대회 논문집*, pp. 1080-1084, 2007.

[5] M. S. Brandstein and H. Silverman, "A practical

methodology for speech source localization with microphone arrays," *Computer Speech and Language*, 11(2): 91-126, 1997.

[6] C. H. Knnapp and G. C. Cater, "The generalized correlation method for estimation of time delay," *IEEE Trnas. on Acoustics, Speech and Signal Processing*, vol. Assp-24, no. 4, 1976.

[7] C. H. Knnapp and G. C. Cater, "The generalized correlation method for estimation of time delay," *IEEE Trnas. on acoustics, speech and signal processing*, Vol. Assp-24, No. 4, 1976.

[8] 황성목, 박윤식, 박영진, "머리전달함수를 이용한 로봇의 화자 위치 추정," *한국소음진동공학회 추계 학술대회 논문집*, pp.637-640, 2005.

저 자 소 개



김 용 은(학생회원)
2005년 전북대학교 전자정보공학부 학사 졸업
2007년 전북대학교 정보통신공학과 석사 졸업
2007년~현재 전북대학교 전자정보공학부 박사과정
<주관심분야 : 통신, 신호처리, 반도체>



조 수 현(학생회원)
2007년 전북대학교 전자정보공학부 학사 졸업
2009년 전북대학교 전자공학과 석사 졸업
<주관심분야 : 통신, 컴퓨터, 임베디드 시스템>



정 진 균(정회원)
1985년 전북대학교 전자공학 학사 졸업
1989년 미국 미네소타 주립대학 전기공학 석사 졸업
1991년 미국 미네소타 주립대학 전기공학 박사 졸업
<주관심분야 : 통신, 컴퓨터, 신호처리, 반도체>