

가려진 마이크로폰을 이용한 음원 위치 추적*

Sound Source Localization using Acoustically Shadowed Microphones

이 협 우** · 육 동 석**
Hyeopwoo Lee · Dongsuk Yook

ABSTRACT

In many practical applications of robots, finding the location of an incoming sound is an important issue for the development of efficient human robot interface. Most sound source localization algorithms make use of only those microphones that are acoustically visible from the sound source or do not take into account the effect of sound diffraction, thereby degrading the sound source localization performance. This paper proposes a new sound source localization method that can utilize those microphones that are acoustically shadowed from the sound source. The experiment results show that use of the acoustically shadowed microphones, which receive higher signal-to-noise ratio signals than the others and are closer to the sound source, improves the performance of sound source localization.

Keywords: sound source localization, SRP-PHAT, microphone array

1. 서 론

인간과 휴머노이드 로봇 사이의 자연스럽고 효율적인 의사소통을 위해서는 시스템이 소리의 발생 위치를 알아야 한다[1]. 정확한 음원의 위치 추적은 로봇이 인간의 얼굴을 바라보게 할 수 있고, 다른 방향에서 들어오는 잡음을 줄일 수 있다[2]. 이것은 인간과 로봇사이의 인터페이스에서 중요한 역할을 하는 화자 인식이나 음성 인식에 응용될 수 있다.

두 마이크로폰 사이의 신호 도달 시간차 추정을 통한 기존의 음원 위치 추적 알고리즘들은 마이크로폰 배열에 음원으로 부터 직접 전파 경로가 존재한다는 것을 가정하고 알고리즘을 수행한다. 하지만, 로봇의 몸통과 같은 구조물이 신호의 입력 경로를 방해하는 제약이 있는 환경에서는 기존 알고리즘의 가정이 맞지 않는 경우가 발생할 수 있다. 이러한 경우, 기존 알고리즘에 전파 경로가 가려진 마이크로폰들을 사용할 수 없고, 빔형성 단계에서 사용 불가능한 마이크로폰의 개수에 비례하여 signal-to-noise ratio(SNR)가 작아지는 문제가 발생 할 수 있다[3]. 따라서 시스템에 장착된

* 이 논문은 2006년도 정부(교육과학기술부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임 (KRF-2006-311-D00822).

** 고려대학교 컴퓨터·전파통신공학과 음성정보처리 연구실

일부의 마이크로폰을 사용할 수 없는 경우, 음원 위치 추적 알고리즘의 성능은 저하될 수 있다.

이 문제를 해결하기 위해 본 논문에서는 음원으로부터 직접 전파 경로가 가려진 마이크로폰에 음파 회절현상을 통하여 도착하는 음성 신호의 이론적 시간차를 구함으로써 시간차를 이용한 음원 위치 추적 알고리즘의 성능 향상을 시도하였다. 본 논문의 구성은 다음과 같다. 이어지는 장에서 기존 음원 위치 추적 기법들을 살펴보고, 3 장에서 음원으로부터 가려진 마이크로폰을 사용하는 방법을 유도한다. 4 장에서는 제안한 알고리즘의 성능을 평가하고, 결론을 맺는다.

2. 음원 위치 추적 기법

마이크로폰 배열 기반의 음원 위치 추적의 방법은 부공간 추정(sub-space estimation) 방법, 시간차 추정(time delay estimation) 방법, 빔형성(beamforming) 방법 등으로 나눌 수 있다.

부공간 추정을 이용한 방법에는 대표적으로 multiple signal classification(MUSIC)을 기반으로 한 알고리즘들이 있다[4]. MUSIC은 입력된 신호에 원하는 음성과 관련이 없는 잡음을 가정하고, cross correlation matrix를 eigen decomposition하여 전체 입력된 신호들을 신호 부공간과 잡음 부공간으로 분리한다. 이들이 서로 수직이라는 원리를 이용하여, 음원의 방향을 예측하기 위해 다음을 계산한다.

$$P(e^{j\omega}) = \frac{1}{\sum_{n=s+1}^N |e^H v_n|^2} \quad (1)$$

여기서 s 는 추정하는 신호의 개수 이고, v_n 는 잡음 고유벡터, e^H 는 탐색하는 방향에 대한 신호 고유벡터의 Hermitian 형태로, 가장 큰 $P(e^{j\omega})$ 값을 갖는 방향을 입력된 음원의 방향으로 예측한다. MUSIC 기법은 높은 해상도를 보이지만 잡음이 있는 실내 환경에서 중요한 요소인 반향을 고려하지 않고 있어 수행 시 다른 음원 위치 추적 방법에 비해 낮은 성능을 보일 수 있다.

시간차 추정 방법 중 가장 대표적인 방법으로 generalized cross correlation(GCC)이 있다[5]. 두 마이크로폰에서 입력받은 신호 사이의 도착 시간차(τ)를 다음과 같이 cross correlation function을 이용하여 구한다.

$$R_{12}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi_{12}(\omega) X_1(\omega) X_2^*(\omega) e^{j\omega\tau} d\omega \quad (2)$$

식 (2)에서 $\psi_{12}(\omega)$ 는 가중치 함수를 나타내고, $X_i(\omega)$ 는 i 번째 마이크로폰을 통해 입력된 신호의 Fourier transform, '*'는 complex conjugate를 나타낸다. 마이크로폰 간의 시간차 하나는 하나의 hyperbolic 형태의 후보 음원 위치를 나타내므로, 특정한 신호의 발생 위치를 가리키기 위해서는 적어도 두 개의 시간차가 필요하다. GCC 방법은 계산 량이 적고 구현이 간단하지만 반향 크기에 따

른 성능의 저하가 큰 단점이 있다.

마지막으로, 빔형성기를 이용하는 대표적인 방법으로 steered response power (SRP) 방법이 있다 [6]. 입의 좌표에 대한 방향으로 filter-and-sum beamformer의 출력 파워($SRP(q)$)를 구하여 가장 큰 값을 갖는 지점을 소리의 발생 위치라고 예측하는 것으로 출력 파워는 다음과 같이 표현된다.

$$SRP(q) = 2\pi \sum_{m=1}^N \sum_{n=1}^N R_{mn}(\tau_{mn}) \quad (3)$$

여기서 q 는 음원의 후보 위치 좌표이고, N 은 마이크로폰 개수, $\tau_{mn} = \tau_m - \tau_n$ 은 두 마이크로폰 사이의 신호 도달 시간차를 나타낸다. $R_{mn}(\tau_{mn})$ 은 앞서 설명한 GCC의 값이다. SRP 알고리즘은 해석상 GCC의 모든 마이크로폰 조합의 합을 이용한 시간차 추정 기법이라 할 수 있다. 특히, 가중치 함수를 $\psi_{ij}(\omega) = 1/|X_i(\omega)X_j^*(\omega)|$ 로 적용하면 입력 신호의 파워를 같게 하고 phase의 차이를 이용하는 것으로, phase transform (PHAT)이라 한다. PHAT 가중치 기법은 반향에 강인하다고 알려져 있고, 이를 적용한 SRP 알고리즘을 SRP-PHAT라 한다. SRP-PHAT는 다른 음원 위치 추적 방법들 보다 좋은 성능을 보이지만 모든 후보 위치에 대하여 파워를 계산하고 위치를 추정하므로 계산 량이 많은 단점이 존재한다.

하지만, 로봇과 같이 외형의 제약으로 인한 마이크로폰의 장착 위치나 개수의 제한이 있는 경우, 발생하는 음원으로부터 직접적인 전파 경로가 없는 마이크로폰들이 존재한다. 위의 방법들은 이와 같은 문제를 고려하지 않았고, 그로 인한 부정확한 시간차 추정이 전체 알고리즘의 성능을 저하시킬 수 있다. 다음 장에서는 이러한 문제에 대한 해결책을 소개한다.

3. 음원으로부터 가려진 마이크로폰의 영향

로봇을 중심으로 360 도의 모든 방향에 대한 음원의 위치를 추적하기 위해서 uniform circular array(UCA)를 장착할 수 있지만, 음원 위치 추적에 있어 로봇의 외형은 마이크로폰 배열에 하나의 제약 조건이 될 수 있다. 예를 들어, 로봇의 몸통 모양과 팔의 위치와 같은 외형은 음원에서 발생하는 신호의 직접 전파 경로를 차단할 수 있다. 이점을 고려하지 않은 기존의 음원 위치 추적 알고리즘들은 직접 전파 경로를 가지는 마이크로폰만을 가정하기 때문에 위치 추적 알고리즘의 성능 저하를 야기할 수 있다.

<그림 1>은 로봇이 원형의 몸통을 가지고 있고 8 개의 마이크로폰을 등 간격 원형 배열형태로 장착한 상태를 위에서 내려다본 것 모습이다. 큰 원은 로봇의 몸통을 나타내고, 작은 점들은 장착된 마이크로폰들을 나타낸다. 신호가 빛금 친 파선 부분인 'Region 1'에서 발생하면, 총 8 개의 마이크로폰 중에 3 개의 마이크로폰만이 음원으로부터 직접 전파된 소리를 받아들이고, 나머지 5 개의 마이크로폰은 로봇 몸통에 의해 전파 경로가 가려지게 된다. 또한, 어두운 부분인 'Region 2'에서 신호 발생 시 4 개의 마이크로폰만이 직접 전파경로를 가지게 된다. 이것은 장착된 마이크로폰에 입력된 자원을 모두 사용하지 못하고 일부의 마이크로폰만을 음원 위치 추적에 활용한다는 것을 의미한다.

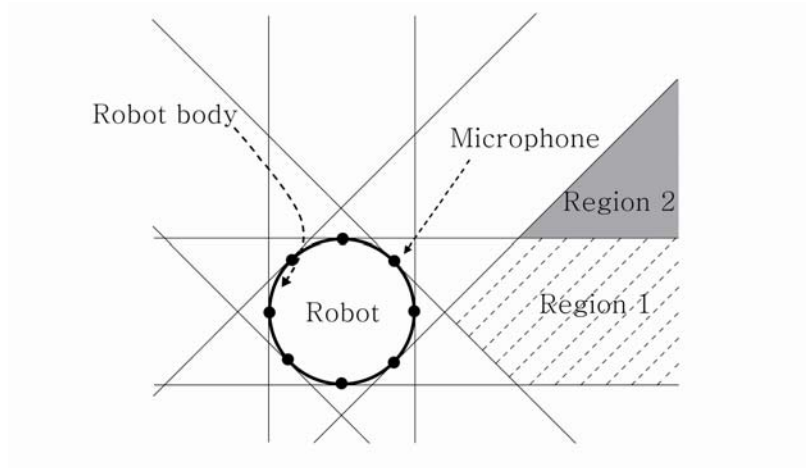


그림 1. 가려진 마이크론 예제

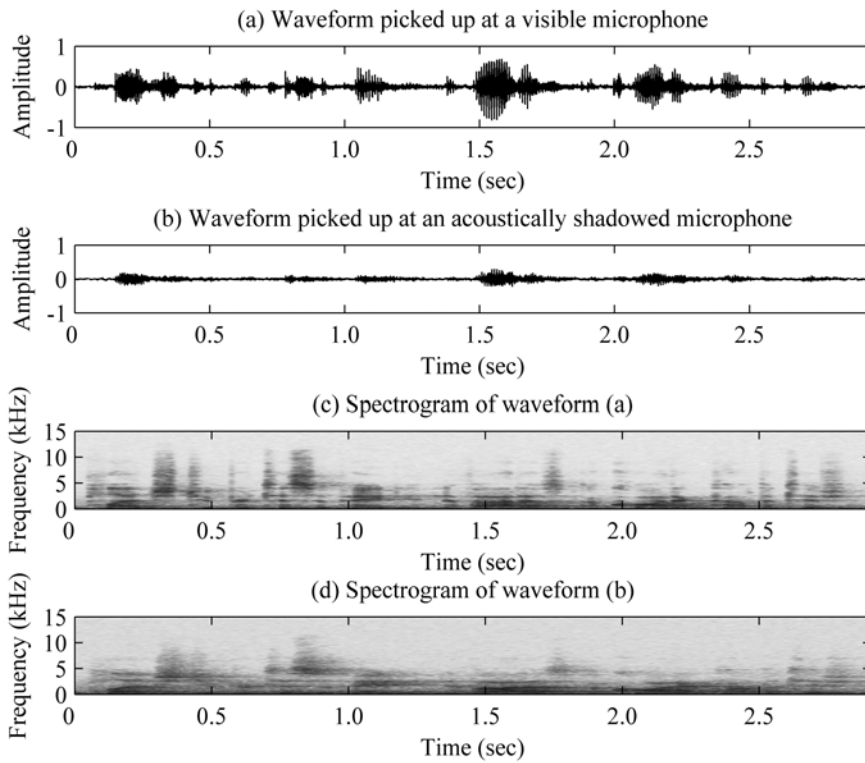


그림 2. 마이크론에 입력된 신호

음원의 위치 추적 성능이 이용 가능한 마이크론의 개수와 빔형성된 신호의 SNR에 비례하여 향상되는 사실에 비추어 보면 직접 전파 경로를 갖지 못하는 마이크론은 위치 추적 알고리즘의

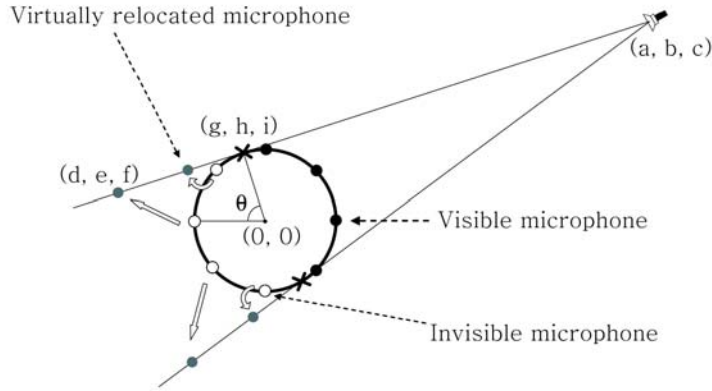


그림 3. Virtual relocation of microphones

성능을 저하 시키는 요인이 될 수 있다. 사용가능한 모든 입력된 신호를 이용하여 음성을 강화하면 높은 SNR의 신호를 얻을 수 있다는 사실을 토대로 직접 전파 경로가 가려진 마이크론도 사용하면, 음원 위치 추적의 성능을 향상 시킬 수 있을 것이다. <그림 2>는 가려진 마이크론과 직접 전파 경로가 있는 마이크론에 입력된 신호를 비교한 것이다. <그림 2(a)>는 직접 전파 경로가 존재하는 마이크론에 의해 입력된 신호이고, (b)는 가려진 마이크론에 의해 입력된 신호이다. (c)와 (d)는 (a)와 (b)의 spectrogram을 나타낸다. 물론, 직접 전파 경로가 있는 (a)와 (c)의 경우 좀 더 강하고 명확한 신호의 특성을 보이지만, (b)와 (d)에서도 여전히 입력된 신호의 특성이 보이므로 이를 이용하고자 한다. 본 연구에서는 가려진 마이크론을 직접 전파 경로를 가지는 곳으로 가상적으로 옮겨 시간차를 추정하는 virtual relocation of microphone(VROM) 방법을 제안한다.

<그림 3>은 <그림 1>의 'Region 2'에서 신호가 발생된 경우를 도식화한 것으로, 신호가 (a, b, c)에서 발생하면 신호는 로봇의 몸통과 'X' 표시된 접점까지 직접 전파 경로를 갖는다. 검은색 점은 음원에서부터 보이는 직접 전파 경로가 있는 마이크론들이고, 흰색 점들은 가려진 마이크론들을 나타내는데, 가려진 마이크론들에도 신호가 입력된다. 신호가 접점 'X' 이후부터 로봇의 표면을 따라 굽어 들어가는 회절 현상에 의해 전파되므로, 가려진 마이크론을 기존의 시간차 추정 기반의 음원 위치 추적 알고리즘에 적용하기 위해서 마이크론의 위치를 직접 전파 경로가 있는 위치로 가상적으로 옮기면 시간차를 이론적으로 추정할 수 있다. 신호가 전파되는 경로의 길이를 유지하고, 음원과 접점 그리고 가상의 마이크론 위치가 일직선이 되는 곳(<그림 3>의 회색 점)으로 가려진 마이크론을 가상으로 이동시킨다. 음원 (a, b, c)와 로봇 몸통의 접점 (g, h, i)의 좌표와 접점과 각각의 가려진 마이크론이 이루는 각도인 θ 를 이용하여, 가상으로 옮긴 마이크론의 좌표 (d, e, f)를 식 (4)를 이용하여 구할 수 있다.

$$A. \begin{cases} g = \frac{ar^2 \pm rb\sqrt{a^2+b^2-r^2}}{a^2+b^2} \\ h = \pm \sqrt{r^2-g^2} \\ i = c \end{cases} \quad B. \begin{cases} d = g \pm \theta h \\ e = \frac{r^2-dg}{h} \\ f = i \end{cases} \quad (4)$$

제안된 VROM 방법을 음원 위치 추적 알고리즘인 SRP-PHAT에 적용한 VROM-SRP-PHAT는 다음과 같이 수행된다.

- 1) 후보 음원 위치 q 를 선택한다.
- 2) 식 (4).A를 이용하여, 접점을 구한다.
- 3) 접점을 기준으로 각도 θ 를 이용하여 가려진 마이크로폰을 찾아내고, 식 (4).B를 이용하여 가상으로 옮긴 마이크로폰의 위치를 계산한다.
- 4) $\tau_{ij} = \frac{|p_i - q| - |p_j - q|}{v}$ 를 이용하여 시간차를 구한다. 여기서 p_i, p_j 는 각 마이크로폰의 공간 좌표이고, v 는 음파의 속도이다.
- 5) Steered response power를 식 (3)을 이용하여 구한다.
- 6) Step 1)에서 5)를 모든 음원의 후보 위치에 대하여 계산하고 그 중 가장 큰 값을 갖는 위치를 음원의 위치로 추정한다.

4. 실험 결과

VROM 방법의 성능을 알아보기 위해 8 개의 마이크로폰을 등 간격 원형 형태로 배열하여 반지름 15 cm 높이 50 cm 플라스틱 통에 장착하였다. 9 x 4.7 x 3 m 크기의 회의실에서 3 번 마이크로폰을 정면으로, 직접 전파 경로를 가지는 마이크로폰을 3 개로 하고(visible: 2, 3, 4 번 마이크로폰, invisible: 1, 5, 6, 7, 8 번 마이크로폰), 스피커와의 거리 1, 3, 5 m에서 48 kHz sampling rate으로 음성 데이터를 녹음하였다. 녹음한 회의실의 반향은 약 0.8 초였다. <그림 4>는 데이터를 녹음한 환경과 마이크로폰의 번호 배치를 나타낸다. 녹음한 데이터를 2048 samples (43 ms) 단위의 프레임으로 나누고 Hamming window를 적용하여 프레임 길이의 반씩 전진하며 실험을 진행하였다.

실험은 크게 세 단계로 나누어진다. 우선 가려진 마이크로폰을 모두 사용한 경우의 성능을 측정하고, 다음으로 사용하는 가려진 마이크로폰의 개수를 달리하여 음원 위치 추적 알고리즘의 성능 개선 기여도를 살펴본다. 그리고, 3 차원 공간에서의 실험 결과도 살펴본다. 우선, 2 차원 평면에서 SRP-PHAT와 VROM-SRP-PHAT의 성능을 비교하였다. <표 1>은 각도(방위각)의 정확도와 위치 측정 오차를 나타내는 root mean squared error (RMSE), 정답과의 거리 오차를 측정된 것이다. 정확도는 예측한 각도와 정답 각도와의 차이가 ± 5 도의 범위 안에 존재하면 정답으로 간주하여 프레임의 개수를 구하고 이를 전체 프레임 수(약 2700 개)로 나누어 계산하였다. 오차를 두는 이유는 스피커 크기와 각도의 실측 정확성을 고려한 것이다. 거리 오차는 예측한 음원과 실제 음원의 거리

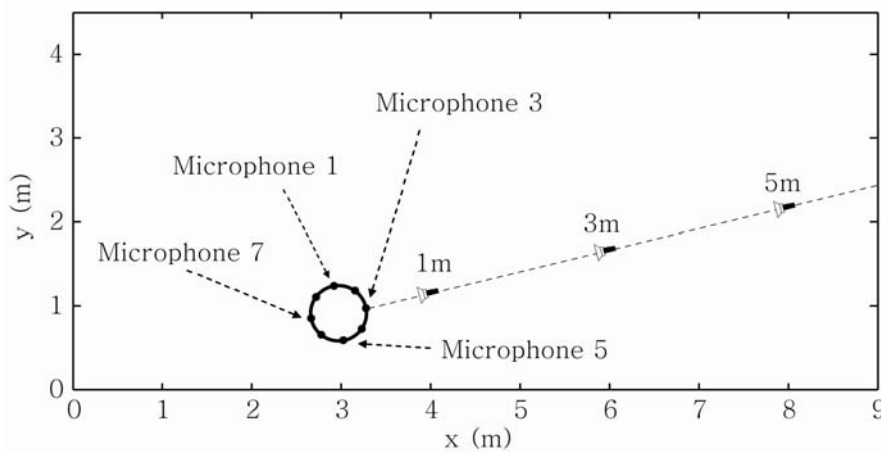


그림 4. 실험 데이터 녹음 환경

다음과 같이 계산되는 RMSE는 음원 위치 추적의 각도 오차를 뜻한다.

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{\alpha}_t - \alpha)^2} \tag{5}$$

식 (5)에서 T 는 프레임의 개수이고, α 는 실제 정답 음원 각도를 나타내며, $\hat{\alpha}_t$ 는 t 프레임에 알고리즘이 예측한 음원의 각도를 뜻한다. 정확도, RMSE, 거리 오차에서 전반적으로 VROM-SRP-PHAT가 높은 성능을 보인다.

표 1. 음원 위치 추적 알고리즘의 성능 비교

거리(m)	SRP-PHAT			VROM-SRP-PHAT		
	정확도(%)	RMSE(각도)	거리오차(m)	정확도(%)	RMSE(각도)	거리오차(m)
1	88.4	12.9	3.9	91.7	10.6	1.4
3	80.8	14.6	2.0	86	11.8	1.8
5	66.6	17.2	1.1	71.4	15.5	1.4

하지만, 다른 세션의 녹음에 대한 실험 결과에서는 VROM-SRP-PHAT가 기존의 SRP-PHAT보다 약 1.5%의 정확도 하락을 보이기도 하였다. 이와 같은 일관되지 않는 실험의 결과는 각각의 가려진 마이크로폰마다 성능에 미치는 영향이 다르다는 것을 의미할 수 있다.

마이크로폰마다의 성능 기여도 분석을 위한 실험으로, 사용하는 가려진 마이크로폰의 개수와 음원 위치 추적 성능을 비교하였다[7]. 회절이나 반향 등의 영향으로 마이크로폰 간의 기여도는 다를 수 있기 때문에, 가려진 마이크로폰의 개수와 위치 추적 알고리즘 성능의 관계를 알아보기 위해 마이크로폰의 개수를 달리하며 실험하였다. 마이크로폰의 번호는 <그림 4>에서 보는 것과 같이 시계 방향으로 1에서 8까지 차례로 부여하였다. 음원이 3번 마이크로폰의 정면에 위치 할 때, 2, 3, 4

번이 보이는 마이크로폰, 1 번과 5 번 마이크로폰은 보이는 마이크로폰에 양쪽으로 인접하였고, 6 번과 8 번은 다음 간격으로 떨어졌으며, 7 번이 음원에서 가장 멀리 떨어진 보이지 않는 마이크로폰이다. <표 2>는 녹음 환경에서 명시한 2, 3, 4 번의 직접 경로가 있는 마이크로폰 3 개를 기준으로, 가려진 마이크로폰의 개수를 증가시켜 가며 VROM-SRP-PHAT를 수행하여 측정된 음원 위치 추적 정확도와 오차 정도를 보여주는 RMSE의 평균을 나타낸다. <표 2>의 처음 세 줄은 <표 1>의 SRP-PHAT와 같고, 마지막 세 줄은 VROM-SRP-PHAT의 결과와 같다. 표에서 '3 visible mics + 1 invisible mic'는 5 가지 경우(2,3,4,1 / 2,3,4,5 / 2,3,4,6 / 2,3,4,7 / 2,3,4,8)의 결과에 대한 평균값이고 '3 visible mics + 2 invisible mics'는 10가지 마이크로폰 조합의 평균을 나타낸다. 이런 평균값들은 약간의 오차를 보여주는 하나, VROM-SRP-PHAT와 큰 차이를 보이지 않는다.

표 2. 마이크로폰 구성에 따른 VROM-SRP-PHAT의 성능 비교

마이크로폰 구성	거리 (m)	정확도 (%)	RMSE (각도)	거리오차 (m)
3 visible mics (SRP-PHAT)	1	88.4	12.9	3.9
	3	80.8	14.6	2.0
	5	66.6	17.2	1.1
3 visible mics + 1 invisible mic	1	91.9	10.0	1.9
	3	82.8	12.7	1.6
	5	69.0	17.0	1.4
3 visible mics + 2 invisible mics	1	91.3	10.6	1.6
	3	82.5	13.8	1.7
	5	66.9	17.9	1.5
3 visible mics + 3 invisible mics	1	91.1	10.9	1.5
	3	82.9	13.8	1.7
	5	66.3	18.3	1.5
3 visible mics + 4 invisible mics	1	91.1	10.5	1.4
	3	83.7	13.3	1.7
	5	66.7	18.0	1.5
3 visible mics + 5 invisible mics (VROM-SRP-PHAT)	1	91.7	10.6	1.4
	3	86	11.8	1.8
	5	71.4	15.5	1.4

사용된 가려진 마이크로폰 개수간의 일관되지 않은 결과는 각각의 마이크로폰이 서로 다른 기여도에서 비롯된 것으로, 각 마이크로폰간의 음원 위치 추적 알고리즘의 기여도를 살펴보기 위해 마이크로폰 조합별로 실험한 결과를 <그림 5>와 <그림 6>에 나타낸다. 마름모 표시의 실선은 음원과의 거리 1 m에서의 결과, 사각형 표시의 긴 점선은 3 m에서의 결과, 삼각형 표시의 점선은 5 m에서의 결과이다. x축은 직접 전파 경로가 있는 2, 3, 4 번 마이크로폰에 추가되어 실험한 가려진 마이크로폰의 번호를 나타낸다. 예를 들어, <그림 5(b)>의 x축 첫 번째 항목 "1, 5"는 보이는 마이크로폰인 2, 3, 4 번 마이크로폰과 1, 5 번의 가려진 마이크로폰을 이용하여 음원 위치 추적을 수행한 결과를 나타낸다. 정확도와 RMSE는 앞의 실험과 같은 방법으로 계산하였다. <그림 5(a)>는 1

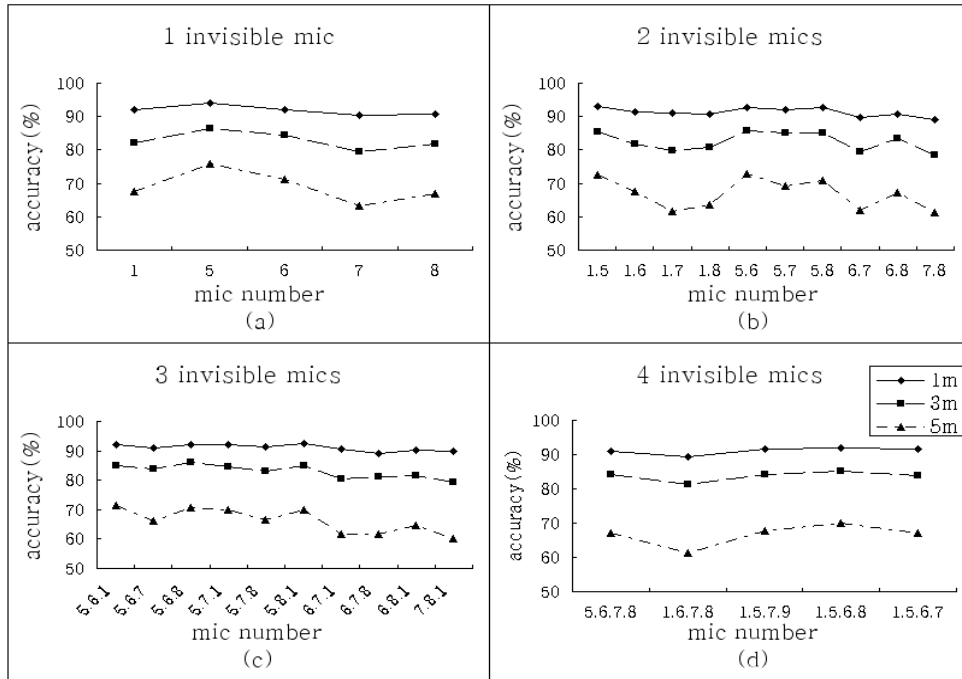


그림 5. 가려진 마이크로폰 개수에 따른 정확도

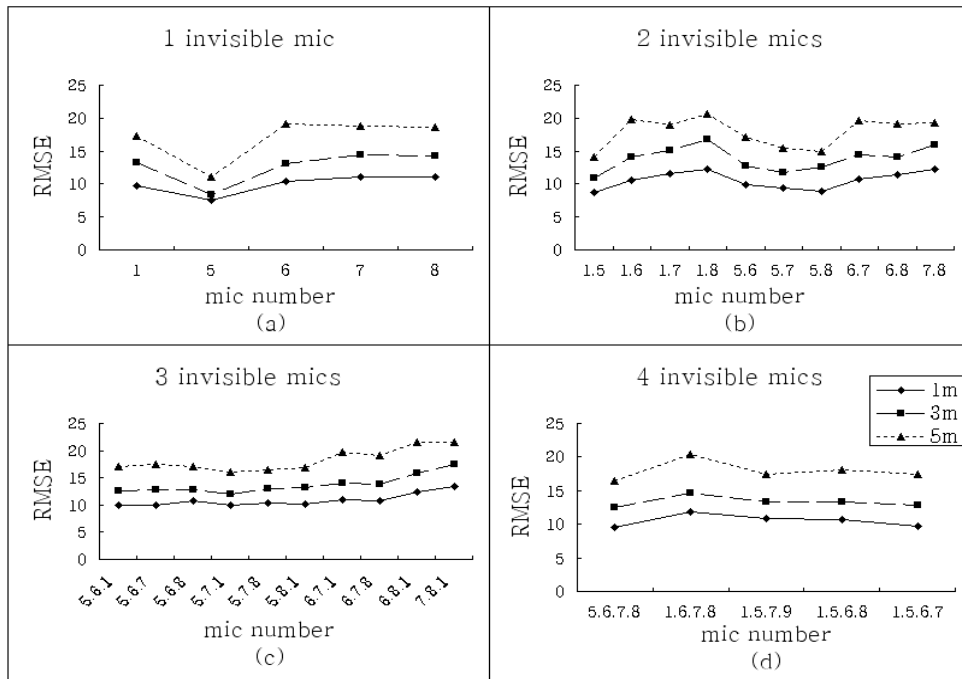


그림 6. 가려진 마이크로폰 개수에 따른 RMSE

개의 가려진 마이크를 추가한 경우의 위치 추적 알고리즘의 정확도이고, (b)는 2 개, (c)는 3 개, (d)는 4 개의 마이크를 추가한 결과를 나타낸다. 이들 중 가려진 5 번 마이크 하나만을 더 사용한 경우, 정확도 94.1%와 RMSE 7.6 도를 보이며 모든 구성 가능한 마이크 조합 중 가장 좋은 성능을 보인다. 전체적으로 1 번이나 5 번 마이크가 포함되는 조합의 경우 좋은 성능을 보인다. 이와 대조적으로, 음원과 가장 멀리 떨어진 7번 마이크를 추가한 경우, 정확도 90.3%와 11.1 도의 RMSE를 보이며 다른 마이크 조합보다 성능을 저하시켰으며, 이 마이크가 포함된 다른 조합에서도 성능의 저하를 보였다. 여기서, 마이크마다의 음원 위치 추적 영향력이 다르다는 것을 알 수 있다. <그림 6>은 RMSE의 결과를 나타낸다. 전체적으로 정확도와 같은 경향을 보인다. 이러한 성능을 보이는 요인으로 판단되는 것은 입력된 신호의 SNR로 <그림 7>에 나타난다. <그림 7(a)>는 가려진 마이크 하나의 SNR을 나타내고, (b)는 x축에 나타난 2 개의 가려진 마이크에 의해 입력된 신호의 평균 SNR을, (c)는 3 개, (d)는 4 개의 평균 SNR을 나타낸다. 각각의 SNR은 <그림 5>의 음원 위치 추적 정확도와 높은 상관관계를 갖는 것을 알 수 있다. 또한, 보이는 마이크와 인접한 마이크에 입력된 신호의 SNR이 높으며 이 신호를 알고리즘에 적용하는 것이 좋은 결과를 보인다.

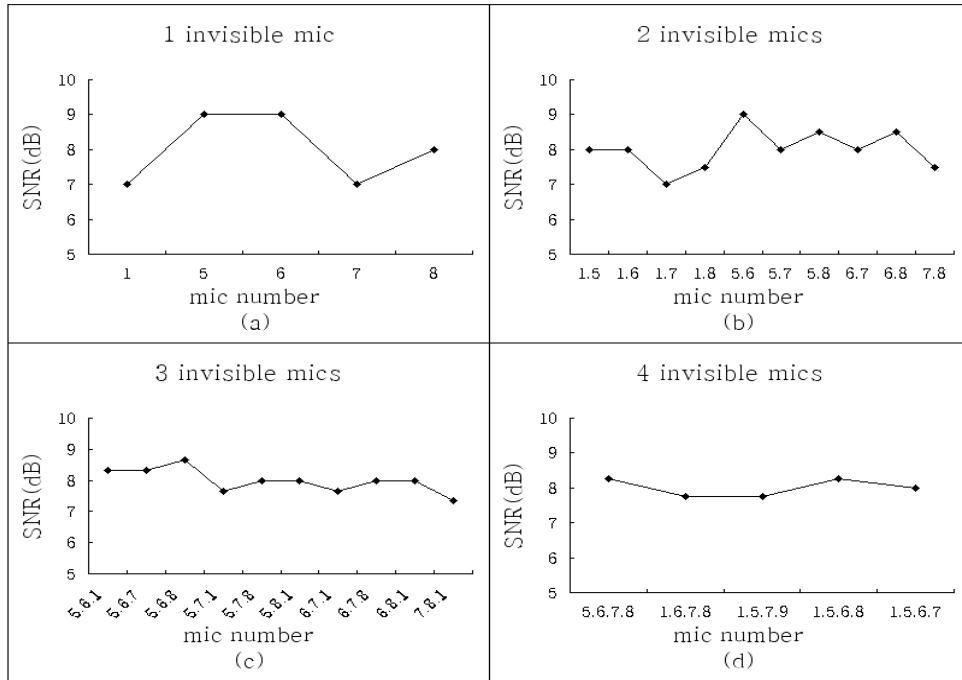


그림 7. 1 m 거리에서 가려진 마이크에 입력된 신호의 평균 SNR

마지막으로 실제 로봇 환경과 같은 3 차원 음원 위치 추적을 위해 같은 방에서 데이터를 녹음하였다. 기존의 원형 마이크론 배열은 방위각만을 찾는 데 적합한 구조이므로, 방위각과 고도각을 동시에 찾기 위해 같은 원형 마이크론 배열을 약 25 cm의 높이 차이로 배열하였다. <표 3>은 두

개의 원형 마이크로폰 배열을 이용한 음원까지의 수평 거리 1 m에 3 가지 높이에 대한 알고리즘 수행 결과를 보여준다. 3 차원의 경우 방위각과 고도각을 예측한 공간 좌표로부터 계산할 수 있다. 2 차원 배열에서는 보이는 마이크로폰이 3 개이지만, 2 개의 원형 배열을 사용한 3 차원 환경에서는 직접 전파경로를 갖는 마이크로폰이 6 개가 된다. 이를 반영하여 가려진 마이크로폰의 개수를 증가시켜 실시한 실험에서는 위아래 짝수개의 마이크로폰을 추가하였고, 그 구성 중 가장 좋은 경우의 성능을 나타내었다. 성능의 측정은 앞에서와 같은 방법을 방위각과 고도각 각각에 적용하였다. 모든 항목에서 가장 성능이 좋은 “6 visible mics + 4 invisible”은 보이는 마이크로폰 6 개와 그와 인접한 SNR이 상대적으로 높은 위아래 4 개의 가려진 마이크로폰을 사용하여 음원의 위치를 추적한 것이다. 2 차원 결과에서와 같이 반향이 있는 3 차원 환경에서 입력된 신호의 SNR과 보이는 마이크로폰과의 상대적 거리가 음원 위치 추적 알고리즘 성능에 영향을 주는 것을 알 수 있다.

표 3. 3차원 환경에서의 알고리즘 성능 비교

알고리즘	높이 (m)	방위각 (%)	고도각 (%)	방위각 RMSE (각도)	고도각 RMSE (각도)	거리오차 (m)
SRP-PHAT	1.1	76.7	74.0	42.2	9.8	2.7
	1.6	84.1	78.2	33.6	3.4	1.8
	1.8	76.1	71.0	39.7	9.9	2.5
6 visible mics + 2 invisible mic	1.1	83.1	81.5	35.6	8.3	1.7
	1.6	87.2	82.5	37.2	3.0	1.2
	1.8	81.0	77.8	41.3	8.5	1.9
6 visible mics + 4 invisible mics	1.1	96.7	83.6	13.3	7.9	1.5
	1.6	97.8	82.5	17.5	3.0	0.5
	1.8	96.4	78.5	19.5	8.2	1.3
VROM- SRP-PHAT	1.1	86.0	75.6	28.5	9.4	1.7
	1.6	89.1	75.8	27.9	3.4	0.8
	1.8	84.0	75.8	29.7	8.9	1.3

5. 결 론

본 연구에서는 로봇의 외형과 같은 마이크로폰 배열 위치의 제약을 해결하기 위해 음원으로부터 직접 전파 경로가 가려진 마이크로폰을 음원에서 보이는 방향으로 위치를 가상으로 옮겨 이론적 시간차를 구하였다. 이렇게 예측한 시간차는 기존의 음원 위치 추적 알고리즘에 직접 사용할 수 있었다. 실험 결과를 종합해보면, 직접 전파 경로가 가려진 마이크로폰의 이론적 시간차를 음원 위치 추적 알고리즘에 사용하는 것이 보이는 마이크로폰만 사용하는 기존의 방법보다 음원 위치 추적 성능 면에서 좋을 수 있다는 것을 알 수 있었다. 하지만 보다 안정적이고 향상된 성능의 음원 위치 추적 알고리즘을 위해서는, 모든 음원으로부터 가려진 마이크로폰을 음원 위치 추적에 사용하기 보다는 상대적으로 직접 전파 경로가 존재하는 마이크로폰과 인접하고 입력된 신호의 SNR이 높은 가려진 마이크로폰을 선택적으로 사용하는 것이 다른 마이크로폰 조합보다 높은 성능 향상을 나타

내는 것을 관찰할 수 있었다. 이것은 잡음, 반향, 회절 등에 의해서 각각의 마이크로폰이 음원 위치 추적에 갖는 영향력이 다르기 때문이라고 해석할 수 있다.

참 고 문 헌

- [1] Valin, J. M., Michaud, F., Rouat, J. & Letourneau, D. 2003. "Robust sound source localization using a microphone array on a mobile robot." in *Proc. IEEE Int. Conf. Intelligent Robots and Systems* 1228-1233.
- [2] Gannot, S., Burshtein, D. & Weinstein, E. 2001. "Signal enhancement using beamforming and nonstationarity with applications to speech." *IEEE Transactions on Signal Processing* 8, 1614-1626.
- [3] Flanagan, J., Surendran, A. & Jan, E. 1993. "Spatially selective sound capture for speech and audio processing." *Speech Communication* 13, 207-222.
- [4] Schmidt, R. O. 1986. "Multiple emitter location and signal parameter estimation." *IEEE Transactions on Antennas and Propagation* 34, 276-280.
- [5] Knapp, C. H. & Carter, G. C. 1976. "The generalized correlation method for estimation of time delay." *IEEE Transactions on Acoustics, Speech, and Signal Processing* 24, 320-327.
- [6] DiBiase, J. 2000. *A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays*. Ph.D. Thesis, Brown University, USA.
- [7] 이협우, 육동석. 2007. "가려진 마이크로폰의 음원 위치 추적 성능 분석." *음성 통신 및 신호처리 학술대회* 229-230.

접수일자: 2008. 8. 8

수정일자: 2008. 8. 26

게재결정: 2008. 9. 2

▲ 이협우

서울특별시 성북구 안암동
고려대학교 정보통신대학 컴퓨터·전파통신공학부 (우: 136-701)
Tel: +82-2-3290-3641 (O)
Fax: +82-2-3290-3641
E-mail: hwlee@voice.korea.ac.kr

▲ 육동석(교신저자)

서울특별시 성북구 안암동
고려대학교 정보통신대학 컴퓨터·전파통신공학부 (우: 136-701)
Tel: +82-2-3290-3202 (O)
Fax: +82-2-3290-3641
E-mail: yook@voice.korea.ac.kr