

2차원 상의 음원위치 추정을 위한 효율적인 영역분할방법

An efficient space dividing method for the two-dimensional sound source localization

김환용, 최홍섭[†]

(Hwan-Yong Kim and Hong-Sub Choi[†])

대진대학교 전자공학과

(Received May 24, 2016; revised June 16, 2016; accepted August 4, 2016)

초 록: 음원의 위치를 찾는 SSL(Sound Source Localization)은 로봇과의 인터페이스, 화상회의, 스마트 자동차 등 여러 분야에서 꼭 필요한 기술이다. 일반적으로 음원의 위치 정보를 활용하는 기술들은 주로 측정 장치에 대한 음원의 각도 정보를 찾아서 이용하고 있다. 그러나 음원의 위치에 대한 각도를 추정할 때 이용하는 사인 역함수의 비선형적인 특성으로 추정된 각도에 오차가 발생하며, 이에 대한 방안으로 마이크가 담당하는 영역을 분할하는 방법이 제안되었다. 본 논문에서는 마이크 어레이 패턴에 따른 영역분할 방법을 제안하고 음원의 위치를 2차원상의 평면 좌표로 특정하는 방법으로 위치 추정 성능을 평가하였다. 실험에서 잡음에 강인한 GCC-PHAT(Generalized Cross Correlation Phase Transform) 방법을 사용했으며, 마이크 어레이의 패턴은 마이크 3개와 4개로 삼각형과 사각형 두 종류로 구성하였으며, 100개의 음성 데이터로 실험한 결과 실제 환경에서는 3개의 마이크 어레이를 사용해서는 영역 분할 해상도가 낮아서 음원의 위치를 정해진 특정 범위내로 추정하는데 실패했으나, 4개 마이크를 이용하여 해상도를 높였더니 위치추정 성공률이 67%로 크게 향상됨을 확인할 수 있었다.

핵심용어: SSL(Sound Source Localization), 마이크 배열, GCC-PHAT, 영역분할방법

ABSTRACT: SSL (Sound Source Localization) has been applied to several applications such as man-machine interface, video conference system, smart car and so on. But in the process of sound source localization, angle estimation error is occurred mainly due to the non-linear characteristics of the sine inverse function. So an approach was proposed to decrease the effect of this non-linear characteristics, which divides the microphone's covering space into narrow regions. In this paper, we proposed an optimal space dividing way according to the pattern of microphone array. In addition, sound source's 2-dimensional position is estimated in order to evaluate the performance of this dividing method. In the experiment, GCC-PHAT (Generalized Cross Correlation PHase Transform) method that is known to be robust with noisy environments is adopted and triangular pattern of 3 microphones and rectangular pattern of 4 microphones are tested with 100 speech data respectively. The experimental results show that triangular pattern can't estimate the correct position due to the lower space area resolution, but performance of rectangular pattern is dramatically improved with correct estimation rate of 67%.

Keywords: SSL (Sound Source Localization), Microphone array, GCC-PHAT, Space dividing method

PACS numbers: 43.60.Ac, 43.60.Vx

1. 서 론

음원의 위치를 추정하는 기술은 인간과 기계의 친화적 인터페이스를 위해 연구되어야 하는 중요한 요소기술로 자리매김하고 있다. 이런 음원 위치 추정

[†]Corresponding author: Hong-Sub Choi (hschoi@daejin.ac.kr)
School of Electrical Engineering, University of Daejin, 1007
Hoguk-ro, Pocheon-si, Gyeonggi-do 11159, Republic of Korea
(Tel: 82-31-539-1903, Fax: 82-31-539-1900)

기술은 휴머노이드 로봇, 가상 회의, 스마트 자동차, 지능형 CCTV 등 여러 분야에 적용할 수 있다. 휴머노이드 로봇은 인간과 의사소통을 위해 음원 위치 추정 기술을 초창기부터 적용해 왔다. 스마트 자동차의 경우 네비게이션에 음성인식 기능과 함께 지향성 마이크를 이용하는 음원 위치 추정 기술을 탑재하여 운전자의 음성만 인식하는 등, 이미 실용적인 단계에 들어서고 있다. 보통 음원 위치 추정은 마이크를 이용하여 음원과 추정기 사이의 각도를 찾는 데, 본 논문에서는 음원의 위치를 2차원상의 좌표로 추정하는 방법을 제안한다.

일반적으로 음원위치 추정방법은 마이크의 개수와 동기화 방법에 따라 사용하는 알고리즘이 달라진다. 위치 추정 알고리즘은 TDOA(Time Difference of Arrival), AOA(Angle Of Arrival), TOA(Time Of Arrival), 방법 등을 사용하며, 이중 음원과 동기화가 필요하지 않고 소량의 마이크센서로 음원의 위치를 찾을 수 있는 TDOA^[1]방법을 많이 이용한다. 이 외에도 빔포밍^[2] 방법이나, 미리 구축한 데이터베이스를 이용한 머리전달함수,^[3] 힐버트변환,^[4] PCA(Principal Component Analysis)^[5] 등을 이용할 수 있다. 본 논문에서는 TDOA 방법을 이용해 지연시간을 구한 뒤, 음원의 위치를 찾는 방법을 이용한다. 음성신호는 각 마이크마다 서로 다른 위상정보를 가지고 입력되는데 이 때 시간영역에서 처리하여 지연시간을 구하는 상호상관함수와, 주파수 영역에서 계산하는 GCC(Generalized Cross Correlation) 방법이 있다. 상호상관함수는 데이터량이 많아질수록 계산량이 제공에 비례하여 늘어나기 때문에 GCC 방법을 많이 이용한다. 그리고 음압의 크기를 백색화하여 위상정보만으로 음원의 위치를 추정할 수 있는데 이때 가중함수를 이용하며 음원의 크기를 백색화하여 지연시간을 계산하는 방법이 GCC-PHAT(Generalized Cross Correlation Phase Transform)라고 한다. PHAT 가중함수를 이용하면 방향효과를 줄이고, 잡음에 강한 특성을 보여 많이 사용한다.

본 논문에서는 TDOA방법을 사용해서 음원의 좌표를 효율적으로 찾기 위해 공간을 분할^[6]하여 좌표를 추정하는 방법을 제안하고 실험을 통하여 성능을 확인하였다.

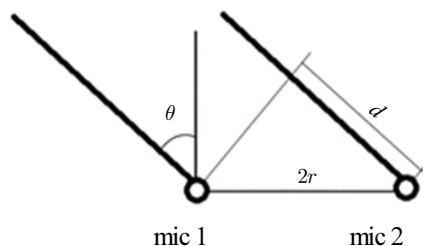


Fig. 1. Angle estimation method using the microphone1 (mic1) and microphone2 (mic2). Distance between microphones is $2r$. d is the delayed distance and θ is the sound source angle.

II. Time Difference of Arrival

2.1 상호상관함수

상호상관함수는 두 신호의 상관 정도를 나타내는 함수이다.

Fig. 1에서 음성 신호를 $s(t)$ 라 하고, 1번 마이크에 들어온 신호를 $x_1(t)$, 2번 마이크에 들어온 신호를 $x_2(t)$, 잡음을 $n(t)$, 두 마이크 사이의 거리를 $2r$, 지연거리를 d 라고 하면 Eq.(1)과 같이 나타낼 수 있다.

$$\begin{aligned} x_1(t) &= s(t) + n(t), \\ x_2(t - \tau) &= s(t - \tau) + n(t). \end{aligned} \quad (1)$$

Eq.(1)에서 τ 값을 계산하기 위해 마이크에 들어온 두 신호의 상호상관함수를 구하면 Eq.(2)와 같이 나타낼 수 있으며

$$R_{x_1x_2}(\tau) = \int_{-\infty}^{\infty} x_1(t)x_2(t - \tau)dt, \quad (2)$$

이 τ 값들 중 가장 큰 상호상관함수 값을 갖는 τ 값이 최종적인 도달 지연시간이 되므로 Eq.(3)을 이용하여 지연시간을 찾을 수 있다.

$$\hat{\tau} = \operatorname{argmax}_{R_{x_1x_2}}(\tau). \quad (3)$$

2.2 Generalized Cross Correlation Phase Transform

GCC방법은 신호의 유사성을 계산하고, 서로간의

위상 차이를 계산하여 지연 시간을 계산하는데 사용하는 알고리즘이다. 상호상관함수의 경우는 시간영역에서 계산하는데 컨볼루션 연산이기 때문에 데이터량에 제곱비례하여 연산량이 증가한다. 그러나 GCC방법은 입력신호를 주파수 영역으로 넘겨 계산하기 때문에 컨볼루션 연산이 곱 연산으로 바뀌어 상호상관함수보다 연산량이 적다. 따라서 실시간으로 처리해야 할 경우에 상호상관함수에 비해 상대적으로 계산량이 적은 GCC방법을 이용하면 실시간 처리가 가능해진다. GCC방법은 Eq.(2)에서 신호 $x_1(t)$ 을 푸리에 변환했을 때 $X_1(\omega)$, 신호 $x_2(t)$ 를 푸리에 변환했을 때 $X_2(\omega)$ 라고 하면 Eq.(2)를 푸리에 변환했을 때 Eq.(4)와 같이 표현할 수 있다.

$$R_{x_1x_2}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X_1(\omega)X_2^*(\omega)e^{i\omega\tau}d\omega. \quad (4)$$

Eq.(4)는 각각의 마이크 센서로 입력받은 신호를 푸리에 변환하여 주파수 영역의 곱 연산으로 바꾼 뒤 역변환하는 것으로 Eq.(2)에 비해 연산량이 적다. Eq.(4)의 방법을 GCC알고리즘이라고 하며, 이때 적절한 가중치 $W(\omega)$ 를 추가하여 Eq.(5)와 같이 계산할 수 있다.

$$R_{x_1x_2}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} W(\omega)X_1(\omega)X_2^*(\omega)e^{i\omega\tau}d\omega,$$

$$W(\omega) = \left| \frac{1}{X_1(\omega)X_2(\omega)} \right|. \quad (5)$$

이때 가중치 $W(\omega)$ 는 신호의 크기를 정규화하여 신호를 백색화하고, 위상차이만을 이용하여 지연 시간을 계산하는 것으로 GCC-PHAT방법이라 한다.

III. 마이크의 물리적 배치에 따른 각도 추정방법

3.1 2개의 마이크를 이용한 각도추정

Fig. 1에서는 마이크 2개를 이용하여 음원의 각도를 찾는 방법을 나타낸다. 그러나 2차원상에서 음원의 각도를 찾기 위해서는 최소 3개 이상의 마이크가

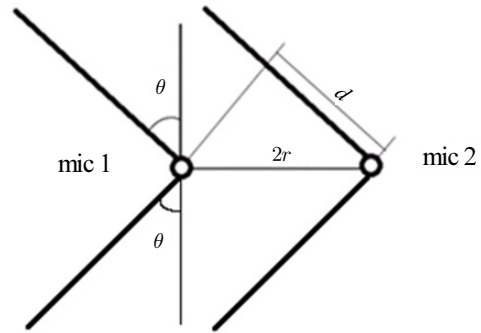


Fig. 2. Front-back confusion phenomenon is present when sound source angle θ is estimated by two microphones. Distance between two microphones is $2r$, and d is the delayed distance.

필요하다.

Fig. 2를 보면 마이크 2개를 이용하면 음원의 각도 2개가 나와 정확한 위치를 찾을 수 없다는 것을 알 수 있다. 일단 두 개의 마이크에서 각도 θ 를 구해야 하는데 Eq.(6)을 이용하면 θ 를 구할 수 있다.

$$\theta = \sin^{-1} \frac{d}{2r} = \sin^{-1} \frac{c\hat{\tau}}{2r}. \quad (6)$$

Eq.(6)을 살펴보면 $2r$ 은 마이크 사이의 거리이고, c 는 음속도로 상수이므로 Eq.(6)은 지연시간 τ 와 θ 의 1대1 매핑인 것을 알 수 있다. 즉, Eq.(2)나 Eq.(5)에서 τ 값에 따른 상호상관값을 구하게 되는데, 이때 Eq.(6)을 이용하여 τ 를 θ 로 바꿔줄 수 있다. 이때 θ 값을

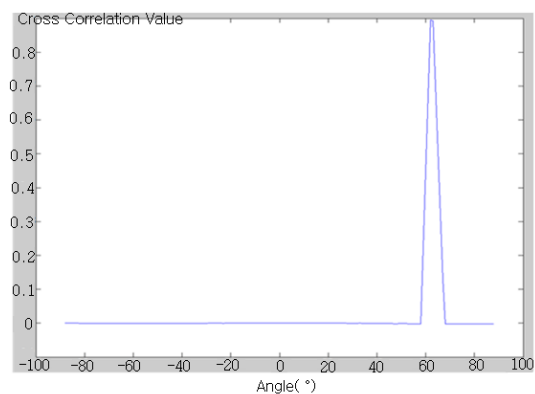


Fig. 3. Relationship between cross correlation coefficient value (x-axis) and sound source angle (y-axis) when sound source is located with 60° respect to reference axis.

-180 ~ 180°로 바꿔주기 위해 -90 ~ 90°까지의 값을 Fig. 2의 앞뒤 혼동현상에 의해 반대쪽에 그대로 값을 추가해주면 된다. 즉, 음원이 30° 존재할 경우에는 θ 가 30°와 150°일 때 상호상관값이 최대값을 가지게 되며, 결국 두 개의 각도 중 한 곳에 음원이 존재한다는 것을 알 수 있다.

Fig. 3은 60°에 음원이 존재할 때 각도에 따른 상호상관값을 나타낸 것이다. -90 ~ 90°까지 나타낸 것인데, 60°에 피크치가 두드러지게 나타나는 것을 확인할 수 있다.

3.2 3개의 마이크를 이용한 각도추정

3개의 마이크를 이용하여 각도를 추정할 경우 여러 가지 방법을 사용할 수 있다. 두 개의 마이크를 이용하여 대략적인 각도를 추정하였으며, 앞뒤 구분만 가능하면 되므로, 각 마이크 쌍에서 상호상관값을 계산했을 때 상호상관값의 피크치의 위치에 따라 대략적인 음원의 위치를 추정한 후, 임의의 마이크 쌍에서 음원의 각도를 추정할 수 있다. 본 논문에서는 Eq.(5)에서 상호상관값에 따른 지연시간 τ 를 Eq.(6)을 이용해서 각도 θ 로 1대1 매핑시킨 뒤, 각 마이크 쌍에서 계산한 θ 에 따른 상호상관값을 모두 합하였을 때 가장 큰 값을 나타내는 각도를 음원의 위치로 추정한다.

Fig. 4는 마이크 센서 3개를 정삼각형으로 배치한

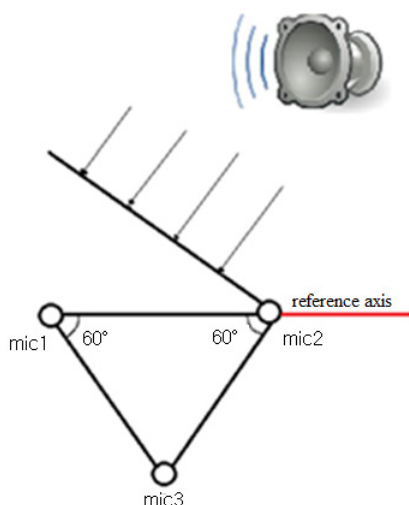


Fig. 4. Layout of 3 microphones and reference axis.

뒤, 기준축을 정한 것이다. 기준축을 정한 이유는 θ 에 따른 상호상관값을 모두 더하기 위해서다. 3개의 마이크 쌍이 존재하므로 각각의 기준축이 모두 다르면 더할 수 없으므로 기준축을 통일하게 된다. Fig. 4에서는 마이크를 정삼각형으로 배치하였으므로 기준축에 맞춰 회전시켜 각각의 θ 에 따른 상호상관값을 더해지게 되면 음원의 각도를 추정할 수 있다. 이때 τ 를 θ 로 매핑할 때는 τ 의 최대값과 최소값을 고려해야 한다. $d = c\tau$ 이고, 지연거리 $d \leq 2r$ 이므로

$$-\frac{2r}{c} \leq \tau \leq \frac{2r}{c} \tag{7}$$

가 성립한다. 여기서 τ 를 θ 로 매핑할 때 결국 $-\frac{2r}{c} = -90^\circ$ 가 되며 $\frac{2r}{c} = 90^\circ$ 가 된다. 이때 Eq.(7)에서 샘플링 주파수를 f_s 라고 하고, 지연 샘플을 n 이라고 하면 $\tau = \frac{n}{f_s}$ 의 관계가 성립한다. 따라서 지연시간이 연속적이지 않으므로 각도로 매핑할 때 가운데 없는 값들은 보간법을 이용하여 매핑할 수 있다.

Fig. 4에서 음성신호가 발생하였을 때 각 마이크 쌍으로부터 각도에 따른 상호상관값을 구한 것이 Fig. 5이다. Fig. 5를 보면 30°부근에 음원이 존재하여 세 개의 마이크 쌍에서 음원이 존재하는 각도를 가리키는 것을 볼 수 있다. 따라서 세 개의 그래프를 모두 더하면 음원이 있는 지점에서 가장 큰 값을 가지게 되어 음원이 있는 위치의 각도를 찾을 수 있다. 3

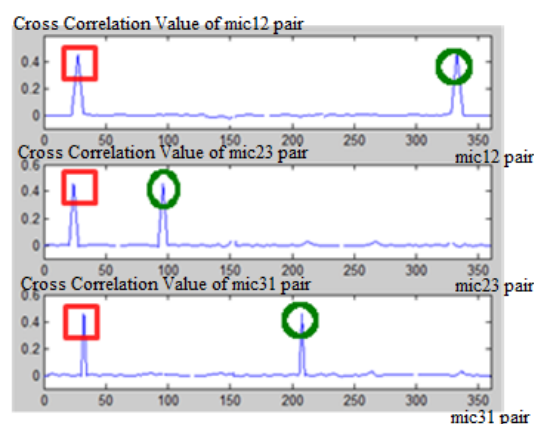


Fig. 5. 3 Relationships between cross correlation coefficient value (x-axis) and sound source angle (y-axis) with each microphone pair from Fig. 4.

개의 상호상관값 모두를 더해서 일정 임계치를 초과할 때만 음원이 존재하는 것으로 하고, 그렇지 않을 경우에는 음원이 존재하지 않는 것으로 가정한다. 그러면 원으로 그려진 곳은 상대적으로 작아지게 되어 음원이 존재하지 않는 것으로 판별하고, 네모가 그려진 곳은 임계치 이상이 되어 음원이 존재하는 것으로 판별할 수 있다. 이 방법을 이용하면 음원의 각도를 추정할 수 있으므로 네모가 그려진 피크치는 남겨두고, 앞뒤 혼동현상에 의해 원으로 그려진 곳의 피크치는 음원이 존재하는 곳이 아니므로 제거할 수 있다.

IV. 음원의 좌표 추정 방법

4.1 직선 방정식을 이용한 좌표 추정

Fig. 4에서 3가지 마이크 쌍의 기준축을 통일하여 음원이 존재하는 각도를 계산했다. 이때 계산한 각도가 두 마이크 사이의 중점으로 계산했다고 하면, 기울기를 알고 한 점을 알고 있으므로 하나의 직선 방정식을 만들 수 있다. 마이크가 3개라면 총 3개의 쌍이 존재하므로 이 중 2개의 쌍만을 이용하면 우리가 원하는 음원의 좌표를 계산할 수 있다. Fig. 6을 보면 마이크 1, 2의 쌍에서 하나의 직선이 음원과 두 마이크 사이를 통과하고, 마이크 2, 3 쌍에서 음원과 두 마이크 사이를 통과하고 있다. 따라서 이 두 직선의 교점을 계산하면 음원의 좌표를 계산할 수 있다.

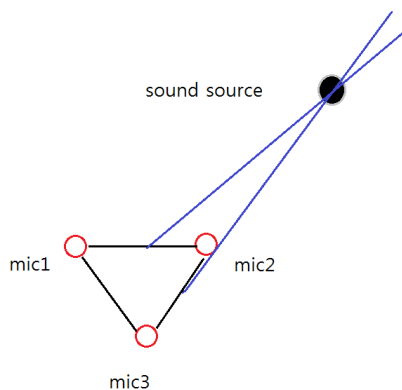


Fig. 6. Sound source coordinates estimation method.

4.2 사인 역함수

Eq.(6)에서 지연시간을 이용해 각도를 계산할 때

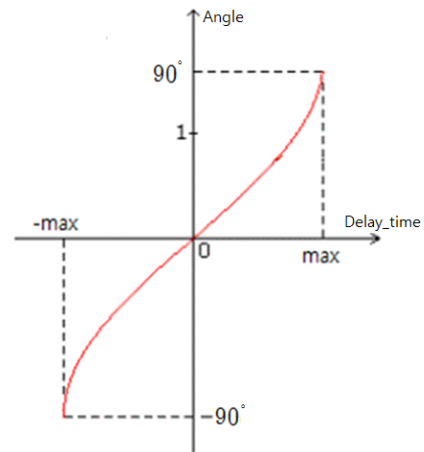


Fig. 7. Inverse sine function graph.

사인 역함수를 이용하는데 사인 역함수가 선형적이지 못한 특성으로 경우에 따라 각도 추정오차가 증폭될 수 있다. Fig. 7에서 가로축이 지연시간이고, 세로축은 각도에 해당한다. 0° 부근에서 완전한 선형은 아니지만 어느 정도 선형적인 특성을 갖는 것을 볼 수 있다. 그러나 각도가 90°나 -90°에 가까워지게 되면 그래프가 비선형적인 특성을 많이 띄게 된다. Fig. 7이 의미하는 것은 지연시간이 세로축의 각도에서 90°나 -90°에 해당하는 부분이면 지연시간에서의 작은 오차에도 추정 각도가 많이 변한다는 것을 뜻한다. 이것은 직선방정식의 교점을 이용해 좌표를 계산할 때 큰 오차를 야기할 수 있으며 이를 고려해 좌표를 추정해야 한다. 이에 대한 대처 방안은 다음 장에서 설명하도록 한다.

V. 공간 영역 분할

5.1 마이크가 3개일 때 공간영역분할

Eq.(6)에서 사인의 역함수를 이용해 직선방정식을 구하는데, 음원의 각도를 추정할 때 오차가 적어지도록 하기 위해 Fig. 7의 사인 역함수 그래프에서 최대한 0도에 가까운 부근에서 음원을 찾도록 한다. 일단 Fig. 8과 같이 마이크를 배치하고, 각도만 찾다고 하였을 때, 음원이 들어오는 대략적인 각도를 알 수 있다면 앞뒤 혼동현상에 의한 오류를 막을 수 있으므로 Fig. 8과 같이 공간을 6등분하여 각도를 추정할 수 있다. 즉, mic12라고 적힌 곳의 영역에서는

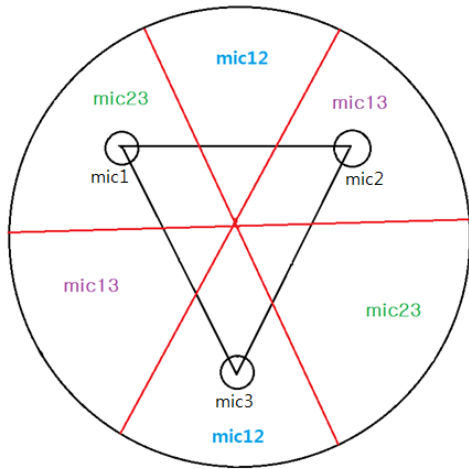


Fig. 8. Space area using 3 microphones (mic1, mic2, mic3) consists of 3 sections of mic12, mic23 and mic13, where mic12 represents microphone pair with mic1 and mic2.

mic12를 이용하여 구한 지연시간을 각도로 매핑했을 때 가장 선형적인 특성을 보이게 된다. 문제는 앞뒤 혼동현상에 의해 음원이 위에서 들어오는지 아래에서 들어오는지를 알 수 없다는 것이다. 그러나 Fig. 5와 같이 공간 좌표로 매핑하여 미리 음원의 각도를 구한 뒤, 앞뒤 혼동현상에 의해 생기는 피크치를 제거하면 음원이 들어오는 방향을 알 수 있다. 그러나 이렇게 음원이 들어오는 방향을 구했을 때는, Eq.(6)의 역함수로 매핑했을 때 오차가 큰 마이크 쌍의 상호 상관값이 존재하여 실제 음원의 각도와 차이를 보인다.

따라서 역함수로 매핑하여 대략적인 음원의 위치를 추정하면 Fig. 8에서 음원이 어느 공간영역에 위치해 있는지를 알 수 있다. 그 후 해당 공간영역에서 역함수로 매핑했을 때 가장 선형적인 특성을 보이

는 마이크 쌍에서 얻은 각도를 선택하면 앞뒤 혼동 현상도 없애고, 좀 더 신뢰할 수 있는 각도를 추정할 수 있다.

5.2 공간 영역 분할을 이용한 좌표 추정

mic23 쌍의 축에 인접한 부근에 존재하는 음원의 좌표를 찾을 때, mic12 쌍과 mic23 쌍을 이용하여 직선방정식을 만들어 음원의 좌표를 찾으면 mic23 쌍에서 보면 거의 -90°에 가깝기 때문에 mic23 쌍에 의한 오차가 커지게 된다. 이런 오차를 줄이기 위해서는 Fig. 6에서와 같이 mic12 쌍과 mic23 쌍을 이용하는 것이 아니라, Fig. 9와 같이 mic12 쌍과 mic31 쌍을 이용하면 오차를 훨씬 더 줄일 수 있다.

좌표를 추정하는 것이 아니라 각도만 추정한다면 Fig. 8과 같이 360°영역을 60°씩 분할하여 각각의 마이크 페어가 커버하는 공간을 -30 ~ 30°로 60°씩 나누면 되지만 이 방법은 교점이 생기지 않아 좌표를 찾을 수 없다. 따라서 교점을 만들기 위해서는 Fig. 8에서 영역을 60°보다 더 크게 나누어 교점이 생기게 하는 수밖에 없다. 즉, 각각의 마이크 쌍이 담당하는 공간이 -60 ~ 60°로 늘어나야 720°를 담당하게 되어 교점이 생겨 좌표를 계산할 수 있다. 하지만 담당해야 되는 영역이 넓어짐으로써 Eq.(6)을 이용하여 각도를 구할 때 생기는 오차가 커지는 것 역시 감안해야 한다. 다음 절에서는 이러한 오차를 줄이기 위해 마이크 4개를 이용한 영역분할을 제안한다.

5.3 마이크가 4개일 때 공간영역분할

마이크가 3개일 때는 담당해야 하는 공간이 -60 ~ 60°로 넓은 영역이 되어 오차가 크게 되는 지점들이 존재한다. 이 오차를 줄이기 위해서는 샘플링 주파수를 높여 해상도를 높이거나, 물리적으로 마이크의 개수를 늘려야 한다. 그러나 마이크의 개수를 늘리면 공간영역을 다시 분할해야 한다.

Fig. 10은 사각형의 중점을 기준으로 90°씩 영역을 나누는 것이다. 원점으로부터 mic1과 mic2까지 연결되어 있는 대각선 안의 영역에 들어오면 mic12를 이용하여 직선방정식을 만든다. 이와 마찬가지로 mic2와 mic3의 영역에 음원이 들어오면 가장 선형적인 특성을 띄며 오차가 적은 mic23을 이용하여 직선 방정식

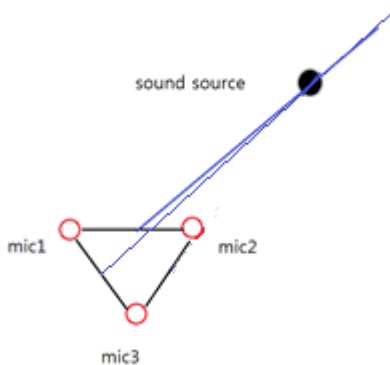


Fig. 9. Example of optimal coordinates estimation.

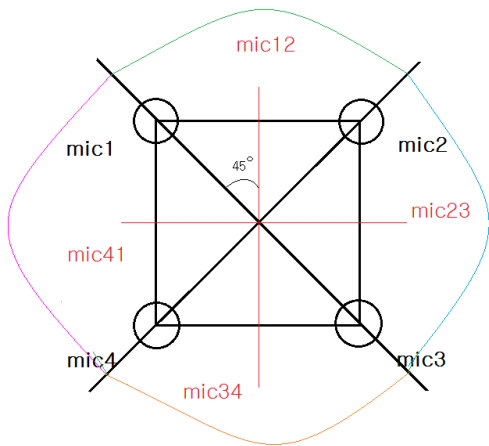


Fig. 10. Space area method 1 using 4 microphones (mic1, mic2, mic3 and mic4) consists of 4 sections (mic12, mic23, mic34 and mic41).

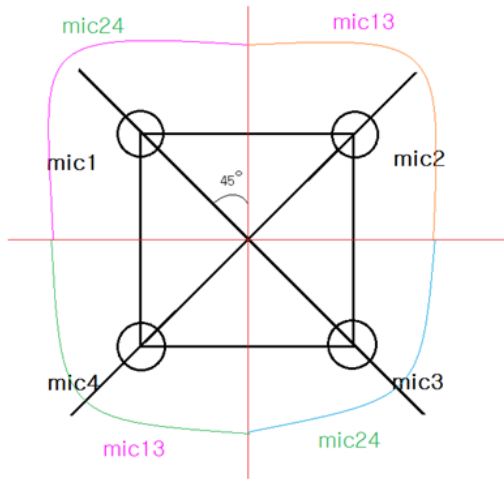


Fig. 11. Space area method 2 using 4 microphones (mic1, mic2, mic3 and mic4) consists of 2 sections (mic13 and mic24).

을 세운다. 이렇게 직선 하나를 이용하면 교점을 찾을 수 없기 때문에 Fig. 11과 같이 360° 영역을 더 나누는데, 이번에는 대각선으로 마주보는 마이크 즉, mic13, mic24를 이용하여 직선방정식을 만드는 것이다. Figs. 10과 11에서 각각의 곡선들은 마이크 페어가 90° 씩 담당하는 영역을 표시한 것이다.

이렇게 되면 각 쌍에서 -45 ~ 45°의 각도에서만 음원의 각도를 구하게 되어 -60 ~ 60°때보다 더 적은 오차를 가지고 음원의 위치를 찾을 수 있다. 이때 마이크가 4개이므로 마이크 쌍이 총 6개가 나오게 되는데, 마이크가 3개일 때 보다 계산량이 늘어나게 된

다. 하지만 본 논문에선 음원의 발생지점이 어느 영역에 속해 있는지만 찾으려 하기 때문에 4개의 마이크 중에서 임의의 마이크 3개를 골라 처리하면 음원이 위치한 영역을 알 수 있다.

VI. 실험 및 결과

6.1 4개 마이크를 이용한 좌표 추정

실험을 위해 마이크 4개를 Fig. 12와 같이 정사각형으로 20 cm씩 간격으로 배치하였다. 음원의 위치는 Fig. 12에서 mic1과 mic3의 중점을 원점 (0,0)으로 하였을 때 가로축으로 60 cm, 세로축으로 170 cm 떨어져 있다.

각각의 마이크와 음원은 지면으로부터 132 cm 상공에 있고, 특정 잡음은 존재하지 않으며 반향이 적은 넓은 강의실에서 실험을 진행하였다. 샘플링 주파수는 44.1 kHz이고, 16 bit로 양자화 하였다. 음원은 녹음된 사운드를 재생해서 사용했으며 스피커의 크기는 가로 10 cm, 세로 11 cm, 높이 10 cm이다. 또한 마이크는 ecm 8000 측정용 콘덴서 마이크를 사용하였고, 멀티채널을 이용하기 위해 외장 사운드카드 M_Audio Fasttrack c600을 이용하여 5 s 길이의 총 100 개의 데이터를 획득하였다. 음원 위치 추정 성공 여부는 추정된 좌표가 음원으로부터 직경 30 cm 반경 이내에 있으면 성공한 것으로 판단하였는데, 이는

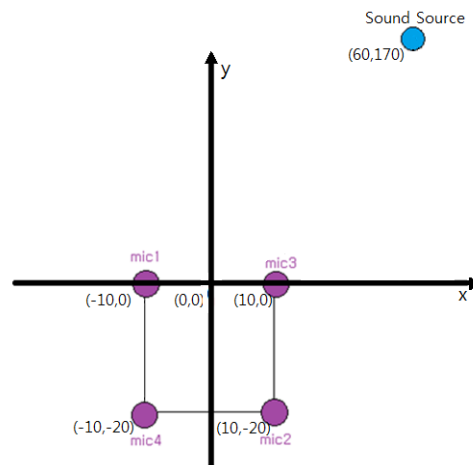


Fig. 12. 4 microphones placement and sound source position with coordinates 60 cm, 170 cm along with origin (0,0).

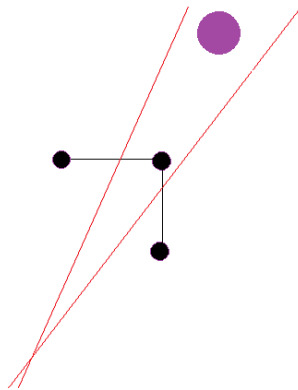


Fig. 13. Example of sound source estimation error due to the big size of sound source. Final estimated position is occurred at the opposite side with respect to the real source position.

음원으로부터 마이크의 중점까지의 거리가 약 180 cm를 조금 넘는데, 음원과 마이크 사이의 거리의 20%를 넘지 않는 범위로 설정한 것이다. 이때 범위는 사용자가 어느 정도를 오차의 범위로 정하는 지에 따라 달라진다. 본 논문에서는 20%를 넘지 않는 30 cm로 하였다. 사용한 음성신호는 직접 녹음한 데이터를 사용하였으며 14 s의 음성을 랜덤하게 재생시켜 데이터를 획득하였다.

데이터를 획득할 때 고려해야 할 사항 중 음원의 크기가 문제될 수 있다. Fig. 13은 마이크 쌍을 이용하여 좌표를 추정할 때 생길 수 있는 오차를 나타낸다. 음원의 크기가 작을수록, 즉 점에 가까울수록 마이크 페어가 정확하게 음원의 좌표를 찾아낼 수 있고, 음원의 크기가 커지면 커질수록 음원이 공간상에서 차지하는 면적이 넓어지기 때문에 정확하게 좌표를 찾아내기 어렵다. 만일 음원이 Fig. 13과 같이 큰 경우에 좌표를 추정하면 직선방정식의 교점이 정 반대쪽에 나타나는 경우가 생길 수 있다. 즉, 각도는 맞게 찾아내지만 좌표는 전혀 엉뚱한 지점에 생기는 경우이다. 또한 실제 환경에서는 음원의 크기 외에 여러 가지 측정오류 사항들이 존재한다. 음원과 마이크가 직선거리상에 존재하지 않아서 생기는 오차도 있고, 또한 온도차에 따른 음속도의 변화, 외부의 잡음으로 인해 신호가 왜곡되는 경우, 마이크의 주파수특성으로 인한 신호의 변형 등에 의한 오차가 생기기도 한다.

Fig. 14는 Fig.10과 Fig.11을 이용하였을 때 공간을

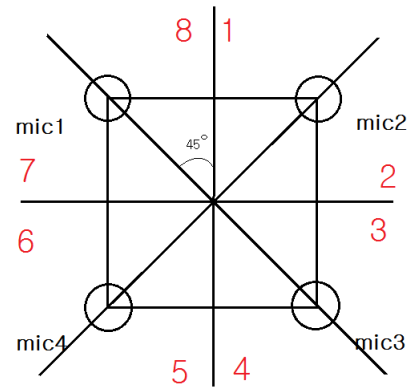


Fig. 14. Space area way with 8 sections.

Table 1. The experiment result.

Space number	1	5
Space distribution rate (%)	88	12

Table 2. Area 5 coordinate distribution.

coordinate	(-25,-92)	(-22,-93)	(-19,-98)	(-10,-42)
distribution rate (%)	25	50	17	8

Table 3. Area 1 distribution.

	success	failure
Recognition rate (%)	76	24

분할하고, 해당 공간을 구분하기 위해 각각의 공간에 번호를 붙인 것이다. 예시 만약 음원이 1번 영역에 있는데, 추정 결과 음원 위치를 다른 엉뚱한 영역으로 좌표를 추정할 경우 1번 영역으로 영역을 특정하였으므로 잘못된 데이터라는 것을 쉽게 판단할 수 있다.

Table 1은 100개의 데이터에 대한 실험 결과를 보여주고 있는데 88개는 1번 영역에 있다고 맞게 판정하지만 12개는 반대 방향인 5번 영역으로 잘못 추정하고 있음을 알 수 있다.

Table 1에서 1번 영역을 벗어난 좌표들의 분포 상황을 Table 2에서 보여준다. Fig. 13에서 예상했던 바와 같이 좌표가 1번 영역의 반대편인 5번 영역에 나타나는 것을 볼 수 있다.

Table 3은 Table 1에서 5번 영역에 생기는 경우를 뺀 나머지 88개 데이터의 위치 추정 결과를 나타낸 것이다. 실제 음원으로부터 30 cm 이내에 있을 경우

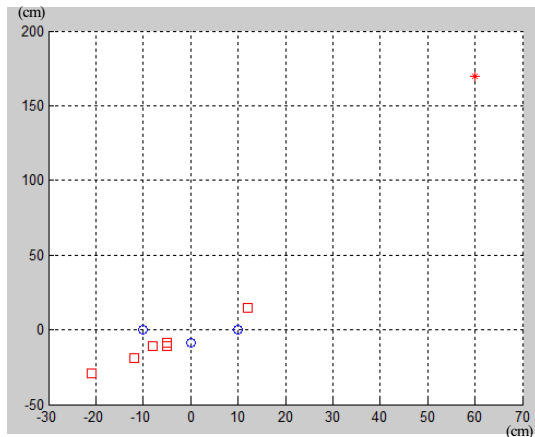


Fig. 15. Results of sound source coordinate estimation using 3 microphones. Asterisked point and circle point represent sound source position and microphone's position respectively. Square point represents position where estimation error is occurred.

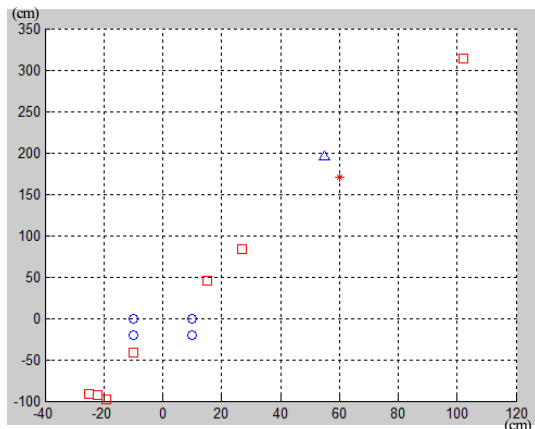


Fig. 16. Results of sound source coordinate estimation using 4 microphones. Asterisked point and circle point represent sound source position and microphone's position respectively. Square point represents estimation error position whereas triangle point means position that is estimated correctly.

에 성공이라고 했을 때, 성공률은 76%이고, 그 외 1번 영역의 엉뚱한 지점을 추정하는 경우가 24%임을 확인하였다.

Fig. 15는 3개의 마이크를 이용한 실험에서 추정된 좌표들의 분포를 나타내고, Fig. 16은 4개 마이크를 이용했을 때 좌표 분포를 나타낸다. 단위는 cm이며 실제 음원과 마이크의 좌표, 추정좌표를 나타낸다. 동그라미 표시는 마이크 위치를 나타내고, 별(*) 표시는 실제 음원의 좌표를, 네모 표시는 음원의 위치

를 잘못 추정했을 때, 세모 표시는 허용 오차 범위내로 추정했을 때를 나타낸다. 실험 결과로 마이크가 3개일 때는 영역 해상도가 낮고, 사인역함수에 의한 오차가 커서 음원의 위치를 허용 범위 내에서 찾지 못하였지만, 4개일 때는 해상도가 높고, 역함수를 이용할 때 오차가 적어 음원의 위치를 추정할 수 있음을 확인하였다.

VII. 결 론

본 논문에서는 마이크 센서를 이용하여 음원의 각도를 찾을 때 이용하는 사인역함수의 비선형성에 의한 오차를 줄이기 위해 공간분할을 이용하여 음원 좌표를 구하는 방법을 제안하였다. 음원의 좌표를 찾는 방법은 먼저 GCC-PHAT 알고리즘을 이용하여 지연시간을 계산한 뒤, 각각의 마이크 쌍으로부터 계산되는 각도를 이용하여 음원이 존재하는 영역을 찾는다. 다음에 앞에서 계산한 각도를 이용하여 직선 방정식을 구하고 교점을 계산하여 음원의 좌표를 추정한다. 이 과정에서 사인의 역함수를 이용하는데 사인 역함수의 비선형적인 구간에서 오차가 커질 수 있다. 따라서 각각의 마이크 쌍 중 직선방정식을 만들었을 때 사인 역함수가 가장 선형적인 특성을 띠는 쌍을 선택하기 위해 마이크가 3개일 때와 4개일 때 각각의 영역을 분할하는 방법을 제안하였다. 실험 결과에 의하면 마이크가 3개일 때는 담당해야 하는 각도가 넓어 오차가 심해 음원 좌표 추정 성능이 거의 나오지 않지만, 마이크가 4개일 때는 해상도가 높아지고, 담당해야 하는 각도가 줄어들어 67%의 성공률을 보였다. 따라서 마이크의 개수가 증가할수록 추정오차가 줄어드는 것을 확인하였고, 마이크가 6개, 8개...로 더 늘어난다면 오차는 더욱 작아질 것으로 예상하고 있지만 현재 사용하고 있는 M_Audio Fasttrack c600은 4채널 외장 사운드카드로 최대 4개의 마이크 채널까지만 지원하기 때문에, 그 이상의 마이크 사용에 의한 성능 비교는 추후 진행할 계획이다. 또한 같은 영역 안에 존재하더라도 마이크가 사각형의 형태로 배치되어 있기 때문에 위치에 따라 성능이 달라지게 된다. 이는 마이크의 수를 늘려서 배치형태를 바꾸면서 마이크의 배치 형태에 따른 중

합적인 성능 분석을 통해 추가적인 연구를 진행할 것이다.

References

1. W. Rhee and J. S. Choi, "Experimental studies for noise source positioning using TDOA algorithm" (in Korean), *KSNVE. Suppl.1(s)* **5**, 108-112 (2006).
2. B. Jang, D. Y. Sim, C. D. Kim, C. B. Lee, and K. H. Cga, "A study for estimating absolute position of sound source using reference microphone and beamforming method" (in Korean), *KSPC. Suppl.1(s)* **10**, 371-374 (1997).
3. S. M. Hwang, *Sound source localization using HRTF database*, (Master's thesis, KAIST, 2005).
4. Y. J. Park, M. W. Lee, S. H. Min, and Y. H. Han, "Sound localization ased on LP residual and hilbert transform for intelligent robot" (in Korean), *KIISS*, **10**, 382-385 (2008).
5. J. S. Koo and H. J. Park, *Sound source localization based on PCA* (in Korean), (Master's thesis, University of Kwangwoon, 2014).
6. Y. E. Kim, S. H. Cho, and J. G. Chung, "Efficient sound source localization system using angle division" (in Korean), *KSPC*, **46**, 114-119 (2009).

저자 약력

▶ 김 환 용 (Hwan-Yong Kim)



2014년~2월: 대진대학교 전자공학과 학사
 2016년~8월: 대진대학교 전자공학과 석사
 졸업
 2016년 6월 ~ 현재: Engis Technology

▶ 최 흥 섭 (Hong-Sub Choi)



1985년~2월: 서울대학교 전자공학과
 공학사
 1987년~2월: 서울대학교 전자공학과
 공학석사
 1994년~2월: 서울대학교 전자공학과
 공학박사
 1995년~현재: 대진대학교 공과대학
 전자공학과 교수