

마이크로폰의 이득 특성에 강인한 위치 추적

최지성, 이지연, 정상배, 한민수
한국정보통신대학교 음성음향정보연구실

A new sound source localization method robust to microphones' gain

Jisung Choi, Jiyeoun Lee, Sangbae Jeong, Minsoo Hahn
Speech and Audio Information Lab, Information and Communication Univ.

dir81@icu.ac.kr, jyle278@icu.ac.kr
sangbae@icu.ac.kr, mshahn@icu.ac.kr

Abstract

This paper suggests an algorithm that can estimate the direction of the sound source with three microphones arranged on a circle. The algorithm is robust to microphones' gains because it uses only the time differences between microphones. To make this possible, a cost function which normalizes the microphone's gains is utilized and a procedure to detect the rough position of the sound source is also proposed. Through our experiments, we obtained significant performance improvement compared with the energy-based localizer.

I. 서론

위치 추적이란 음원이 발생한 곳을 찾는 것을 의미하며 지능형 로봇 또는 보안 시스템에 응용될 수 있다. 지능형 로봇이 실생활에 활용되기 위해서는 주위 상황을 인지하고 판단하는 능력이 필요하다. 따라서 음원이 발생한 곳을 찾는 위치 추적은 로봇으로 하여금, 로봇이 도움을 필요로 하는 사람이나 개체 근처로 이동하여 주위 상황을 판단하고 적절한 대응조치를 취할 수 있도록 해 준다.

과거의 음원 추적 시스템은 주로 노이즈가 없는 환경에서 두 개 이하의 마이크로폰을 사용하여 가까운 거리의 음원을 찾아내도록 연구되어 왔다. 최근에는 3개 이

상의 마이크로폰을 이용하여 사용자의 호출음이나 박수 소리를 입력 받아 인식함으로써 실험용 또는 장난감 로봇, 카메라나 키보드와 같은 입력 장치의 대안으로 활용할 수 있도록 하는 자동 인터페이스 구현을 위한 연구가 진행되고 있다.

기존의 위치 추적 알고리즘은 마이크로폰 별 입력음의 강도의 차이(Intensity Difference)를 이용하여 위치를 측정하였다[1]. 강도 차이 방법은 음원이 발생한 부분에서 에너지가 크게 발생하므로 이를 이용하여 음원 위치를 찾는 것이다. 그러나 많은 경우 각각의 마이크로폰은 서로 다른 특성을 가지고 있다. 따라서 동일한 환경 조건에서 같은 입력을 주었을 경우 각각의 마이크로폰의 특성에 따라서 다른 강도를 가지므로 음원 발생구간이 다르게 찾아지는 경우가 발생하게 된다. 이는 강도 차이에 의한 음원추적을 이용할 경우 큰 오차를 내는 원인이 된다. 따라서 본 논문에서는, 마이크로폰의 이득 특성에 강인한 위치 추적 알고리즘을 제안하였다. 마이크로폰의 이득에 독립적이기 위해서 시간 차이(Time Difference)를 이용한 위치 추적 시스템에서는 각각의 마이크로폰들로 들어온 신호 앞에서 입력된 신호의 크기를 보상해준다.

본 논문은, 1장의 서론에 이어 2장에서는 음원 추적 알고리즘의 개략적인 방법에 대해 언급하고, 3장에서는 제안된 음원 추적 알고리즘에 대해 설명하고 4장에서는 실험 및 결과를 검토한 후, 마지막으로 5장에서는 결론 및 향후 연구계획에 대해 논한다.

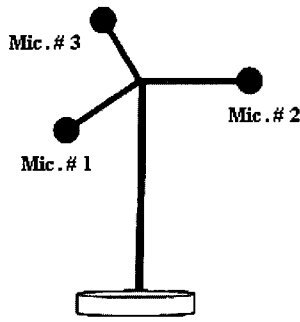


그림 1. 3채널 위치·추적 시스템

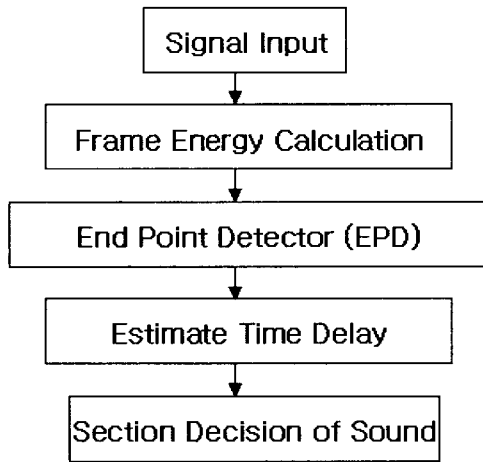


그림 2. 위치 추적 알고리즘 순서도

II. 위치 추적 알고리즘

그림 1은 3개의 마이크로폰을 이용한 위치 추적 시스템을 보여준다. 3개의 마이크로폰은 고도 40cm위에 위치하고 있고 마이크 사이의 간격은 120°이며 원의 반지름은 14.5cm이다. 3개의 마이크로폰에서 음성신호를 입력받은 다음, 신호의 에너지를 계산한다.

$$E_n = \sum_{m=n-N+1}^n x^2(m) \quad (1)$$

에너지를 계산한 후, EPD에서 에너지 정보를 이용하여 신호의 시작점과 끝점을 검출한다. 그런 다음, 각 마이크로폰 간의 시간 지연을 계산하게 되는데 마이크로폰들의 이득에 독립적인 상호상관(cross correlation)방법을 사용하여 최소 지연시간을 계산한다. 이렇게 구한 지연 시간 정보를 이용하여 음원이 발생한 곳을 찾는 것이다.

III. 제안한 알고리즘

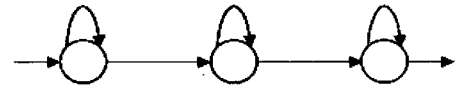


그림 3. EPD system

3.1 End Point Detector (EPD)

음원의 시작점과 끝점을 검출하기 위해서는 먼저 음원의 에너지를 식(1)을 이용하여 계산한다. 에너지는 매 10 msec 단위로 추출되며 분석 구간의 길이는 30 msec 이다. 처음 100 msec까지는 그림 3과 같이 배경 잡음만 존재하는 구간이라고 설정하고 음성구간 검출을 위한 에너지의 문턱값을 구한다. 음성구간의 시작 및 끝을 검출하기 위해서 현재 분석 구간 전 후 각 5 프레임씩 총 10프레임의 에너지가 고려되는데, 10개 중 5개 이상이 정해진 에너지의 문턱값을 넘으면 음성이 시작되었다고 간주하고 5개 이하이면 음성이 끝났다고 판단한다. 이러한 방법을 사용할 경우에 음성이 아님에도 불구하고 큰 에너지를 갖는 짧은 구간의 간섭신호에 강인하게 음성구간을 검출할 수 있다는 장점이 있다.

3.2 Estimate Time Delay

음원이 발생한 구간을 결정하기 위해서 두 신호 사이의 지연 시간을 알아야한다. 지연시간의 계산 방법은 오래 전부터 많이 연구되어왔으며 상호 상관방법이 가장 보편적으로 쓰인다. 다른 방법들과 비교하여 계산적으로 간단하고 능률적이기 때문이다[2]. 본 논문에서는 상호상관방법에서 $J(\tau)$ 를 최소화하는 τ 값을 찾게 된다. 그 때의 τ 값이 지연시간이 된다. 또한 식(2)에서와 같이 α 를 곱해주어 들어오는 신호의 모양을 보고 크기 차이를 보정해 준다. 따라서 마이크로폰의 이득 특성에 강인한 각 마이크로폰 간의 지연 시간을 계산할 수 있다. 식(3)과 같이 0이 될 때의 α 값이 식(4)와 같이 구하여 신호를 보정하게 된다.

$$J(\tau) = \sum_n (\alpha x_1(n-\tau) - x_2(n))^2 \quad (2)$$

$$\frac{\partial J(\tau)}{\partial \alpha} = 2 \sum_n (\alpha x_1(n-\tau) - x_2(n)) x_1(n-\tau) = 0 \quad (3)$$

$$\alpha^* = \frac{\sum_n (x_1(n-\tau) - x_2(n))}{\sum_n x_1(n-\tau)^2} \quad (4)$$

3.3 Section Decision of Sound

발생한 음원의 구간 설정은 식(2)의 시간차 측정에 의해 결정되었다. 마이크로폰 1번과 2번 사이의 시간차는 Delay 12로 표현되며, 2번과 3번 사이는 Delay 23, 1번과 3번 사이는 Delay 13으로 표현된다. 1번 마이크로폰을 0°로 기준을 잡으면 마이크로폰 1번과 2번 사이, 즉 0°~120° 사이에서 음원이 발생하면 채널(1, 2)로 구간으로 설정하게 되고, 마이크로폰 2번과 3번 사이, 즉 120°~240° 사이에서 음원이 발생하면 채널(2, 3)으로, 마이크로폰 3번과 1번 사이, 즉 240°~360° 사이에서 음원이 발생하였을 때는 채널(1, 3)으로 음원이 발생한 구간으로 잡게 된다. 이렇게 구간을 잡았을 때의 문제점은 경계점인 마이크로폰 1번 앞(0°), 2번 앞(120°), 3번 앞(240°)에서 음원이 발생하였을 때, 음원이 발생구간을 채널(1, 2), 채널(2, 3), 채널(1, 3)으로 구간을 설정하게 되므로 정확성이 떨어진다. 이를 보완하기 위해서 경계점에서 음원이 발생하였을 때에 마이크로폰에 도달하는 시간차가 0이면 그림 6와 같이 경계점 구간을 잡도록 하였다.

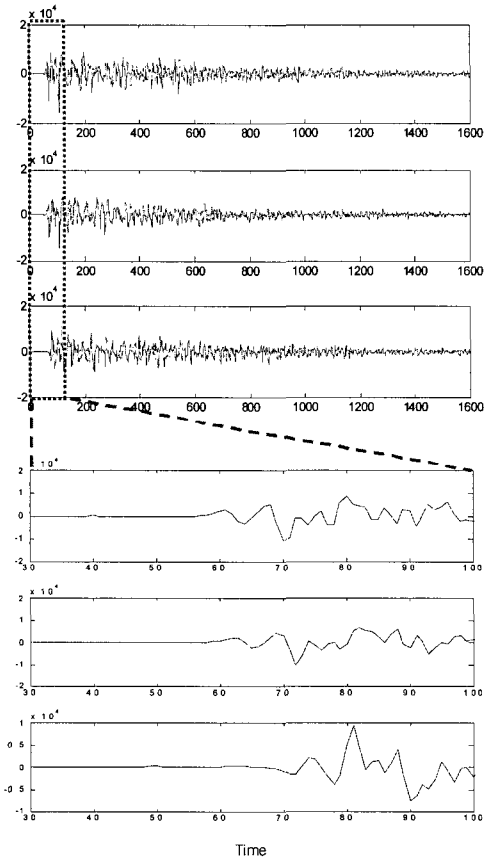


그림 4. 각 마이크로폰에서 획득된 음원 파형 지연 시간의 예 (60°).

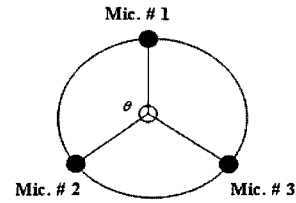


그림 5. 음원 구간 설정 방법

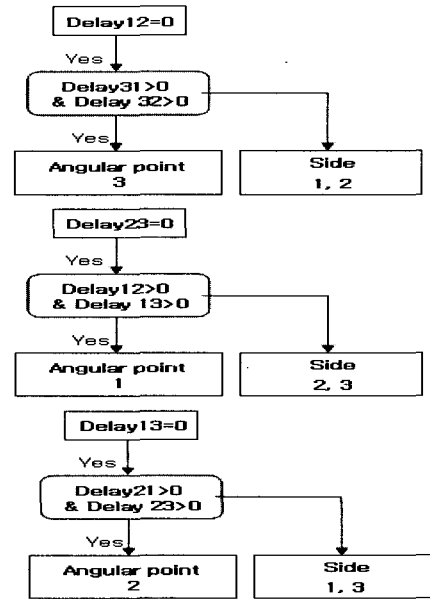


그림6. 경계점 구간 결정 순서도

IV. 실험 및 결과

실험환경은 컴퓨터가 존재하는 조용한 실험실 환경에서 행해졌다. 음원은 박수소리였으며 150cm떨어진 곳에서 발생되었으며, 16 kHz로 샘플링 되고 16 bit로 양자화 되었다. 3개의 마이크로폰을 이용하여 120° 간격으로 각각의 마이크로폰이 존재하며 마이크로폰 1번을 0°로 기준으로 해서 10° 간격으로 마이크로폰의 각도를 변경하여 36개의 데이터로 실험하였다. 그 결과 값은 표1이다. 똑같은 데이터를 가지고 음원을 발생시켜 강도 차이(Time Difference)에 의한 음원 발생 구간을 선택할 경우 각각의 마이크로폰의 이득이 다르기 때문에 음원 발생 구간을 잘못 선택하는 경우가 있다. 표2는 마이크로폰의 이득이 조정되지 않았을 때의 결과 값이다. 본 마이크로폰에서 3번 마이크로폰의 이득이 다른 1번과 2번의 마이크로폰의 이득보다 크기 때문에 구간을 잘못 선택하였음을 볼 수 있다. 그러나 마이크로폰의 이득에 강인한 위치추적을 하였을 경우, 경계점 부근에서 미세한 차이로 인하여 에러가 발생하는 경우

를 제외하고는 마이크로폰의 이득에 상관없이 정확한 구간을 선택하는 것을 알 수 있다. 표3은 표1과 2의 실험을 3번 반복하여 정확도를 측정한 결과이다.

표1. 마이크로폰의 이득에 강인한 위치추적 결과

Degree	Delay12	Delay23	Delay13	Side
0°	11	0	11	1
30°	9	4	13	12
60°	2	9	11	12
90°	-3	12	21	12
120°	-11	11	0	2
150°	-12	6	-6	23
180°	-10	-2	-12	23
210°	-5	-7	-12	23
240°	2	-12	-9	13
270°	7	-12	-5	13
300°	10	-11	-1	13
330°	12	-7	6	13

표 2. 강도차이를 이용한 위치 추적

Degree	Intensity (1,2)	Intensity (2,3)	Intensity (1,3)	Side
0°	49.083	50.732	52.647	32
30°	51.155	51.333	59.847	32
60°	51.315	53.873	55.8488	32
90°	50.324	52.700	56.861	32
120°	93.592	95.236	98.152	13
150°	54.238	50.602	52.192	13
180°	51.059	49.822	54.481	13
210°	93.765	94.686	98.771	32
240°	56.751	52.748	58.389	13
270°	52.560	50.035	56.448	13
300°	50.710	53.617	54.414	32
330°	53.764	56.355	59.130	32

표3. 마이크로폰의 이득에 강인한 위치 추적과 강도 차이를 이용한 위치 추적의 성공률 비교

	성공률
마이크로폰의 이득에 강인한 위치추적	94.5%
강도 차이를 이용한 위치추적	25%

V. 결론 및 향후 연구 계획

본 논문에서는 마이크마다 이득이 다를 경우, 즉 교정이 제대로 되지 않은 멀티마이크로폰 기반 시스템에도 적용이 가능한 위치 추적 방법을 제안하였다. 이를 위해서 마이크의 이득에 독립적인 신호간의 시간 지연

추정 방법을 사용하였고 그를 이용한 효과적인 위치 검출 방법을 정립하였다.

마이크로폰의 이득에 강인한 위치 추적 알고리즘은 채널 선택에 있어서 정확한 결과를 보였다. 구간의 경계점에서도 상당히 정확한 결과를 보였다. 마이크로폰의 이득에 민감한 강도 차이를 이용하여 음원의 위치를 추적하는 경우보다 훨씬 뛰어난 결과 값을 갖는 것을 볼 수 있다. 그러나 경계점 부근에서 미세한 차이로 인하여 에러가 발생하는 것을 볼 수 있는데 이는 시간 차의 해상도를 높이거나 반향을 고려한 위치 추적을 한다면 해결될 것이다. 또한 음원의 높이와 실험 환경이 음원 위치를 판단하는데 큰 영향을 미친다는 연구 보고가 된 바와 같이 잔향 환경과 음원의 고도를 고려하여 개선한다면 좀 더 나은 성능을 기대할 수 있을 것이다 [3]. 향후 4개의 마이크로폰을 사용하여 고도 정보가 포함된 3-D 음원위치 추적 기술을 향후 연구할 예정이다.

본 논문에서 연구하는 위치 추적 알고리즘은 로봇에 적용될 수 있다. 적용될 수 있는 시나리오로는 공공 서비스 대상자가 음성이나 어떤 소리로 로봇을 호출 할 경우 호출자를 찾아가기, 청각 장애자에게 로봇이 어떤 비상벨이나 비정상적인 음향을 감지하여 전달하고 대응 조치하기 등을 들 수 있다.

위치 추적 기술은 인간과 로봇의 상호작용에 유용할 것이며, 3-D 위치 추적의 기반이 되는 역할을 수행할 것이다.

참고문헌

- [1] 이지연, 한민수, "지능형 로봇 웨버"를 위한 음원추적 기술," *대한음성학회 가을 학술대회 발표 논문집*, pp117-120, Nov.2005
- [2] C.H knapp and G.C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust. Speech Signal Process.*, vol 24, no4, pp320-327, Aug. 1976.
- [3] Huang, J., Supaongprapa, T., Terakura, I., Ohnishi, N., and Sugie, N., "Mobile Robot and Sound Localization," *Proceedings of the 1997 IEEE/RSJ International Conference on IROS '97*, Volume: 2, p.7-11 Sept. 1997