호칭 기반을 위한 위치 추적

최지성, 이지연, 정상배, 한민수
한국정보통신대학교 음성음향정보연구실

Sound Source Localization for Spoken name

Jisung Choi, Ji-Yeoun Lee, Sangbae Jeong, Minsoo Hahn
Speech and Audio Information Lab., Information and Communications Univ.
{dir81, jyle278, sangbae, mshahn} @icu.ac.kr

Abstract: 위치 추적 시스템은 음원이 박수소리와 같은 임펄스 형태의 신호보다 호칭기반의 음성 신호일 때 성능이 감소한다. 호칭 기반의 위치 추적은 가정용 로봇에 친근함을 주기 위한 필요하다. 본 논문에서는 3 개의 마이크로폰을 이용하여 호칭 기반을 위한 최적의 위치 추적 알고리즘을 제시한다. 호칭기반의 위치 추적을 위해 시간 영역과 주파수 영역에서 위치 추적 방법을 비교한다. 본 논문에서는 마이크로폰의 이득 특성에 강인한 위치 추적 방법과 계차(difference) 기반의 위치 추적 방법 그리고 주파수 영역에서 GCC-PHAT(generalized cross correlation phase transform)기반의 위치 추적 방법을 사용하여 음원이 음성 신호일 때 최적의 알고리즘을 찾는다. 본 호칭 기반을 위한 위치 추적 시스템은 실시간으로 작동되는 가정용 로봇에 적용될 수 있다.

Keywords: 호칭 기반을 위한 위치 추적, 마이크로폰의 이득 특성에 강인한 위치 추적, 계차 기반, GCC-PHAT 기반

Ⅰ. 서론

위치 추적이란 음원이 발생한 곳을 찾는 것을 의미한다. 위치 추적은 지능형 로봇, 화상 회의 시스템, 음성 인식 시스템 등에서 사용하는 중요한 기술이다[1]. 지능형 로봇은 위치 추적 시스템을 이용하여 공공장소나 가정에서 주위 상황을 인지하고 판단하여 도움을 필요로 하는 근처로 이동하여 적절한 대응조치를 취할 수 있도록 해 준다. 화상회의 시스템에서의 위치 추적 시스템은 말하는 화자에게 자동으로 초점을 맞춘다[2].

과거의 음원 추적 시스템은 주로 노이즈가 없는 환경에서 두 개의 마이크로폰을 사용하여 가까운 거리의 음원을 찾도록 연구되어 왔다. 최근에는 3개 이상의 마이크로폰을 이용하여 신호를 입력 받아 인식함으로써 실험용 또는 장난감 로봇, 카메라나 키보드와 같은 입력 장치의 대안으로 활용할 수 있도록 하는 자동 인터페이스 구현을 위한 연구가 진행되고 있다.

위치 추적 기술은 일반적으로 세가지로 분류할 수 있다. 첫 번째는 강도 차이(intensity difference)를 이용한 방법, 두 번째는 도착 지연 시간(time delay of arrival) 이용한 방법, 그리고 세 번째는 빔포밍(beam-forming)을 이용한 방법이 있다[3][4]. 이 중 도착 지연 시간을 이용한 방법은 다른 방법들에 비해 계산의 간단하고 정확성이 높기 때문에 가장 널리 쓰이고 있다[5].

기존의 위치 추적 시스템에서는 박수소리 같은 임펄스 형태의 신호를 입력 받아 시간 영역에서 지연 시간에 기반하여 마이크로폰의 이득 특성에 강인한 위치 추적을 수행하였다[6]. 그러나 마이크로폰의 이득 특성에 강인한 위치 추적 방법은 박수 같은 임펄스 형태에서는 $\pm 7^{\circ}$ 의 정확한 결과를 보이나 호칭기반의 음성 신호일 때는 성능이 감소 한다. 지능형 서비스 로봇에서는 로봇에 친근함을 주기 위해 호칭 기반의 위치 추적이 필수적이다. 따라서 본 논문에서는 호칭 기반의 위치 추적을 위해 시간 영역에서 마이크로폰의 이득 특성에 강인한 위치 추적과 계차(difference) 기반의 위치 추적방법 그리고 주파수 영역에서 GCC-PHAT (generalized cross correlation phase transform)기반의 위치 추적 방법을 비교하여 음원이 음성 신호일 때 최적의 위치 추적 알고리즘을 찾는다.

본 논문은 1장의 서론에 이어 2장에서는 개략적인 호칭 기반을 위치 추적 알고리즘의 내용에 대해 언급하고, 3장에서는 호칭 기반을 위한 3가지 알고리즘들에 대해 각각 설명한 후, 4장에서는 실험 및 결과를 검토 후 호칭 기반을 위한 최적의 알고리즘을 찾고, 5장에서 결론 및 향후 연구계획에 대해 논한다.

Ⅱ. 위치 추적 알고리즘

그림 1은 시간 지연에 기반한 위치 추적 알고리즘의 순서도이다. 먼저 들어온 신호의 에너지를 이용하여 신호의 음성 구간을 검출한 후, 호칭 기반을 위한 음원 추적을 위해서 3가지 방법의 알고리즘을 각각 사용한다. 첫 번째로 마이크로폰의 이득 특성에 강한 위치 추적 알고리즘을, 두 번째는 계차 기반의 위치 추적 알고리즘을 사용하여 시간 영역에서의 신호간의 지연 시간을 측정하고 마지막으로 GCC-PHAT기반의 위치 추적 알고리즘을 사용하여 주파수 영역에서 신호간의 지연 시간을 구한다. 이렇게 각각의 알고리즘에서 구한 지연 시간 정보를 이용하여 음원의 발생한 각도를 측정한다.

1. 음원 검출

음원의 시작점과 끝점을 검출하기 위해서는 먼저 음원의 에너지를 식(1)을 이용하여 계산한다[7].

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

식(1)은 N개의 분석 구간에 대해 분석구간 시작부터 n 번째까지의 샘플에 대한 에너지를 구하는 것이다. 에너

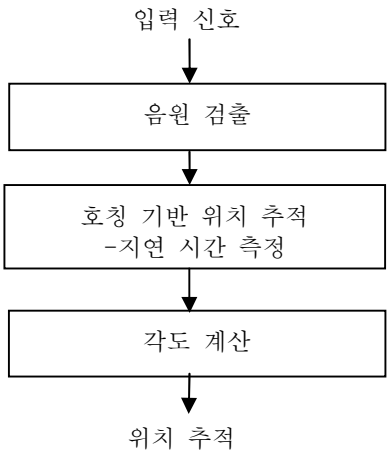

그림 1. 위치 추적 시스템 순서도

지는 10 msec 단위로 추출되며 분석 구간의 길이는 30 msec 이다. 음원구간의 시작 및 끝을 검출하기 위해서, 처음 100 msec까지는 잡음만 존재하는 구간이라고 설정하고 에너지의 문턱 값을 구한다. 그리고 현재 분석 구간 전 후 각 5 프레임씩 총 10프레임의 에너지가 고려되는데, 10개 중 5개 이상이 정해진 에너지의 문턱 값을 넘으면 음원이 시작되었다고 간주하고 5개 이하이면 음원이 끝났다고 판단한다. 이러한 방법을 사용할 경우, 음원이 아님에도 불구하고 큰 에너지를 갖는 짧은 구간의 간섭신호에 강인하게 음원구간을 검출할 수 있다는 장점이 있다.

2. 각도 계산

3개의 마이크로폰은 각각 120° 간격으로 원 안에 위치 한다. 음원이 위치한 각도의 계산을 위해서 먼저 마이크로폰에 들어오는 음원의 파장은 평면파라고 가정한다. 그림 2에서 원으로 들어오는 음원의 각도를 θ라고 하면 θ는 식 (5)에 의해 구해진다.

$$\theta = \cos^{-1}\left(\frac{\Delta t * \upsilon}{\ell * S}\right) - 30^{\circ} \qquad (2)$$

Δt는 식(2)를 통해 구해진 두 개의 마이크로폰 사이의 시간 지연이며 υ는 음원의 속도이고 ℓ은 두 개의 마이크로폰 사이의 거리이며 S는 음원의 샘플링율이다.

Ⅲ. 호칭 기반 위치 추적을 위한 지연 시간 측정

1. 마이크로폰의 이득 특성에 강인한 위치 추적 알고리즘

음원이 발생한 각도를 측정하기 위해서 두 신호 사이의 지연시간을 알아야 한다. 지연시간의 계산 방법은 오래 전부터 많이 연구되어왔으며 상호 상관도를 이용한 방법이 가장 보편적으로 쓰인다. 그 이유는 다른 방법들과 비교하여 계산적으로 간단하고 능률적이기 때문이다.

도착 시간 지연에 기반한 위치 추적 방법은 강도 차이

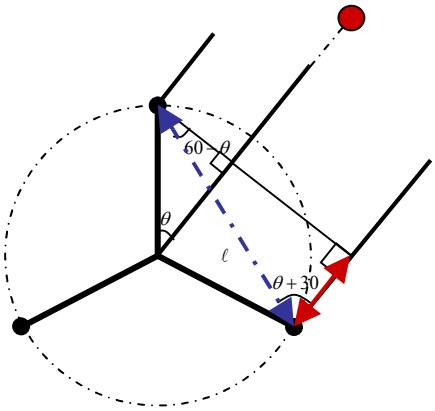

그림 2. 각도 계산

에 기반한 위치 추적과 마찬가지로 마이크로폰의 이득 특성에 따라서 음원의 각도가 잘못 측정 되는 문제점을 가지고 있다. 마이크로폰 간의 시간 지연 시간 측정을 위해 상호상관(cross correlation) 방법을 사용하여 되는 데 마이크로폰 마다 이득 특성이 다르기 때문에 신호간의 상호상관성을 잘못 측정하기 때문이다[6]. 따라서 마이크로폰의 이득 특성을 보상함으로써 시간 지연에 기반한 위치 추적 방법을 사용하였다. 마이크로폰의 이득 특성에 강인한 위치 추적 시, 마이크로폰의 이득에 독립적인 상호상관방법을 사용하여 마이크로폰 간의 정확한 최소 지연 시간을 계산할 수 있다. $J(\tau)$는 비용함수이며 본 논문에서는 상호상관방법을 이용하여 $J(\tau)$을 최소화하는 지연시간 τ을 찾는다. 식(3)과 (4)를 이용하여 마이크로폰의 이득으로부터 입력신호를 보상해 줄 수 있는 α 값을 구할 수 있다. 식(4)를 통해 구한 α 값을 식(2)에 대입시켜 보상해준 신호의 비용함수 $J(\tau)$가 최소화되는 τ을 구하게 되는데, 이때의 $J(\tau)$ 값을 최소화하는 τ 값이 신호간의 지연시간이다. 따라서 α에 의해서 마이크로폰의 이득 특성에 강인한 각 마이크로폰 간의 지연 시간을 계산할 수 있다.

$$J(\tau) = \sum_{n} (\alpha x_i(n-\tau) - x_j(n))^2 \qquad (3)$$
$$(i, j) = (1,2),(2,3),(1,3)$$

$$\frac{\partial J(\tau)}{\partial \alpha} = 2\sum_{n} (\alpha x_i(n-\tau) - x_j(n)) x_i(n-\tau) = 0 \qquad (4)$$

$$\alpha^* = \frac{\sum_{n} x_i(n-\tau) - x_j(n) x_j}{\sum_{n} x_i(n-\tau)^2} = 0 \qquad (5)$$

2. 계차 기반의 위치 추적 알고리즘

계차 기반의 위치 추적은 식 (6)과 같이 음성 신호의 현재 샘플과 앞 샘플의 차를 이용하는 것이다.

$$new\_x_i(n) = x_i(n) - x_i(n-1), i = 1,2,3 \quad (6)$$

계차기반의 위치 추적은 음성 신호의 피크 점을 더 잘 찾을 수 있을 뿐만 아니라 각각의 마이크로폰의 이득 특성에 상관없이 각 신호의 시간 지연에 의한 위치 추적을 할 수 있다. 따라서 앞의 마이크로폰의 이득특성에 강인한 위치 추적에서 사용된 식(2)의 상호상관방법과 다르게 마이크로폰의 이득을 보상하지 않고도 식(6)에 의한 계차 신호들을 식(7)의 상호상관에 의해서 시간 지연을 구할 수 있다.

$$J(\tau) = \sum_n (x_i(n-\tau) - x_j(n))^2 \quad (7)$$

3. GCC-PHAT(generalized cross correlation phase transform)기반의 위치 추적 알고리즘

GCC(generalized cross correlation)은 주파수영역에서의 상호상관 방법이다[8]. $x_1(n)$ 과 $x_2(n)$ 을 각각 첫 번째와 두 번째 마이크로폰에서 받은 신호라고 하면 식(8)를 통해 주파수 영역에서의 $x_1(n)$ 와 $x_2(n)$ 사이의 지연시간을 구할 수 있다. 식 (8)에서는 $x_1(n)$ 과 $x_2(n)$ 를 푸리에 변환(the furrier transform)한 $X_1(\omega)$ 와 $X_2(\omega)$ 을 이용한다.

$$R_{x_1x_2}(n) = \frac{1}{2\pi}\int_{-\infty}^{\infty} W(\omega)X_1(\omega)X_2^*(\omega)e^{j\omega n}d\omega \quad (8)$$

$W(\omega)$ 는 주파수 가중 함수로서 $X_1(\omega)X_2^*(\omega)$ 의 역수이며 이 가중 함수를 PHAT(phase transform)라고 한다[1]. PHAT를 가중함수로 사용하는 GCC방법을 GCC-PHAT라고 한다.

$$W(\omega) = \frac{1}{\left|X_1(\omega)X_2^*(\omega)\right|} \quad (9)$$

GCC-PHAT 기반의 지연 시간은 식(10)에 의해서 구할 수 있다.

$$D = \arg\max R_{x_1x_2}(n) \quad (10)$$

V.실험 및 결과

1. 실험환경

호칭 기반을 위한 위치 추적 실험은 조용한 실험실 환경에서 수행하였다. 그림 3는 3개의 마이크로폰을 이용한 위치 추적 시스템 환경을 보여준다. 3개의 마이크로폰은 높이 0.5m위에 로봇의 머리 부분에 위치하고 있고 마이크로폰간의 각도는 120° 이며 마이크로폰 간의 거리는 0. 32m이다. 음원은 로봇의 이름인 “ 웨버” 를 사용하였으며, 각 각의 마이크로폰으로부터 2m 떨어진 곳에서 발생시켜 실험하였다. 각각의 마이크로폰에서 수집된 음원은 16 kHz로 샘플링 되고 16 bit로 양자화 되었다. 음원의 위치를 마이크로폰 1번을 0° 로 기준으로 해서 30° 간격으로 변경하여 실험하였다.

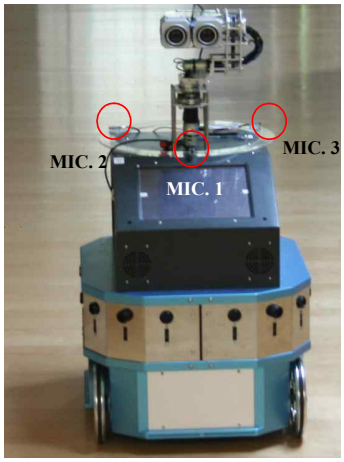

그림 3. 시스템 구성도

2. 결과

표 1은 마이크로폰 이득에 강인한 시간 지연 방법을 사용했을 때의 위치 추적 결과이다. 최대 에러가 31.49° 이며 이는 임펄스 형태의 신호보다 “ 웨버” 인 음성 신호일 때 신호의 피크 점이 뚜렷하지 않기 때문에 이와 같은 에러가 생긴다. 표 2은 계차 기반의 위치 추적 결과이며 최대 오차는 6.66°로 마이크로폰 이득에 강인한 위치 추적보다 최대 오차가 24.83°만큼 감소한 것을 볼 수 있다. 표 3은 GCC-PHAT 기반의 위치 추적 결과이며 최대 오차가 12.66°로 계차 기반의 위치 추적 방법과 약 6° 정도 높은 오차를 보였다. 평균 오차를 비교해보면, 마이크로폰의 이득 특성에 강인한 위치 추적의 평균 오차는 13.55°, 계차기반의 위치 추적의 평균 오차는 2.33°, GCC-PHAT기반의 위치 추적의 평균 오차는 7°로 계차 기반의 위치 추적이 호칭 기반을 위한 위치 추적에 최적의 알고리즘임을 볼 수 있다.

표 1. 마이크로폰의 이득특성에 강인한 시간 지연 방법

Location of sound source	Azimuth	Error rate
0°	31.49°	31.49°
30°	44.84°	14.84°
60°	54.72°	5.28°
90°	120°	30°
120°	139.26°	19.26°
150°	159.94°	9.94°
180°	181.95°	1.95°
210°	228.05°	18.05°
240°	241.8°	1.8°
270°	282.13°	12.13°
300°	298.70°	1.30°
330°	316.50°	16.50°

표 2. 계차 기반의 시간 지연 방법

Location of sound source	Azimuth	Error rate
0°	3.80°	3.80°
30°	26.86°	3.14°
60°	60°	0°
90°	91.54°	1.54°
120°	119.10°	0.9°
150°	146.94	3.06°
180°	180.0°	0°
210°	216.30	6.30°
240°	240.	0°
270°	263.34	6.66°
300°	297.39	2.61°
330°	330.02	0.02°

표 3. GCC-PHAT 기반의 시간 지연 방법

Location of sound source	Azimuth	Error rate
0°	12.65°	12.65°
30°	25.35°	4.65°
60°	52.16°	7.84°
90°	93.06°	3.06°
120°	107.34°	12.66°
150°	142.15°	7.85°
180°	172.16°	7.84°
210°	207.12°	2.88°
240°	227.34°	12.66°
270°	266.82°	3.18°
300°	296.09°	3.91°
330°	334.84°	4.84°

V.결론

본 논문에서는 호칭 기반의 위치 추적을 위해서 마이크로폰의 이득 특성에 강인한 위치 추적과 계차 기반의 위치 추적 그리고 GCC-PHAT기반의 위치 추적 결과를 비교하였다. 음원이 음성이었을 때 계차기반의 위치 추적 알고리즘이 호칭 기반을 위한 위치 추적에 적합한 알고리즘이라는 것을 보였다.

마이크로폰의 이득 특성에 강인한 위치 추적 알고리즘은 음원이 음성이었을 때 평균오차가 13.55°의 결과를 보였다. 계차 기반의 위치 추적의 평균 오차는 마이크로폰의 이득 특성에 강인한 위치 추적보다 평균 오차가 11.22° 감소된 2.33°의 결과를 보였다. 마지막으로 GCC-PHAT기반의 위치 추적의 평균 오차는 계차 기반의 위치 추적 평균 오차보다 4.67° 더 높았다. 즉 시간 영역에서 계차 기반의 음원 추적이 시간 영역에서의 마이크로폰의 이득 특성에 강인한 위치추적과 주파수 영역에서의 GCC-PHAT기반의 위치 추적보다 호칭 기반을 위한 위치 추적에 가장 적합함을 볼 수 있다.

향후, 본 알고리즘은 실시간으로 작동되는 지능형 서비스 로봇에 사용될 것이다. 또한, 음원의 높이와 실험 환경이 음원 위치를 판단하는데 큰 영향을 미친다는 연구 보고가 된 바와 같이 잔향 환경과 음원의 고도를 고려하여 음원의 위치를 추적한다면 좀 더 나은 성능을 기대할 수 있을 것이다[9].

본 논문의 호칭 기반을 위한 위치 추적은 고도정보가 포함된 3-D 위치 추적의 기반의 되는 역할을 수행할 것이다.

참고문헌

[1] L. R. Rabiner and B. H. Juang, Fundamental of speech recognitions, Prentice Hall, 1993

[2] H. Wang and P. Chu, " Voice source localization for automatic camera pointing system in videoconferencing," *in Proc. OCASSP*, vol. 1, pp. 197-190, 1997

[3] 이지연, 한민수, " 지능형 로봇 " 웨버" 를 위한 음원추적 기술," *대한음성학회 가을 학술대회 발표논문집*, pp117-120, Nov.2005

[4] M. Brandstein and D. Ward, Microphone Array, Springer, 2001

[5] M. Brandstein and H. Silverman, " A practical methodology for speech source localization with microphone arrays," *Comput.*, Speech Lng., vol. 11, no. 2, pp. 91-126, 1997

[6]최지성, 한민수, " 마이크로폰의 이득 특성에 강인한 위치 추적," *한국지능로봇 하계종합 학술대회*, 2006

[7] L.R.Rabiner, R.W.Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, 1978

[8] C.H knapp and G.C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust. Speech Signal Pocess.*, vol 24, no4, pp320-327, Aug. 1976.

[9] 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