

지능형 로봇 “웨버”를 위한 음원 추적 기술

이지연* 한민수* 지수영** 조영조**
* 한국정보통신대학교 음성음향정보연구실
** 한국전자통신연구원 지능형로봇연구단

Sound Localization Technique for Intelligent Service Robot “WEVER”

Ji-Yeoun Lee*, Minsoo Hahn*, Su-young Ji**, Young-Jo Cho**

* Speech and Audio Information Lab, Information and Communication Univ.

** Intelligent Robot Research Division, Electronics and Telecommunications Research Institute

jyle278@icu.ac.kr, mshahn@icu.ac.kr,
chisy@etri.re.kr, youngjo@etri.re.kr

Abstract

This paper suggests an algorithm that can estimate the direction of the sound source in real time. Our intelligent service robot, WEVER, is used to implement the proposed method at the home environment. The algorithm uses the time difference and sound intensity information among the recorded sound source by four microphones. Also, to deal with noise of robot itself, the kalman filter is implemented. The proposed method takes shorter execution time than that of an existing algorithm to fit the real-time service robot. The result shows relatively small error within the range of ± 7 degree.

I. 서론

지능형 로봇은, 네트워크 기반을 활용하여 정보 도우미 및 공공 도우미로서의 역할을 함에 따라 주위 상황을 인지하고 판단하는 능력이 필요하다. 따라서 음원 발원처 위치 추적은 로봇으로 하여금, 로봇의 도움이 필요한 사람이나 개체 근처로 이동하여 주위 상황을 판단하고 적절한 대응조치를 취하게 한다. 과거, 음

원 추적은 정숙한 환경에서, 근거리에서, 두개의 마이크로폰을 사용하는 시스템으로 주로 연구되어왔다. 최근 실험용 또는 장난감 로봇에 탑재하여, 인간과 로봇의 자연스러운 상호 작용을 목적으로, 사용자의 호출 음성이나 박수소리를 입력받아 인식함으로써 카메라나 키보드와 같은 입력 장치의 대용으로 활용할 수 있도록 하는 자동 인터페이스 구현을 위한 연구가 진행되고 있다.

음원 추적 시스템은 잡음이 없거나 비교적 조용한 실험실 환경에서는 좋은 성능을 나타낸다. 그러나 이를 실용화하여 실제 현장에서 사용할 경우에는 여러 가지 잡음 요인들에 의하여 시스템 성능이 현저히 저하된다. 로봇에 실장 할 경우, 가장 문제가 되는 것이 로봇 본체의 잡음이다. 따라서 본 논문에서는 로봇의 몸체 둘레에 가장 음원 감지가 잘되는 네 개 마이크로폰들의 위치를 선정하고, 각 마이크로폰이 감지하는 음원의 강도 및 시간차를 분석하는 한편 로봇 본체 잡음을 감소시킨 후 음원의 방향을 계산한다. 음원의 위치가 인지되면 로봇은 그 위치로 이동하는 동시에 실시간으로 음원 위치를 지속적으로 확인하고 갱신하여 시스템의 신뢰도를 높인다.

본 논문은, 제 2장에서 제안된 음원 추적 알고리즘에 대해 설명하고, 제 3장에서 실험환경 및 성능 평가 결과에 관해 언급하고, 마지막으로 제 4장에서 결론을

맞는다.

II. 음원 추적 알고리즘

실제 상황에서 로봇이 음원을 듣고 방향을 설정하고 그 위치로 움직이기 위해서는 로봇의 실행시간을 고려해야 한다. 신호의 강도를 이용하여 강도가 큰 순서대로 두 개의 마이크를 결정하고, 그 결정된 채널 사이에서 음원이 발생되었다고 가정한다. 그 다음, 그 채널에 대해서만 시간차를 계산한다. 그러므로 모든 마이크에 대해 시간차를 계산하는 기존의 방법보다 실행시간 면에서 빠른 방법이라고 할 수 있다.

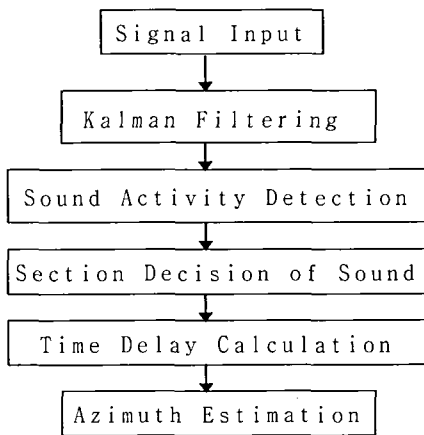


그림 1. 음원 추적 알고리즘 순서도

1. 칼만 필터링

로봇에 실장 할 경우, 가장 문제가 되는 것이 로봇 본체의 잡음이다. 마이크의 입력으로 우리가 원하는 신호뿐만 아니라 로봇의 진동, 모터 구동소리, CPU, 팬 소리 등이 동시에 들어가기 때문에, 학습 시에 우리가 원하는 신호만으로 생성한 기준 패턴과는 다른 형태가 된다.

본 논문에서 사용하는 로봇의 경우, 로봇의 뒤쪽에 위치하는 마이크의 잡음 레벨이 로봇 앞쪽에 위치하는 마이크보다 더 크다. 왜냐하면 모터, CPU, 팬, 초음파 기기들이 로봇의 뒤쪽에 위치하기 때문이다. 따라서 여러 개의 마이크들은 위의 기기의 위치에 따라 서로 다른 잡음 레벨을 가진다. 그래서 동시에, 실시간으로, 마이크 각각에 대해서 칼만 필터를 적용했다.

2. 음원 추출

식(1)의 단시간 에너지를 이용하여 음원을 검출한다. 정확한 음원 구간을 검출을 하기 위한 에너지 문턱 값

은 여러 번의 실험을 통해 결정된 실험값으로 설정하였다.

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

문턱값 : $10 \times \log_{10}(E_n)$

3. 발생 음원 구간 설정

발생한 음원의 구간 설정은 식(1)의 단 구간 에너지를 이용하였다. 그것을 이용하여 사용된 네 개의 마이크 중에서, 강도가 가장 큰 순서대로 두 개의 마이크가 먼저 선택된다. 그리고 그 두 개의 마이크 사이의 구간에서 음원이 발생되었다고 가정되어진다.

예를 들어서, 그림 2와 같이, 우선 네 개의 마이크들의 에너지를 계산한다. 마이크 1번과 4번의 강도가 가장 크다고 결정되면, 그 사이의 구간에서만 음원이 발생되었다고 가정하고 음원 발생 구간을 결정한다.

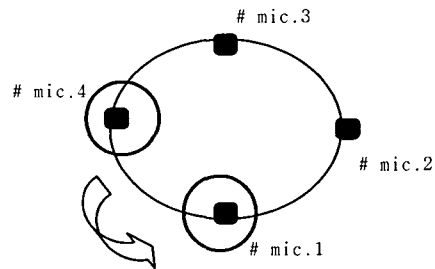


그림 2. 음원 구간의 설정 방법

4. 시간 지연 계산

음원이 발생한 방위각을 결정하기 위해서 두 신호 사이의 지연 시간을 알아야한다. 지연시간의 계산 방법은 과거 오래전부터 많이 연구되어왔다. 그러나 상호 상관(cross correlation)방법이 가장 보편적으로 쓰인다. 다른 방법들과 비교하여 계산적으로 간단하고 능률적이기 때문이다[3].

두 마이크 사이의 지연시간은 식(2)와 같이 정규화된 상호 상관 방법 (the normalized cross-correlation method)에 의해서 계산되어진다.

$$\bar{\tau} = \arg \max \{ \overline{\phi_{ij}}(d) \} \quad (2)$$

(i, j) = (0, 1), (1, 2)

$$\overline{\phi_{ij}} = \lim_{l \rightarrow \infty} \frac{\sum_{k=-l}^l x_1(k)x_2(k+d)}{\sqrt{\sum_{k=-l}^l x_1^2(k) \sum_{k=-l}^l x_2^2(k)}}$$

$$k = 0, 1, 2N-1$$

d : value from -10,000 to 10,000

위의 식에서 $x_1(k)$ 과 $x_2(k)$ 는 신호 강도에 의해 결정된 두 개의 마이크로폰으로 들어오는 신호이다. 이때, 각각 마이크로폰간의 도착 지연 시간은 식(3)과 같은 조건을 만족해야 한다.

$$|\overline{\tau}| \leq D/C \quad (3)$$

D : 두개의 마이크로폰 사이의 거리

C : 소리 속도

5. 방위각 추측

그림 3은 본 논문에서 사용한 로봇의 마이크로폰 배열을 보여준다. 이 그림은 로봇을 위쪽에서 내려다보았을 때의 그림이다. x(mm)라고 쓰인 윗부분이 로봇의 앞쪽이고, 그 아랫부분이 로봇의 뒤쪽 모습이다. 각각 마이크로폰들의 위치는 다음과 같다. 이것은 (x축, y축, z축)을 의미한다. 즉 마이크로폰 1번, 2번, 3번, 그리고 4번의 위치는 (65, 142, 315), (-233, 176, 291), (-419, -117, 291), (7, -174, 291)이다. 그리고 위의 좌표를 삼각 함수 공식에 응용하여 두개의 마이크로폰 사이의 각도와 거리를 계산한다.

그림 3에서, d_n 과 α_n , ($n = 1, 2, 3, 4$) 은 각각 거리와 각도를 의미한다.

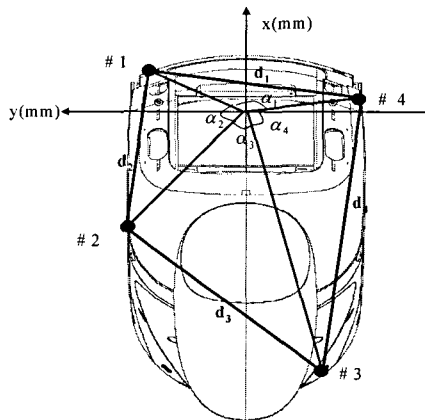
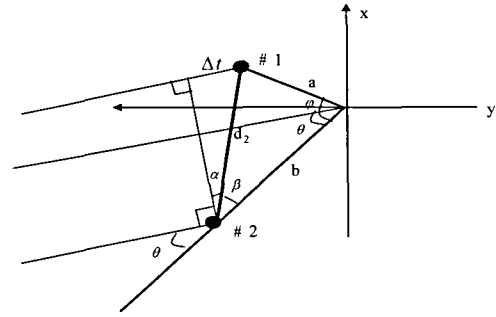


그림 3. 마이크로폰들의 배열

그림 4는 마이크로폰 1번과 2번사이의 방위각 추측의

예를 보여준다.



a : 15.62 cm, b : 29.2 cm

d_2 : 29.99 cm, θ : 방위각

Δt : 시간 차이(time difference)

그림. 4 방위각의 계산

그림 4에서, 마이크로폰의 중심에서 음원까지의 거리가 b보다 훨씬 길다고 가정할 때, 마이크로폰에 들어오는 음원의 파장은 평면파라고 가정할 수 있다.

이 때, 지연 시간은 식(4)로써 근사화 할 수 있다. 따라서 식 (4)를 가지고 현재 음원이 발생한 방위각을 계산할 수 있다. 또한 식(4)를 응용하여, 모든 채널사이에서의 방위각을 계산할 수 있다.

$$\theta = 180 - 90 - \sin^{-1}\left(\frac{\Delta t \times v_s}{S_{\text{samplingrate}} \times d_2}\right) - \sin^{-1}\left(\frac{a}{d_2} \sin \varphi\right)$$

v_s : 22°C 에서 345 m/s (4)

$S_{\text{samplingrate}}$: 16 kHz

위 식(4)에서 계산된 방위각은 현재 음원의 구간에 따라서 최종 방위각으로 맞추어진다. 그림 3에서 마이크로폰 1~2구간은 0도~77.53도, 2~3 구간은 77.53도~130.2도, 3~4 구간은 130.2도~206.9도, 4~1구간은 206.9도~360도까지의 방위각범위 내에서, 현재 음원의 위치에 따라, 최종 방위각 각도가 결정된다.

III. 실험환경과 결과

1. 실험환경

실험은 집안 환경과 비슷하게 꾸며놓은 실험실에서 행해졌다. 음원은 박수소리였으며 4m에서 발생되었다. 또한 그것은 16 kHz로 샘플링 되고 16 bit로 양자화 되었다.

그림 5는 이 실험에서 음원 추적을 위해 사용한 로

빛을 보여준다. 그림 5에서 보이듯이, 여덟 개의 마이크로폰들이 로봇 몸통에 같은 높이를 가지고 위치되어 있다. 본 실험에서는 360도를 포함하면서 적당한 위치에 있는 네 개의 마이크로폰을 선택하여 이용하였다.



마이크로폰
그림 5. 로봇 "웨버"

음원은 마이크로폰들의 높이보다 약간 높은 위치에서 30도의 간격을 두면서, 즉 12개의 위치에서 3번씩 발생되었다. 로봇의 키가 약 1m 정도 되기 때문에, 음원은 항상 마이크로폰들의 위치보다 조금 더 높은 위치에서 발생될 수밖에 없다.

2. 방위각 추측 결과

본 실험에서는 각각의 위치에서 실시간으로 3번씩 녹음된 박수소리 음원에 대해, 도착 지연 시간을 구하고 방위각으로 환산하여, 평균하여 원래의 각도와 비교하여 오차를 계산하였다. 표 1은 실험 성능을 보여준다.

표 1에 나타난 바와 같이 본 논문의 음원 추적 알고리즘은, 오차범위 7도 안에서 거의 만족스러운 결과를 나타내었음을 확인할 수 있었다.

표1. 방위각 추측 결과

Original degree	Average degree (in 4m)	Average error
0	0.00	0.00
30	36.19	6.19
60	55.27	4.73
90	86.52	3.48
120	124.66	4.66
150	144.06	5.94
180	183.69	3.69
210	216.24	6.24
240	244.66	4.66
270	272.45	2.45
300	292.05	7.95
330	336.14	6.14

IV. 결론

본 논문은 네 개의 마이크로폰을 사용하여, 음원의 강도와 시간 지연 정보에 기반을 둔 음원 추적 기술을 제안하였다. 그리고 칼만 필터가 로봇 본체의 잡음에 대처하기 위해 구현되었다. 이것은 정확한 지연 시간의 추측을 위해 필요하다.

음원 추적 알고리즘은 오차범위 7도 안에서 거의 정확한 결과를 보였다. 그러나 음원의 높이와 실험 환경이 음원 위치를 판단하는데 큰 영향을 미친다는 연구 보고[3]가 된 바와 같이 잔향, 매아리 환경과 음원의 고도를 고려하여 개선한다면 좀 더 나은 성능을 기대할 수 있을 것이다.

본 논문에서 연구하는 음원 추적 알고리즘의 실제 적용 시나리오로는, 공공 서비스 대상자가 음성이나 박수로 로봇을 호출 할 경우 호출자를 찾아가기, 물건이 떨어져 파열음을 낼 경우 그곳으로 이동하여 상황 판단 및 대응조치하기, 비상벨이나 비정상적인 음향을 감지하여 그곳으로 이동한 후 공공 서비스 조정실에 비디오 및 기타 정보를 네트워크로 전달하고, 조정실의 명령을 되받아 수행하는 것 등을 들을 수 있다.

음원 추적 기술은 인간과 로봇의 상호작용에 유용할 것이며, 음원 분류와 사용자 확인 기술의 기반을 마련하는 역할을 수행할 것이다.

참고문헌

[1] Sangbae Jeong, Minsoo Hahn, "Car noise reduction by Kalman filtering," *ICSP-2001*, Daejeon, Korea(ROK)

[2] Nguyen, D., Aarabi, P., Sheikholeslami, Ai, "Real-time sound localization using field-programmable gate arrays," *Acoustics, Speech, and Signal Processing 2003*, Volume: 2 ,6-10 April

[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

[4] Huang, J., Ohnishi, N., and Sugie, N., "Spatial Localization of Sound Sources : Azimuth and Elevation Estimation," *Proceedings of the IEEE Conference on IMTC/98* , Volume: 1 , p.18-21 May 1998