논문 2007-02-80

지능형 서비스 로봇의 주의집중을 위한 시간지연 기반 실시간 음원추적 기술개발

(TDOA-Based Sound Source Localization to Control Intelligent Service Robot's Attention)

배 경 숙, 이 재 연, 곽 근 창, 윤 호 섭
(Kyung-Sook Bae, Jae-Yeon Lee, Keurn-Chang Kwak, Ho-Sub Yoon)

Abstract : 인간과 상호작용하는 로봇의 경우, 자연스러운 움직임과 행동은 사람에게 친근감을 제공한다. 그러한 자연스러운 움직임과 행동의 기반이 되는 요소기술들 중 하나로 음원추적을 꼽을 수 있다. 본 논문에서는 3개의 마이크를 장착한 로봇에서의 시간지연기반 음원추적 시스템에 대해 소개한다. 개발된 음원추적 시스템을 지능형 서비스 로봇 WEVER-R2에 적용한 결과, 에러를 ±15도 이내에서 93% 이상의 성공률을 보였다.

Keywords : 지능형 로봇, 음원추적, TDOA, 주의집중

I. 서론

최근 지능형 서비스 로봇(Intelligent Service Robot)에 관한 연구가 활발히 진행되고 있다. 특히, 인간로봇 상호작용(HRI) 기술은 다양한 의사소통 채널을 통해 인간과 로봇이 자연스럽게 상호작용할 수 있는 지능형 서비스 로봇의 핵심적인 기술이다. 일반적으로 인간로봇 상호작용 기술은 로봇카메라로부터 취득한 영상정보를 근거로 하는 비디오 기반 상호작용 기술과 로봇에 부착된 마이크로부터 취득한 음성정보에 근거한 오디오 기반상호작용 기술로 나뉠 수 있으며 얼굴인식, 표정인식, 제스쳐인식, 음성인식, 화자인식, 음원추적 등이 그 대표적인 예이다.

이렇듯 다양한 HRI 기술 중 음원 추적 기술은 로봇의 주의집중을 위해 사용자와의 상호작용을 시작하는 가장 첫 번째 단계로 호출에 응답하도록 하는 로봇의 가장 기본적인 기능을 구현하는 필수 기술이다. 이와 관련한 국내 연구동향을 살펴보면, Choi[3]은 오디오와 비디오 정보를 융합한 시청각 기반의 음원 추적 기술을 수행하였다. 먼저 음원추적을 수행한 후, 추적오차를 보상하기 위해 얼굴 검출을 이용하여 호출자의 위치를 파악한다. Park[4]은 휴머노이드 로봇에서 세 개의 마이크를 이용하여 음원추적을 수행하고 마찬가지로 얼굴검출을 통해 추적오차를 보상하였다. 한편, 국외 연구동향은 시청각 기반 음원추적 뿐만 아니라 에코와 반향에 강인한 알고리듬에 관한 연구가 많이 진행되고 있으며, 음원추적 기술을 기반으로 한 음원분리 연구도 활발히 진행되고 있다. Huang[6]은 에코와 반향에 대처하기 위한 인간 청각시스템의 선행 효과를 기반으로 한 모델기반 음원추적(model-based sound localization) 시스템을 개발하였다. Valin[7],[8]은 8개의 마이크를 이용하여 steered beam-former를 통한 3차원 공간에서의 강인한 음원추적을 수행하였다. Nakadai[9]는 머리의 양 측면에 한 쌍의 마이크와 내부 잡음 제거를 위해 머리 안쪽에 다른 한 쌍의 마이크를 장착하여 SIG 휴

* 교신저자(Corresponding Author)
논문접수 : 2007. 12. 7. 채택확정 : 2008. 2. 1.
배경숙 : 한국전자통신연구원 지능형로봇연구단 인간로봇상호작용 연구팀
이재연 : 한국전자통신연구원 지능형로봇연구단 인간로봇상호작용 연구팀
곽근창 : 조선대학교 제어계측로봇 공학과
윤호섭 : 한국전자통신연구원 지능형로봇연구단 인간로봇상호작용 연구팀
※ 본 연구는 정보통신부 및 정보통신연구진흥원의 IT 신성장동력 핵심기술 개발사업의 일환으로 수행하였음.[2005-S-033-03, URC를 위한 내장형 컴포넌트 기술개발 및 표준화]

머노이드 로봇에 음원추적 기술을 적용하였다. Yamamoto[10]은 ASIMO에서 동시에 발생된 혼합된 음성을 가지고 음원추적 및 분리기술과 음성인식 기술을 결합하여 잡음과 에코환경에서 강인하면서도 정확한 성능을 보여주었다.

음원을 예측하는 방법 중 소리의 도착 지연시간(TDOA: Time Delay of Arrival)을 이용한 방법은 계산이 간단하고 정확성이 높기 때문에 가장 널리 사용되고 있다. 본 논문에서는 3개의 마이크를 장착한 로봇, Wever-R2에서의 시간지연기반 음원추적 시스템에 대해 소개한다. 또 성능평가를 위해 일반 가정환경과 같은 테스트베드에서 음원추적용 데이터베이스를 구축하였다. 이 데이터베이스는 음원거리(1-5m), 호출각도(0-360도, 45도 간격)에 의해 구축되었으며, FOV(Filed of View)의 범위와 평균 추적오차에 의해 성능이 평가된다.

II. 도착지연시간(TDOA)

소리의 도착 지연시간을 이용한 음원추적 방법은 그림 1에서와 같이 음원 S로부터 나오는 소리의 신호가 음원으로부터 멀리 위치한 마이크2 보다 가까이 위치한 마이크1에 먼저 도달한다는 기본 원리를 이용한 것이다. 이때 생기는 신호의 도달시간 차와 마이크간의 거리, 소리의 속도를 이용하여 음원의 방향을 추정할 수 있다. 일반적으로 음속은 기온이 섭씨 15도 일 때 340.0m/s로 알려져 있고 마이크간의 거리는 쉽게 측정이 가능하다. 따라서 정확한 음원추적을 위해서는 정확한 신호의 도달시간차를 구하는 것이 중요하다.

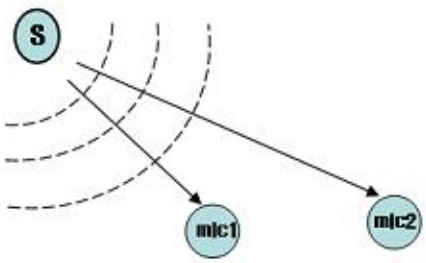


그림 1. 음원과 마이크간 거리에 따른 신호의 도착 지연

Fig 1. The sound signal is delayed longer in reaching the far microphone

III. TDOA 기반 음원추적 시스템

1. 지연 시간 측정

마이크 간의 지연시간 측정을 위한 방법으로는 영교차(zero crossing) 방법과 상호상관(cross correlation)도를 이용한 방법 등 여러 가지 방법이 있다. 그 중 상호상관도를 이용한 방법이 비교적 계산 량이 적고 좋은 성능을 보이기 때문에 가장 널리 쓰이고 있다.

두 마이크에서 받은 신호를 각각 $x_i(n)$, $x_j(n)$라고 할 때, 두 신호 사이의 상호상관도는 다음 식에 의해 얻어지고 이 식을 최소화 하는 τ_{ij}가 두 마이크 i,j 간의 도착지연시간이 된다.

$$R(\tau) = \sum_{n}(x_i(n-\tau)\times x_j(n))^2 \quad (1)$$

도착 지연시간 τ_{ij}은 디지털 신호의 샘플수로 표현이 되며 D_{ij}을 두 마이크 간의 거리, C를 음속이라 할 때 τ_{ij}는 다음 조건을 만족해야 한다.

$$|\tau_{ij}| \le \frac{D_{ij}}{C} \quad (2)$$

본 음원추적 시스템은 마이크간 거리 0.32m, 신호는 16kHz 음속 349.09m/s 환경에서 개발되었다. 따라서 $-14.66 \le \tau_{ij} \le 14.66$ 조건을 만족해야 한다. 표 1은 이와 같은 개발환경에서 지연시간의 그라운드 트루스를 보여준다.

표 1. 지연시간의 그라운드 트루스

Table 1. Ground Truth of Time Delay

각도	τ_{ij}
0°	14.666705
30°, -30°	12.701739
45°, -45°	10.370928
60°, -60°	7.333355
90°, -90°	0.000005
120°, 120°	-7.333347
135°, -135°	-10.370921
150°, -150°	-12.701735
180°	-14.666705

보다 정확한 τ 를 구하기 위해서 음성시그널에 W사이즈의 윈도우를 씌워 윈도우 내에 있는 시그널에 대해서 지연시간을 측정한 뒤, W/2 이동하여 다시 지연시간을 측정한다. 이 같은 방법으로 음성시그널이 끝날 때까지 반복하여 지연시간을 누적하여 히스토그램화 한다. 그중 가장 큰 누적값을 갖는 τ_{max}와 그 이웃하는 값을 이용하여 보다 정확한 지연시간을 구한다. 식(3)은 최종지연시간을 결정하는 수식이다.

$$\tau = \frac{\sum_{j=-1}^{1} (\tau_{max} + j)h[\tau_{max} + j]}{\sum_{j=-1}^{1} h[\tau_{max} + j]} \quad \text{식(3)}$$

2. 각도 계산

3개의 마이크는 그림 2과 같이 각각 120° 간격으로 원형으로 위치한다. 음원이 위치한 각도를 계산하기 위해서 먼저 마이크에 들어오는 음원의 파장은 평면파라고 가정한다.

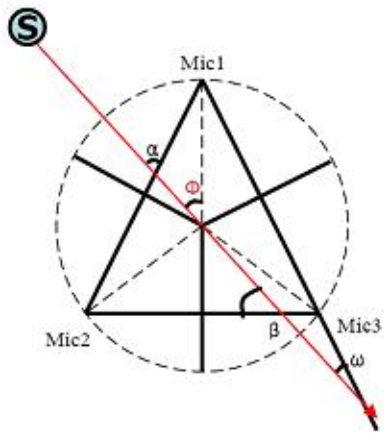


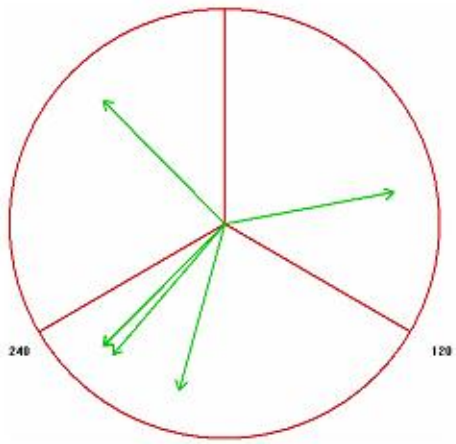
그림 2. 각도 계산
Fig 2. Azimuth Estimation

Mic_i과 Mic_j 사이의 지연시간을 τ_{ij}이라고 하면, 각 마이크 구간에서 측정된 지연시간 $\tau_{12}, \tau_{23}, \tau_{13}$을 식(4)에 각각 대입하여 그림 3의 α, β, ω를 구할 수 있다. C는 음속, D는 마이크간 거리, F는 샘플링 레이트를 나타낸다.

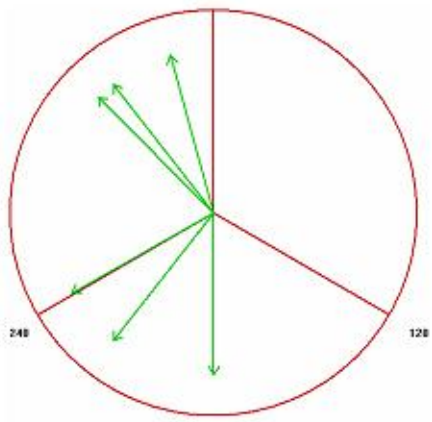
$$azimuth = \cos^{-1}\left(\frac{\tau \times C}{D \times F}\right) \quad \text{식(4)}$$

그러나 두 마이크를 기준으로 음원이 정면에 있는지 후면에 있는지 알 수 없으므로 α, β, ω 뿐만 아니라 $-\alpha, -\beta, -\omega$도 고려해야한다. $\pm\alpha$, $\pm\beta$, $\pm\omega$는 삼각법에 의해 Φ에 관한 식으로 다음과 같이 표현된다.

$$\Phi_{\alpha 1} = \alpha - 30, \quad \Phi_{\alpha 2} = -\alpha - 30$$
$$\Phi_{\beta 1} = \beta + 90, \quad \Phi_{\beta 2} = -\beta + 90$$
$$\Phi_{\omega 1} = \omega + 30, \quad \Phi_{\omega 2} = -\omega + 30$$



(a) $\tau_{12}, \tau_{23}, \tau_{13}$가 모두 옳은 값인 경우 (-165도)



(b) $\tau_{12}, \tau_{23}, \tau_{13}$ 중 하나가 잘못된 값인 경우 (-60도)

그림 3. 후보각
Fig 3. Candidate azimuths

이와 같이 후보각들이 추출되면 추출된 후보각 중 차이가 가장 작은 두 개를 선택하여 그 평균을 음원의 방향으로 결정한다. 그림 3(a) 같이 모든 구간에서의 시간지연이 정확히 계산되었다면 여섯 개의 후보각 중 세 개는 밀집되어 있다. 반면, 그림 3(b)와 같이 τ_{12}, τ_{23} τ_{13} 중 하나가 잘못된 값일 경우에도 가장 가까운 두 개를 선택하게 되므로 신뢰할 수 있는 추적각도를 얻을 수 있다.

Ⅲ. 실험 및 결과

1. 데이터베이스

실제 로봇환경에서 음원추적 시스템의 성능을 평가하기 위해 가정환경처럼 꾸며진 테스트 베드에서 음원추적용 데이터베이스를 구축하였다. 그림 4는 ETRI 지능형로봇연구단에서 제작된 네트워크기반 지능형 서비스 로봇인 WEVER-R2이고 세 개의 마이크가 120도 간격으로 배치되어 있다. 음원 취득에는 ETRI에서 개발된 다채널 보드인 MIM(Multimodal Interface Module)을 사용하였다. 음원추적용 데이터베이스는 로봇호출 음성인 "웨버"를 사용하였으며 0-360도 범위에서 45도 간격으로 1-5m 마다 3번씩 두 명이 발성하여 총 240set의 음성샘플을 수집하였다. 음성샘플들은 160kHz, 16bit, mono, PCM 형식으로 취득되었다.

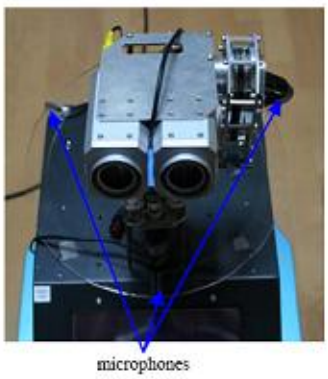


그림 4. 로봇 플랫폼, WEVER-R2
Fig 4. Robot Platform, WEVER-R2

2. 실험 결과

본 절에서는 개발한 음원추적 시스템의 성능에 대해 설명한다.

그림 5은 음원추적 오차 허용율을 ±0에서 ±40도까지 증가시켰을 때의 음원추적 성공률을 나타낸다. 파란 선은 음성 시그널 전체에 대해 지연시간을 측정했을 때의 결과이고 노란 선은 음성 시그널을 300샘플 단위의 세그먼트로 나누어 각 세그먼트에서 지연시간을 측정하여 음원추적을 수행했을 때의 결과이다. 제안한 세그먼트 단위의 지연시간 측정을 이용한 음원추적의 성능이 더 좋음을 알 수 있다.

그림 6은 세그먼트 사이즈에 따른 음원추적 성능을 나타낸다. 세그먼트의 사이즈를 300 샘플로 했을 때 가장 좋은 성능을 보였다.

그림 7과 같이 음원의 거리에 따른 성능 변화는 크지 않았는데 쓰인 데이터는 사람이 직접 발성한 음성을 취득한 데이터로써 사람은 거리가 멀어질수록 더욱 크게 호출하는 경향이 있고 일반 가정환경은 반향이 크지 않기 때문인 것으로 해석된다.

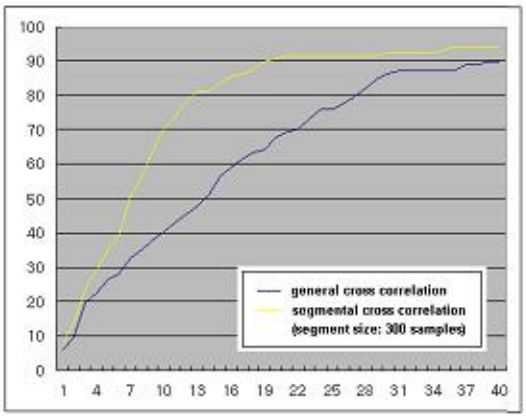


그림 5. 허용 오차율에 따른 성능비교
Fig 5. Performance Comparison between general and segmental cross correlation

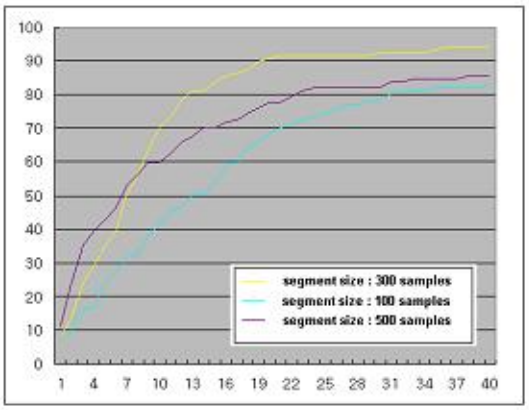


그림 6. 세그먼트 사이즈에 따른 성능비교
Fig 6. Performance with segment size

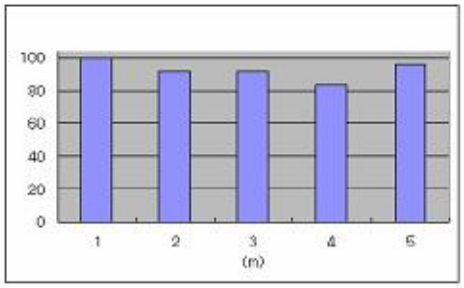


그림 8. 거리에 따른 음원추적 성능
Fig 7. Performance with distance

Ⅳ. 결 론

본 논문은 지능형 서비스 로봇의 주의집중을 위한 시간지연 기반 실시간 음원추적 기술을 개발하고 실제의 가정환경과 같은 테스트베드에서 성능을 평가하였다. 보다 정확히 지연시간을 계산하기 위해 음성신호 전체의 상호상관도를 이용하는 것보다 여러 개의 세그먼트로 분할 한 뒤, 각 세그먼트의 상호상관도를 이용한 방법이 보다 좋은 성능을 나타냈다.

이러한 음원추적 기술은 근거리 및 원거리에서 얼굴검출 및 인식 방법과 융합하여 효과적으로 시청각기반 음원추적 기술로 발전시킬 수 있다.[10]

참고문헌

[1] O. Deniz, M. Castrillon, J. Lorenzo, C. Guerra, D. Hernandez, "CASIMIRO: A RobotHead for Human-Computer Interaction", Proceedings the 2002 IEEE. Int. Workshop in Robot and Human Interactive Communication, 2002.

[2] J. Huang, T. Supaongprapa, I. Terakura, F. Wang, N. Ohnishi, and N. Sugie, "A model based sound localization system and its application to robot navigation," Robotics and Autonomous Systems, pp. 199-209, 1999.

[3] Jiyeoun Lee, and Minsoo Hahn, "Sound Localization Technique for Intelligent Service Robot "Wever", Proceddings of the KSPS conference, pp. 117-120, Nov.2005.

[4] Jisung Choi, Jiyeoun Lee, Sangbae Jeong, Keunchang Kwak, Suyoung Chi, Minsoo Hahn, "Multimodal Sound Source Localization for Intelligent Service Robot," International Conference on Ubiquitous Robots and Ambient Intelligence, 2006.

[5] C. H knapp and G .C. Carter, "The generalized correlation method for estimation of time delay," IEEE Trans. Acoustic, Speech Signal Processing, Vol.24, No.4, pp. 320-327, 1976.

[6] M. Brandstein and D. Ward, Microphone Array: Signal Processing Techniques and Applications, Springer-Verlag, New York, 2001.

[7] 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.

[8] G. C. Carter, A. H. Nuttall, and P.G. Cable, "The smmothed coherence transform(SCOT)," Proceedings of the IEEE, vol.61, pp. 1497-1498, 1973.

[9] J. Huang, T. Supaongprapa, I. Terakura, F. Wang, N. Ohnishi, and N. Sugie, "Mobile Robot and Sound Localization," Proceedings of the 1997 IEEE/RSJ International Conference on IROS '97, Vol. 2, pp. 7-11, 1997.

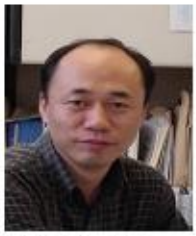
[10] D. H .Kim, J. Y. Lee, E .Y .Cha and Y.J.Cho, "Face identification using mutliple combination strategy for human robot interaction," Proc. of the 16th IFAC world congress in Prague, C2ech Republic, 2005.

저 자 소 개

배 경 숙

2002년 : 숙명여자대학교 컴퓨터과학과 학사.
2004년 : 숙명여자대학교 컴퓨터과학과 석사.
현재 : 한국전자통신연구원 지능형로봇연구단 인간로봇상호작용연구팀 연구원.
관심분야: 영상처리, 패턴인식, 화자인식, 음원추적.
Email: pavin@etri.re.kr

이 재 연



1984년 : 서울대학교 공과대학 제어계측학과 학사.
1986년 : KAIST 전기 및 전자공학과 석사.
1996년 : 일본 東년海(Tokai)대학 광공학 박사.
1986년~현재 : 한국전자통신연구원 인간로봇상호작용연구팀 책임연구원.
관심분야: HRI, 컴퓨터비젼, 패턴인식
Email: leejy@etri.re.kr

곽 근 창

2002년 : 충북대학교 전기공학과 공학박사 졸업.
2002년 : 충북대학교 BK21 사업단 연구원.
2003년 : Univ. of Alberta 박사 후 연구원.
2005년 : 한국전자통신연구원 선임연구원.
현재 : 조선대학교 제어계측로봇공학과 전임강사.
관심분야: 로봇 비젼 및 청각시스템, 생체정보처리
Email: kwak@chosun.ac.kr

윤 호 섭

1989년 : 숭실대학교 전자계산학과 학사.
1991년 : 숭실대학원 전자계산학과 석사.
2003년 : KAIST 전자계산학 박사.
1991년 : KIST 시스템공학연구소 선임연구원.
1998~현재 : 한국전자통신연구원 인간로봇상호작용연구팀 책임연구원 및 팀장.
관심분야: HRI, 영상처리, 생체인식, 패턴인식
Email: hsyoon@etri.re.kr