지능형 로봇을 위한 GCC-PHAT 기반 음원추적 기술의 성능분석

Performance analysis of GCC-PHAT-based sound source localization for intelligent robots

박 범 철[1] · 반 규 대[2] · 곽 근 창[3] · 윤 호 섭[4]

Park Beom-Chul[1] · Ban Kyu-Dae[2] · Kwak Keun-Chang[3] · Yoon Ho-Sup[4]

Abstract In this paper, we present a Sound Source Localization (SSL) based GCC (Generalized Cross Correlation)–PHAT (Phase Transform) and new measurement method of angle with robot auditory system for a network-based intelligent service robot. The main goal of this paper is to analysis performance of TDOA and GCC-PHAT sound source localization method and new angle measurement method is compared. We use GCC-PHAT for measuring time delays between several microphones. And sound source location is calculated by using time delays and new measurement method of angle. The robot platform used in this work is wever-R2, which is a network-based intelligent service robot developed at Intelligent Robot Research Division in ETRI.

Keywords : Sound source localization, GCC-PHAT, Intelligent robot

1. 서 론

오늘날 지능형 서비스 로봇 (Intelligent Service Robots)에 대한 관심이 날로 더해가고 있는 가운데 지능형 로봇의 기술 개발에 대한 중요성이 나날이 부각되고 있다. 지능적인 로봇을 개발하기 위해서는 많은 기반 기술들이 필요하며, 특히 카메라와 마이크로폰으로부터 얻어진 영상 및 음성정보를 이용하여 인간과 로봇이 자연스럽게 교감하는 인간-로봇 상호작용 (Human-Robot Interaction) 기술이 많이 연구되고 있다[1]. 이러한 기술 가운데 음원이 발생한 곳을 찾아내는 음원 추적 (Sound Source Localization) 기술은 로봇이 공공장소나 가정에서 주위 상황을 인지하고 판단하여 도움을 필요로 하는 근처로 이동하여 적절한 대응조치를 취할 수 있도록 함으로써 주의집중을 수행할 수 있다 [2]. 현재 음원 추적 기술은 일반적으로 시간영역 혹은 주파수영역에서의 분석을 통한 연구가 많이 진행되고 있다. 대표적으로 널리 사용되는 방법에는 강도 차이를 이용한 방법[3], TDOA (Time Delay of Arrival) 방법[4]과 GCC-PHAT [5], 빔포밍 (beam-forming) 방법[6] 등이 있다. TDOA을 이용한 방법은 계산이 간단하고 비교적 정확성이 좋아 가장 널리 쓰이고 있는 반면[7], GCC-PHAT은 잡음이나 반향환경에서 좋은 특성을 보이고 있다.

본 논문에서는 로봇환경에 좋은 특성을 보이는 주파수영역 분석방법인 GCC-PHAT기반의 음원 추적 방법의 성능을 TDOA의 변형된 형태인 계차 기반 음원추적 방법과 비교한다. 또한 지연시간을 이용해서 각도를 추정할 때 기존의 경험적인 구간선택 방법 없이 추정된 각도들의 정보를 이용하여 신뢰도 있는 추적각도를 얻는다. 성능분석을 위해 일반 가정 환경과 같은 테스트베드에서 WEVER-R2 로봇을 통하여 음원 추적용 데이터 베이스를 구축한다. 이 데이터베이스는 음원 거리(1~5m), 호출 각도 (2채널: 0~180도, 30도 간격, 3채널: 0~360도, 45도 간격), 마이크 채널 수(2, 3개)에 의해 구축되었으며, FOV(Field of View)의 범위와 평균추적오차에 의해 성능이 비교 된다.

※ 본 연구는 정보통신부 및 정보통신연구진흥원의 IT 신성장동력 핵심기술 개발사업의 일환으로 수행하였음. [2005-S-033-03, URC 를 위한 내장형 컴포넌트 기술개발 및 표준화]

1 과학기술연합대학원대학교 컴퓨터소프트웨어 및 공학과 석사과정 (E-mail : parkbc@etri.re.kr)

2 과학기술연합대학원대학교 컴퓨터소프트웨어 및 공학과 석박사 통합과정 (E-mail : kdban@etri.re.kr)

3 조선대학교 전자정보공과대학 제어계측공학과 전임강사 (E-mail : kwak@chosun.ac.kr)

4 한국전자통신연구원 지능형로봇 연구단 인간로봇상호작용 연구팀 팀장 (E-mail : yoonhs@etri.re.kr)

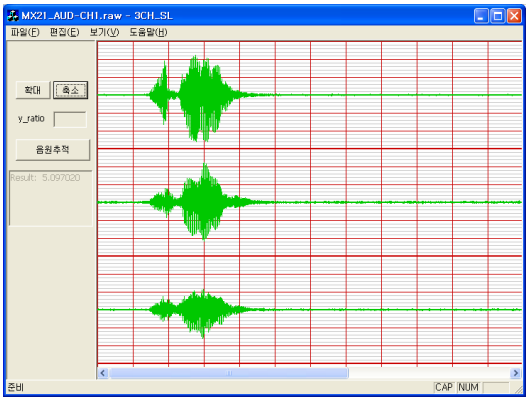

그림 1. 3개의 마이크에 들어온 음성 파형 (1m, 0도 발성)

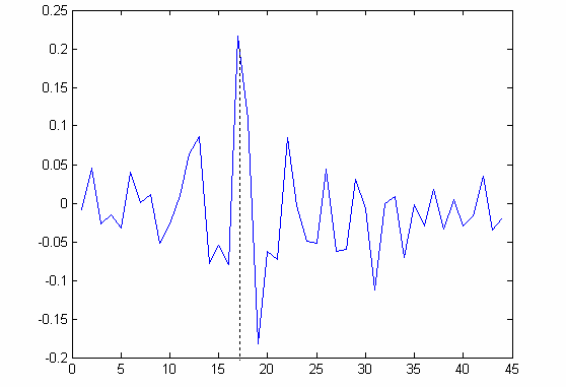

그림 2. 지연시간의 측정

2. GCC-PHAT 기반 음원 추적 방법

2.1 GCC-PHAT

GCC는 주파수영역에서의 상호상관 방법이다. 주파수 영역에서의 GCC-PHAT 기반의 음원 추적은 잡음 환경과 반향 환경에서 큰 장점을 가지고 있다.

그림 1은 로봇에 설치된 세 개의 마이크로부터 받아진 음성파형을 보여주며, 이 음성파형을 이용하여 마이크 간의 지연시간을 측정하게 된다.

GCC-PHAT 방법을 간략히 기술하면 다음과 같다. 두 개의 마이크에서 받은 신호 $x_1(n)$ 과 $x_2(n)$ 사이의 상호상관도는 다음 식에 의해 얻어진다.

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

$W(\omega)$ 는 주파수 가중 함수로써 $X_1(\omega)X_2^*(\omega)$ 의 역수이며 이 가중 함수를 PHAT [8]이라고 한다. PHAT은 시간지연을 추정함에 있어서 각 주파수의 상대적인 중요성을 결정하는 주파수에 종속된 가중치 된 함수이며, 식은 다음과 같이 표현 된다.

$$W(\omega) = \frac{1}{|X_1(\omega)X_2^*(\omega)|} \quad (2)$$

식 (1)을 통해서 구해진 $R_{x_1x_2}(n)$ 를 통하여 마이크 사이의 최종적인 지연 시간은 다음 식과 같이 구해 질 수 있다.

$$\tau = \arg\max R_{x_1x_2}(n) \quad (3)$$

그림 2의 가로축은 지연시간을 나타내며 세로축은 $R_{x_1x_2}(n)$ 를 나타낸다. 마이크 사이의 지연시간은 $R_{x_1x_2}(n)$ 값이 최대를 나타내는 가로축의 값이다.

2.2 추적각도 계산

다음은 GCC-PHAT에 의해 얻어진 여러 각도들의 정보를 이용하여 신뢰할 수 있는 추적각도를 추정하는 방법을 소개한다. 먼저 그림 3은 음원의 파장이 평면파 라고 가정할 경우 3개의 마이크가 그림과 같이 배치되어 있을 경우 2개 마이크 사이의 각도를 측정하는 방법을 보여주고 있다. 여기서 d 는 마이크간의 거리, v 는 음속도이며, τ_{12} 는 채널 1과 채널2간의 지연시간이다. 마이크 사이 세 개의 각도는 다음의 식 (4), (5), (6)으로 표현 할 수 있다.

$$\alpha = \cos^{-1}\left(\frac{v\tau_{12}}{d}\right) \quad (4)$$

$$\beta = \cos^{-1}\left(\frac{v\tau_{23}}{d}\right) \quad (5)$$

$$\gamma = \cos^{-1}\left(\frac{v\tau_{13}}{d}\right) \quad (6)$$

그림 4는 위의 식에서 구해진 3개의 마이크 사이 각도와 기하학적 특성을 이용하여 6개의 각도를 얻을 수 있는 방법을 보여주고 있다.

6개의 각도는 다음의 식으로 구해진다.

$$\Phi_{\alpha 1} = \alpha - 30,\ \Phi_{\alpha 2} = -\alpha - 30$$
$$\Phi_{\beta 1} = \beta + 90,\ \Phi_{\beta 2} = -\beta + 90 \quad (7)$$
$$\Phi_{\gamma 1} = \gamma + 30,\ \Phi_{\gamma 2} = -\gamma + 30$$

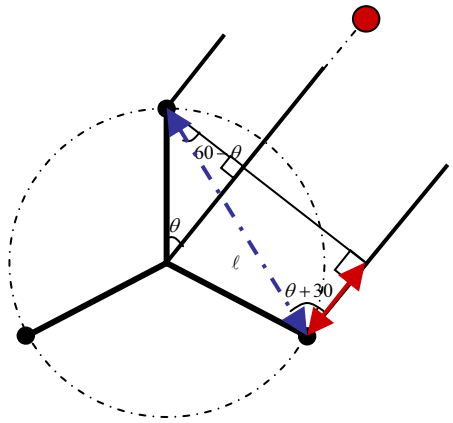

그림 3. 음원에 대한 마이크 사이의 각도측정

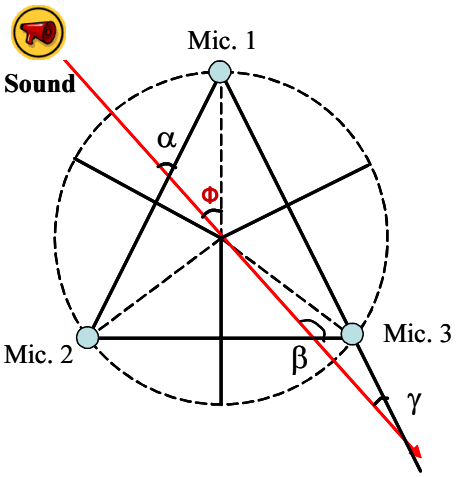

그림 4. 6개의 각도 계산 방법(3채널인 경우)

여기서 3개의 각도가 아닌 6개의 각도를 구하게 되는 이유는 음원이 들어오는 위치에 따라 α, β, γ 의해 추정되는 추적 각도는 각각 2개이기 때문이다.

최종적인 추적 각도는 6개의 각도 중 3쌍이 한 방향으로 모인다고 가정하여 이들 6개 각도 중 가장 적은 오차를 보이는 두 개의 각도를 얻어 평균값을 계산한다. 이렇게 함으로써 잘못 추적된 각도 값을 얻을지라도 신뢰할 수 있는 추적각도를 얻을 수 있다.

3. 음원추적용 데이터베이스

실제의 로봇환경에서 음원 추적 알고리즘의 성능을 평가하기 위하여 가정 환경처럼 꾸며진 테스트 베드에서 음원추적용 데이터베이스를 구축하였다. 그림 5는 ETRI지능형로봇연구단에서 제작된 네트워크기반 지능형 로봇인 WEVER-R2 이다. 또한, 그림 6은 마이크로폰이 3개일 경우 이들의 배치(120도 간격)를 보여주고 있다. 그림 6에서 보여진 마이크로폰은 ETRI에서 개발된 다채널 보드인 MIM (Multimodal Interface Module) 음원보드를 사용하였다. 음원추적용 데이터 베이스는 로봇호출 음성인“ 웨버” 를 사용하였으며, 3채널일 경우 두 명(M1, M2)에 의해 0~360도 범위에서 45도 간격으로 1~5m에서 각 m마다 3번씩 발성하여 총 120 set의 음성 샘플들을 각각 구축되었다 (CH3_M1, CH3_M2). 또한 2채널인 경우에는 0~180도에서 30도 간격으로 총 105 set의 음성샘플들을 각각 구축하였다 (CH2_M1, CH2_M2).

음성샘플들은 16kHz, 16bit, mono, PCM 형식으로 취득되었다. 그림 4는 로봇정면인 0° , 1m에서 로봇을 호출한 음성샘플의 예를 보여주고 있다.

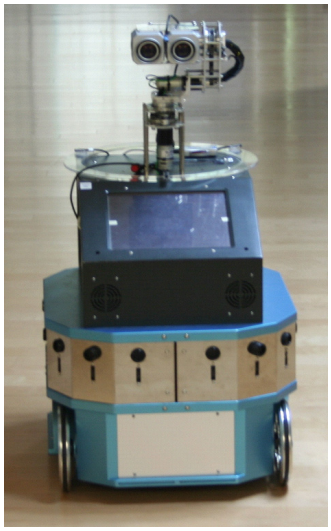

그림 5. ETRI의 지능형 로봇 “WEVER-R2”

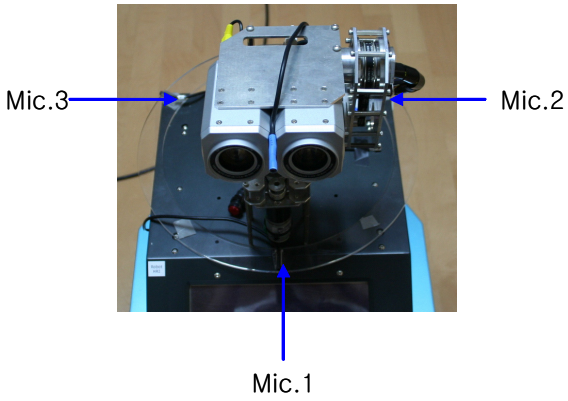

그림 6. 마이크로폰의 배치 (3채널인 경우)

표 1. CH3_M1 DB에 대한 성능비교

	FOV±10		FOV±15	
	성공률 (%)	평균 오차	성공률 (%)	평균 오차
TDOA (type 1)	42.5	3.3	52.5	5.0
TDOA (type 2)	66.6	4.2	75	5.15
GCC-PHAT (type 2)	88.3	2.9	98.3	3.8

표 2. CH3_M2 DB에 대한 성능비교

	FOV±10		FOV±15	
	성공률 (%)	평균 오차	성공률(%)	평균 오차
TDOA 변형(type 1)	36.7	3.8	45	5.3
TDOA 변형 (type 2)	62.5	3.9	69.2	4.4
GCC-PHAT (type 2)	89.2	4.2	94.2	4.6

4. 실험 및 결과

본 절에서는 음원추적용 데이터베이스를 이용하여 주파수영역의 GCC-PHAT과 시간영역의TDOA를 비교하였다. 여기서, TDOA는 계차기반의 음원추적 방법으로 경험적인 구간선택방법(type1)과 제안된 추적각도 추정방법(type2)이 비교 된다. 여기서 계차기반의 음원추적은 음성신호의 현재 샘플과 앞 샘플과의 차로 새로운 신호를 만들어서 사용 된다. 성능지표로써 추적성공률과 추적성공한 경우의 평균추적오차를 고려하였다. 추적성공률은 음원추적 후 호출자에게 다가가기 위해 얼굴검출과 연계해서 사용되어지기 때문에 FOV(±10, ±15)를 고려하였다. WEVER-R2에 부착된 로봇카메라는 FOV±24를 나타내고 있다.

먼저 3채널일 경우의 실험결과는 표1과 2와 같이 추적성공률과 이들에 대한 추적평균오차에 대한 성능을 비교하고 있다. 표에서 보는 바와 같이 CH3_M1과 CH3_M2 DB 모두 GCC-PHAT 기반(type2) 음원추적방법이 TDOA보다 좋은 성능을 나타내는 것을 알 수 있다. 또한 TDOA는 거리가 멀어질수록 성능저하가 나타나지만, GCC-PHAT의 경우는 5m까지 성능의 차이가 없음을 확인하였다.

표 3. CH2_M1 DB에 대한 성능비교

	FOV±10		FOV±15	
	성공률 (%)	평균 오차	성공률 (%)	평균 오차
TDOA변형 (type 2)	30.5	3.4	53.3	7.4
GCC-PHAT (type 2)	81.9	8.4	86.7	8.0

표 4. CH2_M2 DB에 대한 성능비교

	FOV±10		FOV±15	
	성공률 (%)	평균 오차	성공률 (%)	평균 오차
TDOA변형 (type 2)	25.7	3.5	46.7	7.5
GCC-PHAT (type 2)	81.0	8.2	84.8	8.0

2채널의 경우는 제안된 추적각도 추정방법(type2)만을 비교하였다. 표3과 4는 추적성공률과 이들에 대한 추적평균오차에 대한 성능 비교를 보여주고 있다. 표에서 보는 바와 같이 GCC-PHAT 기반 음원추적이 TDOA변형인 계차기반 음원추적 보다 35% (FOV±15)정도 더 좋은 성능을 나타내는 것을 알 수 있다.

5. 결 론

본 논문은 시간영역과 주파수영역의 대표적인 TDOA와 GCC-PHAT기반 음원추적 방법을 실제의 가정환경과 같은 테스트 베드에서 지능형 로봇에 적용하여 성능을 비교 및 분석하였다. 실제 가정환경에서 시간영역에서 구한 지연시간보다 주파수영역으로 변환하여 구한 지연시간이 더욱 정확한 정보를 가지고 있다는 것을 알 수 있었다. 또한 여러 개의 추정된 각도 정보로부터 신뢰성 있는 추적각도 방법의 유용성을 확인하였다. 이러한 음원추적 기술은 근거리 및 원거리에서 영상정보인 얼굴검출과 인식방법을 이용하여 효과적으로 시청각기반 음원추적 기술[9,10]로 발전시킬 수 있다.

참고문헌

[1] O. Deniz, M. Castrillon, J. Lorenzo, C. Guerra, D.Hernandez, M.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", Proceedings 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 Arrays: 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 smoothed 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 , p.7-11, 1997.

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

박 범 철

2006 서울산업대학교 전자정보공학과 (공학사)

2006~현재 과학기술연합대학원대학교 컴퓨터소프트웨어 및 공학 석사과정

관심분야 : 인간로봇상호작용, 패턴인식

반 규 대

2005 충북대학교 전기전자 및 컴퓨터공학부 (공학사)

2006~현재 과학기술연합대학원대학교 컴퓨터소프트웨어 및 공학 석사 박사 통합 과정

관심분야 : 인간로봇상호작용, 패턴인식

곽 근 창

1996 충북대학교 전기공학과 (공학사)

1998 충북대학교 전기공학과 (공학석사)

2002 충북대학교 전기공학과 (공학박사)

2003~2005 Dept, Electrical and Computer Engineering, University of Alberta, Post-Doc.

2005~2007 한국전자통신연구원 선임연구원

2007~현재 조선대학교 전자정보공과대학 제어계측로봇공학과, 전임강사

관심분야 : 인간로봇상호작용, 계산지능, 생체인식

윤 호 섭

1991 KIST SERI 인공지능부 연구원

1991 석사

1998 ETRI 컴퓨터소프트웨어 연구소 영상정보처리연구팀

2003 박사

2003~현재 ETRI 지능형로봇연구단 인간로봇상호작용 연구팀

관심분야 : 인간로봇상호작용, 영상처리, 패턴인식