

머리전달함수를 이용한 로봇의 화자 위치 추정 Detection of Speaker Position for Robot Using HRTF

황성목† · 박윤식* · 박영진**

Sungmok Hwang, Younsik Park and Youngjin Park

Key Words : Sound Source Localization(음원 위치 추정), Head Related Transfer Function(머리전달함수), Phase Criterion(위상 척도), Interaural Transfer Function(양이간 전달함수)

ABSTRACT

We propose a sound source localization method using the Head-Related-Transfer-Function (HRTF) to be implemented in a given platform. HRTFs contain not only the information regarding proper time delays but also phase and magnitude distortions due to diffraction and scattering by the shading object. Therefore, a set of HRTFs for any given platform provides a substantial amount of information as to the whereabouts of the source. In this study, we introduce new phase criterion in order to find the sound source location in accordance with the HRTF database empirically obtained in an anechoic chamber with the given platform. Using this criterion, we analyze the estimation performance of the proposed method in a household environment.

1. 서론

인간은 몇 가지 유용한 능력을 선천적으로 가지고 있다. 청각, 시각 등의 이른바 오감이 그 대표적인 예라 할 수 있다. 본 연구는 이 중에서 청각과 밀접한 연관이 있다. 인간은 3 차원 공간상에서 소리를 들었을 때, 소음원의 대략적인 위치를 본능적으로 알 수 있다. 인간에게 두 귀가 있듯이 로봇은 여러 개의 음향센서를 가질 수 있기 때문에 로봇 역시 이러한 능력을 가질 수 있다. 이러한 기술을 음원 위치 추정기술이라 한다. 이는 각 마이크로폰에 도달한 음성 신호를 이용하여 공간상에 존재하는 물리적인 음원의 위치를 추정하는 기술이다. 이를 통해 로봇이 인간의 위치를 파악하고 그 방향으로 머리를 돌리거나 이동하는데 사용됨으로써 로봇이 인간의 명령을 인식하고 수행할 준비가 되어 있음을 보여주는 로봇과 인간과의 효과적인 상호작용의 한 단계를 이루는 중요기술 중의 하나이다. 또한, 청각을 이용한 음원 위치 추정 기술은 시각 정보와 혼합을 통해 보다 빠르고 정확한 추정 성능을 지닐 수 있는 상호보완적인 관계를 이룬다.

현재 음원 위치 추정을 위해서 도착지연시간 방법(Time Delay Of Arrival)⁽¹⁾, 빔형성 방법 (beamforming)⁽²⁾, 고해상도 스펙트럼 방법(spectral analysis)⁽³⁾, 회전형 구조와 적응형 지연시간 추정방법⁽⁴⁾, Ray Tracing Formula⁽⁵⁾ 등이 개발되어 사용되고 있다. 그런데 로봇과 같이 음원과 센서의 상대적인 위치가 실시간으로 변하고, 센서의 부착위치가 플랫폼의 형상에 의해 제약을 받는 시스템의 경우 상호상관계수(cross-correlation)에 기반한 도착지연시간 방법이 가장 널리 사용되고 있다. 그러나 이런 방법들은 음원에서 방사된 음파가 마이크로폰에 도달하기까지 플랫폼의 형상에 의한 반사, 산란, 회절 등의 영향을 고려하지 못한다는 단점이 있다. 그러므로 보다 정확한 음원 위치추정을 위해서는 플랫폼의 형상을 고려할 필요가 있고, 본 연구에서는 대상 플랫폼의 머리전달함수(Head Related Transfer Function: HRTF)를 사용함으로써 이런 문제점을 해결하고 보다 정확한 음원 위치 추정이 가능한 방법을 제안한다.

2. 대상 플랫폼의 머리전달함수

완전한 구체와 거리가 무한한 점음원(ideal point source)을 가정했을 때, 주파수 영역의 이론적인 머리전달함수는 이미 19C 말 Lord Rayleigh 에 의해 밝혀졌다^(6,7,8). 머리전달함수는 본 연구의 핵심내용이 되므로 먼저 플랫폼의 머리전달함수를 측정하고

† 책임저자; 한국과학기술원
E-mail : tjdahr78@kaist.ac.kr
Tel : (042) 869-3060, Fax : (042) 869-8220
* 한국과학기술원
** 한국과학기술원

분석하는 것이 선행되어야 한다. 이에 무향실 (anechoic chamber)에서 음원의 거리, 수평각, 고도각을 달리하면서 플랫폼의 각 센서 위치의 머리전달 함수를 측정하였다. 그림 1.은 화자의 위치 추정을 위해 플랫폼에 부착되어 있는 센서의 위치를 나타낸다.

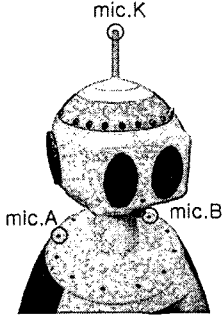


그림 1. 화자 위치 추정을 위한 마이크 배치

본 논문에서는 수평각 추정을 위해 mic.A 와 mic.B 를 사용하고 동시에 고도각 추정을 위해 mic.K 를 같이 사용한다. 음원의 거리가 1m 인 경우, 고도각이 0°인 수평면상에서 음원의 반시계 방향으로 이동함에 따른 각 마이크 머리전달 함수의 unwrapped phase 는 그림 2., 그림 3. 과 같다. mic.K 의 경우에는 음원의 수평각이 달라지더라도 음원과 마이크 사이의 상대적인 위치는 거의 동일하므로 HRTF phase 는 거의 일치한다. 그러나 mic.B 의 경우에는 음원이 반시계 방향으로 90°까지 이동함에 따라 마이크와 음원 사이의 거리가 가까워 지므로 수평각에 따라 phase 가 달라진다. 특히 phase 의 기울기 즉, group delay 가 수평각이 증가할수록 작아짐을 보이는데 이는 음원에서 방사된 음파가 각 마이크에 도달하기까지의 지연시간에 대한 정보를 담고 있기 때문이다.

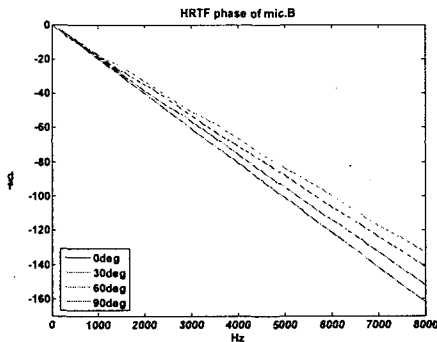


그림 2. mic.B 의 HRTF phase plot

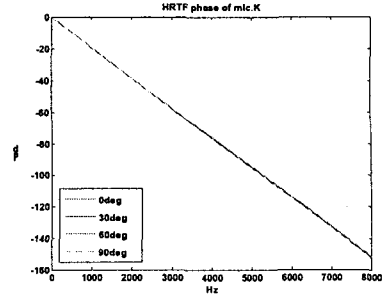


그림 3. mic.K 의 HRTF phase plot

수평각의 경우와 같이 고도각이 달라짐에 따라 음원과 마이크의 상대적인 위치가 변하게 되고, 플랫폼 형상에 의한 영향이 달라지기 때문에 고도각에 따라 각 마이크 HRTF 의 unwrapped phase 는 달라진다.

3. 위치 추정 알고리즘

앞에서 살펴본 것과 같이 HRTF 는 음원의 수평각, 고도각에 전적으로 의존한다. HRTF 의 이러한 성질을 이용하여 본 연구에서는 다음과 같은 위상 척도(phase criterion)를 이용하여 음원 위치를 추정하는 방법을 제안한다.

$$e_{phase} = \int \gamma^2(f) \{ \theta(f) - \theta_{HRTF}(f) \}^2 df \quad (1)$$

γ^2 : coherence between two mic. outputs

θ : ITF phase

θ_{HRTF} : phase difference between two mic. HRTFs

노이즈가 없는 이상적인 환경에서는 두 마이크간의 전달함수(사람이나 더미헤드의 경우에는 Interaural Transfer Function: ITF 라 부른다)의 위상은 각 마이크 HRTF phase 의 차이가 된다. 따라서 ITF 의 위상 (θ)을 산출한 후, 이를 각각의 음원의 위치에 대해 미리 측정된 HRTF 의 위상차(θ_{HRTF}) 데이터베이스

와 관심 주파수 영역에서 비교하여 e_{phase} 를 최소화 하는 θ_{HRTF} 를 찾음으로써 음원의 위치를 추정할 수 있다. 이때, coherence function 을 가중치함수 (weighting function)로 사용함으로써 각 주파수 별 가중치를 달리한다. Coherence 가 좋다는 것은 각 마이크로폰에 들어온 신호의 선형관계가 만족된다는 것이고, 비연관된(uncorrelated) 노이즈의 영향이 적다는 것이므로, 각 주파수별 coherence 를 가중치로 사용함으로써 측정잡음 등의 비연관된 노이즈의 영

향을 줄일 수 있다. 또한 로봇은 화자의 음성으로부터 화자의 위치를 추정해야 하므로 사람의 대략적인 음성 주파수 영역(voice frequency band)인 300Hz ~ 4kHz 를 관심 주파수 영역으로 선정했다.

4. 무향실 실험을 통한 검증

수평각 추정을 위해 수평면 상에서 식 (1)을 이용하여 실제 음원 위치에 대해 보유하고 있는 HRTF 데이터베이스를 이용하였을 때의 mic.A 와 mic.B 사이의 위상차도(phase criterion)를 그림 4.에 나타내었다. 실제 음원 위치 부근의 HRTF 를 이용했을 때 위상차도의 값이 적게 나오는 것을 확인할 수 있다. 그러나 실제 음원 위치와 거의 대칭되는 각도에서도 위상차도의 값이 작게 나오는데, 이는 위상차만을 이용하면 앞-뒤 구분이 모호해짐을 의미한다. 수평각 추정과 마찬가지로 수평각 30° 인 경우, 고도각 변화에 따른 위상차도를 그림 5.에 나타내었다.

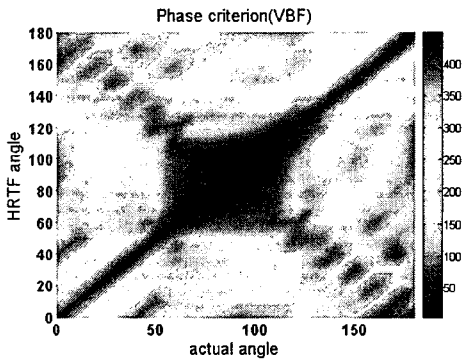


그림 4. 수평각 추정을 위한 mic.A 와 mic.B 의 위상차도

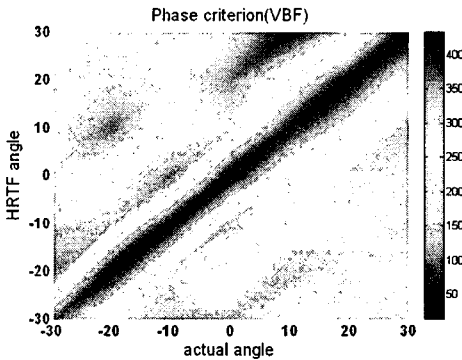


그림 5. 수평각이 30° 인 경우, 고도각 추정을 위한 mic.K 와 mic.B 의 위상차도

결과에서 알 수 있듯이 수평각 추정과 달리 실제 음원의 고도각에 해당하는 HRTF 각도에서 뚜렷하게 위상차도가 작아짐을 볼 수 있다. 이는 양이간 축(interaural axis)을 기준으로 로봇의 앞-뒤는 거의 대칭인데 반해, 로봇의 위-아래는 대칭이 아니기 때문에 나타나는 결과이며, 로봇 플랫폼의 형상에 의한 음장의 왜곡 현상을 HRTF 가 반영하고 있기 때문으로 해석된다.

5. 실사용 환경에서의 실험

실제 로봇이 사용되는 환경은 무향실과는 달리 주변잡음, 반사, 간향 등의 노이즈가 존재하는 환경이므로 이런 환경에서 제안하는 방법의 성능을 검증하기 위해 실사용 환경에서의 실험이 필요하다. 그림 6. ~ 그림 11. 은 실제 사용환경에서 로봇과 약 1m 의 거리를 둔 화자가 서있는 경우(고도각 약 20°), 의자에 앉은 경우(고도각 약 0°), 바닥에 앉은 경우(고도각 약 -20°)에 대해서 수평각을 변화시키면서 음성 신호를 방출했을 때, 로봇이 화자의 수평각과 고도각을 추정한 결과이다. 각 위치별 15 회씩 반복하여 실험하였으며, 그림에서 점의 크기는 그 위치를 추정한 횟수에 비례한다.

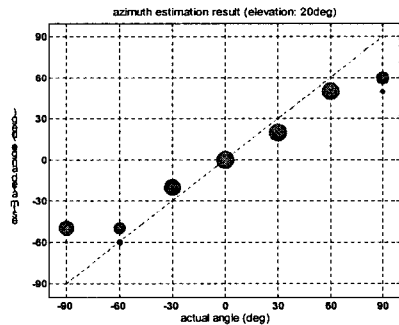


그림 6. 고도각 20° 인 경우 화자의 수평각 추정

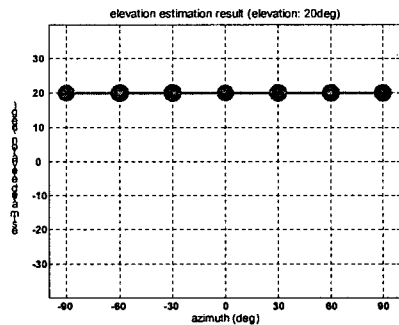


그림 7. 고도각 20° 인 경우 화자의 고도각 추정

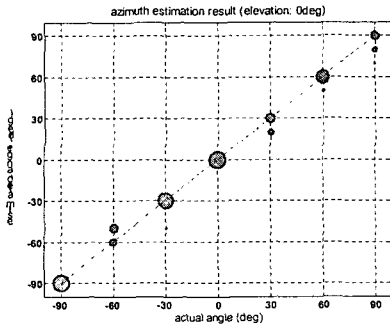


그림 8. 고도각 0° 인 경우 화자의 수평각 추정

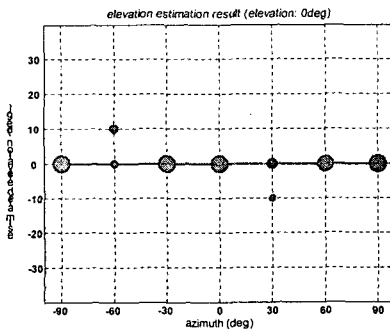


그림 9. 고도각 0° 인 경우 화자의 고도각 추정

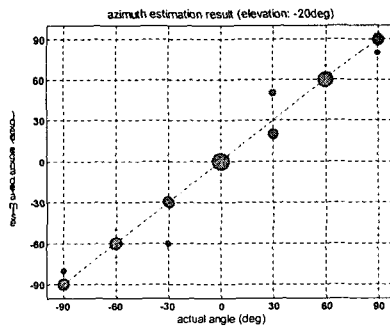


그림 10. 고도각 -20° 인 경우 화자의 수평각 추정

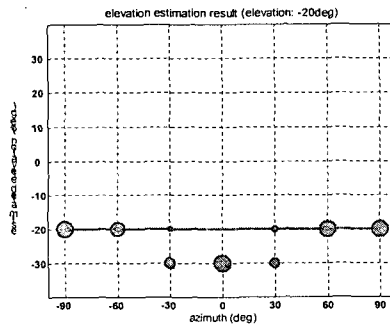


그림 11. 고도각 -20° 인 경우 화자의 고도각 추정

위 결과에서 고도각이 20°인 경우에 수평각이 ± 90°일 때 추정 에러가 가장 큰 것을 확인할 수 있다. 이는 화자가 이 위치에 있는 경우 mic.A 혹은 mic.B가 플랫폼에 의해 가려지게 되어 마이크로 들어가는 신호의 크기가 작아지게 되고, 상대적으로 노이즈의 영향을 많이 받기 때문으로 해석된다. 그러나 전반적으로 수평각은 약 20°의 오차, 고도각은 약 10°이내의 오차로써 화자의 위치를 추정함을 확인할 수 있다.

6. 결론

본 연구에서는 로봇의 화자 위치 추정을 위해 미리 측정된 플랫폼 HRTF를 이용하는 방법을 제안하였다. 이를 위해 먼저 대상 로봇 플랫폼의 HRTF를 무향실에서 측정하여 데이터베이스를 구축하고, 수평각과 고도각에 따른 HRTF의 특징을 분석하였다. 그리고 HRTF의 위상(phase)정보를 이용한 위상척도(phase criterion)를 사용함으로써 실제 사용환경에서 3차원 공간상의 화자의 위치를 추정할 수 있음을 보였다.

후기

이 연구는 산업자원부의 21C 프론티어사업(인간기능생활지원지능로봇사업)과 과학재단의 국가지정연구실사업(3차원 입체음향 기술), BK 21 사업의 지원을 받아 수행되었습니다.

참고문헌.

- (1) M. S. Brandstein & H. F. Silverman, "A robust method for speech signal time-delay estimation in reverberation rooms," Proc. ICASSP-97, vol. 1, pp. 375-378, April, 1997.
- (2) M. Wax & T. Kailath, "Optimum localization of multiple sources by passive array," IEEE Tran. On Acoustics, Speech and Signal Processing, vol. 31, no.5, pp. 1210-1217, Oct, 1983.
- (3) R. Schmidt, "A signal subspace approach to multiple emitter location and spectral estimation," Ph. D thesis, Stanford University, Stanford, CA, USA, MUSIC, 1981.
- (4) Doh-Hyoung Kim, Youngjin Park, "Development of Moving Sound Source Localization System", Technical Acoustica, 2005 (accepted).
- (5) Woodworth, R. S. and Schlosberg, G., 1962, Experimental Psychology.
- (6) J. W. Srtutt, "On the acoustic shadow of a sphere," Philos. Trans. R. Soc. London, Ser. A 203, 87-89., 1904.
- (7) J. W. Srtutt, The Theory of sound (Dover, New york), 2nded., vol. 1 and 2., 1945.
- (8) R. O. Duda & W. L. Martens, "Range dependence of the response of a spherical head model," Journal of Acoustic Society of America, 104 (5), November, 1998.