

휴머노이드 로봇을 위한 공간상의 화자 위치 측정

김진성, 김의현, 김도익, 유범재
한국과학기술연구원 인지로봇연구원

Spatial Speaker Localization for a Humanoid Robot

Jinsung Kim, Ui-Hyun Kim, Doik Kim, and Bum-Jae You
Center for Cognitive Robotics Research, Korea Institute of Science and Technology

Abstract - 과학기술의 발전과 함께 인간형 서비스 로봇에 대한 관심이 고조되고 있다. 서비스 로봇의 핵심 중의 하나는 인간과의 상호작용이라 할 수 있다. 자연스러운 상호작용을 위하여, 화자를 바라보고, 깨끗한 음성신호를 얻는 과정에서 음원의 위치 측정은 필연적이다. 본 논문은 마이크로폰에 도달되는 동일 신호의 도착시간지연(Time Delay of Arrival: TDOA) 특성 행렬을 정의하고, 이를 이용하여 공간상의 화자 위치 측정 방법론을 제안하였다. 휴머노이드 로봇의 머리에 마이크로폰 배열을 구성하였고, 실제 시스템을 통한 실험을 통하여 방향 감지 및 높이 구분을 실행하였다.

1. 서 론

과학기술의 발전과 함께, 인간의 삶이 풍요로워 짐에 따라 인간의 삶의 질 향상에 관련된 시스템의 연구가 다양하게 이루어지고 있다. 최근 들어 지능형 공간의 실현을 위해 위치 추적이나 음성 및 영상 정보를 이용하여 화자 혹은 청자의 위치를 추적하고 이 정보를 이용하여 사용자에게 적합한 서비스를 제공하기 위한 연구가 진행되고 있다. 사용자에게 일방적인 도움을 주거나 정보를 제공하는 단 방향적 로봇이 아닌 사용자와 로봇간의 커뮤니케이션이 가능한 인간 친화형 서비스 로봇 개발에 관심이 집중되고 있다[3]. 따라서, 인간 친화적이고 원활한 상호작용을 위하여 청각 시스템의 연구는 필연적이라 할 수 있다. 음원의 위치 측정(Sound Source Localization)이란, 음원에서 발생하는 음파를 마이크로폰으로 측정하고, 측정된 신호를 이용하여 음원의 거리, 고도와 방위각을 측정해 내는 것으로 정의된다. 이에 대한 주된 방법은 두 마이크로폰 각각에 도달하는 신호의 강도차(Interaural Level Difference: ILD)와 시간차(Interaural Time Difference: ITD)를 이용하여 음원의 위치를 측정하는 방법이 있다. 강도차를 이용한 방법은 다수의 마이크로폰에 입력되는 신호 중 가장 큰 에너지를 가지고 있는 신호의 방향을 음원의 방향으로 결정하는 방법으로서, 시스템이 간단해지는 장점이 있지만, 공간상의 위치 측정 능력이 떨어지고, 많은 수의 마이크로폰을 필요로 하기 때문에 효율적이지 못하다. 시간차를 이용하는 방법은 한 쌍의 마이크로폰에 입력되는 신호의 도착 시간차, 즉 도착시간지연을 이용하여 위치를 측정하는 것이다. 마이크로폰에 도착되는 신호의 시간차를 두 신호의 상관성(Cross-Correlation)에 적용하여, 공간상의 음원 위치를 측정한다. 이론적으로는 3개의 지연시간 추정치만으로도 3차원 공간에서 유일한 음원의 위치를 결정할 수 있지만, 음원의 위치를 계산하기 위하여 비선형 수식을 풀어야 하기 때문에 정확한 음원 위치를 추정하기가 어렵고, 따라서 좀 더 많은 수의 도착시간지연 추정치를 이용하게 된다[2,3].

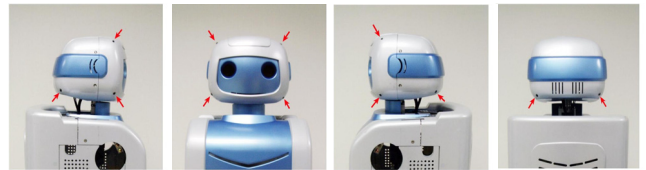
본 논문에서는 휴머노이드 로봇의 두상에 대한 마이크로폰 배열을 제안한다. 이에 대한 방위각 및 높이의 도착시간지연 분포를 살펴보고, 여기에 적합한 도착시간지연 특성 행렬을 제안한다. 도착시간지연 특성 행렬이란, 임의의 위치에서 발생된 입력신호에 대하여, 목적에 맞게 배열된 센서(마이크로폰) 쌍 마다 발생한 도착지연시간 값을 성분으로 구성된 행렬로서 일종의 데이터베이스이다. 도착시간지연 데이터베이스를 이용한 방법이 [2]에서 소개되었으나, 본 논문은 휴머노이드 로봇의 시청각 통합 시스템에 적합하도록 구체화된 도착시간지연 특성 행렬을 제안하고, 이를 기반으로 화자 위치의 방위각 및 높이 구분 방법론을 실제 휴머노이드 로봇의 청각 시스템으로 구현하는데 중점을 두고자 한다.

2. 본 론

2.1 휴머노이드 로봇의 두상에 대한 마이크로폰 배열

마이크로폰 배열은 음원의 위치 측정을 위한 기본적인 작업으로서 음원 측정 시스템의 목적에 적합하게 설계하여야 한다. 본 연구의 궁극적인 목적은 휴머노이드 로봇에 인간 친화적인 청각 시스템을 구현하는 것이다. 형태적인 면에서 보면, 청각 시스템은 인간과 마찬가지로 머리의 좌우 양측에 마이크로폰이 한 개씩 있어야 한다. 실제로 마이크로폰을 휴머노이드 머리의 좌우측 귀부분에 각각 1개씩 설치하여 실행하는 경우가 있다. 그러나, 마이크로폰 2개로는 인간과 같이 공간상의 음원

위치 측정을 하는 기능을 발휘하기에는 어려움이 있다. 본 논문의 마이크로폰 배열은 인간의 형태적인 면에서 좀 부족하더라도 기능적인 면을 보장하고자 마이크로폰을 6개를 사용한다. 비록 인간의 형태와 유사한 2개의 마이크로폰을 사용하지는 않지만, 그림 1과 같이, 마이크로폰 3개를 두상의 좌우 측 귀 위치에 설치하였다. 형태 및 기능적인 면에서 인간과 유사한 청각구조로 가기위한 중간 단계로 볼 수 있겠다.



〈그림 1〉 휴머노이드 로봇의 두상에 장착된 마이크로폰 배열

2.2 상호파워스펙트럼을 이용한 도착시간지연 추정

음원의 위치 측정 과정 중에서, 마이크로폰쌍에 대한 도착시간지연 값의 정확한 검출은 매우 중요하다. 주변 잡음 및 음향 반향 등의 장애 요소들의 영향을 최소화하여 가급적 음원 위치 추정 오차를 최소화 할 수 있는 다양한 알고리즘이 제안되고 있다. 대표적인 알고리즘으로 마이크로폰 수신 신호간의 상관관계를 이용한 상호상관관계(Cross-Correlation) 방법과 신호의 상관성에 가중치를 부과하는 일반화된 상호상관관계(Generalized Cross-Correlation : GCC) 방법 등이 있다[1]. 본 논문은 GCC 함수의 가중함수에 따라 다양한 방법 중의 하나로, 상호파워스펙트럼(Cross-Power Spectrum Phase : CSP) 방법을 이용하여 TDOA를 구한다.

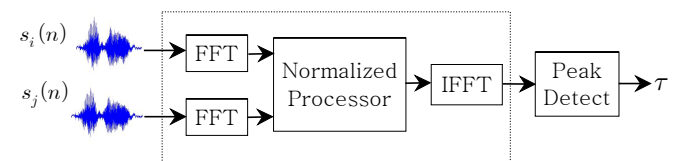
i 번째 마이크로폰으로 입력된 신호를 $s_i(n)$ 이라 할때, 상호파워스펙트럼 계수는 FFT(Fast Fourier Transform)와 IFFT(Inverse Fast Fourier Transform)를 이용하여 다음과 같이 정의 된다:

$$CSP_{ij}(k) = IFFT \left[\frac{FFT[s_i(n)] \cdot FFT[s_j(n)]^*}{\|FFT[s_i(n)]\| \cdot \|FFT[s_j(n)]\|} \right] \quad (1)$$

상호파워스펙트럼 계수를 구하여 최대값을 갖는 k 를 검출하여 도착시간지연 τ 를 추정한다:

$$\tau = \arg \max (CSP_{ij}(k)). \quad (2)$$

일련의 과정을 정리하면 그림 2와 같다.



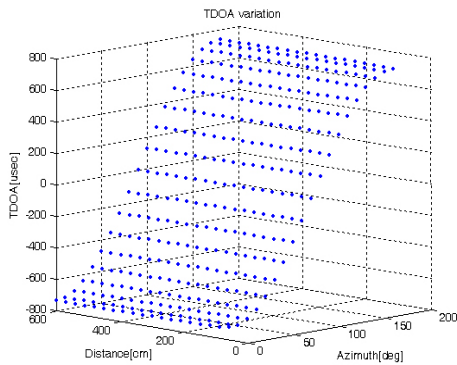
〈그림 2〉 상호파워스펙트럼을 이용한 도착시간지연 검출

2.3 도착시간지연 특성행렬을 이용한 시스템 구현

본 논문은 검출된 도착시간지연의 값들로 구성된 데이터베이스로서 도착시간지연 특성행렬을 제안한다. 본 절에서는 복잡한 비선형 수식 및 근사 수식을 이용하지 않고, 도착시간지연 특성 행렬을 이용하여 방향 감지 및 높이 구분을 하는 간단한 방법을 제안한다. 본 연구는 휴머노이드 로봇의 시청각 시스템 통합을 가정한다. 일반적으로 카메라의 시야각(Horizontal Field of View, HFOV)은 50°를 넘는다. 따라서, 휴머노이드

로봇의 시청각 통합 시스템에 가장 간단하고 적절하도록, 청각 시스템에 의한 방향 감지 분해능을 10°로 하여 시스템을 구현한다.

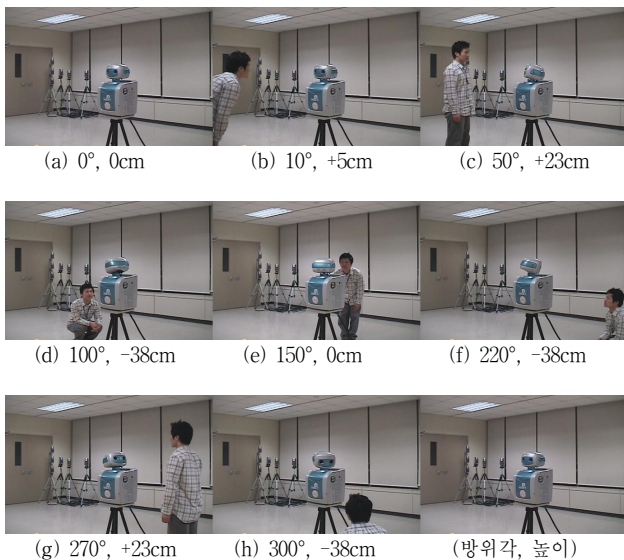
제안된 마이크로폰 배열에 대한 도착시간지연 값은, 음원의 거리보다는 방위각에 대하여 변화율이 크고, 거리가 멀어질수록 변화율이 작아진다. 이는 그림 2를 통하여 알 수 있다.



〈그림 2〉 임의 방위각 음원의 거리에 대한 도착시간지연 분포

본 논문은 이러한 도착시간지연 분포를 이용하여 도착시간지연 특성 행렬을 제안한다. 마이크로폰의 성능에 대한 유효거리를 고려하여, 휴머노이드 로봇의 음원 위치 측정 반경을 약 5m 이내로 가정한다. 음원방위각에 대한 거리별 도착시간지연 분포를 기반으로 하여, 2m 지점의 도착시간지연 값을 기준값으로 선택하여 도착시간지연 특성 행렬을 구성하였다. 도착시간지연 특성행렬의 행은 음원의 방위각 및 높이를 의미하고, 열은 방위각 및 높이 구분에 적절하게 지정된 마이크로폰 쌍의 순서를 의미한다. 특성 행렬의 각 성분은 각 방위각 및 높이에 대한 해당 마이크로폰 쌍의 도착시간지연 값을 나타낸다. 임의의 위치에서 음원이 발생하면, 각 방위각 및 높이에 대한 도착시간지연의 오차 절대값의 최소값을 갖는 행을 선택하여 음원의 방위각 및 높이 구분을 하게된다.

실험 환경은 일반 사무실 공간으로서 에어컨 및 컴퓨터의 잡음이 55dB을 유지하는 환경에서 이루어졌다. 휴머노이드 두상에 의한 입력 신호의 변형 및 반향을 인정하고 실험을 하였다. NI4472 DAQ보드, Aphex-207 프리앰프와 MKE 2P-C 마이크로폰을 사용하였다. 음원의 위치추정을 위한 전처리 과정으로 음성구간 검출방법[4]을 사용하여, 음성 신호에 대한 화자에게만 반응하도록 하였다. 각 10° 방위각(36회) 별 높이(상중하, 3번)에서 20번씩 반복실험을 하여, 총 2,160번(36×3×20)을 실시 하였다. 결과는 그림 3과 표 1과 같다.



〈그림 3〉 임의 방위각 및 높이의 음원에 대한 위치추정

〈표 1〉 음원의 위치 측정 결과

방위각 0°~360°	높이		
	상(+20cm)	중(0cm)	하(-20cm)
97.27%	94.25%	99.61%	95.63%

3. 결론 및 고찰

본 논문은 임의의 방위각 감지 및 높이 구분을 위한 마이크로폰 배열을 휴머노이드 로봇의 두상에 제안하였다. 이에 적합한 도착시간지연 특성행렬을 제안하고, 정의하여 실제 휴머노이드 로봇의 청각 시스템에 적용하였다. 논문은 복잡한 비선형 수식 및 근사 수식을 사용하지 않고, 미리 정의된 도착시간지연 특성 행렬을 이용하여, 방위각 감지 및 높이 구분이 가능한 실제 휴머노이드 로봇을 구현하였다는 것에 의미를 둔다.

【참고 문헌】

- [1] Charles H. Knapp and G. Clifford Carter, "The Generalized Correlation Method for Estimation of Time Delay", IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-24, No.4, pp.320-327, 1976
- [2] Jie Huang, Katsunori Kume, Akira Saji, Masahiro Nishihashi, Teppei Watanabe and Willian L. Martens, "Robotic Spatial Sound Localization and Its 3-D sound Human Interface", Proc. of the First International Symposium on Cyber Worlds, pp. 191-197, 2002
- [3] 송민규, "시청각 정보를 이용한 개인 추적 방법 연구", 전북대학교 전자정보통신공학과 석사학위 논문, 2006
- [4] J. Sohn, N. S. Kim, and W. Sung, "A Statistical Model-Based Voice Activity Detection," IEEE Signal Processing Letters, vol. 6, no. 1, pp. 1-3, 1999.