

도착시간지연 특성 행렬을 이용한 음원의 위치 측정

김진성^{1,2}, 김의현¹, 유범재¹

¹한국과학기술연구원 인지로봇연구단, ²고려대학교 전기공학과

Sound Source Localization using Time Delay of Arrival Feature Matrix

Jinsung Kim^{1,2}, Ui-Hyun Kim¹ and Bum-Jae You¹

¹Center for Cognitive Robotics Research, Korea Institute of Science and Technology

²Department of Electrical Engineering, Korea University

1 Abstract - 최근들어 유비쿼터스 환경에서 인간 친화적 인터페이스를 갖는 서비스 로봇에 대한 연구가 매우 활발하다. 인간과 자연스러운 상호작용을 위하여, 화자를 바라보고, 깨끗한 음성신호를 얻는 과정에서 음원의 위치 측정은 필수적이다. 본 논문은 마이크로폰에 도달되는 동일 신호의 도착 시간지연(Time Delay of Arrival: TDOA) 특성 행렬을 정의하고, 이를 이용하여, 음원의 위치 측정 시스템을 구현한다. 휴머노이드 로봇에 구현된 실제 시스템을 통하여 방향 감지 및 높이 구분이 잘 이루어짐을 확인한다.

1. 서 론

인간의 삶이 풍요로워짐에 따라 인간의 삶의 질 향상에 관련된 시스템의 연구가 다양하게 이루어지고 있다. 그 예로써 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. 최근에는 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. 음원에서 발생하는 음파를 마이크로폰으로 측정하고, 측정된 신호를 이용하여 음원의 거리, 고도와 방위각을 알 수 있는데, 이것을 음원의 위치 측정(Sound Source Localization)이라고 한다. 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].

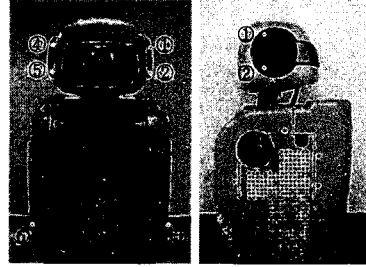
본 논문은 휴머노이드 로봇에 장착된 마이크로폰 배열에 대한 방위각 및 높이의 도착시간지연 분포를 살펴보고, 도착시간지연 특성 행렬을 제안한다. 이를 이용하여 음원의 방향 감지 및 높이 구분을 실제 시스템에 적용하여 실행한다.

2. 본 론

음원의 위치 측정의 과정은 간단히 두 가지 단계로 정리할 수 있다. 첫 번째로 두 개 이상의 마이크로폰에 수신되는 음원 신호로부터 TDOA를 추정해 내는 것이고, 그 다음 단계는 추정된 TDOA를 이용하여 배열된 마이크로폰 위치를 기하학적으로 해석하여 음원의 위치를 측정해내는 것이다.

2.1 마이크로폰 배열

음원의 위치 측정을 위한 기본적인 단계로 마이크로폰 배열을 음원 측정 시스템의 목적에 적합하게 설계하여야 한다. 본 연구의 궁극적인 목적은 휴머노이드 로봇에 인간 친화적인 청각 시스템을 구현하는 것이다. 인간 친화적이라는 의미는 형태적인 면과 기능적인 면에서 인간과 가장 유사하다는 의미로 해석될 수 있다. 형태적인 면으로 볼 때 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 연구에서 사용한 마이크로폰 배열은 인간의 형태적인 면에서 좀 부족하더라도 기능적인 면을 보강하고자 마이크로폰을 6개를 사용한다. 마이크로폰 3개씩 좌우 측 귀 위치에 설치하면, 형태 및 기능적인 면에서 어느 정도 인간과 유사한 구조가 될 수 있겠으나, 음원의 높이를 측정하는 기능적인 면을 고려하여 마이크로폰 2개는 좌우 몸통 부분에 설치하였다. 형태 및 기능적인 면에서 인간과 유사한 청각구조로 가기 위한 중간 단계로 볼 수 있겠다. 구체적인 마이크로폰 배열의 모습은 그림 1과 같다.



〈그림 1〉 휴머노이드 로봇의 마이크로폰 배열

2.2 상호파워스펙트럼을 이용한 TDOA 추정

음원의 위치 측정 과정 중에 TDOA의 정확한 검출은 매우 중요한 의미를 가진다. 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.

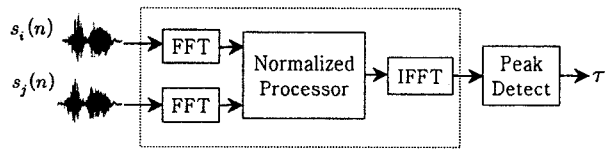
2번째 마이크로폰으로 입력된 신호를 $s_2(n)$ 이라 할때, CSP 계수는 FFT(Fast Fourier Transform)와 IFFT(Inverse Fast Fourier Transform)를 이용하여 다음과 같이 정의 된다:

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

CSP 계수를 구하여 최대값을 갖는 k 를 검출하여 TDOA τ 를 추정한다:

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

일련의 과정을 정리하면 그림 2와 같다.



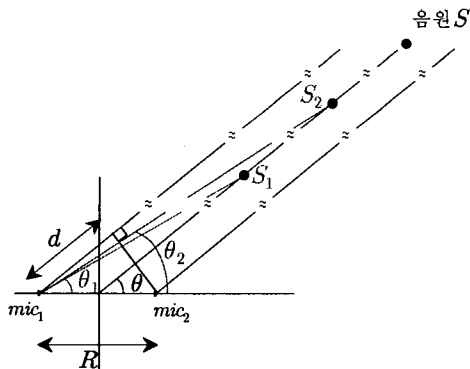
〈그림 2〉 CSP 방법을 이용한 피크 검출

2.3 TDOA 특성 행렬

검출된 TDOA를 가지고 음원의 위치 측정을 하는 방법은 다양하다. 본 절에서는 복잡한 비선형 수식 및 근사 수식을 사용하지 않고, TDOA를 이용하여 방향 감지 및 높이 구분을 하는 간단한 방법을 제안한다. 본 연구는 휴머노이드 로봇의 시청각 시스템 통합을 가정하여, 휴머노이드 로봇의 방향 감지 분해능을 10°로 하고, 방위각은 전방 기준 ±90°에 국한한다.

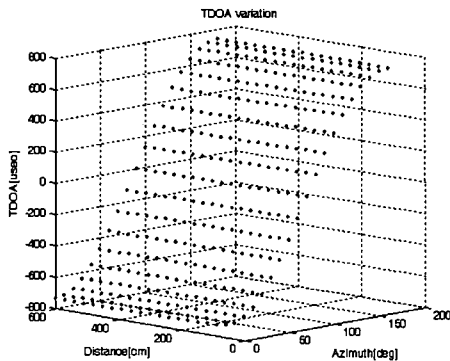
제안된 마이크로폰 배열은 그림 1과 같이 정면을 기준으로 좌우 대칭을 이루며, 머리에 2쌍과 몸통에 1쌍이 있다. 일반적으로 그림 3과 같이 음원이 마이크로폰과 매우 멀리 떨어져 있다는 가정 하에 각 마이크로폰에 입력되는 신호 역시 평행하게 도달된다고 가정한다. 음파의 속도 c 는 상온에서 약 340m/s 이므로, 동일한 음원으로부터 2개의 마이크로폰에 음이 도달되는 데는 약간의 시간차 $\tau (=d/c)$ 가 발생한다. 이로부터 방위각 θ 는 다음과 같은 관계식을 통해 구해진다:

$$\theta \approx \cos^{-1} \left(\frac{rc}{R} \right). \quad (3)$$



〈그림 3〉 기하학적 관계의 방위각 추정

그러나, 근사화 시키지 않은 실제 방위각을 고려해보자. 예를 들면, 그림 3의 마이크로폰 간격 R 이 30cm 이고, 음원 S 의 방위각 θ 가 45° 인 경우, 음원의 거리를 1m로 가정하면, θ_1 은 39.52° 가 되고, 5m로 가정하면, θ_2 는 43.81° 가 된다는 사실을 알 수 있다. 1번과 4번 마이크로폰의 방위각 별 음원 거리에 대한 TDOA를 조사하여 분포를 살펴보면, 그림 4와 같다.

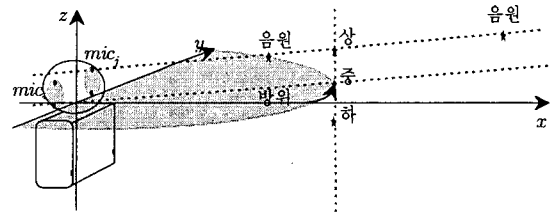


〈그림 4〉 임의 방위각 음원의 거리에 대한 TDOA 분포

그림 4는 음원의 거리보다는 방위각에 대해서 TDOA의 변화율이 큼을 보여주고, 거리가 멀어질수록 TDOA 변화율이 작아짐을 알 수 있다. 본 논문은 이러한 TDOA 분포를 이용하여 TDOA 특성 행렬을 제안하고, 시스템을 위해 필요한 마이크로폰 쌍의 개수이다. 구체적으로는 1번과 4번, 2번과 5번, 6번과 3번, 4번과 2번 그리고 5번과 1번 마이크로폰 쌍을 의미한다. 방위각 TDOA 특성행렬의 각 성분으로는 해당되는 마이크로폰 쌍으로부터 구해진 2m 지점의 TDOA 값이 들어간다. 높이에 대한 위치 측정은 고차 측정의 초기 단계로 거리 1~1.5m 지점의 높이 상중하에 대한 측정으로 국한한다. 거리가 2m인 지점에서 휴머노이드 로봇의 키 보다 적어도 20cm 이상 높이에 있어야 고각이 5° 이상 나오게 된다. 따라서, 본 실험은 휴머노이드 로봇의 신장이 약 150cm라 가정하고, 약 1~1.5m 거리에서 사람의 키가 180cm, 150cm, 120cm인 경우를 가정하여 상중하 높이 구분하는 것에 제한하여 시스템을 구현했다. 높이 TDOA 특성행렬은 사이즈가 57×8 이고, 행의 개수 57은 방위각별 높이를 상중하로 나누어 구분된 (19×3) 크기이다. 열의 개수 8은 수직 높이 구분에 적합한 마이크로폰 쌍이다. 설명한 2가지 TDOA 특성행렬을 사용하여 방위각 검지 및 높이 구분을 실행한다.

2.4 실험 결과

일반 사무실 공간에서 휴머노이드 로봇에 마이크로폰을 설치하여 음원 위치 측정 실험을 하였다. 휴머노이드 두상에 의한 입력 신호의 변형 및 방향을 인정하고 실험을 하였다. NI4472 DAQ보드, Aphex-207 프리앰프와 MKE 2P-C 마이크로폰을 사용하였다. 그림 5와 같이 방위각은 방위각 TDOA 특성 행렬을 이용하여 검지하고, 높이 상중하는 방위각을 먼저 검지한 후, 높이 TDOA 특성 행렬을 이용하여 구분하였다.



〈그림 5〉 음원의 방위각 검지 및 높이 구분

방위각 검지 실험은 5m 이내의 각 m마다 10° 간격으로 5회씩 총 475회 음원을 발생시켜 실험하였다. 표1은 방위각 별 25회(m별 5회)씩 실험결과를 정리하였다.

	0°	20°	40°	60°	80°
성공율	100%	8%	96%	100%	96%
평균오차	0°	18.4°	0.4°	0°	0.4°
	100°	120°	140°	160°	180°
성공율	100%	100%	100%	4%	100%
평균오차	0°	0°	0°	19.2°	0°

〈표 1〉 음원의 거리에 대한 방위각 검지 결과

실험 결과 0° , $30^\circ \sim 150^\circ$, 180° 구간은 비교적 정확한 방위각을 검지하나, $10^\circ \sim 20^\circ$ 와 $160^\circ \sim 170^\circ$ 구간은 정확한 방위각을 검지하지 못하고, 각각 주로 0° 와 180° 값을 선택하였다. 그 이유는 마이크로폰 배열 구조상 $10^\circ \sim 20^\circ$ 와 $160^\circ \sim 170^\circ$ 구간이 방위각 별 TDOA 차이가 약 $10 \sim 20\mu s$ 로서 각각 0° 및 180° 의 TDOA 값과 차이가 매우 적기 때문이다. 본 시스템의 샘플링율이 16KHz로서 한 개의 샘플이 추정할 수 있는 최소 TDOA는 $62.5\mu s$ 이므로 구별에 한계가 있다. 샘플링율은 높았다면 향상된 결과를 가져올 것이라 예상 할 수 있으나, 16KHz 이상의 샘플링율은 고려하지 않았다.

높이 구분 실험은 1~1.5m 지점에서 방위각 별(19개) 상중하 높이에서 5회씩 총 285회 음원을 발생시켜 실험하였다. 본 시스템은 방위각을 검지한 후, 높이 구분을 하는 구조이므로, 효율성을 위하여 선택된 방위각에 따라 높이 구분을 하기에 적합한 마이크로폰 쌍을 선택한다. 표2는 방위각 별로 상중하 5회씩 총 15회를 정리한 결과이다.

		0°	20°	40°	60°	80°
방위각 검지	성공율	100%	6.7%	93.3%	100%	100%
	평균오차	0°	18.6°	0.7°	0°	0°
높이 구분	성공율	100%	93.3%	100%	100%	100%
	오류회수	0회	1회	0회	0회	0회
		100°	120°	140°	160°	180°
방위각 검지	성공율	100%	93.3%	100%	6.7%	100%
	평균오차	0°	0.7°	0°	18.6°	0°
높이 구분	성공율	100%	100%	93.3%	93.3%	100%
	오류회수	0회	0회	1회	1회	0회

〈표 2〉 음원의 방위각에 대한 높이 구분 결과

방위각 검지율은 표1과 비슷한 결과를 나타냈다. 그러나 상중하 높이 구분에서 방위각이 20° 이내의 오차가 발생하더라도, 상중하 구분은 비교적 정확하였다. 그 이유는 방위각 $0^\circ \sim 20^\circ$ 와 $160^\circ \sim 180^\circ$ 검지 오류에 불구하고, 높이 TDOA 특성행렬의 높이별 TDOA 차이가 크기 때문이다.

3. 결론 및 고찰

본 논문은 휴머노이드 로봇의 방위각 검지 및 높이 구분을 위한 마이크로폰 배열을 제안하고, TDOA 특성행렬을 정의하여 사용하였다. 본 논문은 복잡한 비선형 수식 및 근사 수식을 사용하지 않고, 정의된 TDOA 특성행렬 방식을 이용하여, 방위각 검지 및 높이 구분이 가능한 실제 시스템을 구현하였다는 것에 의미를 둔다.

$10^\circ \sim 20^\circ$ 와 $160^\circ \sim 170^\circ$ 구간의 방위각 검지 부정확성은 마이크로폰 배열에 기인한 것이다. 또한 시각 시스템에 의해 보정이 가능하다고 하나, 10° 간격의 방위각 분해능은 음원의 거리가 멀어짐에 차이를 크게 느낄 수 있으므로 개선되어야 할 점이다. 음원 위치를 3차원 좌표형태로 측정하는 것이 앞으로의 연구과제이다.

〈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 William 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