

論文

반향음과 잡음 환경을 고려한 실시간 소리 추적 시스템

기창돈*, 김강호**, 이택진**

Real-Time Sound Localization System For Reverberant And Noisy Environment

Chang-don Kee*, Ghang-ho Kim** and Taik-jin Lee**

ABSTRACT

Sound localization algorithm usually adapts three step process: sampling sound signals, estimating time difference of arrival between microphones, estimate location of sound source. To apply this process in indoor environment, sound localization algorithm must be strong enough against reverberant and noisy condition. Additionally, calculation efficiency must be considered in implementing real-time sound localization system. To implement real-time robust sound localization system we adapt four low cost condenser microphones which reduce the cost and total calculation load. And to get TDOA(Time Differences of Arrival) of microphones we adapt GCC-PHAT(Generalized Cross Correlation-Phase Transform) which is robust algorithm to the reverberant and noise environment. The position of sound source was calculated by using iterative least square algorithm which produce highly accurate position data.

초 록

소리를 이용한 위치 추적은 마이크로폰을 이용하여 신호를 수집하고 수집된 신호로부터 마이크로폰 간의 신호 도달 시간차를 추정된 뒤 추정된 시간차를 이용하여 소리의 발생 위치를 추정하는 과정을 거치게 된다. 실내 환경에서 이를 활용하기 위해서는 잡음과 반향음에 대한 강건성을 확보해야만 하는 제약이 따른다. 특히 실시간으로 구현하기 위해서는 계산의 효율성까지 고려되어야 한다. 본 논문에서는 네 개의 저가 콘텐서 마이크로폰을 이용하여 비용적인 측면과 계산량에서의 효율성을 모두 추구하였다. 네 개의 마이크로폰을 이용하여 마이크로폰 간의 소리 도달 시간차를 구하는 계산량을 줄였고 GCC-PHAT(Generalized Cross Correlation-Phase Transform) 알고리즘을 이용해서 강건성을 높였으며 iterative least square 방식을 이용하여 높은 정확도의 위치 데이터를 얻을 수 있었다.

Key Words : Sound Localization(음원위치추적), Microphone Arrays(마이크로폰 어레이), GCC-PHAT(가중치를 고려한 일반화된 상호상관법), TDOA(도달시간지연)

†2009년 12월 1일 접수 ~ 2010년 2월 23일 심사완료

* 정회원, 서울대학교 기계항공공학부

교신저자, E-mail : kee@snu.ac.kr

서울시 관악구 신림동 산56-1

** 정회원, 서울대학교 기계항공공학부 대학원

I. 서 론

소리를 이용한 위치 추적은 활용도가 높아 음성인식, 사람의 추적, 로봇의 위치 추적 등 다양한 분야에서 연구되고 있다[1]. 소리를 이용한 위

치 추적 시스템 관련 연구가 활발한 이유는 다양한 종류의 대상에 적용할 수 있고 실내 환경에서 이용이 가능하다는 점과 센티미터 수준의 높은 정확도를 얻을 수 있다는 점 때문이다. 소리를 위치 추적 시스템에 이용할 경우 사람, 로봇 등 소리를 발생할 수 있는 모든 대상에 대해서 전파 등의 특별한 신호 발생 장치를 장착하지 않고도 위치를 추적할 수 있으며 이에 따라 시스템의 구성이 간단하다는 장점을 가지고 있다. 실내에서 소리를 이용하여 위치를 추적할 경우 전파를 이용한 다른 위치 추적 시스템에 비해서 다중경로 오차(M용한tipath)에 의한 오차가 적게 작용하기 때문에 정확도가 월등히 높고 강건한 장점을 가지고 있다. 반향음과 잡음이 존재하는 일반적인 실내 환경에서 소리를 이용하여 대상의 위치를 추적할 경우에도 수 센티미터 급의 위치 정확도를 얻을 수 있기 때문에 다른 위치 추적 시스템에 비해서 높은 정확도를 가질 수 있다. 본 논문에서는 소리를 이용한 위치 추적의 장점을 이용하여 실시간으로 소리의 위치를 추적하는 시스템을 구성하였다. 이를 위해서 반향음과 잡음이 존재하는 일반적인 실내 실험 환경에서 네 개의 저가형 콘텐서 마이크를 설치하고 대상의 위치를 추적하는 알고리즘을 제안하고 검증하는 연구를 수행하였다. 기존의 연구에서는 소리의 위치를 추적하기 위해서 다수의 마이크를 설치하고 GCC-PHAT (Generalized Cross Correlation - Phase Transform) [2][3][5], SRP(Steered Response Power) [8], SRP-PHAT(Steered Response Power - Phase Transform) [1] 알고리즘을 이용하여 소리가 발생한 지점과 각 마이크 사이의 거리를 TDOA(Time Differences Of Arrival) 방식으로 구하고 이를 바탕으로 소리가 발생한 지점의 위치를 구하는 알고리즘을 적용하고 있다. 본 논문에서는 잡음과 반향음에 강한 GCC-PHAT 방식을 이용하여 각 마이크 간의 소리 도달 시간차를 구하고 이를 바탕으로 최적의 위치를 구할 수 있도록 iterative least square 방식을 이용하여 소리 발생 위치를 구하는 알고리즘을 제시하였으며 실험을 통하여 결과를 검증하였다.

II. 위치 추적 알고리즘과 테스트

2.1 소리를 이용한 위치 추적 개요

소리를 이용한 위치 추적 알고리즘은 기본적으로 그 위치를 정확히 알고 있는 다수의 마이크를 배열을 설치한 뒤 소리가 발생한 지점과 마



그림 1. 소리를 이용한 위치 추적 알고리즘

이크로폰과의 직선거리를 측정하고 이를 바탕으로 마이크 배열에 대한 소리 발생 지점의 상대적인 위치를 추정하는 방식이다.

2.2 소리의 모델링과 거리 추정

소리를 이용한 위치 추적을 하기 위해서는 첫 번째로 마이크와 소리 발생지점과의 상대적인 거리 측정, 두 번째로 측정된 상대적인 거리를 이용해서 마이크와 소리 발생 위치 계산 과정을 거치게 된다. 이를 위해서 본 장에서는 소리의 모델링과 이를 이용한 거리 측정 알고리즘을 살펴보도록 한다. 우선 두 마이크를 기준으로 동일한 소리가 수신되었을 경우 이의 모델링은 다음과 같다[6].

$$\begin{aligned} x_1[n] &= h_1[n] * s[n] + n_1[n] \\ x_2[n] &= h_2[n] * s[n - \tau] + n_2[n] \end{aligned} \quad (1)$$

x_1, x_2 는 마이크 1과 2에 수신된 신호, s 는 발생한 소리, h_1, h_2 는 반향음과 마이크의 임펄스 반응, n_1, n_2 는 각 마이크 채널의 잡음, τ 는 두 마이크 간의 소리가 도달하는 상대적인 시간차를 나타낸다.

두 마이크 간에 도달하는 소리의 시간차는 마이크 간에 수집된 신호간의 상관함수를 정의하고 이를 최대로 하는 $\hat{\tau}$ 로 구할 수 있다. 두 신호의 상관함수는 다음과 같다[6].

$$R_{12} = \sum_{n=0}^{N-1} x_1[n] x_2[n - \tau] \quad (2)$$

위의 상관함수는 계산량을 줄이고 주파수 대역에서의 조작을 위해서 FFT(Fast Fourier Transform)를 이용하는 방식으로 계산할 수 있다[6].

$$\begin{aligned} R_{12}(\tau) &= \sum_{k=0}^{N-1} X_1[k] X_2[k]^* e^{j2\pi k\tau/N} \\ \hat{\tau} &= \arg \max_{\tau} R_{12} \end{aligned} \quad (3)$$

식(3)에서 X_1, X_2 는 x_1, x_2 의 Fourier transform이고 $X_1 X_2^*$ 는 x_1, x_2 의 cross spectrum을 나타낸다. 위의 수식을 통해서 두 마이크의

로 수신된 신호의 cross correlation을 구할 수 있지만 음성 신호의 경우 저주파수 대역의 신호가 고주파수 대역의 신호보다 강하게 작용하고 있어 correlation peak의 범위가 넓어지는 문제점이 발생한다. 이를 해결하기 위해서 whitened cross correlation 알고리즘을 적용하여 주파수의 파워를 normalize하도록 한다. 이는 식(4)와 같다.

$$R_{12}(\tau) = \sum_{k=0}^{N-1} \frac{X_1[k]X_2[k]^*}{|X_1[k]||X_2[k]^*|} e^{j2\pi k\tau/N} \quad (4)$$

Whitened cross correlation을 이용하게 되면 모든 주파수대역의 크기가 표준화하게 되어 cross correlation peak의 범위가 좁아지게 된다. 이는 더욱 정확하게 소리와 마이크로폰 사이의 직선거리를 구할 수 있는 것을 의미한다.

2.3 위치 추정 알고리즘

마이크로폰을 이용해서 마이크로폰으로부터 도달한 소리의 상대적인 시간차를 구하게 되면 이를 바탕으로 소리의 위치를 계산해 낼 수 있다. 기존의 위치 계산 알고리즘은 두 마이크로폰에 도달하는 소리의 상대적인 시간차와 두 마이크로폰 사이의 기저 벡터를 이용해서 기저벡터와 소리의 발생 위치의 각을 구하는 방식을 이용하였다. 이 알고리즘은 가까운 거리에 대해서 선형화에 따른 오차가 발생하는 문제점이 있다. [6][7][9] 본 논문에서는 이를 해결하기 위해서 GNSS 측위에서 많이 이용하는 iterative least square 알고리즘을 수정해서 이용하였다. 이를 설명하기 위해 마이크로폰과 음원과의 기하 구조를 살펴보자.

그림 2는 하나의 마이크로폰과 하나의 소리 발생 지점의 관계를 표시한 것이다. i번째 마이크로폰의 벡터를 MIC_i 로 두고 소리의 발생 지점을 S 로 나타내었다. MIC_i 의 좌표는 $M_i = (x_i, y_i, z_i)$ 로 표시할 수 있고 S 의 좌표는 $S = (x, y, z)$ 로 표시할 수 있다. $e_i = (e_{ix}, e_{iy}, e_{iz})$ 는 식(6)과 같이 MIC_i 와 S 가 이루는 벡터의 단위 벡터이다. 이를

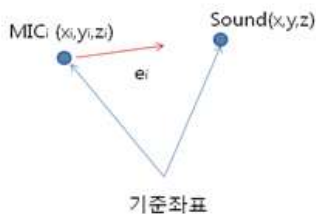


그림 2. 소리와 마이크로폰의 기하 배치에서의 관계

이용하여 4개의 마이크로폰을 이용한 경우에 소리의 위치를 수식화 하면 (5)와 같다. c 는 소리의 속도, τ 는 소리가 마이크에 도달하는데 걸린 시간이다. 수식(5)와 (6)에서 마이크로폰의 위치와 마이크로폰과 소리 사이의 위치 단위벡터가 모두 포함되어 있는데 초기의 소리 위치를 임의로 가정하고 단위벡터를 구성한 뒤 계산된 소리의 위치를 이용해서 단위벡터를 다시 업데이트하는 iteration 방식으로 위치를 구할 수 있다. 결론적으로 사용자의 위치와 초기 소리 도달 시간 값은 H 행렬을 pseudo inverse 하여 구하게 된다.

$$\begin{bmatrix} e_1 & -1 \\ e_2 & -1 \\ e_3 & -1 \\ e_4 & -1 \end{bmatrix} \begin{bmatrix} S \\ c\tau_1 \end{bmatrix} = \begin{bmatrix} e_1 \cdot M_1 \\ e_2 \cdot M_2 \\ e_3 \cdot M_3 \\ e_4 \cdot M_4 \end{bmatrix} - \begin{bmatrix} c(\tau_1 - \tau_1) \\ c(\tau_2 - \tau_1) \\ c(\tau_3 - \tau_1) \\ c(\tau_4 - \tau_1) \end{bmatrix} \quad (5)$$

$$Hx = y$$

$$x = (H^T H)^{-1} H^T y$$

$$e_i = \frac{M_i - S}{|M_i - S|} \quad (6)$$

2.4 시스템의 개요

본 논문에서 구성한 실시간 위치 추적 시스템은 네 개의 마이크로폰을 실내 천장에 설치하여 신호를 수집하도록 하였다. 실내 환경에서는 마이크로폰을 천장에 배치하는 것이 현실적인 방안이기 때문이다. 신호는 16kHz, 8bit의 분해능으로 수집하였고 네 개의 마이크로폰으로부터 초당 64kByte에 해당하는 데이터가 생성된다. 실험의 편의를 위해 소리는 스피커를 통해 발생하도록 하였다. 설치된 마이크로폰의 배치 좌표는 (x,y,z) 형식으로 아래와 같고 단위는 미터 기준이다. 실험환경에 쓰인 방의 크기는 대략 가로 7.2m, 세로 5.8m, 높이 3.2m 의 직사각형 구조이다.

$$\begin{aligned} Mic1 &= (0.47, 0, 3.2), & Mic2 &= (0, 3.27, 3.2) \\ Mic3 &= (5.69, 3.23, 3.2), & Mic4 &= (5.42, 0, 3.2) \end{aligned}$$



그림 3. 실내 실험 환경

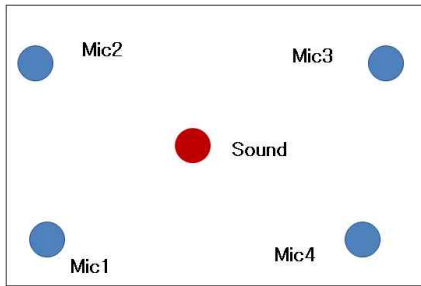


그림 4. 마이크로폰의 실내 배치 구조

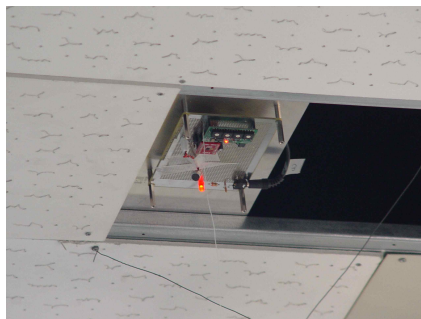


그림 5. 마이크로폰과 FM 송신모듈

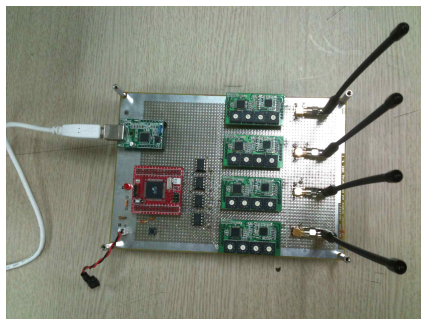


그림 6. 중앙 ADC 보드

실내 천장에 4개의 마이크로폰을 설치하고 수집된 마이크로폰의 신호는 무선 FM 송신기를 통해서 중앙 ADC 보드로 전송되도록 하였다. 그림 5는 실제 설치된 무선마이크로폰과 음성 신호를 전송하기 위한 FM 무선 송신 모듈이다. 설치된 모듈에는 신호의 증폭 정도를 자동으로 제어하기 위한 AGC 모듈이 포함되어 있다.

중앙 ADC 보드는 각 마이크로폰 모듈에서 FM 송신 모듈을 통해 전송된 신호를 FM 수신 모듈로 수신하여 이를 ADC 칩을 통해 ADC한다음 이를 USB를 통해서 PC로 전송하도록 하였다. USB 모듈을 이용한 이유는 ADC로 샘플링 데이터는 초당 16kHz로 4개의 채널을 통해서 8bit로 샘플링되기 때문에 64kByte의 고용량 데

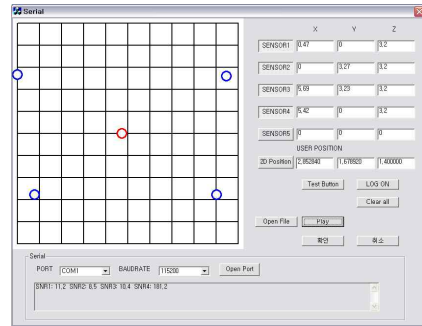


그림 7. 실시간 위치 추적 프로그램

이터 전송량이 가능하여야 한다. 이를 위해서 고속의 데이터 전송이 가능한 USB모듈을 이용하였다.

실시간 위치 추정 프로그램에서는 연속적으로 신호를 입력받아 신호의 SNR(Signal to Noise Ratio), 신호 도달 시간차(TDOA), 소리의 위치를 계산하도록 하고 이를 그래프에 표시하도록 작성되었다.

2.5 신호 수집

네 개의 마이크로폰을 이용해서 스피커에서 출력되는 신호를 측정 한 결과는 그림 8와 같다. 측정된 신호는 최대 크기의 절대값이 1이 되도록 양자화시켰으며 AGC(Automatic Gain Control)를 통해 실제 발생된 소리의 크기에 상관없이 수집된 신호는 일정한 크기를 유지하도록 하였다.

수집된 신호의 시간대별 주파수 스펙트럼은 그림 9와 같다. 그림에서 붉은색으로 표시된 부분이 강한 신호 특성을 나타내는 부분이고 초록색으로 갈수록 약한 신호 특성을 나타내는 부분이다. 시간에 따라 512샘플 단위로 신호의 주파수를 분석하였으며 음성 신호가 잡음 신호 보다 강한 경우 저주파가 크게 나타나고 잡음 신호가

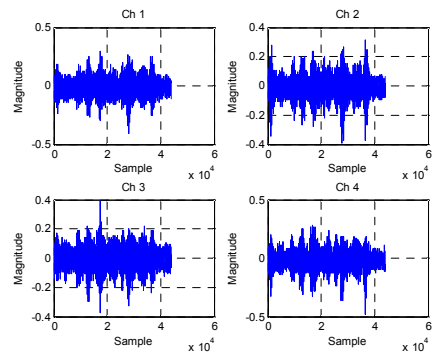


그림 8. 수집된 소리 신호

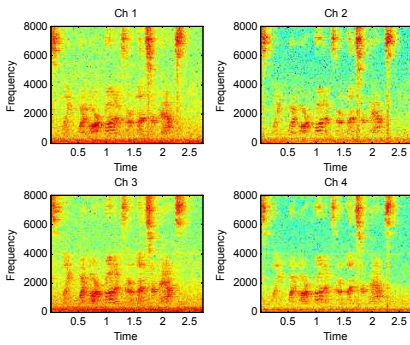


그림 9. 수집된 신호의 주파수 스펙트럼

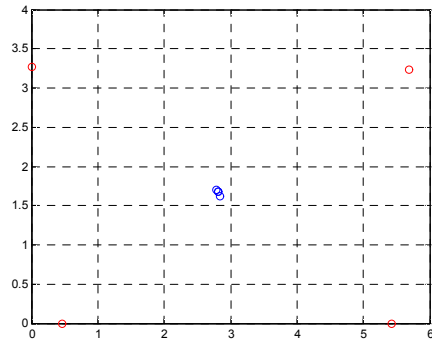


그림 11. 위치 추정 결과 그래프

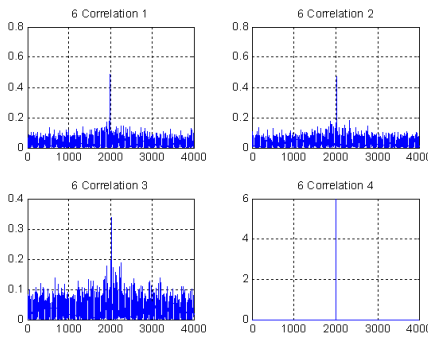


그림 10. Whiten Cross Correlation 결과

음성신호보다 강한 경우 고주파수에 신호가 크게 나타나는 특성을 보인다. 이는 측정된 잡음이 주파수 영역에서 백색잡음의 특성을 보이기 때문이다.

측정된 신호를 이용해서 신호간의 Cross Correlation을 구한 결과는 그림 10과 같다. Correlation을 구할 때 기준 신호는 마이크론에 들어오는 신호의 감도가 모두 비슷하기 때문에 4번 마이크론으로 임의적으로 정하였으며 그래프 해석의 편의를 위해서 Correlation 결과값을 모두 절대값으로 표시하였다. 그래프에서 peak이 나타나는 지점이 마이크론 간의 상대적인 신호 도달 시간차를 나타낸다. 4번 마이크론의 신호는 자기 자신과의 상관 결과이기 때문에 다른 그래프보가 결과 값이 크게 나타났다.

2.6 위치 추정 계산 및 결과

측정된 신호를 바탕으로 위치를 추정한 결과 그래프는 그림 11과 같다. SNR이 일정한 값 이상일 때만 위치를 구하였으며 마이크를 모두 천장에 배치하여 각 마이크와 음원간의 e_i 벡터에서 z 값이 비슷하게 나오기 때문에 H 매트릭스

의 inverse를 구할 때 singular가 발생한다. 그래서 직접적으로 3차원 위치를 구할 수 없고 소리의 z 좌표는 알고 있다는 가정 하에 실제 z 좌표값을 이용하여 계산을 수행하였다. 2차원이 아닌 3차원 위치를 계산으로 구하기 위해서는 추가적인 마이크의 설치 필요하지 않고 마이크 4개 중에 적어도 하나를 높이가 다른 지점에 설치하면 된다. 앞의 z 좌표 값을 알고 있다는 가정을 통해서 측정된 소리의 2차원 평균 위치는 $S=(2.80, 1.68, 0.04)$ 이고 STD 값은 $(0.018, 0.028, 0)$ 으로 계산되었고 이는 실제 음원의 위치인 $(2.78, 1.68, 0.04)$ 에 센티미터급 오차에 근접하는 정확한 값을 확인할 수 있었다.

III. 결 론

본 논문에서는 실시간으로 소리의 위치를 추적하기 위해서 네 개의 마이크론을 설치하고 신호를 수집한 뒤 GCC-PHAT 알고리즘을 이용하여 마이크론 간의 상대적인 신호 도달 시간차를 추정하고 이를 바탕으로 iterative least square 알고리즘으로 소리의 위치를 추정하는 알고리즘을 제시하였다. 제시된 알고리즘은 실제 실험을 통해서 추정된 위치와 실제 위치 간의 정확도를 비교하였고 이를 통해 실제 실내 환경에서 센티미터급의 정확도를 얻을 수 있었다. 실제 실험에 이용된 마이크론이 저가형의 콘텐츠 마이크론임 점과 네 개만 이용했다는 점을 고려하면 이는 실내 환경에서 본 논문의 결과를 충분히 활용할 수 있다는 결론을 얻을 수 있다.

참고문헌

1) M. Brandstein and D. Ward(Eds), "Microphone Arrays", Signal Processing

Techniques and Applications, Springer, 2001.

2) C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay", *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 24, no. 4, Aug 1976, pp. 320~327.

3) M. S. Brandstein and H. F. Silverman, "A robust method for speech signal time-delay estimation in reverberant rooms", *Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing*, 1997, pp. 375~378.

4) Y. Chan, R. Hattin, J. Plant, "The least squares estimation of time delay and its use in signal detection", *Acoustics, Speech and Signal Processing, IEEE Transactions on*, Vol. 26, No. 3, 29 January 2003, pp. 217~222.

5) A. Badali, J.-M. Valin, F. Michaud, and P. Aarabi, "Evaluating Real-time Audio Localization Algorithms for Artificial Audition in Robotics", http://people.xiph.org/~jm/papers/iros09_badali.pdf.

6) J.-M. Valin, F. Michaud, J. Rouat, and D. Letourneau, "Robust sound source localization using a microphone array on a mobile robot", *Proceedings International Conference on Intelligent Robots and Systems*, 2003, pp. 1228~1233.

7) Michael Brandstein, Michael S. Brandstein, John E. Adcock, Harvey F. Silverman, "Microphone-Array Localization Error Estimation with Application to Sensor Placement", *Acoustical Society of America*, 1996, pp. 3807~3816.

8) J.P. Dmochowski, J. Benesty, S. Affes, "A Generalized Steered Response Power Method for Computationally Viable Source Localization", *Audio, Speech, and Language Processing, IEEE Transactions on*, Volume 15, Issue 8, Nov, 2007, pp. 2510-2526.

9) K. Takahashi, H. Yamasaki, "Audio-visual sensor fusion system for intelligent sound sensing, Multi sensor Fusion and Integration for Intelligent Systems", 1994. *IEEE International Conference on MFI '94*, Oct 1994, pp. 493-500.