

조정 응답 파워 방법과 결합된 generalized cross correlation with phase transform 음원 위치 추정

Generalized cross correlation with phase transform sound source localization combined with steered response power method

김영준,¹ 오민재,¹ 이인성[†]

(Young-Joon Kim,¹ Min-Jae Oh,¹ and In-Sung Lee^{1 †})

충북대학교 전자통신공학과

(Received June 19, 2017; revised July 10, 2017; accepted September 28, 2017)

초 록: 본 논문에서는 잔향과 잡음이 존재하는 실제 환경을 모델링하여 두 개의 마이크로폰을 이용한 음원 위치 추정의 정확성을 향상시키는 방법을 제안하였다. 입력신호에 VAD(Voice Activity Detection)를 적용하여 묵음 구간을 제외한 음성 구간만을 사용하였고, 샘플링 주파수의 제한으로 인한 측정 범위를 벗어나는 프레임은 업샘플링(up-sampling)을 통해 지연시간을 다시 추정하였다. 여기서 계산된 도착 지연 시간은 Time-table을 참조해 주변 후보 위치의 지연 값들과의 비교로 최대 파워 값을 갖는 지연 시간을 선택하여 음원 위치의 정확도를 높였다. 또한 프레임간의 상관성을 이용하여 연속된 음성 프레임의 경우 큰 추정 차가 발생하는 곳을 찾아 주변 프레임의 평균값으로 대체함으로써 음원의 위치 추정 성능을 향상시켰다.

핵심용어: 마이크로폰 어레이, GCC-PHAT (Generalized Cross Correlation with Phase Transform), 음원 위치 추정, 도착 지연 시간, 음성 구간 검출

ABSTRACT: We propose a methods which is reducing direction estimation error of sound source in the reverberant and noisy environments. The proposed algorithm divides speech signal into voice and unvoice using VAD. We estimate the direction of source when current frame is voiced. TDOA (Time-Difference of Arrival) between microphone array using the GCC-PHAT (Generalized Cross Correlation with Phase Transform) method will be estimated in that frame. Then, we compare the peak value of cross-correlation of two signals applied to estimated time-delay with other time-delay in time-table in order to improve the accuracy of source location . If the angle of current frame is far different from before and after frame in successive voiced frame, the angle of current frame is replaced with mean value of the estimated angle in before and after frames.

Keywords: Microphone array, GCC-PHAT (Generalized Cross Correlation with Phase Transform), Sound source localization, TDOA (Time Difference of Anival) , VAD (Voice Activity Detection)

PACS numbers: 43.60.Jn, 43.60.Bf

1. 서 론

최근 마이크로폰 어레이를 이용한 음원의 위치 추정 및 음질 향상을 위한 연구가 꾸준히 진행되고 있

다. 화자의 위치를 파악하는 지능형 로봇 시스템,^[1,2] 화상회의 시스템,^[3,4] CCTV와 연동된 보안 시스템^[5,6] 등 다양한 분야에서 음원의 위치 추정 방법이 사용되고 있다. 그러나 잔향이나 잡음이 존재하는 사용 환경이 음원의 위치 추정 성능을 감소시키고, 단말기의 크기에 따라 마이크로폰 어레이를 구성하는 마이크로폰 개수가 제한되어 진다.

[†]Corresponding author: In-Sung Lee (inslee@chungbuk.ac.kr)
Department of Radio Cmmunication Engineering, Chungbuk National University, 1 Chungdae-ro, Seowon-Gu, Cheongju, Chungbuk 28644, Republic of Korea
(Tel: 82-43-261-3223)

마이크로폰 어레이를 이용한 음원의 위치 추정 방법 중 각 마이크로폰에 도착하는 신호의 도착 지연 시간을 보상하고, 그 신호들의 합을 통하여 위치를 추정하는 빔포밍 방법 등은 많은 연산량이 사용되고, 우수한 성능을 위해서는 많은 수의 마이크로폰이 요구된다.^[7] 또한, MUSIC(Multiple Signal Classification) 알고리즘^[8] 등과 같이 각 마이크로폰에 수신된 신호들의 상관 행렬의 스펙트럼 분석을 기반으로 하는 방법은 협대역 신호에 대하여 고안된 방법으로 광대역 신호인 음성 신호를 다루기 위한 부가적 연산과 고유치 분해에 따른 많은 연산량이 요구된다. 반면 GCC-PHAT(Generalized Cross Correlation with Phase Transform) 방법은 TDOA(Time-Difference of Arrival) 기반으로, 지연시간 추정에 따라 발생하는 쌍곡선의 교차점을 통해 음원의 위치를 추정하는 방법으로,^[9,10] 비교적 간단하여 연산량이 적고, 최소 두 개의 마이크로폰으로도 측정이 가능하여 다양하게 사용되고 있다.^[11]

본 논문에서는 잔향과 잡음이 존재하는 환경을 모델링하여 두 개의 마이크로폰을 이용한 음원 위치 추정의 정확성을 향상시키는 방법을 제안한다. 샘플링 주파수 제한으로 인한 측정 범위를 벗어나는 프레임은 업샘플링을 이용하여 도착 지연 시간을 재추정하였다. 추정된 지연 시간은 미리 구해진 Time-table을 참조하여 주변 후보 위치의 지연 값들을 토대로 입력 신호를 지연 시간만큼 딜레이시킨 후 합을 구해 각각의 파워 값을 비교함으로써, 최대 파워 값을 갖는 지연 시간을 선택하였으며, 프레임간의 상관성을 이용하여 연속된 음성 프레임의 경우 큰 추정 차가 발생하는 곳을 찾아 주변 프레임의 평균값으로 대체하여 음원 위치 추정의 정확성을 향상시켰다. 또한 마이크로폰에 입력되는 신호를 음성과 비음성 구간으로 구분

하여 위치 추정 알고리즘을 수행함으로써 음원 위치 추정 성능 저하의 원인을 감소시켰다.

II. TDOA 기반의 음원 위치 추정 알고리즘

2.1 TDOA를 이용한 각도 추정 방법

마이크로폰 어레이를 구성하는 각 마이크로폰 간 신호의 도착 지연 시간 추정 방법에서 음원 위치 추정은 지연 시간을 추정하는 단계와 음원의 위치를 찾는 단계로 나누어진다. Fig. 1은 한 개의 음원과 두 개의 마이크로폰이 있을 경우 지연 시간의 발생을 나타내고 있다. 음원으로부터 음속으로 신호가 입력될 때 MIC1보다 MIC2에 신호가 먼저 입력된다. 이때 MIC2에 신호가 입력된 시점부터 MIC1에 신호가 입력될 때까지의 거리(d)에 해당하는 시간이 지연된다. 도착 지연 시간을 구하게 되면 삼각법에 의하여 음원의 각도를 추정하게 된다. 지연 시간 τ 는 Eq. (1)과 같이 신호가 더 이동해야 하는 거리인 d 를 음속 $c(340m/s)$ 로 나눈 값이다. 이를 음원의 각도에 관하여 Eq. (2)와 같이 정리하면 최종적으로 음원의 위치를 추정하게 된다.

$$\tau = \frac{d}{c} = \frac{MD \times \cos\theta}{c}. \quad (1)$$

$$\theta = \cos^{-1} \frac{c\tau}{MD}. \quad (2)$$

2.2 TDOA 추정을 위한 샘플링 주파수와 마이크로폰 간의 간격

TDOA 추정에 있어서 샘플링 주파수와 마이크로폰 어레이 사이의 관계는 변환된 이산 신호의 측정 가능한 범위를 결정하는데 이용된다. Fig. 1의 MD 는 음원의 위치가 0° 혹은 180° 에 위치하였을 경우 마이크로폰 사이의 최대 거리차가 된다. 이 최대값은 지연 시간 추정의 값의 유효한 범위를 나타내는데 이용된다. Eq. (3)은 샘플링 주파수와 마이크로폰 사이의 거리가 주어졌을 때 측정 가능한 최대 지연 샘플을 의미한다.

샘플링 주파수가 16 kHz이고 마이크로폰 사이의 거리가 0.5m인 경우 Eq. (3)에 의하여 최대 측정 가능 지연 샘플은 23이 되며, 이에 따라 측정 가능 범위는 -23~23

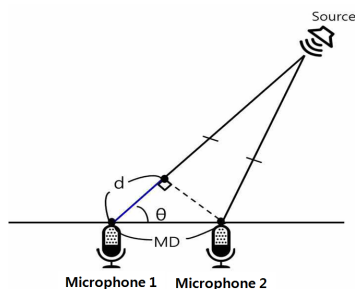


Fig. 1. Microphone array and sound source.

으로 총 47개의 지연 값을 가지게 되며, 각각 0°~180°를 47 등분하여 매핑된다.

$$DelaySample_{max} = \frac{f_s \times MD}{c}. \quad (3)$$

3.3 GCC(Generalized Cross-Correlation)

시간 영역의 상호상관함수는 주파수 영역의 상호 파워 스펙트럼 $G_{x_1x_2}(\omega)$ 와 푸리에 변환 관계를 갖는다. 실내 공간에서 음원 $s(t)$ 와 잡음 $n(t)$ 에 의해 i 번째 마이크론에 입력되는 신호 $x_i(t)$ 와 $x_i[n]$ 은 Eqs. (4)와 (5)로 모델링 된다.

$$x_i(t) = s(t - \tau_i) * h_i(t) + n_i(t), \quad i = 1, 2. \quad (4)$$

$$x_i[n] = s[n - \tau_i] * h_i[n] + n_i[n], \quad i = 1, 2. \quad (5)$$

이때, h 는 실내 임펄스 응답을 나타낸다. 위와 같이 모델링된 신호의 상관 함수를 상호 파워 스펙트럼과 푸리에 변환 관계를 이용하여 표현하면 Eq. (6)과 같이 표현된다. 이상 신호 형태의 마이크론 출력신호를 N -포인트 DFT(Discrete Fourier Transform)로 구한 이산 스펙트럼 $X_1[k]$, $X_2[k]$ 를 곱하여 Eq. (7)과 같이 상호 파워 스펙트럼을 구할 수 있다.

$$R_{x_1x_2}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G_{x_1x_2}(\omega) e^{j\omega\tau} d\omega. \quad (6)$$

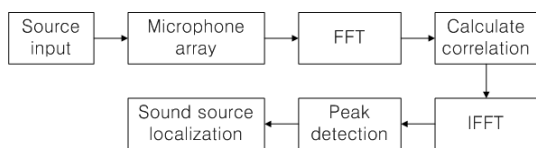


Fig. 2. TDOA block diagram of GCC.

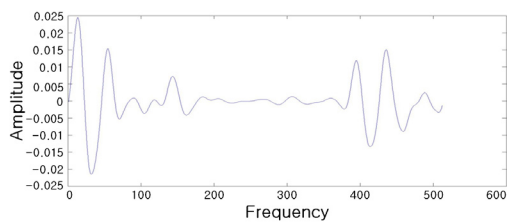


Fig. 3. Peak detection using GCC.

$$R_{x_1x_2}(m) = \sum_{k=0}^{N-1} X_1[k] X_2[k] * e^{j \frac{2\pi mk}{N}}. \quad (7)$$

지연 시간 τ 은 $R_{x_1x_2}(\tau)$ 를 최대가 되게 하는 값을 의미한다. Fig. 2는 GCC 함수의 TDOA 추정 과정을 나타낸 블록도이고, Fig. 3은 GCC 함수의 결과를 나타내고 있다. Fig. 3에서 알 수 있듯이 피크값이 존재하지만, 퍼짐 현상을 확인할 수 있다.

GCC 함수는 Eqs. (8)과 (9)와 같이 PHAT 프리필터 $\Psi(\omega)$ 를 사용하여 계산한다.^[12] 프리필터는 잔향에 따른 다중 경로로 인한 상관 함수의 피크 값 퍼짐 현상과 잡음 간섭으로 인한 피크 값의 부정확성을 억제하기 위하여 적용된다.

$$R_{x_1x_2}^{(g)}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Psi(\omega) G_{x_1x_2}(\omega) e^{j\omega\tau} d\omega. \quad (8)$$

$$R_{x_1x_2}(m) = \sum_{k=0}^{N-1} \Psi[k] X_1[k] X_2[k] * e^{j \frac{2\pi mk}{N}}. \quad (9)$$

$$\Psi(\omega) = \frac{1}{|G_{x_1x_2}(\omega)|}. \quad (10)$$

GCC 함수의 다양한 프리필터 중 PHAT는 잔향에 의한 퍼짐 현상을 억제하는데 우수한 성능을 가지지만, 신호 대 잡음비가 낮은 주파수 영역을 강조하기 때문에 잡음에 취약하다. 반면에 다른 필터들은 잡음의 영향을 줄이는 특성을 가지고 있지만, 잔향에는 취약한 성능을 가진다. 따라서 잔향이 존재하는 실내 환경에 적용하는 경우 PHAT가 일반적으로 가장 많이 사용되고 있다.

III. TDOA의 정확성 향상 방법

GCC-PHAT 함수를 이용하여 TDOA를 구하는 방법은 추정에 이용되는 데이터가 많을수록 정확성이 향상되지만, 음원의 위치의 변화를 실시간으로 추적하기에는 적절하지 않다. 따라서 제안하는 알고리즘은 실시간으로 음원의 위치를 추정하며, 적은 계산량으로 정확한 위치를 추정하기 위한 알고리즘을 제안한다. 실시간 위치 인식 가능한 저계산량 GCC-PHAT

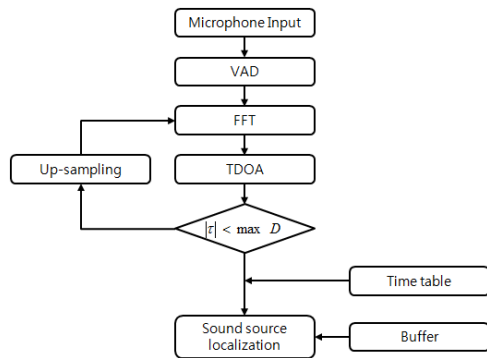


Fig. 4. Proposed algorithm.

알고리즘에 위치 인식 정확도를 높이기 위해 업샘플링 방법, 조정응답 파워 방법, VAD(Voice Activity Detection) 기능을 결합하는 방법을 제안한다. 업샘플링 방법을 사용하여 위치 인식 분해능을 높였고 위치 인식 정확도를 더욱 높이기 위해 조정 응답 파워 방법을 사용하여 GCC-PHA에서 예측된 지연 예측 값의 주변 9개의 후보 중 응답 파워 값을 최대로 갖는 지연 값을 최종 예측 지연으로 결정한다. 또한 각 마이크로폰에 입력된 신호에 VAD를 적용하여 음성 구간을 검출하고, 검출된 음성 구간의 신호를 프레임 단위로 나누어 TDOA 추정을 수행한다. 실제 음성이 존재하지 않는 묵음 구간에서 위치 인식을 수행하지 않고 음원이 존재하는 프레임에서만 위치 인식 알고리즘을 수행함으로써 위치 인식의 정확도를 높인다.

음성 신호는 프레임 단위로 FFT를 취한 후 GCC-PHAT 알고리즘을 수행한다. GCC-PHAT를 이용하여 측정 가능 범위를 벗어난 지연 시간이 검출되면, 해당 프레임을 업샘플링하여 측정 가능 범위를 증가시켜 다시 TDOA 추정을 수행한다. 검출된 음원 위치는 프레임의 연속성을 고려하여 버퍼에 저장된 음원 위치 추정 값과의 비교를 통하여 최종 음원 위치를 추정한다. Fig. 4는 제안하는 알고리즘을 나타내고 있다.

3.1 업샘플링(Up-Sampling)

샘플링 주파수는 측정 가능한 지연 샘플의 최대값과 관련이 있다. 이 최대값에 의한 측정 범위 $-\max \sim \max$ 의 각 지연 샘플들은 측정하고자 하는 각도 범위인 $0^\circ \sim 180^\circ$ 에 각각 매핑된다. 즉, 측정 가능 범위의 지연 샘플들은 각도의 해상도를 나타낸다. TDOA를

Table 1. Delay sample and angle mapping.

Delay sample	Angle(°)	Delay sample	Angle(°)
0	90	-1	92.43
1	87.56	-2	94.87
2	85.12	-3	97.32
⋮	⋮	⋮	⋮
22	20.77	-22	159.22
23	12.17	-23	167.82

Table 2. Delay sample and angle mapping after up-sampling.

Delay sample	Angle(°)	Delay sample	Angle(°)
0	90	-1	91.21
1	88.78	-2	92.43
2	87.56	-3	93.65
⋮	⋮	⋮	⋮
46	12.17	-46	167.82
47	2.88	-47	177.13

추정하는데 있어서 입력 신호 데이터의 증가는 신호의 유사성을 비교함에 있어서 더 정확한 측정 값을 가질 뿐만 아니라, 측정 가능한 최대 지연 샘플의 증가로 인한 각도에 대한 해상도를 증가시킨다.

오디오 신호의 업샘플링 과정은 오디오 데이터 사이에 영의 값을 넣어 데이터 수를 늘리고 저주파통과 필터를 통과하여 얻는다.

본 논문에서의 실험 환경 조건인 16 kHz의 샘플링 주파수에서는 최대 지연 샘플 값이 23으로 총 47개의 지연 샘플들을 가지고 있다. 이때, 업샘플링하여 샘플링 주파수를 2배 증가시키면 최대 지연 샘플 값이 47로 증가하게 되고, -47 ~ 47 범위의 지연 샘플들을 가지게 된다. Table 1은 16 kHz의 샘플링 주파수를 사용하였을 때 지연 샘플 값에 따른 주파수 매핑 상황을 나타내고, Table 2는 2배로 업샘플링하였을 때의 지연 샘플 값과 그에 따른 주파수 매핑을 나타낸다. 0에 가까운 지연샘플일수록 각도의 변화가 적고 $\pm \max$ 에 가까운 지연샘플은 각도의 변화가 더 크다.

3.2 조정 응답 파워 방법을 사용한 페루프 예측

마이크로폰의 위치와 샘플링 주파수를 알고 있을 경우 임의의 위치에서 발생된 음원에 의한 시간 지연을 미리 구할 수 있다. 고정된 마이크로폰 어레이의

위치와 임의의 음원 위치를 각각 $(x_{mi}, y_{mi}), (x_s, y_s)$ 으로 표현하면, Eq. (11)을 이용하여 음원의 위치에 따른 지연 값을 미리 구할 수 있다.

$$d_k = \frac{\sqrt{(x_{m1} - x_{sk})^2 + (y_{m1} - y_{sk})^2}}{\sqrt{(x_{m2} - x_{sk})^2 + (y_{m2} - y_{sk})^2}} \quad (11)$$

GCC-PHAT를 이용해 구해진 값이 Eq. (11)에 의해 계산된 Time-table의 후보 위치 중 하나와 매칭이 되고, 그 값을 감싸고 있는 주변의 지연 값과 추정된 τ 를 이용하여 입력 신호를 지연시켜 DAS(Delay and Sum)을 하였을 경우 최대 파워 값을 나타내는 τ 를 최종 위치로 추정한다. GCC-PHAT 얻어진 τ 는 기존 SRP(Steered Response Power)-PHAT 방식과 유사하게 후보 위치들 추정하는 방식으로, 모든 후보 위치의 추정으로 인한 많은 계산량을 후보 위치의 한 부분만을 추정 영역으로 정하여 그 계산량을 줄이고, 잡음이나 잔향에 의한 추정의 정확도 저하를 주변 값들과의 비교를 통하여 최적의 위치를 추정함으로써 정확도를 향상시킬 수 있다. Fig. 5는 측정하고자 하는 실내 환경에서 임의의 음원 위치별 값과 각도를 나타내고 파란원이 GCC-PHAT에 의해 계산된 값일 때, 빨간 테두리에 둘러싸인 부분이 비교 추정될 부분이다.

3.3 음성 구간 검출

VAD 방법에는 영교차율(Zero Crossing Rate), 선형

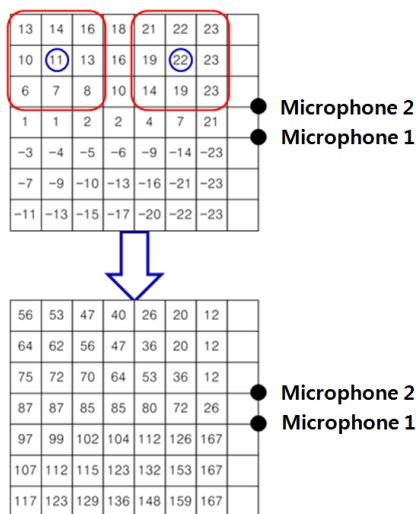


Fig. 5. Time-table.

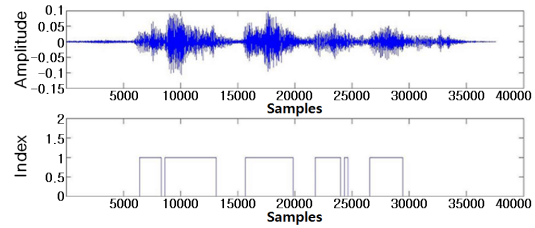


Fig. 6. SNR 20 dB, TDOA estimation result of GCC-PHAT.

예측 부호화(Linear Predictive Coding, LPC) 계수, 프레임의 에너지를 활용하는 양적 특징들을 이용하는 방법, 우도비(Likelihood Ratio), 엔트로피 등과 같이 통계적인 특징들을 이용하는 방법들이 있다.^[13]

제안하는 알고리즘에서는 Eq. (12)를 이용하여 양적 특징을 활용한 방법 중 프레임의 에너지를 측정하여 음성과 비음성 구간을 구분하였다.

$$E_k = \frac{1}{L} \sum_{m=1}^L |x_k(m)| \quad (12)$$

k 는 프레임 인덱스, L 은 프레임 길이를 나타낸다.

본 논문에서는 프레임 에너지의 임계값을 $0.01J$ 로 정의하여, 임계값보다 높으면 음성 구간, 낮으면 비음성 구간으로 판별하였다. 전체 신호에 대하여 프레임 별 음성 구간을 Fig. 6에 나타내었다.

IV. 성능 평가

4.1 실험 환경

4.1.1 입력 신호 모델링

두 개의 마이크로폰과 각각 1개의 음원과 잡음, 그리고 잔향이 존재하는 $5\text{ m} \times 5\text{ m}$ 내 공간을 모델링하였다. 실내 공간에서 음원과 잡음의 위치가 서로 다르기 때문에, 음원과 잡음에 대한 실내 충격 응답 함수를 각각 적용하여 마이크로폰에 입력되는 신호를 생성하였다.^[14]

Fig. 7은 마이크로폰 1에 입력된 신호를 나타낸다. 음원과 잡음 각각에 대하여 직접음과 간접음을 모두 적용하였다. Fig. 8은 두 마이크로폰에 입력된 신호의 지연 시간을 보여준다. 실험에 사용된 음원은 16 kHz

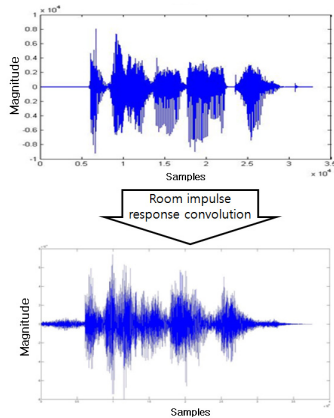


Fig. 7. Microphone input signal generation process.

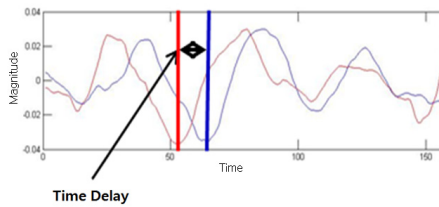


Fig. 8. Delay time of input signal.

의 한국 남성, 한국 여성, 미국 남성 및 Babble Noise이다. 실험에 있어서 DFT Point 개수는 512, 프레임 길이는 320을 사용하였다.

4.1.2 음원 위치 설정

두 개의 마이크론 사이의 거리를 0.5 m로 고정하

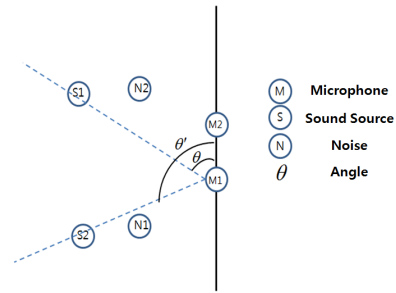


Fig. 9. Test environment.

고, 음원 및 잡음의 위치를 변경하며 실험을 진행하였다. 마이크론 기준에서 음원이 60°, 2 m의 거리에 위치할 때 잡음은 135°, 1.41 m 거리에, 음원이 120°, 2 m의 거리에 위치할 때 잡음은 45°, 1.41 m에 위치하였다. Fig. 9는 실험을 위한 마이크론 어레이, 음원, 잡음의 배치를 나타낸다.

4.2 성능 평가

본 논문에서는 위의 실험 환경에서 기존의 GCC-PHAT와 기존 GCC-PHAT 방법에 VAD를 적용하여 음성 구간에서만 음원의 위치 추정, 제안하는 방법에 대한 각각의 실험을 통하여 성능을 평가하였다. Table 3은 마이크론 간의 거리 0.5 m일 때 각각의 실험 결과를 보여준다.

지연 샘플당 매칭되는 각도의 수 제한에 따라 추정하고자 하는 각도의 ± 10° 이내의 결과를 성공, 각도

Table 3. MD:0.5m, SNR:20 dB, comparison of experimental results.

		GCC-PHAT		Applied VAD		Proposed method	
		60°	120°	60°	120°	60°	120°
Korean male	Success	46/117	45/117	25/46	31/49	38/46	43/49
	Failure	58/117	56/117	17/46	12/49	8/46	6/49
	Etc	13/117	17/117	4/46	6/49	0/46	0/49
	Rate	39 %	38 %	54 %	63 %	84 %	87 %
Korean female	Success	39/117	45/117	27/50	32/54	44/50	49/54
	Failure	54/117	37/117	18/50	14/54	6/50	4/54
	Etc	24/117	35/117	5/50	8/54	0/50	1/54
	Rate	33 %	38 %	54 %	59 %	88 %	90 %
Foreign male	Success	37/117	42/117	24/37	28/38	31/37	33/38
	Failure	63/117	58/117	12/37	7/38	4/37	4/38
	Etc	17/117	17/117	1/37	3/38	2/37	1/38
	Rate	31 %	35 %	64 %	73 %	83 %	86 %

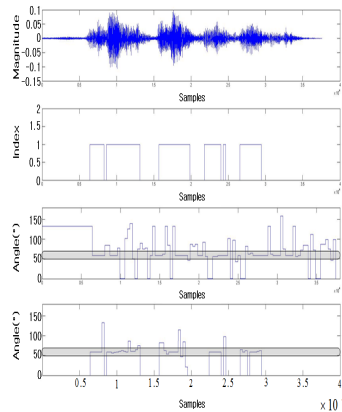


Fig. 10. Performance comparison between GCC-PHAT and proposed algorithm.

차이가 많이 나는 결과는 실패, Eq. (3)에 의한 추정 지연 시간 범위를 넘는 결과는 기타로 분류하였다.

Table 3의 GCC-PHAT와 VAD가 적용된 방법의 결과에서 알 수 있듯이 GCC-PHAT 성능은 비음성 구간에서의 추정 실패에 큰 영향을 받는다. 그 이유는 음원의 프레임 별 처리에 있어서 각 프레임마다 TDOA 추정 값들이 결정되는데, 음원의 위치 정보를 표현하지 않는 잡음이나 묵음과 같은 비음성 구간이 상당부분 차지하면서 음원 위치 추정의 전체적인 정확성에 영향을 주기 때문이다. 이러한 문제점을 해결하기 위하여, 본 논문에서는 VAD를 적용하여 음원의 위치 정보를 담고 있는 음성 구간에서만 TDOA를 추정하여 성능을 향상시켰다. 또한 음성 구간에서도 측정 범위를 벗어난 TDOA 값을 추정하는 프레임은 업샘플링 과정을 통하여 TDOA 값을 다시 추정하고 Time-table을 이용하여 음원의 위치를 최종적으로 추정함으로써 Table 3의 내용과 같이 성능을 향상시켰다.

Fig. 10은 한국 남성의 음원이 60°에 위치하는 상황에서 기존 GCC-PHAT의 결과와 제안하는 방법의 결과를 나타낸 것이며, 그림의 음영 부분은 실험에서 오차 범위로 한 ±10° 범위인 50° ~ 70°를 나타낸 것이다. Fig. 10은 순차적으로 마이크로폰에 입력된 음성 및 잡음 신호, 음성 구간 판별 결과, 기존 알고리즘의 위치 추정 결과, 제안하는 방법의 위치 추정 결과이다. Fig. 10을 비교하였을 때 제안하는 방법이 비음성 구간의 영향을 적게 받아 기존의 방법보다 균일하게 음원의 위치를 추정할 것을 알 수 있다.

V. 결 론

본 논문에서는 GCC-PHAT를 이용한 음원 위치 추정 알고리즘의 잔향 및 잡음에 의한 정확성 저하 특성을 개선하고자 GCC-PHAT 알고리즘에 3가지 방법을 추가하는 방법을 제안하였다. 업샘플링 방법, VAD 적용, 조정 파워 방법을 사용한 여러 후보 지연 값 중의 최적의 지연 값 결정의 방법을 GCC-PHAT에 적용하였다. 프레임 별로 입력 신호를 나누어 실시간으로 위치를 추정하였으며, 정확성의 향상을 위해 VAD를 적용하여 음원의 정보를 담고 있는 음성 구간에서만 TDOA를 추정하였다. 또한 마이크로폰 특성에 따른 샘플링 주파수에 의한 측정 범위를 초과하는 신호는 업샘플링을 통해 최대 측정 범위를 넓혀 측정 가능하도록 하였으며, 미리 구해진 Time-table의 주변 지연 값들과의 파워 값을 비교하여 음원의 위치를 최종적으로 추정하였다.

기존 GCC-PHAT 방법의 경우 비음성 구간에서의 미 없는 TDOA를 추정함에 따라 그 정확도가 떨어지고, 음성 구간에서도 음원 위치 추정의 오차가 발생하는 반면, 제안하는 알고리즘의 경우 음원의 정보를 담고 있는 음성 구간에서만 TDOA를 추정함에 따라 위치 추정의 정확성이 향상되었고, 음성 구간에서의 추정 오차도 기존의 알고리즘보다 비음성 구간의 영향을 적게 받아 그 편차가 적은 것을 확인할 수 있었다. 실험 결과 기존 알고리즘의 전체 추정 평균 정확도 35.7%, 음성구간에서의 평균 정확도 61.2%를 제안하는 방법에서는 평균 정확도 86.3%로 각각 50.6%, 25.1%의 성능을 향상시켰다. 이후 제안하는 알고리

즘의 성능 개선을 위하여, 마이크론 간 거리에 따른 성능 저하와, 여러 레벨의 SNR에 따른 음원 위치 추정 성능을 개선하고자 한다.

감사의 글

이 논문은 2016 년도 충북대학교 연구년제 사업의 연구비 지원에 의하여 연구되었음(This work was financially supported by the Research Year of Chungbuk National University in 2016).

References

1. Y. E. Kim and J. G. Chung, "The method of elevation accuracy in sound source localization system" (in Korean), *IEEK* **2**, 24-29 (2009).
2. K. Nakadai, T. Lourens, H. G. Okuno, and H. Kitano, "Active audition for humanoid," *AAAI/IAAI*, 832-839 (2000).
3. B. Kwon, G. Kim, and Y. Park, "Considering microphone positions in sound source localization methods: in robot application," *Robot & Human Interactive Communication*, **7**, 1050-1054 (2007).
4. W. Kelleman, "A self-steering digital microphone array," *Acoustics, Speech, and Signal Processing (ICASSP)*, **5**, 3581-3584 (1991).
5. J. Stachurski, L. Netsch and R. Cole, "Sound Source localization for video surveillance camera," *Advanced Video and Signal Based Surveillance (AVSS)*, 93-98 (2013).
6. M. Omologo and P. Svaizer, "Acoustic event localization using a crosspower-spectrum phase based technique," *Acoustics, Speech, and Signal Processing (ICASSP)*, **2**, 273-276 (1994).
7. A. Johansson and S. Nordholm, "Robust acoustic direction of arrival estimation using Root-SRP-PHAT, a realtime implementation," *Acoustics, Speech, and Signal Processing (ICASSP)*, **4**, 933-936 (2005).
8. J. Wang, Y. Zhao, and Z. Wang, "A MUSIC like DOA estimation method for signals with low SNR," *Global Symposium on Millimeter-Waves(GSMM)*, 321-324 (2008).
9. J. Baszun, "Passive sound source localization system," *Zeszyty Naukowe Politechniki Bialostockiej. Informatyka*, 5-16 (2011).
10. M. S. Brandstein and H. F. Silverman, "A robust method for speech signal time-delay estimation in reverberant rooms," *Acoustics, Speech, and Signal Processing (ICASSP)*, **1**, 375-378 (1997).
11. J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, *Robust Localization in Reverberant Rooms, Microphone Arrays* (Springer, Berlin Heidelberg, 2001), chapter 8, pp. 157-180.
12. Abad Sorbet. Marta, "Comparison of different methods for time delay estimation," *UPNA 10-15* (2010).
13. G. H. Lee, Y. J. Lee, and M. N. Kim, "Voice activity detection algorithm using wavelet band entropy ensemble analysis in car noisy environments" (in Korean), *Journal of Korea Multimedia Society*, **16**, 1005-1017 (2013).
14. A. Bertrand and G. Bernardi, *Audio processing: lab sessions, Session 1: Introduction to the acoustic simulation environment*, 2015.

저자 약력

▶ 김 영 준 (Young-Joon Kim)



2004년 2월: 충북대학교 전기전자및 컴퓨터 공학부 학사
2008년 2월: 충북대학교 전파공학과 석사
2009년 2월: 충북BIT연구중심대학육성사업단 전임 연구원
2009년 3월 ~ 현재: 충북대학교 전파통신 공학과 박사 수료

▶ 오 민 재 (Min-Jae Oh)



2014년: 충북대학교 정보통신과(학사)
2014년 ~ 현재: 충북대학교 전파통신 공학과(석사)
<관심 분야> 음성/ 오디오 신호처리

▶ 이 인 성 (In-Sung Lee)



1983년 2월: 연세대학교 전자공학과 학사
1985년 2월: 연세대학교 전자공학과 석사
1992년 12월: Texas A&M University 전기공학과 박사
1986년 5월 ~ 1987년 7월: 한국통신 연구개발단 전임 연구원
1993년 2월 ~ 1995년 9월: 한국전자통신연구원 이동통신기술연구단 선임 연구원
1995년 5월 ~ 현재: 충북대학교 전자정보대학 교수