

순환 확률분포를 이용한 다중 음원 방향 추정

Direction Estimation of Multiple Sound Sources Using Circular Probability Distributions

남 승 현, 김 용 호
(Seung-Hyon Nam, Yong-Hoh Kim)

배재대학교 전자공학과

(접수일자: 2011년 5월 25일; 수정일자: 2011년 7월 4일; 채택일자: 2011년 7월 21일)

본 논문에서는 주기성을 갖는 순환 확률분포를 이용하여 $0^\circ \sim 360^\circ$ 범위의 다중 음원의 방향을 추정하는 기법을 제안한다. 음원의 방향 정보를 담고 있는 마이크로폰 간의 위상차는 확률분포의 혼합물로 간주될 수 있으며, 음원 방향은 이 확률분포의 혼합물에 적용된 로그-우도함수 (log-likelihood function)를 최대화함으로써 추정된다. 주기성을 갖는 데이터의 분석에 von Mises 확률분포가 널리 활용된다는 사실은 잘 알려져 있지만, 본 논문에서는 기존의 Gaussian이나 Laplacian 확률분포에 2π 모듈로 (modulo) 연산을 적용함으로써 $0^\circ \sim 360^\circ$ 범위의 주기성을 갖는 순환 확률분포를 정의하고 이를 방향 추정에 활용한다. 순환 확률분포의 혼합물에 대한 로그-우도함수를 최대가 되게 하는 음원의 방향은 EM (Expectation-Maximization) 알고리즘을 이용하여 추정된다. 다양한 반향 환경에서의 실험 결과 Laplacian 확률분포가 von Mises나 Gaussian 확률분포보다 우수한 성능을 제공함을 확인할 수 있다.

핵심용어: 방향 추정, 순환 확률분포, 로그-우도함수, Expectation-Maximization

투고분야: 음향 신호처리 분야 (1,2)

This paper presents techniques for estimating directions of multiple sound sources ranging from 0° to 360° using circular probability distributions having a periodic property. Phase differences containing direction information of sources can be modeled as mixtures of multiple probability distributions and source directions can be estimated by maximizing log-likelihood functions. Although the von Mises distribution is widely used for analyzing this kind of periodic data, we define a new class of circular probability distributions from Gaussian and Laplacian distributions by adopting a modulo operation to have 2π -periodicity. Direction estimation with these circular probability distributions is done by implementing corresponding EM (Expectation-Maximization) algorithms. Simulation results in various reverberant environments confirm that Laplacian distribution provides better performance than von Mises and Gaussian distributions.

Keywords: Direction estimation, Circular probability distribution, Log-likelihood function, Expectation-Maximization

ASK subject classification: Acoustic Signal Processing (1,2)

I. 서론

음원의 방향 추정은 로봇, 화상회의, 음원분리, 환경감시 등 다양한 분야에 적용될 수 있는 요소 기술의 하나이다. 음원의 방향 추정은 신호가 다수의 마이크로폰들에 인가될 때 발생하는 신호의 도래시각차 (Time Difference of Arrival: TDOA)를 이용하여 추정된다. TDOA는 마이크로폰 신호 간의 GCC (Cross-correlation)로부터 계산

되는데 동시에 여러 음원이 존재할수록, 반향이 커질수록, 분산 잡음이나 센서 잡음이 클수록 정확도가 저하되는 경향이 있다. 이러한 문제들을 보완하기 위하여, GCC-PHAT, SRP 빔형성기법, 사람의 청각특성을 이용한 방법, subspace 방식들이 제안되었다 [1-6].

방향을 추정하는 또 다른 방법은 우도함수 (likelihood function)를 최대화하는 것이다 [7]. 이는 음원의 방향이 반향이나 잡음 등의 영향을 받아 특정한 확률분포를 따른다는 가정에 기인한 것이다. 여러 개의 음원이 동시에 존재하는 경우 우도함수는 여러 개의 독립적인 확률분포 함수가 혼합된 형태로 정의된다. 가장 널리 활용되는 것

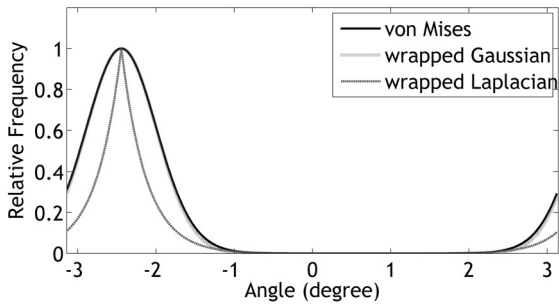


그림 1. 2π 주기를 갖는 순환 확률분포
Fig. 1. 2π -periodic circular probability distributions.

은 GMM (Gaussian Mixture Model)이지만 [7] 꼬리가 긴 특징을 갖는 LMM (Laplacian Mixture Model)을 적용한 사례도 있다 [8]. 이때 개별적인 음원들의 혼합 정도를 사전에 알 수 없으므로 우도함수의 최대화는 통상적으로 EM (Expectation-Maximization) 알고리즘을 통해 계산된다. 혼합된 확률분포를 이용한 음원의 방향 추정은 각각의 주파수-시간에서 음원의 혼합 정도 (responsibility)를 계산할 수 있으므로, 음원 분리에 적용할 때 soft-decision을 사용할 수 있다는 장점이 있다.

마이크로폰 어레이가 2개의 센서로 구성된 경우 2차원 평면에서 180° 전방의 방위각 추정이 가능하고 3개의 마이크로폰으로 구성된 2차원 평면 어레이를 사용하면 360° 전체 방위각 범위에서 추정할 수 있다.

우도함수를 이용한 방향 추정에서 고려될 부가적인 문제는 3개 이상의 마이크로폰으로 구성된 평면 어레이를 이용하여 360° 범위의 음원 방향을 추정하는 경우 확률분포가 순환 특성을 가져야 한다는 점이다. von Mises 확률분포는 2π 주기를 갖도록 고안되었기 때문에 이런 용도에 널리 활용되었다 [9]. 그러나 기존의 Gaussian과 Laplacian 확률분포 함수를 이용하여 360° 전 방향을 추정을 하려면 확률분포 함수가 2π 주기를 갖도록 변형되어야 한다. 이 논문에서는 이렇게 변형된 확률분포 함수를 각각 wrapped Gaussian, wrapped Laplacian 확률분포 함수라고 부르고 (그림 1), 2π 주기성을 갖는 이들 세 가지 확률분포를 통칭하여 순환 확률분포 (circular probability distribution)라고 부르기로 한다.

Laplacian 확률분포는 긴 꼬리를 갖는 특징으로 인해 간섭신호나 outlier에 의해 영향을 덜 받는 강점이 있다. 이러한 특징은 다중 음원 환경에서 상호 음원 간의 영향력을 배제할 수 있는 가능성을 제공한다.

이 논문에서는 먼저 3개의 마이크로폰으로 구성된 정삼각형 마이크로폰 어레이의 구조와 공간 에일리어싱 문

제를 기술한다. 다음 2π 주기성을 갖는 von Mises 확률분포를 소개하고, Gaussian과 Laplacian을 주기 데이터에 적용하기 위해 필요한 변형에 대하여 기술하며, 이들 순환 확률분포 함수들을 이용한 음원의 방향 추정을 위한 EM 알고리즘을 도출한다. 마지막으로 컴퓨터 모의실험을 통해 순환 확률분포들의 성능을 비교한다.

II. 마이크로폰 어레이의 공간 에일리어싱

본 논문에서는 그림 2와 같이 3개의 마이크로폰으로 구성된 정삼각형 구조의 어레이를 고려한다. 마이크로폰 사이의 간격은 d 이고 음원의 방향은 마이크로폰 1을 기준으로 시계방향으로 양의 각도를 갖는다고 가정된다. τ_{ij} 를 마이크로폰 i 와 j 간의 시간지연이라고 하면, 음원의 방향 θ 는 마이크로폰 2, 3을 기준으로 할 때 다음의 관계를 갖는다.

$$\tau_{12} = \frac{d \sin(\theta + 2\pi/3)}{c} \quad (1a)$$

$$\tau_{23} = \frac{d \sin(\theta)}{c} \quad (1b)$$

$$\tau_{31} = \frac{d \sin(\theta - 2\pi/3)}{c} \quad (1c)$$

여기서 c 는 소리의 속도이다. 시간지연의 범위는 $|\tau_{ij}| < d/c = \tau_{\max}$ 로 제한된다.

시간지연을 계산하는 방법은 마이크로폰 신호 간의 cross-correlation이 최대가 되는 시점을 계산하는 것이며, 실제적으로 이는 주파수 영역에서 두 신호 간의 위상차를 계산함으로써 효과적으로 수행된다. 이제 마이크로폰 신호의 단구간 푸리에 변환 (short-time Fourier transform)을 다음과 같이 정의하자.

$$X_i(n, \omega) = F\{w(m)x_i((n-1)L+m)\}, \quad (2)$$

$$i = 1, 2, 3, \quad m = 0, 1, \dots, N-1$$

여기서 L 은 신호의 블록 크기 (혹은 hop 크기), $x_i((n-1)L+m)$ 은 마이크로폰 i 에서 수집된 n 번째 프레임 신호, $w(m)$ 은 길이 N 의 윈도우 함수이다. 그러면 시간-주파수 (n, f) 에서 두 마이크로폰 i, j 신호간의 위상차는 마이크로폰 i, j 간의 시간지연 τ_{ij} 와 다음의 관계를 갖는다.

$$\arg\left\{\frac{X_i(n,f)}{X_j(n,f)}\right\} = 2\pi f\tau_{ij} + 2\pi l, \quad (3)$$

$$i \neq j, l = 0, \pm 1, \pm 2 \dots$$

여기서 마이크로폰 신호 간의 위상차는 항상 $[-\pi, \pi]$ 의 범위로 제한되므로, 마이크로폰 간의 거리가 $d > \frac{c}{2f}$ 이면 공간 에일리어싱으로 인해 위상차와 시간지연의 일대일 관계가 성립하지 않는다.

에일리어싱의 문제를 해결하기 위해 다양한 방법이 제안되었다. 한 방법은 특정 시간지연 τ^* 에 해당하는 방향의 신호가 인가되었을 때 주파수 f 에서 식 (3)을 만족하는 모든 l 값을 수용하도록 고안된 특별한 stencil 필터를 활용하는 것이다 [10]. 또 다른 방법은 정수 l 을 별도로 고려하지 않고 대신 $-\tau_{\max}$ 부터 $+\tau_{\max}$ 의 시간지연 범위를 여러 개의 구간으로 나누고 각각의 구간에 독립적인 Gaussian 확률분포를 배정함으로써 측정된 위상차로부터 최적의 시간지연을 추정하였다 [7]. 그러나 이 방식은 구간의 수만큼 많은 Gaussian 확률분포함수를 배정함에 따라 복잡도가 증가되는 단점이 있다. 최근에는 정수 l 을 잠재변수로 정의하기도 하였다 [11]. 만약 τ_{ij} 의 값을 대략적으로 알고 있다면, 정수 l 은 가능한 여러 값 중의 하나가 될 수밖에 없다는 점을 이용한 것이다. 이 방식을 적용하면, 식 (3)과 (1)로부터 마이크로폰 신호 간의 위상차 $\arg\left\{\frac{X_i(n,f)}{X_j(n,f)}\right\}$ 는 방향 $\theta(n,f)$ 로 일대일 매핑될 수 있다.

그림 2의 정삼각형 어레이는 세 쌍의 마이크로폰으로부터 3개의 지연시간을 계산할 수 있으며, 이들은 좌표 변환을 통해 다음과 같이 마이크로폰 어레이의 중심을 기준으로 하는 시간지연으로 변환된다 [12].

$$\begin{pmatrix} \tau'_1(n,f) \\ \tau'_2(n,f) \\ \tau'_3(n,f) \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{2}{3}} \sin(\frac{-\pi}{6}) & \sqrt{\frac{2}{3}} & \sqrt{\frac{2}{3}} \sin(\frac{7\pi}{6}) \\ \sqrt{\frac{2}{3}} \sin(\frac{2\pi}{3}) & 0 & \sqrt{\frac{2}{3}} \sin(\frac{-2\pi}{3}) \\ \sqrt{\frac{1}{3}} & \sqrt{\frac{1}{3}} & \sqrt{\frac{1}{3}} \end{pmatrix} \begin{pmatrix} \tau_{12}(n,f) \\ \tau_{23}(n,f) \\ \tau_{31}(n,f) \end{pmatrix}$$

$$= \begin{pmatrix} a \sin \theta \\ a \cos \theta \\ 0 \end{pmatrix} \quad (4)$$

이로부터 관찰된 음원의 방향은 다음과 같이 산출된다.

$$\theta(n,f) = \arctan\left\{\frac{\tau'_1(n,f)}{\tau'_2(n,f)}\right\}, \quad -\pi < \theta(n,f) \leq \pi \quad (5)$$

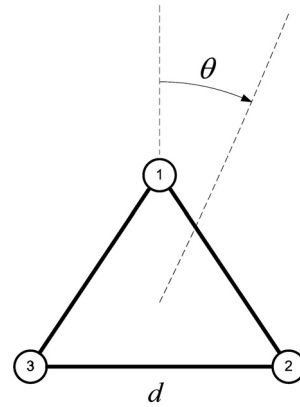


그림 2. 정삼각형 마이크로폰 어레이
Fig. 2. Equilateral triangular microphone array.

식 (5)로부터 산출된 $\theta(n,f)$ 를 모든 시간-주파수에서 누적하면 GCC-PHAT 히스토그램을 얻을 수 있다. 히스토그램은 해당 방향에 음원이 존재할 가능성을 상대적으로 보여준다. 히스토그램을 저대역 통과필터로 부드럽게 처리한 다음 피크값의 위치를 추정하면 음원의 방향을 추정할 수 있다 [6,12].

III. 순환 확률분포 함수를 이용한 방향 추정 알고리즘

3.1. 주기 데이터를 위한 순환 확률분포

그림 1에서 볼 수 있듯 음원의 방향은 2π 주기를 갖기 때문에 주기 데이터에 적합한 확률분포 함수를 사용하여야 한다. 이런 목적으로 널리 활용되는 확률분포는 von Mises 확률분포로서 다음과 같이 정의된다 [9].

$$V(\theta_{nf}|\theta_k, m_k) = \frac{1}{2\pi I_0(m_k)} \exp\{m_k \cos(\theta_{nf} - \theta_k)\} \quad (6)$$

여기서 I_0 는 제1 유형의 0차 베셀함수 (the zeroth-order modified Bessel function of the first kind)이고, θ_k 와 m_k 는 각각 평균과 집중도 (concentration)이다. von Mises 확률분포의 주기성은 $\cos(\theta_{nf} - \theta_k)$ 에서 보장된다.

일반적으로 Gaussian이나 Laplacian 확률분포는 주기성을 갖지 않지만, von Mises 확률분포와 유사하게 데이터를 2π 모듈로 (modulo)로 정의하면 주기성을 갖는 데이터를 적절히 표현할 수 있다 [13].

$$\tilde{\theta}_k^{nf} = \text{mod}(\theta_{nf} - \theta_k + \pi, 2\pi) - \pi \quad (7)$$

이렇게 정의된 방향정보는 Gaussian과 Laplacian 확률분포에 동일하게 적용될 수 있다. 본 논문에서는 이들을 각각 wrapped Gaussian과 wrapped Laplacian 확률분포라고 부르기로 하고 다음과 같이 정의한다.

$$N(\theta_{nf}|\theta_k, \sigma_k^2) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left\{-\frac{(\tilde{\theta}_k^{nf})^2}{2\sigma_k^2}\right\} \quad (8)$$

$$L(\theta_{nf}|\theta_k, \lambda_k) = \frac{1}{2\lambda_k} \exp\left\{-\frac{|\tilde{\theta}_k^{nf}|}{\lambda_k}\right\} \quad (9)$$

3.2. 순환 확률분포를 이용한 방향 추정

일반적으로 K 개의 확률분포로 구성된 혼합물은 다음과 같이 표현될 수 있다.

$$p(\theta_{nf}) = \sum_{k=1}^K \pi_k P_k \quad (10)$$

여기서 $\theta_{nf} = \theta(n, f)$ 를 단순히 표기한 것이며, P_k 는 k 번째 음원 방향 θ_k 에 위치한 확률분포 함수를 의미하고, π_k 는 k 번째 확률분포의 가중치이다. 이제 z_k^{nf} 이 시간-주파수 (n, f) 의 신호가 음원 k 에 속함을 나타내는 0과 1의 값을 갖는 잠재변수 (latent variable)라고 정의하고, 잠재변수 벡터의 모든 집합을 $\mathbf{Z} = \{[z_1^{nf}, \dots, z_K^{nf}]^T\}_{nf}$ 로 정의하자. 또한 $\boldsymbol{\mu} = [\mu_1, \dots, \mu_K]^T$ 를 확률분포 함수의 파라미터 (hyper-parameter)의 벡터로 정의하고, $\boldsymbol{\theta} = \{\theta_{nf}\}$ 를 모든 시간-주파수에서의 방향 관찰 값의 벡터라고 가정하면, $(\boldsymbol{\theta}, \mathbf{Z})$ 에 대한 우도함수는 다음과 같이 표현할 수 있다.

$$p(\boldsymbol{\theta}, \mathbf{Z}|\boldsymbol{\pi}, \boldsymbol{\mu}) = \prod_{n,f} \prod_{k=1}^K \pi_k^{z_k^{nf}} P_k^{z_k^{nf}} \quad (11)$$

식 (11)의 우도함수는 잠재변수를 알지 못하므로 직접 최대화할 수 없다. 대신 EM 알고리즘을 적용하여 반복적으로 최대화한다. 이를 위해, 보조함수 (auxiliary function)를 다음과 같이 $\log\{p(\boldsymbol{\theta}, \mathbf{Z}|\boldsymbol{\pi}, \boldsymbol{\mu})\}$ 의 잠재변수에 대한 평균으로 정의한다.

$$Q = E_{\mathbf{Z}}\{\log p(\boldsymbol{\theta}, \mathbf{Z}|\boldsymbol{\pi}, \boldsymbol{\mu})\} = \sum_{n,f} \sum_{k=1}^K \gamma_k^{nf} \{\ln \pi_k - \ln P_k\} \quad (12)$$

여기서 γ_k^{nf} 는 잠재변수 z_k^{nf} 의 평균으로 정의되며 EM 알고리즘의 E-스텝에서 다음과 같이 계산된다.

$$\gamma_k^{nf} = \frac{\pi_k P_k}{\sum_{j=1}^K \pi_j P_j} \quad (13)$$

EM의 M-스텝에서는 $\sum_k \pi_k = 1$ 조건을 활용하여 식 (12)의 보조함수 Q 를 최대화하는 파라미터를 계산한다. von Mises, wrapped Gaussian, wrapped Laplacian 확률분포 혼합모델에서 파라미터 π_k 에 대한 식은 다음과 동일하게 주어진다.

$$\pi_k = \frac{1}{NF} \sum_{n,f} \gamma_k^{nf} \quad (14)$$

여기서 N 과 F 는 각각 수집된 신호의 블록수와 주파수 개수를 의미한다.

이외의 다른 파라미터들에 대한 표현식은 각 모델에 따라 달리 유도된다. 여기서는 상세한 유도과정은 생략하고 유도과정에서의 주요 사항만을 언급하고 결과를 기술하기로 한다.

3.2.1. vMMM (von Mises Mixture Model)

von Mises에서 식 (12)의 보조함수를 최대화하는 평균 θ_k 은 식 (12)를 θ_k 에 대하여 미분함으로써 다음과 같이 얻어진다.

$$\theta_k = \arctan\left(\frac{\sum_{n,f} \gamma_k^{nf} \sin \theta_{nf}}{\sum_{n,f} \gamma_k^{nf} \cos \theta_{nf}}\right) \quad (15a)$$

그러나 파라미터 m_k 는 직접 계산이 불가능하며 다음을 만족하는 값을 산술적으로 계산하여야 한다 [9].

$$A(m_k) = \frac{I_1(m_k)}{I_0(m_k)} = \frac{\sum_{n,f} \gamma_k^{nf} \cos(\theta_{nf} - \theta_k)}{\sum_{n,f} \gamma_k^{nf}} \quad (15b)$$

여기서 $I_1(m_k)$ 는 제 1 유형의 1차 베셀함수 (the first-order modified Bessel function of the first kind)이다.

3.2.2. wGMM (wrapped Gaussian Mixture Model)

Gaussian의 경우 θ_k 의 추정값은 모듈로 특성으로 인해 von Mises에서 적용한 방법처럼 직접적으로 계산할 수 없다. 대신 기존의 평균값에 대한 편차를 갱신해나가는 방법으로 계산한다. 현재의 추정값 θ_k 는 이전의 추정값

θ_k^{old} 와 오차 $\tilde{\theta}_k$ 의 합으로 정의할 수 있으므로 식 (7)을 다음과 같이 표현될 수 있다.

$$\begin{aligned} \tilde{\theta}_k^{nf} &= \text{mod}(\theta_{nf} - \theta_k^{old} - \tilde{\theta}_k + \pi, 2\pi) - \pi \\ &= \theta_k^{old} + \text{mod}(\theta_{nf} - \tilde{\theta}_k + \pi, 2\pi) - \pi \end{aligned} \quad (16)$$

이제 식 (16)을 이용하면, 식 (12)의 보조함수를 최대화하는 Gaussian에 대한 파라미터 갱신식은 다음과 같이 표현될 수 있다 [13].

$$\theta_k = \theta_k^{old} + \frac{\sum_{n,f} \gamma_k^{nf} \tilde{\theta}_k^{nf}}{\sum_{n,f} \gamma_k^{nf}} \quad (17a)$$

$$\sigma_k^2 = \frac{\sum_{n,f} \gamma_k^{nf} (\tilde{\theta}_k^{nf})^2}{\sum_{n,f} \gamma_k^{nf}} \quad (17b)$$

3.2.3. wLMM (wrapped Laplacian Mixture Model)

Laplacian의 경우에는 앞서 언급한 Gaussian에서 적용한 바와 같이 식 (16)을 적용함으로써 식 (17a)와 유사한 형태의 갱신식을 유도할 수 있지만, Laplacian의 경우 식 (9)에서 나타난 바와 같이 절대값 표현을 포함하고 있어 $\theta_k = \theta_{nf}$ 에서 미분이 가능하지 않다는 문제가 발생한다. 이러한 문제는 [8]에 제안된 방법을 활용함으로써 해결할 수 있으며, 결과적으로 다음과 같은 파라미터 갱신식을 얻을 수 있다.

$$\theta_k = \theta_k^{old} + \frac{\sum_{n,f} \gamma_k^{nf} \tilde{\theta}_k^{nf}}{\sum_{n,f} \frac{1}{|\tilde{\theta}_k^{nf}|}} \quad (18a)$$

$$\lambda_k = \frac{\sum_{n,f} \gamma_k^{nf} |\tilde{\theta}_k^{nf}|}{\sum_{n,f} \gamma_k^{nf}} \quad (18b)$$

IV. 실험 및 결과고찰

제안된 알고리즘들의 성능을 비교 평가하기 위하여 다음과 같이 모의실험을 실시하였다. 실험에 사용된 임펄스 응답은 MATLAB 패키지인 Roomsim [14]을 이용하여 생성되었다. 실험에서 설정한 환경은 $6\text{ m} \times 9\text{ m} \times 3.5\text{ m}$ 크기의 방 중앙에 마이크로폰 어레이가 놓여져 있으며 마이크로폰 사이의 간격은 $d = 5\text{ cm}$ 이고 음원들은 마이

크로폰 어레이로부터 1 m 거리에 떨어져 있다. 사용된 음성신호는 16 kHz로 샘플링되었으며 5초 길이의 분량이 다. 다중 음원의 경우 각 음원 파일의 에너지가 동일하도록 정규화된 다음 임펄스 응답과 콘볼루션 합성하였다. 센서 잡음을 고려하기 위해 백색잡음을 추가하여 SNR이 15 dB가 되도록 하였다. 음원의 개수는 사전에 알고 있다고 가정하였으며, EM 알고리즘의 방향 초기값은 GCC-PHAT 히스토그램의 결과를 적용하였다. EM 알고리즘의 최대 반복횟수는 50으로 설정하였다.

각 순환 확률분포에 대한 성능은 서로 다른 음성 파일로 구성된 10개의 세트에 대한 결과로부터 다음과 같이 RMSE (Root Mean Squared Error)를 계산함으로써 평가되었다.

$$\text{RMSE} = \sqrt{\frac{1}{10} \sum_{m=1}^{10} \{\theta_k^* - \theta_k(m)\}^2} \quad (19)$$

4.1. 단일 신호에 대한 성능 비교

반향조건이 각각 $T60 = 50\text{ ms}$, 100 ms , 200 ms 인 조건에서 단일 음원에 대한 RMSE를 비교하였다. 그림 3는 $T60 = 200\text{ ms}$ 일 때 음원의 위치에 따른 성능을 보여준다. 그림 3에서 볼 수 있듯, wLMM이 가장 우수하고 다음 vMMM이며 wGMM이 가장 저조하다. 음원의 방향에 따라 RMSE 값이 달라지는 것은 방 구조에 따른 영향으로

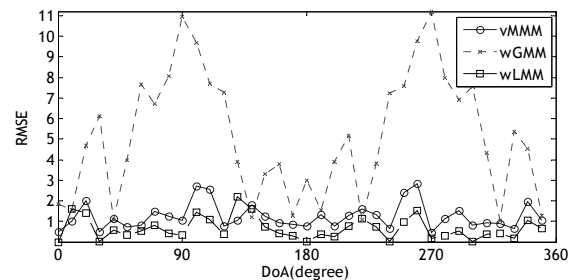


그림 3. 단일 음원에 대한 순환 확률분포들의 방향 추정 성능 ($T60 = 200\text{ ms}$)
Fig. 3. Performance of circular probability distributions for a single source ($T60 = 200\text{ ms}$).

표 1. 단일 음원에 대한 순환 확률분포들의 평균 RMSE
Table 1. Average RMSE of circular probability distributions for a single source over all directions under different reverberant conditions.

확률분포	반향조건		
	50 ms	100 ms	200 ms
vMMM	1.22	1.22	1.24
wGMM	4.60	4.68	5.12
wLMM	0.40	0.52	0.65

판단된다. 흥미로운 점은 wGMM이 vMMM이나 wLMM 보다 방의 구조에 영향을 많이 받는다는 점이다. wGMM의 경우 음원이 90°와 270°에 위치할 때 가장 저조한 성능을 보인다.

표 1은 각각의 반향조건 조건에서 RMSE 값을 방향에 대하여 평균한 것이다. 반향이 증가함에 따라 RMSE는 모든 순환 확률분포에서 공통적으로 점차 증가한다. vMMM의 경우 반향에 따른 RMSE의 증가는 아주 작지만, RMSE 값은 항상 wLMM에 비하여 2배 이상 크다. wGMM의 RMSE 값은 모든 반향조건에서 vMMM과 wLMM의 것에 비하여 현저히 크다.

4.2. 다중 신호에 대한 성능 비교

다중음원에 대한 성능을 평가하기 위하여 4개의 음원이 50°, 170°, 250°, 295°의 위치에 동시에 존재하는 경우를 고려하였다. 대부분의 음원에서 음성 구간은 서로 심하게 겹친다. 반향은 $T_{60}=100\text{ ms}$ 로 설정하였다. 표 2는 잡음이 없는 환경에서 서로 다른 10개의 음원 파일 세트에 대한 실험결과를 평균한 것이다. 표 2에서 확인할 수 있는 바와 같이, 4개의 음원이 동시에 존재하는 경우에도 wLMM은 vMMM이나 wGMM에 비하여 월등하게 우수한 성능을 보여준다. 주목할 만한 사실은, vMMM과

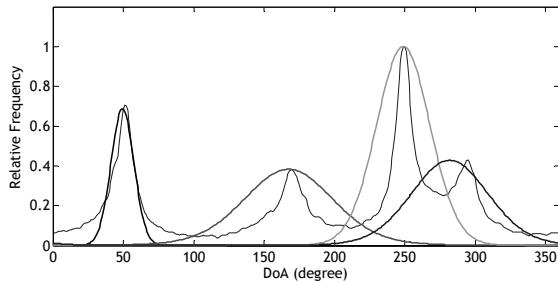


그림 4. 4개의 동시 음원에 대한 vMMM의 방향 추정 ($T_{60}=100\text{ ms}$, 잡음 없는 경우): GCC-PHAT 히스토그램 (가는선), 추정된 von Mises 확률분포 (굵은선)

Fig. 4. Direction estimation of vMMM for four simultaneous sources ($T_{60}=100\text{ ms}$, 잡음없는 경우): GCC-PHAT histogram (thin line) and estimated von Mises distributions (thick line).

표 2. 4개의 동시 음원에 대한 순환 확률분포들의 RMSE ($T_{60}=100\text{ ms}$, 잡음없는 경우)

Table 2. RMSE of circular probability distributions for four simultaneous sources ($T_{60}=100\text{ ms}$, no noise).

확률분포 \ 음원위치	50°	170°	250°	295°	평균
vMMM	1.82	2.32	1.09	6.61	2.96
wGMM	2.08	2.43	1.34	6.30	3.04
wLMM	0.67	2.86	0.64	0.76	1.23

wGMM의 경우 295°에 위치한 음원에 대한 방향 추정 정확도가 현저히 떨어진다는 점이다. 이는 그림 4에서 볼 수 있는 바와 같이 295°과 250°의 음원이 비교적 근접하여 있기 때문으로 해석된다. 이에 반해, wLMM는 정확하게 295°에 위치한 음원의 방향을 추정하였다. 이는 wLMM이 긴 꼬리를 가지고 있는 확률분포의 특성으로 인해 outlier에 강인한 성능을 보이기 때문이다.

다중 음원 환경에서 센서잡음이나 배경잡음의 영향을

표 3. 4개의 동시 음원에 대한 순환 확률분포들의 RMSE ($T_{60}=100\text{ ms}$, $\text{SNR}=15\text{ dB}$)

Table 3. RMSE of circular probability distributions for four simultaneous sources ($T_{60}=100\text{ ms}$, $\text{SNR}=15\text{ dB}$).

확률분포 \ 음원위치	50°	170°	250°	295°	평균
vMMM	1.03	1.96	1.54	10.99	3.88
wGMM	1.07	2.54	1.48	13.19	4.57
wLMM	0.23	0.88	0.61	1.75	0.87

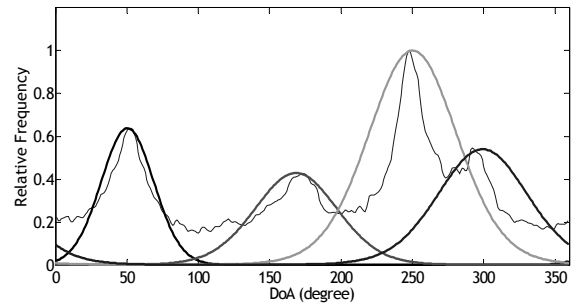


그림 5. 4개의 동시 음원에 대한 vMMM의 방향 추정 ($T_{60}=100\text{ ms}$, $\text{SNR}=15\text{ dB}$): GCC-PHAT 히스토그램 (가는선), 추정된 von Mises 확률분포 (굵은선)

Fig. 5. Direction estimation of vMMM for four simultaneous sources ($T_{60}=100\text{ ms}$, $\text{SNR}=15\text{ dB}$): GCC-PHAT histogram (solid line) and estimated von Mises distributions (thick line).

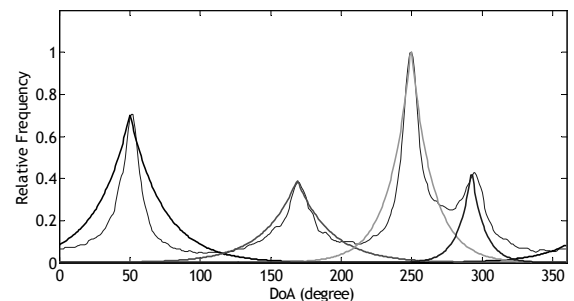


그림 6. 4개의 동시 음원에 대한 wLMM의 방향 추정 ($T_{60}=100\text{ ms}$, 잡음 없는 경우): GCC-PHAT 히스토그램 (가는선), 추정된 wrapped Laplacian 확률분포 (굵은선)

Fig. 6. Direction estimation of wLMM for four simultaneous sources ($T_{60}=100\text{ ms}$, 잡음없는 경우): GCC-PHAT histogram (thin line) and estimated wrapped Laplacian distribution (thick line).

살펴보기 위하여 SNR이 15 dB일 때 성능을 표 3에 비교하였다. 표 2와 3에서 확인할 수 있는 흥미로운 점은, vMMM이나 wGMM의 경우 잡음이 없는 환경에서의 성능이 잡음이 있는 환경에서의 성능보다 더 저조하다는 것이다. 이러한 점을 더 자세히 확인하기 위하여 잡음이 없는 환경 (그림 4)과 잡음이 있는 환경에서 (그림 5) vMMM의 추정 결과를 확인하였다. 그림 4는 295°의 음원 추정이 250°의 음원에 영향을 받아 확률분포 함수의 중앙이 실제 음원의 위치보다 왼쪽으로 이동되었음을 보여준다. 반면에, 그림 5은 295°의 음원 추정이 250°의 영향보다는 잡음의 영향을 받아 실제 음원의 위치보다 오른쪽으로 이동하였음을 알 수 있다. 이러한 결과는 잡음의 유무와 다중 음원의 존재 등이 서로 상호 작용을 일으켜 성능에 영향을 준다는 것을 의미한다. 반면에 표 2와 3을 비교하면 wLMM의 경우에는 이러한 간섭 현상이 아주 적다는 것을 확인할 수 있다. 그림 6에서 확인할 수 있듯 wLMM 확률분포가 음원과 잡음 간의 상호 간섭에 vMMM이나 wGMM보다 훨씬 강인하다는 사실에 기인한다.

V. 결론

본 논문에서는 다중 음원의 방향 추정을 위해 주기성을 갖는 순환 확률분포인 von Mises, wrapped Gaussian, wrapped Laplacian을 기술하고 이들을 이용한 방향추정 알고리즘을 도출하였다. 다양한 환경에서의 음원의 방향 추정 성능을 비교한 결과, 방향 추정은 방의 구조, 반향 정도, 잡음의 유무와 음원들 간의 상호 작용 등에 따라 달라지는 것을 확인할 수 있었다. von Mises와 wrapped Gaussian 확률분포는 이러한 환경적 인자에 민감하게 반응하는 반면에, wrapped Laplacian은 이러한 환경적 인자에 매우 강인하다는 사실을 확인할 수 있었다. 또한 wrapped Laplacian은 모든 환경에서 von Mises와 wrapped Gaussian에 비하여 훨씬 우수한 방향 추정 정확도를 제공한다.

감사의 글

이 논문은 2008년도 정부 (교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임 (KRF-2008-521-D00342).

참고 문헌

1. J. Benesty, J. Chen, and Y. Huang, *Microphone array signal processing*, Springer, 2008.

2. C. H. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 24, pp. 320-327, 1976.
3. R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, pp. 276-280, 1986.
4. H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 33, pp. 823-831, 1985.
5. L. A. Jeffress, "A place theory of sound localization," *J. Comparative Physiol. Psychol.*, vol. 41, no. 1, pp. 35-39, 1948.
6. P. Aarabi, "Self-localizing dynamic microphone arrays," *IEEE Trans. Syst., Man, Cybern. C*, vol. 32, no. 4, pp. 474-484, 2002.
7. M. I. Mandel, D. P. W. Ellis, and T. Jebara, "An EM algorithm for localizing multiple sound sources in reverberant environments," in *Adv. Neural Info. Process. Syst.*, B. Sch Ikopf, J. Platt, and T. Hoffman, Eds. Cambridge, MA: MIT Press, pp. 953-960, 2007.
8. N. Mitianoudis and T. Stathaki, "Batch and online under-determined source separation using laplacian mixture models," *IEEE Trans. on Audio, Speech, and Lang. Proc.*, vol. 15, pp. 1818-1832, 2007.
9. C. M. Bishop, *Pattern recognition and machine learning*, Springer, 2006.
10. C. Liu, B. C. Wheeler, Jr, R. C. Bilger, C. R. Lansing, and A. S. Feng, "Localization of multiple sound sources with two microphones," *J. Acoust. Soc. Amer.*, vol. 108, no. 4, pp. 1888-1905, 2000.
11. N. T. Thom, and S. H. Nam, "An expectation-maximization method for the permutation problem in frequency-domain blind speech separation," in *Proc. of ICASSP2010*, 2010.
12. Y. Hioka, M. Matsuo, and N. Hamada, "Multiple-speech-source localization using advanced histogram mapping method," *Acoustical Science and Technology*, vol. 30, no. 2, 2009.
13. P. Smaragdis, and P. Boufounos, "Position and trajectory learning for microphone arrays," *IEEE Trans. on Speech and Audio Proc.*, Jan, 2007.
14. D. R. Campbell, K. J. Palomaki, and G. J. Brown, "Roomsim, a matlab simulation of shoebox room acoustics for use in teaching and research," in <http://media.paisley.ac.uk/~campbell/Roomsim/>, 2008.

저자 약력

•남 승 현 (Seung-Hyon Nam)

1980년: 서강대학교 전자공학과 (학사)
 1987년: 알라바마 헨츠빌 대학교 컴퓨터및전기공학과 (석사)
 1992년: 텍사스 에이앤엠 대학교 전기공학과 (박사)
 1993년 ~ 현재: 배재대학교 전자공학과 교수
 ※ 주관심 분야: 음성 및 오디오 신호처리, 신호 분리, 어레이 신호처리

•김 용 호 (Yong-Hoh Kim)

1975년: 연세대학교 물리학과 (학사)
 1987년: 아이오와 대학교 컴퓨터및전기공학과 (석사)
 1999년: 충남대학교 전자공학과 (박사)
 1989년 ~ 현재: 배재대학교 전자공학과 교수
 ※ 주관심 분야: 통신이론