

영교차점과 캡스트럼 전처리 기술을 이용한 반향환경에서의 음원방향 추정*

박용진(서강대), 이수연(서강대), 박형민(서강대)

<차 례>

- | | |
|------------------------|---|
| 1. 서론 | 3.1. 직선 경로 신호 구간 검출 |
| 2. 관련 연구 소개 | 3.2. 영교차점을 이용한 음원방향 추정
에 직선 경로 신호 구간 검출 적용 |
| 2.1. 캡스트럼 전처리 기술 | 4. 실험 및 결과 |
| 2.2. 영교차점을 이용한 음원방향 추정 | 5. 결론 |
| 3. 제안한 알고리즘 설명 | |

<Abstract>

Zero-Crossing-Based Source Direction Estimation Using a Cepstral Prefiltering Technique

Yong-Jin Park, Soo-Yeon Lee, Hyung-Min Park

To estimate directions of multi-sound sources, we consider an approach based on zero crossings which provided more robust results to diffuse noise than the conventional cross-correlation-based method [6][7]. In reverberant environments, the performance of source direction estimation can be improved by using signal components through direct paths from sources to microphones. Since a cepstral prefiltering technique [8] removes the effect of reverberation, we propose a source direction estimation method which can find out intervals of the direct-path components by comparing original and cepstral-prefiltered envelopes. Simulations demonstrate that the proposed method can improve the performance of source direction estimation in reverberant environments.

* Keywords: Source direction estimation, Zero crossings, Cepstral analysis.

1. 서 론

인간에게 있어서 음원방향 추정은 여러 다른 소리들과 잡음이 있는 환경에서도 특정한 소리를 찾아내고 추적할 수 있는 중요한 능력이다. 이러한 음원방향 추정은 로봇이 자신을 부르는 사용자를 바라보게 하여 의사소통 능력을 향상시키고 사용자 친밀도를 높이거나, 휴대전화 핸즈프리 기능의 감도 개선 등 원거리 인터페이스 성능 향상에 필수적인 기술이다. 인간은 음원방향 추정을 위해서 두 귀에서 감지되는 신호의 차이를 이용한다고 알려져 있다. 특히 두 귀에서 감지하는 신호의 시간차(interaural time difference: ITD)와 신호의 세기차(interaural intensity difference: IID)가 중요한 요소이다[1]. ITD는 1.5 kHz 이하의 주파수에서 주로 이용되고, 또한 높은 주파수 신호에 대해서는 포락선(envelope)을 검출한 후 사용할 수 있으며, IID는 보다 높은 주파수의 신호에 주로 사용된다. 음원방향 추정을 위하여 인간의 청각 시스템에 기반한 많은 계산 모델들이 개발되어 왔으며[2]-[5], 최근에 각 주파수 대역에서 신호들의 영교차점을 비교하여 ITD를 계산하는 방법이 제안되었다[6][7]. 음원방향 추정을 위한 대표적인 방법인 교차상관(cross-correlation) 값을 이용한 ITD 계산 방법과 비교하여, 이 방법은 잡음이 있는 환경에서 음원방향 추정 성능이 뛰어나고 계산이 간단하다는 장점을 보였다.

비록 영교차점을 이용한 ITD 계산이 확산잡음환경에서는 좋은 성능을 보이나, 실제 환경에 적용하기 위해서는 반향을 고려해야 된다. 본 논문에서는 영교차점을 이용하며 반향신호에 대해 적절하게 대응할 수 있는 방법을 제시한다. 반향환경에 대한 일반적인 채널 필터는 직선 경로(direct path) 성분, 초기 반사(early reflection) 성분, 지연 반사(late reflection) 성분 이렇게 세 가지 성분으로 구분할 수 있다. 여기서 직선 경로에 의한 신호는 음원의 방향에 상응하는 ITD를 얻는데 필요하지만, 두 반향 성분은 원하는 ITD를 얻는데 오히려 방해가 된다[8][9]. 캡스트럼 전처리 방법은 캡스트럼 영역에서 반향 성분을 추정하고 이를 차감한 후 원래 신호를 복원함으로써 반향환경에서 교차상관값을 이용한 음원방향 추정 성능을 개선하게 된다[8]. 그러나 캡스트럼 전처리 방법을 적용하더라도 여전히 반향 성분이 남아있기 때문에 성능 저하가 나타난다.

본 논문에서는 캡스트럼 전처리된 신호를 직접 음원방향 추정에 사용하지 않고, 직선 경로에 의한 신호가 많이 포함되어 있는 구간을 검출하는데 사용한다. 즉, 캡스트럼 전처리된 신호는 원 신호에 비하여 반향 성분이 상당부분 제거된 신호이므로, 각 대역에서 원 신호의 포락선과 캡스트럼 전처리된 신호[8]의 포락선을 비교하여 두 포락선 크기의 차이가 상대적으로 작은 구간을 직선 경로에 의한 신호가 많이 포함되어 있는 구간으로 검출할 수 있다. 이 방법을 통해서 남아있는 반향 성분을 고려하지 않고 직선 경로에 의한 신호가 많이 포함되어 있는 구간만을 선택적으로 사용하여 음원방향을 추정함으로써 반향환경에서 성능을 향상시킬

수 있다.

지연 반사 성분은 상당히 오랜 시간동안 지속되어 다음 신호의 직선 경로 성분과 자주 겹치게 되는데, 이 지연 반사 성분이 음원의 방향에 해당하는 ITD값 추정을 방해한다. 특히, 지연 반사 성분은 에너지 자체가 크지 않아 위의 포락선 비교만으로는 파악하기 어렵다. 일반적으로 지연 반사 성분은 연속적으로 다른 경로를 통해서 신호가 들어오기 때문에 이 성분은 이웃한 신호라 할지라도 보통 다른 ITD를 나타내게 된다. 이 특성을 이용하여 지연 반사 성분에 의한 영향의 정도를 파악하기 위하여, 원래 일정 구간 ITD들의 분산으로부터 확산잡음이 포함된 정도를 측정하는 신호 대 잡음비(signal-to-noise ratio: SNR) 추정 방법[6]을 활용한다.

ITD는 배경잡음에 취약하고, IID는 반향환경에서 취약하다고 알려져 있다[10]. 신호 대 잡음비 추정 방법을 이용하면 배경잡음에 둔감한 음원방향 추정이 가능하기 때문에, 이 논문에서는 ITD만을 고려한 배경잡음과 반향환경에 둔감한 음원방향 추정 알고리즘을 제안한다. ITD라는 용어가 두 귀 신호처리 분야에서 주로 쓰이는 용어이지만, 여기에서는 두 개의 마이크 간의 거리를 공간 엘리어싱(spatial aliasing)을 피하기 위해 인간의 귀보다 훨씬 가까운 거리로 설정한다. 또한 가까운 마이크를 이용하게 되면 두 개의 마이크 사이의 왜곡이 작기 때문에 신뢰성이 높아지는 장점을 가지게 된다.

2. 관련 연구 소개

2.1. 켈스트럼 전처리 기술 (Cepstral Prefiltering Technique)

확산잡음을 무시할 수 있다면 음원 신호 $s(t)$ 와 음원과 마이크 사이의 채널 $h(t)$ 를 이용하여 마이크에 입력되는 신호를 다음과 같이 나타낼 수 있다.

$$x(t) = h(t) * s(t). \tag{1}$$

여기서 $*$ 는 컨볼루션(convolution) 연산을 나타낸다. 이 신호의 복소수 켈스트럼은 다음과 같이 나타낼 수 있다[11].

$$\hat{x}(k) = F^{-1}\{\log X(w)\} = F^{-1}\{\log[H(w)S(w)]\} = \hat{h}(k) + \hat{s}(k). \tag{2}$$

여기서 $X(w)$, $H(w)$, $S(w)$ 는 각각 $x(t)$, $h(t)$, $s(t)$ 의 푸리에 변환 신호를 나타내며, $F^{-1}\{\cdot\}$ 는 역 푸리에 변환을 나타낸다. 또, $\hat{h}(k)$, $\hat{s}(k)$ 는 각각 $h(t)$, $s(t)$ 의 켈스트럼 값을 나타낸다.

주파수 영역에서 일반적인 채널 필터 $H(w)$ 는 다음과 같이 minimum-phase component (MPC)와 all-pass component (APC) 두 부분으로 나누어 생각할 수 있다 [11].

$$H(w) = H_{ap}(w) H_{min}(w). \quad (3)$$

여기서 APC에 해당하는 $H_{ap}(w)$ 는 음원과 마이크 사이의 시간지연과 관계있는 성분이지만, MPC에 해당하는 $H_{min}(w)$ 는 반향 성분에 의해서 음원 신호가 왜곡되도록 영향을 미치는 성분이다[8]. 식 (3)을 cepstrum 변환하면 다음과 같이 나타낼 수 있다.

$$\hat{h}(k) = \hat{h}_{ap}(k) + \hat{h}_{min}(k). \quad (4)$$

여기서 $\hat{h}_{ap}(k)$ 와 $\hat{h}_{min}(k)$ 는 대칭성을 이용하여 다음과 같이 구할 수 있다[11].

$$\hat{h}_{ap}(k) = \begin{cases} \hat{h}(k), & k < 0, \\ 0, & k = 0, \\ -\hat{h}(-k), & k > 0. \end{cases} \quad (5)$$

$$\hat{h}_{min}(k) = \begin{cases} 0, & k < 0, \\ \hat{h}(k), & k = 0, \\ \hat{h}(k) + \hat{h}(-k), & k > 0. \end{cases} \quad (6)$$

식 (2)와 식 (4)를 이용하여 마이크 신호의 cepstrum을 다음과 같이 표현할 수 있다.

$$\hat{x}(k) = \hat{h}_{ap}(k) + \hat{h}_{min}(k) + \hat{s}(k). \quad (7)$$

식 (7)에서 보는 바와 같이 전체 마이크 신호의 cepstrum에서 채널 cepstrum의 MPC 부분 $\hat{h}_{min}(k)$ 만을 추출, 제거하여 음원 신호가 반향 성분으로 왜곡되는 것을 방지하여 음원과 마이크 사이의 시간지연을 정확히 구할 수 있고, 마이크 간의 시간지연 차이로부터 음원의 방향을 정확히 추정할 수 있다[8]. 이 때, 마이크 신호로부터 cepstrum을 구하는 과정은 프레임 단위로 이루어지는데, 일반적으로 음원 신호는 비정상(non-stationary) 특성으로 인해서 그 cepstrum의 MPC는 프레임마다 많이 달라지지만, 음원과 마이크의 위치는 급격하게 바뀌지 않으므로 채널 cepstrum의 MPC는 프레임마다 별 차이가 없다. 따라서 음원 신호 cepstrum의 MPC 평균값이 0이라고 가정하면 마이크 신호 cepstrum의 MPC를 프레임별로 귀환식을 통해 평균을 내어 채널 cepstrum의 MPC를 구할 수 있다. 이렇게 구한 채널 cepstrum

럼의 MPC $\hat{h}_{min}(k)$ 를 마이크 신호의 캡스트럼 $\hat{x}(k)$ 에서 차감하여 채널 캡스트럼의 MPC 성분을 제거한 후 역 캡스트럼 변환을 취하면 반향 성분이 제거된 마이크 신호를 얻을 수 있다[8]. 위 캡스트럼 전처리는 16 ms의 프레임 단위로 이루어지지만, 반향 성분이 제거된 마이크 신호를 얻은 후 다음 과정부터는 프레임에 상관없이 표본단위로 순차적으로 신호를 처리하게 된다.

2.2. 영교차점을 이용한 음원방향 추정

영교차점을 이용한 음원방향 추정의 전체적인 과정은 다음 네 개의 단계로 나눌 수 있다.

2.2.1 주파수영역 분석

두 개의 마이크로 입력된 신호는 각각 인간의 달팽이관과 같은 주파수 특성을 갖는 감마톤 필터뱅크(Gammatone filterbank)를 통과하게 된다[4]. 이 필터뱅크는 중심주파수가 100 Hz부터 2 kHz까지 ERB 스케일에서 선형적으로 증가하는 19개의 채널로 구성되어 있다. 실험에 사용된 감마톤 필터뱅크는 다음 수식과 같다[5].

$$g_i = t^3 \exp(-2\pi b_i t) \cos(2\pi f_i t) u(t) \tag{8}$$

여기서 $i, b_i, f_i, u(t)$ 는 각각 채널 번호, 대역폭에 비례하는 필터 감쇠율, 중심 주파수, 계단함수를 나타낸다. 참고로 포락선은 위 식의 cosine 항을 complex sinusoid로 변경한 후 출력신호의 크기를 계산하여 구한다[12].

2.2.2. ITD 계산

영교차점은 2.2.1절의 필터뱅크를 통과한 입력신호들로부터 검출된다. 필터뱅크를 통과한 신호의 정확한 영교차점은 영점을 지나기 직전과 직후의 표본값을 선형적으로 이어서 영점과 교차되는 시간을 찾는다. 영교차점에 기반한 ITD를 계산하기 위해서 한 마이크 입력신호의 영교차점을 다른 마이크 신호와 비교해서 가장 근접한 영교차점 사이의 시간차를 계산하게 된다[6][13]. 이 시간차는 음성이 두 개의 마이크에 도달하는데 걸리는 최대 시간차보다 작으면 유효하다고 보며, 이보다 큰 시간차 정보는 모두 무시하게 된다. 위에 언급한 바와 같이 두 마이크의 거리를 표본화 주파수까지 공간 엘리머싱을 피할 수 있을 만큼 가깝게 설정한다. 이 때, ITD의 최대값은 입력신호 최대 주파수의 반주기 이하가 되므로 제일 작은 영교차점의 시간차가 음원의 방향에 상응하는 ITD값이 된다.

2.2.3. 신호 대 잡음비 계산

반향이 없는 경우에 특정 음원에서 마이크에 도달하는 신호는 그 음원의 방향에 해당하는 일정한 ITD를 나타내지만, 확산잡음은 특정한 음원을 상징하지 않기 때문에 ITD가 영교차점마다 제각각이다. 따라서 일정 구간의 ITD들에 대한 분산을 이용하여 해당 신호에서 신호 대 잡음비를 다음 식과 같이 추정할 수 있다[6].

$$\widehat{SNR}_i(j) = \begin{cases} 10\log_{10} \frac{1}{w_i^2 S_i^2(j)}, & \text{if } w_i^2 S_i^2(j) < 1, \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

여기서 w_i 는 감마톤 필터뱅크의 i 번째 채널의 중심주파수를 나타내고, $S_i^2(j)$ 는 j 번째 ITD를 기준으로 한 분산값을 나타낸다. 큰 분산값을 가진 ITD는 식 (9)에 따라서 추정한 신호 대 잡음비가 낮게 되어 확산잡음에 큰 영향을 받은 것으로 볼 수 있으므로, 신호 대 잡음비가 크게 추정된 ITD만을 주로 음원방향 추정에 사용하게 된다[6].

2.2.4. 가중치를 반영한 히스토그램(Histogram)을 이용한 음원방향 추정

2.2.3절에서 정의된 신호 대 잡음비가 일정값 이상되는 ITD에 대하여 음원의 방향에 해당하는 방향각으로 변환한다. 이 때, 신호 대 잡음비가 크게 추정된 ITD에 대한 반영률을 상대적으로 높이기 위해서 각 ITD에 상응하는 방향각에 추정된 신호 대 잡음비를 누적하는 히스토그램을 구성한다. 이러한 과정은 필터뱅크의 모든 채널에 대해 반복되며, 이렇게 구성된 히스토그램에서 첨두(peak)값들이 음원의 방향에 해당한다[6].

3. 제안한 알고리즘 설명

3.1. 직선 경로 신호 구간 검출

반향환경에서 음원방향 추정을 하기 위해서는 음원과 마이크 사이의 직선 경로를 통해 들어오는 신호를 구분하여 이 신호로 음원의 방향에 상응하는 ITD를 계산해야 한다. 직선 경로를 통해 들어온 신호들을 많이 포함한 구간은 감마톤 필터뱅크를 통과한 원신호의 포락선과 캡스트럼 전처리된 신호의 포락선을 비교하여 검출할 수 있다. 캡스트럼 전처리된 신호는 음향 채널의 MPC 성분을 캡스트럼 영역에서 차감함으로써 반향신호 성분이 제거되기 때문에[8], 반향신호가 포함된

구간에서 캡스트럼 전처리된 포락선은 원신호의 포락선보다 작은 값을 가지므로 다음과 같이 직선 경로를 통해 들어오는 신호가 많이 포함된 구간을 검출할 수 있다.

$$\frac{E_i(t)}{E_{C_i}(t)} < 0.4 * m\left(\frac{E_i(t)}{E_{C_i}(t)}\right) \ \& \ E_i(t) > 0.01 \quad (10)$$

여기서 $m(\cdot)$ 는 중간(median)값을 나타내며, $E_i(t)$ 와 $E_{C_i}(t)$ 는 각각 20 ms 구간에서 평균을 취한 원신호와 캡스트럼 전처리를 한 신호의 i 번째 필터뱅크 출력의 최대크기가 1로 정규화된 포락선이다. 두 번째 조건은 일정정도 이상의 에너지를 포함한 신호 구간을 찾기 위한 것이다. 이를 통해서 직선 경로를 통한 신호가 많이 포함된 구간을 대략적으로 검출한 후, 그 구간의 시작과 끝 지점을 정확히 찾아야 한다. 직선 경로는 보통 반사 경로에 비하여 짧아서 마이크에 가장 먼저 도착하기 때문에, 상대적으로 반사 성분에 영향을 덜 받는 직선 경로를 통한 신호 구간을 검출하기 위해서는 음성이 시작하는 부분을 찾는 것이 효과적이다. 따라서 식 (10)에 의해서 대략적으로 검출된 구간으로부터 이전 10 ms 범위 안에서 포락선 값이 최소가 되는 지점을 시작점으로 하며, 반대로 끝점은 검출 구간으로부터 이후 10 ms 범위 안에서 포락선 값이 최대가 되는 지점으로 정한다.

3.2. 영교차점을 이용한 음원방향 추정에 직선 경로 신호 구간 검출 적용

3.1절에서 검출된 직선 경로를 통한 신호가 많이 포함된 구간에만 국한하여 ITD를 계산하게 되면 반향환경에서도 적절한 음원방향 추정이 가능하다. 그러므로 영교차점을 이용한 음원방향 추정 알고리즘[6]에서 2.2.1절의 감마톤 필터뱅크 출력 신호를 구한 후, 캡스트럼 전처리를 이용하여 검출된 직선 경로를 통한 신호가 많이 포함된 구간에 대해서만 2.2.2절의 ITD를 계산한다.

일반적으로 지연 반사 성분은 상당한 시간동안 지속되어 자주 뒤따라오는 신호의 직선 경로 성분과 겹치게 된다. 특히, 캡스트럼 전처리를 이용하여 직선 경로를 통한 신호가 다수 포함된 구간을 검출하였다 하더라도 지연 반사 성분은 에너지 자체가 크지 않아 포락선 비교만으로는 파악하기 어려운 반면에, 이렇게 포함된 반향 신호들은 음원의 방향과 다른 ITD를 유발시킬 수 있기 때문에 배제할 필요가 있다. 지연 반사 성분은 보통 연속적으로 다른 경로를 통해서 입력되므로 이 신호에 대한 ITD도 계속 변하게 된다. 이 특성을 이용하여 원래 확산잡음에 의한 영향을 파악하기 위해 제안된 신호 대 잡음비 추정 방법[6]을 지연 반사 성분에 의해 신호가 얼마나 왜곡되었는지를 측정하기 위하여 활용할 수 있다.

따라서 2.2.3절의 신호 대 잡음비 계산을 통해 지연 반사 성분에 영향을 받지 않는 ITD를 선별한 후, 이를 가중치로 하는 2.2.4절의 히스토그램으로부터 음원의

방향에 해당하는 방향각을 찾을 수 있다.

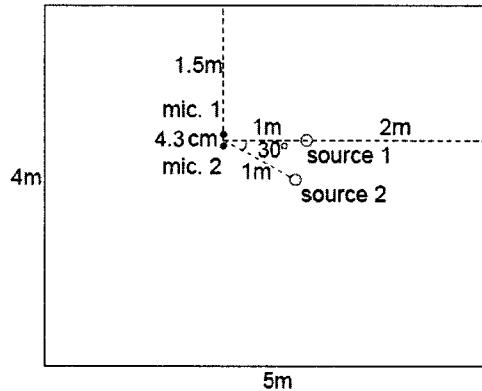
4. 실험 및 결과

반향환경에서 제안한 방법을 통한 음원방향 추정을 수행할 때 신호 대 잡음비를 계산하는데 있어서 직선 경로 신호 구간이 일반적으로 매우 짧기 때문에 현재 ITD와 앞, 뒤로 2개씩 이웃한 ITD를 선택하여 총 5개의 ITD로 신호 대 잡음비를 계산하였다. 또한 지연 반사 성분의 영향을 배제하기 위하여 신호 대 잡음비가 30 dB 이상 되는 ITD들만 선별하여 히스토그램에 사용하였으며, 히스토그램의 해상도는 1° 로 설정하였다. 실험에서 특정 음원으로부터 마이크에 도달하는 반향신호는 그 음원 신호와 image 방법을 이용하여 생성한 음원에서 마이크까지의 음향 반향에 대한 채널 필터[14]의 콘볼루션을 통해서 구하였고, 특정 마이크로 입력되는 반향신호는 모든 음원에서 그 마이크에 도달하는 반향신호를 모두 더하여 생성하였다. <그림 1>은 사용한 음원, 마이크, 방에 대한 정보를 나타내고 있다. 직육면체 방의 반향 시간 RT_{60} 는 0.3초, 0.5초, 0.7초로 설정하였으며, 음원은 TIMIT 데이터베이스[15]의 연속된 문장을 사용하였다. 마이크간의 거리는 4 kHz까지 공간 엘리머싱을 피하기 위해 4.3 cm로 설정하였기 때문에, 일반적인 표본화 주파수에서 특정 음원으로부터 두 마이크까지 도달하는 소리의 상대적인 시간지연을 고려할 때, 하나의 표본 시간차라고 하더라도 상당히 큰 음원방향 차이에 해당하게 된다. 따라서 표본 시간차를 이용하여 음원방향에 대해 만족할 만한 해상도를 얻기 위해서는 높은 표본화 주파수를 사용하여야 하기 때문에, 음원의 표본화 주파수를 1,024 kHz로 높여서 1,024 kHz에서 생성된 음향 채널 필터를 통과시켜 음원방향에 대한 해상도를 확보한 뒤 16 kHz로 낮추어 사용하였다.

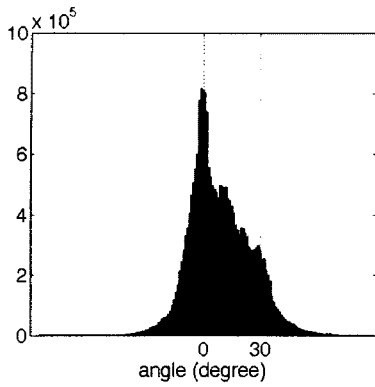
<그림 2>는 위에 제시한 조건으로 실험을 하였을 때 히스토그램 결과를 나타내고 있다. <그림 1>에서 보인 조건과 같은 0° 와 30° , 두 개의 음원에 대해서 <그림 2>의 (a), (c), (e)는 음원과 마이크의 직선 경로 신호 구간을 따로 선택하지 않는 기존의 영교차점을 이용한 음원방향 추정 알고리즘[6]에 해당하는 결과이며, <그림 2>의 (b), (d), (f)는 제안한 방법으로 직선 경로 신호 구간을 선택했을 때의 결과이다. 이 결과를 비교해보면 직선 경로 신호 구간 검출을 통한 방법이 기존 방법에서는 볼 수 없었던 두 개의 첨두값을 확연히 나타내고 있음을 확인할 수 있으며, 그 각도가 음원의 방향인 0° 와 30° 에 해당하는 것을 알 수 있다. 실제로 음원 방향에 해당하는 첨두값을 찾기 위해서는 먼저 기준(중앙) 각도를 하나씩 옮겨가면서 앞뒤 두 각도를 포함한 이웃한 다섯 각도에 대해서 평균을 구한 조금 더 부드러운 히스토그램을 구한다. 그리고 가장 큰 첨두값으로 규준화한 국부 최대치가 일정 기준(여기서는 0.15)을 넘는 첨두들에 해당하는 각도를 음원방향으로

얻게 된다.

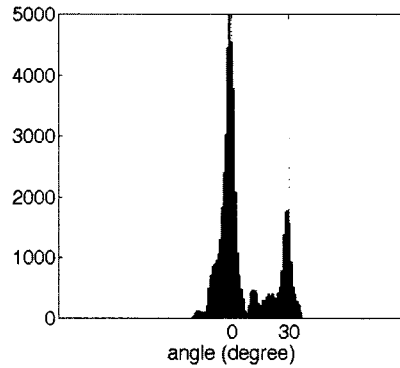
다양한 환경에서 제안한 방법이 적용 가능함을 보이기 위해서 실험 조건을 달리하여 실험을 반복하였다. 반향 시간 RT_{60} 이 0.3초, 0.5초, 0.7초에서 각각 250개의 신호에 대하여 실험하였는데, 그 음원은 TIMIT 데이터베이스에서 5명의 화자로부터 발생된 각 10개의 문장을 사용하였고, 마이크에서 음원의 방향은 0° , 15° , 30° , 45° , 60° 로 바뀌가면서 실험하였다. 마이크 신호는 위의 실험과 동일한 방법을 통해서 생성하였다. <그림 2>의 결과를 통해서 제안한 방법이 다수 개의 음원 방향을 동시에 추정할 수 있음을 보였기 때문에, 여기에서는 반향 시간, 음원 방향, 음원 신호를 변경하였을 때 사용한 파라미터의 둔감성을 확인하기 위하여 하나의 음원에 대해 실험하였다. <그림 3>은 허용오차를 3° , 5° , 10° 로 하였을 때의 음원 방향 추정을 성공한 비율을 보여 주고 있다. 제안한 방법을 적용했을 경우에 직선 경로 신호 구간을 선택하지 않는 기존 방법에 비하여 음원방향 추정 성공비율이 최소 15% 이상 향상됨을 확인할 수 있었다.



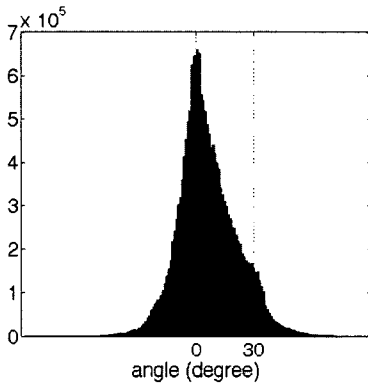
<그림 1> 직육면체 방에서의 음원과 마이크의 위치에 대한 정보. 방의 높이는 3 m, 음원과 마이크의 높이는 1.1 m로 설정하였다



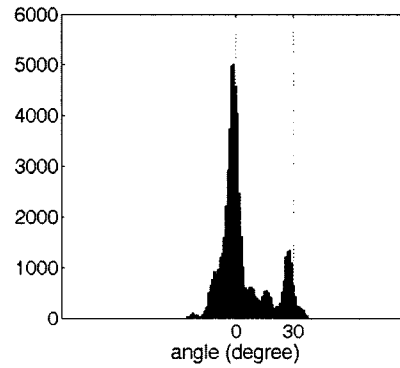
(a)



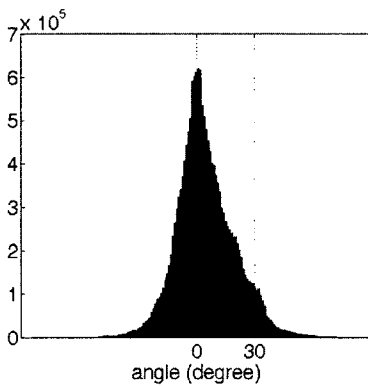
(b)



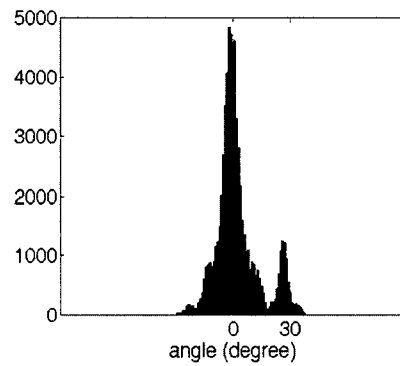
(c)



(d)

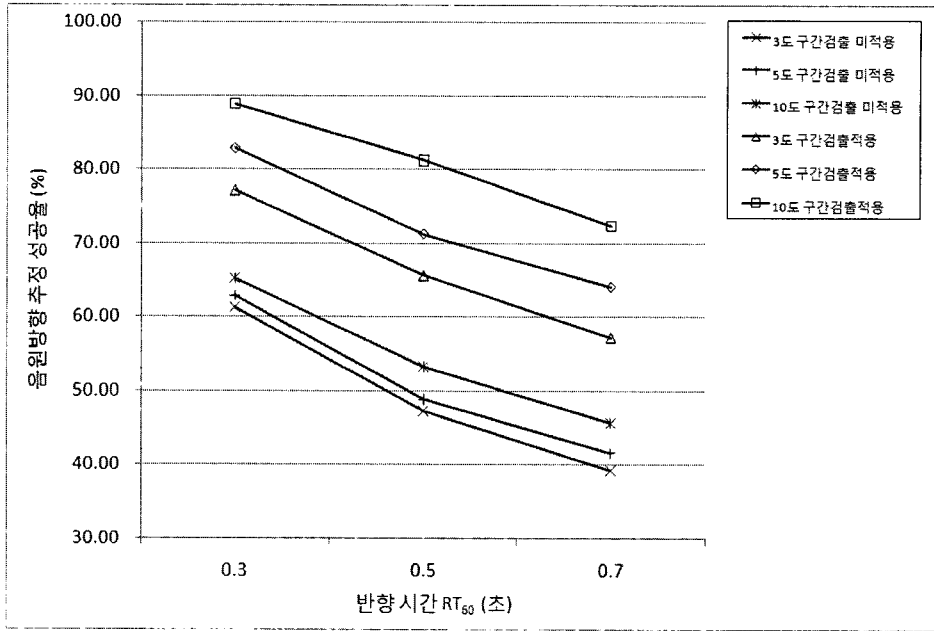


(e)



(f)

<그림 2> <그림 1>환경에서의 음원방향 추정 히스토그램; (a) RT_{60} 가 0.3초일 때 구간검출 미적용, (b) RT_{60} 가 0.3초일 때 구간검출 적용, (c) RT_{60} 가 0.5초일 때 구간검출 미적용, (d) RT_{60} 가 0.5초일 때 구간검출 적용, (e) RT_{60} 가 0.7초일 때 구간검출 미적용, (f) RT_{60} 가 0.7초일 때 구간검출 적용



<그림 3> 3°, 5°, 10° 오차 허용 시 음원방향 추정 성공률

5. 결론

본 논문은 음원방향 추정 방법에 관하여 다루었으며, 특히 실세계 응용을 위해서 필수적인 반향환경을 직접적으로 고려하였다. 즉, 반향 성분이 상당부분 제거된 캡스트림 전처리된 신호의 포락선과 원래 신호의 포락선 크기를 비교하여 음원방향 추정에 적합한 음원과 마이크의 직선 경로를 통한 신호를 주로 포함하는 구간을 검출함으로써 일반적인 반향환경에서도 적용이 가능한 음원방향 추정 알고리즘을 제안하였다. 특히 이 방법은 직선 경로 신호 구간을 검출하는 추가적인 계산 부담이 있지만, 반향환경에서 성능이 크게 떨어지는 기존 방법들의 문제점을 개선할 수 있다. 제안한 방법이 반향환경에서 음원방향 추정 성능을 개선시켰지만, 직선 경로 신호 구간 검출을 위해서 식 (10)에 사용된 파라미터들을 체계적으로 결정할 수 있는 방법, 그리고 교차상관법 등 기존 음원방향 추정 방법으로의 확장에 관한 연구가 향후에 진행되어야 할 것이다.

참 고 문 헌

- [1] J. W. Strutt, "On our perception of sound direction", *Philosophical Magazine*, Vol. 13, pp. 214-232, 1907.
- [2] L. A. Jeffress, "A place theory of sound localization", *Journal of Comparative and Physiological Psychology*, Vol. 41, No. 1, pp. 35-39, 1948.
- [3] R. M. Stern, H. Colburn, "Theory of binaural interaction based on auditory-nerve data. IV. A model of subjective lateral position", *Journal of the Acoustical Society of America*, Vol. 64, No. 1, pp. 127-140, 1978.
- [4] G. J. Brown, M. P. Cooke, "Computational auditory scene analysis", *Computer Speech and Language*, Vol. 8, No. 4, pp. 297-336, 1994.
- [5] M. P. Cooke, *Modeling Auditory Processing and Organization*, Cambridge University Press, 1993.
- [6] Y.-I. Kim, R. M. Kil, "Estimation of interaural time differences based on zero-crossings in noisy multisource environments", *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 15, No. 2, pp. 734-743, 2007.
- [7] Y.-I. Kim, R. M. Kil, "Sound source localization based on zero-crossing peak-amplitude coding", *Proc. Interspeech*, pp. 477-480, 2004.
- [8] A. Stephenne, B. Champagne, "A new cepstral prefiltering technique for estimating time delay under reverberant conditions", *Signal Processing*, Vol. 59, No. 3, pp. 253-266, 1997.
- [9] J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization", *Journal of the Acoustical Society of America*, Vol. 107, No. 1, pp. 384-391, 2000.
- [10] B. Shinn-Cunningham, N. Kopco, T. Martin, "Localizing nearby sound sources in a classroom: binaural room impulse responses", *Journal of the Acoustical Society of America*, Vol. 117, No. 5, pp. 3100-3115, 2005.
- [11] A. V. Oppenheim, R. W. Schaffer, *Discrete-Time Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1975.
- [12] M. P. Cooke, *Modeling Auditory Processing and Organization*, Cambridge University Press, Cambridge, U.K., 1993.
- [13] H.-M. Park, R. M. Stern, "Spatial separation of speech signals using continuously-variable masks estimated from comparisons of zero crossings", *Proc. ICASSP*, pp. 1165-1168, 2006.
- [14] J. B. Allen, D. A. Berkley, "Image method for efficiently simulating small-room acoustics", *Journal of the Acoustical Society of America*, Vol. 65, No. 4, pp. 943-950, 1979.
- [15] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, "DARPA TIMIT acoustic phonetic continuous speech corpus", National Institute of Standards and Technology (NIST), Gaithersburgh, MD, Tech. Rep. NISTIR 4930 99-02, 1993.

접수일자: 2008년 8월 11일

게재결정: 2008년 9월 1일

▶ 박용진(Yong-Jin Park)

주소: 121-742 서울특별시 마포구 신수동 1번지 서강대학교

소속: 서강대학교 전자공학과

전화: 02) 711-8916

E-mail: lovecor@sogang.ac.kr

▶ 이수연(Soo-Yeon Lee)

주소: 121-742 서울특별시 마포구 신수동 1번지 서강대학교

소속: 서강대학교 전자공학과

전화: 02) 711-8916

E-mail: syeonl@sogang.ac.kr

▶ 박형민(Hyung-Min Park) : 교신저자

주소: 121-742 서울특별시 마포구 신수동 1번지 서강대학교

소속: 서강대학교 전자공학과

전화: 02) 705-8916

E-mail: hpark@sogang.ac.kr