

# 음향 채널의 ‘성감’ 특성을 이용한 반향환경에서의 화자 위치 탐지\*

조지원(서강대), 박형민(서강대)

## <차 례>

- |                                      |                 |
|--------------------------------------|-----------------|
| 1. 서론                                | 4. 모의실험 환경 및 결과 |
| 2. 기존 화자 위치 탐지 알고리즘                  | 4.1. 실험 환경      |
| 2.1. 기존 화자 위치 탐지 알고리즘 소개             | 4.2. 실험 결과      |
| 2.2. 적응 고유값 분할 방법의 문제점               | 5. 결론           |
| 3. 음향 채널의 ‘성감’ 특성을 이용한 화자 위치 탐지 알고리즘 |                 |

## <Abstract>

### Speaker Localization in Reverberant Environments Using Sparse Priors on Acoustic Channels

Ji-Won Cho, Hyung-Min Park

In this paper, we propose a method for source localization in reverberant environments based on an adaptive eigenvalue decomposition (AED) algorithm which directly estimates channel impulse responses from a speaker to microphones. Unfortunately, the AED algorithm may suffer from whitening effects on channels estimated from temporally correlated natural sounds. The proposed method which applies sparse priors to the estimated channels can avoid the temporal whitening and improve the performance of source localization in reverberant environments. Experimental results show the effectiveness of the proposed method.

\* Keywords: Source localization, Time delay estimation, Sparse prior.

## 1. 서 론

인간은 두 귀를 이용하여 관심을 가지고 있는 특정 화자의 위치를 파악하고, 그 소리에 집중하여 멀리 떨어져 있는 화자로부터 필요한 정보를 파악할 수 있다. 그러나 현재까지 알려진 대부분의 화자 위치 탐지 알고리즘은 반향이 심하지 않은 환경에서는 화자의 위치를 정확히 파악할 수 있지만 상당한 반향이 수반되는 일반적인 실내 환경에서 화자의 위치를 안정적으로 파악하지 못한다. 화자의 위치 탐지가 가능해지면 휴대전화 핸즈프리 기능의 감도를 크게 개선할 수 있고, 멀리 떨어져 있는 로봇을 편리하게 부를 수 있어서 로봇의 활용도가 높아지며, 불편한 헤드셋 마이크를 사용하지 않고도 음성인식이 가능해지는 등 화자와 마이크의 거리가 먼 경우에도 효과적인 통신 또는 의사소통이 가능해져 대부분의 기기에서 사용자의 이용 편의성이 크게 개선될 것이다.

화자에서 소리가 발생하여 다수의 마이크에 도달하는 데에는 소리의 경로에 따라서 마이크마다 소요 시간이 다르기 때문에, 마이크에 도달하는 신호의 시간차를 정확히 추정한다면 화자와 마이크 사이의 거리, 방향 등을 결정할 수 있다. 이러한 특성을 이용하여 두 개 또는 그 이상의 마이크에 도달하는 신호의 시간차를 추정하여 화자의 위치를 탐지하게 된다. 그러나 반향이 있는 환경에서는 화자에서 마이크까지 직접적으로 도달하는 직선 경로 신호 외에 반향을 통해서 실제 화자의 위치와 다른 방향에서 마이크에 도달하는 신호도 발생하게 되어 이 반향 성분이 화자 위치 탐지를 방해하는 요소로 작용하게 된다. 본 논문에서는 이 반향 성분에 둔감한 화자 위치 탐지 알고리즘을 제안하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서 기존의 화자 위치 탐지 알고리즘에 대해서 살펴보고, 특히 상대적으로 반향 환경에 유리한 적응 고유값 분할 방법(adaptive eigenvalue decomposition method)[1]을 분석한다. 3장에서는 음향 채널의 ‘성김(sparsity)’ 특성을 이용하여 반향 환경에서 화자 위치 탐지에 적합한 알고리즘을 제안한다. 4장에서 모의실험 환경과 결과를 보이고 5장에서 결론을 맺는다.

## 2. 기존 화자 위치 탐지 알고리즘

### 2.1. 기존 화자 위치 탐지 알고리즘 소개

화자의 위치를 탐지하기 위한 알고리즘은 대체적으로 교차상관(cross-correlation) 값에 기반한 방법과 화자에서 마이크까지의 음향 채널 필터를 직접 구하는 두 가지로 구분할 수 있다[1]-[6]. 교차상관값에 기반한 방법은 다음과 같이 일반화된 교차상관함수의 최대값에 해당하는 시간차를 구하는 방법이다.

$$\tau_{ij} = \operatorname{argmax}_m R_{ij}(m), \quad (1)$$

$$R_{ij}(m) = \sum_{k=0}^{K-1} \Phi(k) X_i(k) X_j^*(k) e^{j2\pi mk/K}. \quad (2)$$

여기서,  $K$ 는 이산푸리에변환(discrete Fourier transform)의 길이이고,  $\Phi(k)$ 는 주파수 성분별 가중치를 나타내며,  $X_i(k)$ 는  $i$ 번째 마이크 입력 신호의 이산푸리에변환에서  $k$ 번째 주파수 성분 값을 의미한다. 이 방법은 간단하면서도 반향이 없는 환경에서 마이크에 도달하는 신호간의 시간차를 비교적 정확하게 찾아낼 수 있어서 가장 보편적으로 사용되는 방법이지만, 반향이 있는 환경에서는 반향 성분으로 인해서 교차상관값이 교란되어 그 성능이 크게 저하되는 문제점을 가지고 있다[2].

이와는 반대로 화자에서 마이크까지의 음향 채널 필터를 직접 구한 후, 그 필터에서 직선 경로의 시간지연값을 비교하여 그 차이를 시간차로 추정하는 방법이 있다. 이 방법은 반향을 채널 필터에 온전히 고려할 수 있기 때문에, 반향이 있는 일반적인 실내 환경에서 화자의 위치를 탐지함에 있어서 반향 성분에 상대적으로 둔감하게 영향을 받는다. 이 음향 채널 필터를 직접 구하기 위한 다양한 방법들 중에서도 특히 적응이 가능한 고유값 분할 방법은 잡음이 있는 반향 환경에서 기존 방법들 중에 가장 뛰어난 성능을 나타내고 있어서 각광을 받고 있다[5].

고유값 분할 방법은 다음 비용함수를 최소화하여 미지의 채널 필터를 추정한다[1][7][8].

$$J(n) = \sum_{i=1}^{M-1} \sum_{j=i+1}^M \frac{e_{ij}^2(n)}{\| [\mathbf{h}_1^T \mathbf{h}_2^T \dots \mathbf{h}_M^T] \|^2}. \quad (3)$$

식 (3)에서 화자에서  $i$ 번째와  $j$ 번째 마이크까지 음향 채널사이의 상관관계에 기반한 오차신호는 다음과 같다.

$$e_{ij}(n) = \mathbf{x}_i^T(n) \mathbf{h}_j - \mathbf{x}_j^T(n) \mathbf{h}_i, \quad i, j = 1, 2, \dots, M. \quad (4)$$

식 (3)과 식 (4)에서  $\mathbf{x}_i(n) = [x_i(n) x_i(n-1) \dots x_i(n-L+1)]^T$ 는  $i$ 번째 마이크 입력 신호 벡터이고,  $\mathbf{h}_i = [h_i(0) h_i(1) \dots h_i(L-1)]^T$ 는 화자와  $i$ 번째 마이크 사이의 추정채널이다. 또,  $M$ 과  $L$ 은 각각 마이크의 개수와 추정 채널 필터의 길이이다[1].

이 때, 비용함수를 최소로 하는 채널 필터를 얻기 위한 수식을 단순화시킨 최소평균제곱법(LMS algorithm)은 다음과 같다[1][7][8].

여기서,  $\mu$ 는 학습률을 나타낸다. 추정 채널 필터  $\{\mathbf{h}_i, i = 1, 2, \dots, M\}$ 의 모든 계수 값

$$\hat{\mathbf{h}}_i(n+1) = \hat{\mathbf{h}}_i(n) - \mu \sum_{j=1}^M e_{ji}(n) \mathbf{x}_j(n), \quad i = 1, 2, \dots, M. \quad (5)$$

이 0이 되지 않도록 하기 위해 필터 벡터  $[\mathbf{h}_1^T \ \mathbf{h}_2^T \ \dots \ \mathbf{h}_M^T]$ 는 갱신할 때마다 놈(norm)이 1을 갖도록 정규화한다.

채널 필터가 추정된 후,  $i$ 번째와  $j$ 번째 마이크사이의 시간차는 직선 경로 시간 지연의 차이에 의해서 다음과 같이 결정된다[2].

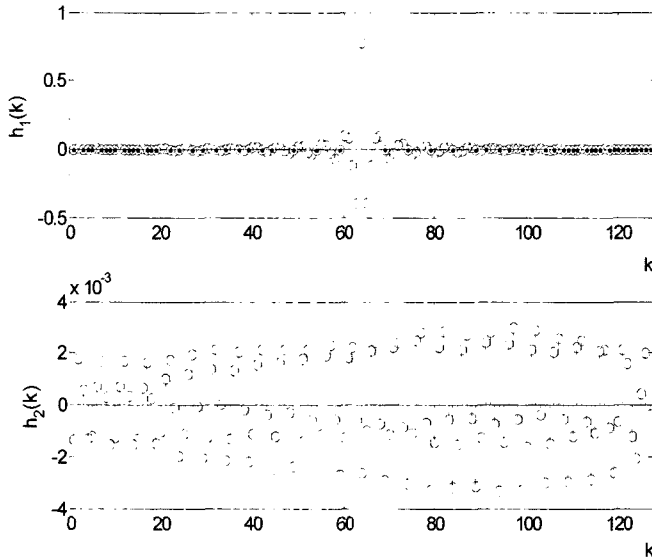
$$\tau_{ij} = \arg \max_m |h_j(m)| - \arg \max_l |h_i(l)|, \quad i, j = 1, 2, \dots, M. \quad (6)$$

## 2.2. 적응 고유값 분할 방법의 문제점

일반적으로 음향 채널 필터는 매우 많은 필터 계수로 구성되어 있다. 적응 고유값 분할 방법은 식 (3)의 비용함수를 최소화하여 채널을 추정하기 때문에, 많은 수의 채널 계수를 정확히 결정하기 위해서는 매우 많은 데이터를 필요로 하지만, 실제적으로 특정 위치의 화자가 말한 데이터의 양이 너무 적은 경우가 많다. 또, 비용함수는 채널들 간의 상관관계에 기초한 오차신호를 척도로 하기 때문에, 채널들 간의 상대적인 값에 의존한다. 따라서 더 적은 비용함수를 제공하는 추정 채널이라고 해서 꼭 원래 채널에 더 가깝다고 할 수 없다. 더구나 이런 경우에, 식 (4)의 오차신호를 줄임에 있어서 두 convolution 결과가 같아지도록 하는 것 외에도 각 convolution 결과의 절대치가 전체적으로 작아지도록 할 수도 있다. 특히, 음성은 이웃한 표본값 사이의 상관도가 매우 높기 때문에 필터 계수를 추정할 때, 후자의 기여도가 상당히 크고 상관도가 높은 신호에 대해서 convolution 결과의 절대치를 작게 하기 위해 추정된 채널이 자주 백색화 현상을 겪게 된다. 이 백색화 현상은 실제로 적응 고유값 분할 방법을 음성 신호에 적용한 <그림 1>의 결과에서 볼 수 있듯이 추정 채널에서 최대값과 그 외 큰 값들의 차이가 크지 않아서 직선 경로에 해당하는 시간지연을 정확히 파악하기가 어렵다.

## 3. 음향 채널의 '성감' 특성을 이용한 화자 위치 탐지 알고리즘

본 논문에서는 데이터의 양이 채널 필터를 추정하기에 충분하지 않을 때, 추가적인 정보를 사용하여 필터 계수를 더 정확하게 추정하는 방법을 고려해보고자 한다[9]. 전형적인 음향 채널은 <그림 1>과 같은 백색화 현상을 겪기보다 <그림 2>의 (a)와 같이 일부를 제외한 대부분의 채널 계수들이 0에 가까운 값으로 채워



<그림 1> 두 개의 마이크에 입력된 음성 신호에 고유값 분할 방법을 적용하여 추정된 채널 결과

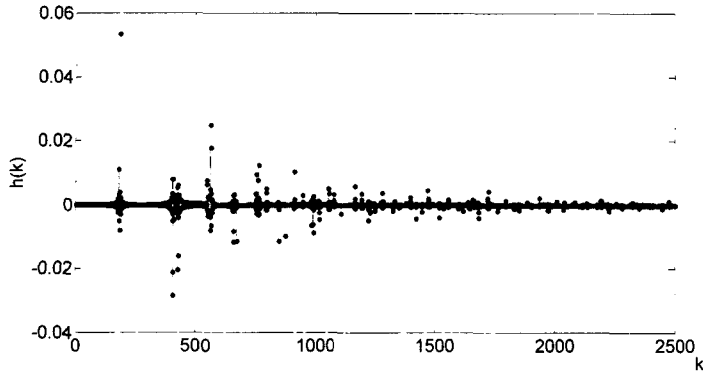
져 있어서 <그림 2>의 (b)에서 예상할 수 있다시피 큰 kurtosis 값을 갖는 ‘성김’ 확률 분포를 가지게 된다. 더욱이, 화자 위치 탐지는 채널 필터에서 직선 경로에 해당하는 시간지연값에만 관련되어 있다. 따라서 채널 필터를 추정하는 동안에 필터 계수들이 ‘성김’ 분포를 가진다는 특성을 반영한다면 화자 위치 탐지를 위한 시간차 추정이 더욱 효과적으로 이루어질 것으로 기대할 수 있다.

다음은 채널 필터의 분포와 관계된 식으로서 ‘성김’ 특성을 반영하는데 활용할 수 있다[11].

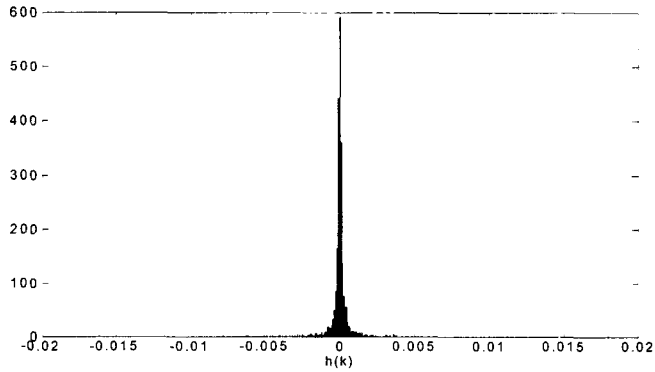
$$\log p(\mathbf{h}_i) = \sum_{k=0}^{L-1} G(h_i(k)) + \text{const}, \quad i = 1, 2, \dots, M. \quad (7)$$

식 (7)에서  $G(\cdot)$ 은 정규분포보다 더 성김 특성을 갖는 super-Gaussian 분포(정규 분포보다 더 큰 kurtosis 값을 갖는 분포)에 대한 확률 밀도 함수의 대수(logarithm) 값을 나타낸다. 여기에서 단순한 수식화를 위해서 채널 필터 계수들이 서로 독립이고 동일한 분포를 가진다는 가정을 적용하였다.

Super-Gaussian 분포로서 일반적으로 사용가능한 간단한 확률밀도함수로서 라플라스(Laplace) 분포를 고려해 본다면, 이 ‘성김’ 특성 적용을 최대도 하기 위한 채널 필터의 갱신 식은 다음과 같다.



(a)



(b)

<그림 2> 음향 채널의 전형적인 특성 (표본화 주파수: 44100 kHz); (a) 필터 모양, (b) (a)에서 나타낸 필터의 히스토그램

$$\Delta h_i(k) \propto \frac{\partial \log p(\mathbf{h}_i)}{\partial h_i(k)} = \frac{\partial \sum_{l=0}^{L-1} G(h_i(l))}{\partial h_i(k)} \propto \frac{\partial \sum_{l=0}^{L-1} -|h_i(l)|}{\partial h_i(k)} = -\text{sgn}\{h_i(k)\}, \quad (8)$$

$i = 1, 2, \dots, M.$

따라서 ‘성감’ 특성을 이용한 화자 위치 탐지 알고리즘은 필터 갱신을 위해서 식 (8)과 식 (5)를 결합한 다음 수식을 사용하여 적은 데이터로도 더 정확한 화자 위치 탐지가 가능하도록 한다.

$$\hat{\mathbf{h}}_i(n+1) = \hat{\mathbf{h}}_i(n) - \mu_1 \sum_{j=1}^M e_{ji}(n) \mathbf{x}_j(n) - \mu_2 \text{sgn} \left\{ \hat{\mathbf{h}}_i(n) - \mu_1 \sum_{j=1}^M e_{ji}(n) \mathbf{x}_j(n) \right\}, \quad (9)$$

$i = 1, 2, \dots, M.$

여기서,  $\mu_1$ 과  $\mu_2$ 는 각각 기존 고유값 분할 방법에 적용되는 학습률과 '성감' 특성을 이용한 필터 계수 학습률을 나타내며, 이들의 상대적인 값에 따라 제안한 알고리즘에서 각각의 비중이 결정된다. 이 알고리즘은 기존 고유값 분할 방법을 통해 미지의 채널 필터를 추정하는 과정에서, 이웃한 표본값 사이의 상관도가 높은 신호에 대하여 백색화를 위해 계수값들의 크기가 커질 때 '성감' 특성을 이용한 필터 계수 학습 부분으로 이를 조정하게 된다. 필터 갱신 후, 고유값 분할 방법에서처럼 필터 벡터,  $[\mathbf{h}_1^T \mathbf{h}_2^T \dots \mathbf{h}_N^T]$ 의 놈(norm)이 1이 되도록 정규화하며, 최종적으로 필요시에 식 (6)을 이용하여 음원의 위치를 결정하게 된다.

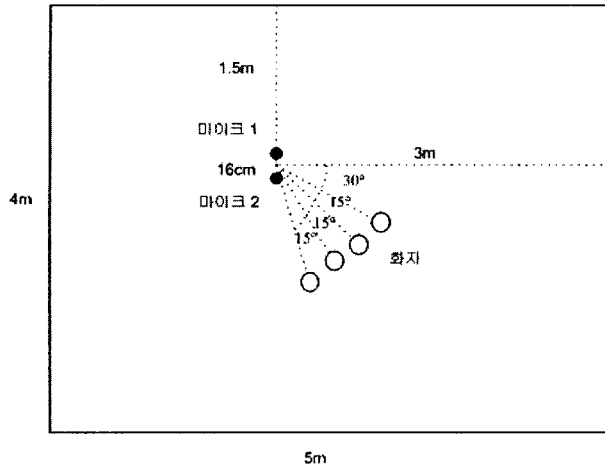
## 4. 모의실험 환경 및 결과

### 4.1. 실험 환경

제안한 화자 위치 탐지 알고리즘의 성능을 평가하기 위해 반향 환경에서 16 cm 거리의 두 마이크에 대한 모의실험을 수행하였다. 반향 신호는 화자 신호와 화자에서 마이크 사이의 채널 필터를 convolution하여 합성하였다[10]. 화자 신호는 TIMIT 데이터베이스[12]의 연속된 문장을 사용하였고, 채널 필터는 직육면체 공간을 가정하고 이미지 방법을 사용하여 화자와 마이크의 위치에 대해 생성한다[9]. <그림 3>은 채널 필터를 생성하기 위한 화자와 마이크의 위치를 나타낸다. 화자와 두 마이크의 중심사이의 거리는 1 m이다. 반향 시간  $RT_{60}$ 은 0.5초, 0.7초로 설정하여 실험하였다. 합성된 반향 신호는 약 25초 정도의 분량에 해당하며, 표본화 주파수는 44.1 kHz이다. 추정 채널로는 128 표본에 해당하는 길이를 사용하였고, 채널 계수는 시간영역에서 한 채널의 중간 한 계수만을 제외하고 나머지는 모두 0으로 초기화하였다.

### 4.2. 실험 결과

<그림 4>와 <그림 5>의 (a)는 각각 기존의 적용 고유값 분할 방법을 이용한 추정 채널과 라플라스 확률밀도를 사용하여 제안한 방법의 추정 채널을 보여준다. 화자의 위치는  $45^\circ$ 이고 반향 시간  $RT_{60}$ 은 0.7초이다. <그림 5>의 (a)의  $h_1(k)$ ,  $h_2(k)$ 에 나타내었듯이 제안한 알고리즘에 의해 추정된 채널에서는 하나의 필터 계수 값을 제외하고 나머지는 모두 거의 0의 값을 나타내어 최고값에 해당하는 직선 경로의 시간지연을 쉽게 파악할 수 있다. 또한, 식 (6)의  $\tau_{12}$ 로부터 계산할 수 있는 채널간의 최고값 차이(peak difference)에서 화자의 방향에 상응하는 시간지연



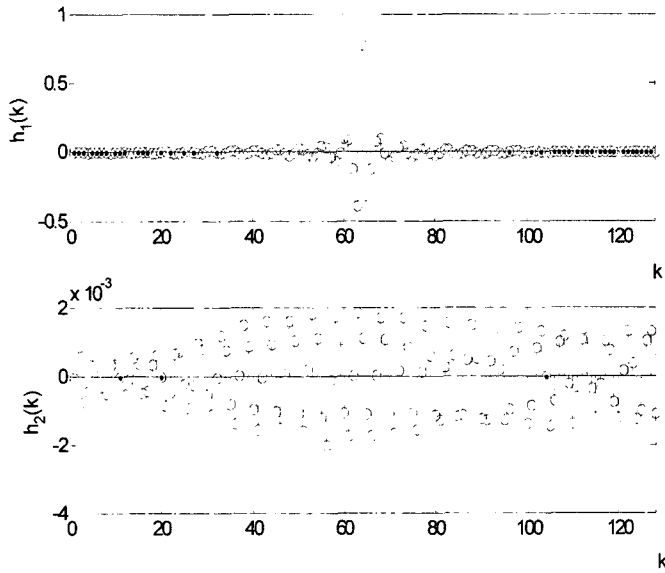
<그림 3> 채널 필터를 생성하기 위한 화자와 마이크의 배치 (직육면체 공간의 높이를 3 m, 화자와 마이크의 높이를 모두 1.1 m로 설정하였음)

차이(14 표본 주기)가 발견됨을 쉽게 알 수 있다. 그러나 <그림 4>의 (a)의  $h_1(k)$ ,  $h_2(k)$ 에서 알 수 있다시피 기존의 적응 고유값 분할 방법에 의해 추정된 채널은 백색화 현상을 겪기 때문에 많은 필터 계수 값이 0에서 많이 벗어나게 된다. 특히 필터 계수 값이 모두 0으로 초기화된  $h_2(k)$ 에서 최고값과 다른 큰 값들 사이에 차이가 거의 존재하지 않아 직선 경로의 시간지연을 파악하기 어렵다. 따라서 채널간의 최고값 차이를 이용하는 화자의 방향도 부정확해진다.

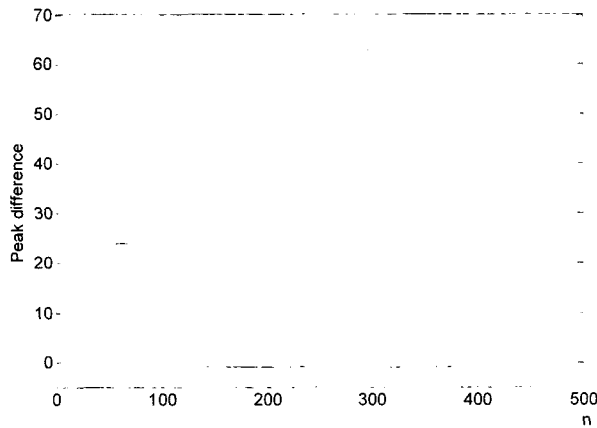
<그림 4>의 (b)는 기존 방법의 입력 신호에 대한 학습 회수에 따른 추정 채널간의 최고값 차이를 보여주는데, 각 추정 채널 최고값의 시간지연이 학습이 진행되는 동안 계속 변하기 때문에 이 최고값 차이도 변하며, 그 값도 화자의 방향에 상응하는 시간지연과 다른 값을 알 수 있다. 이에 반해 제안한 방법의 추정 채널간의 최고값 차이를 나타내는 <그림 5>의 (b)에서는 학습 초기부터 화자의 방향에 해당하는 시간지연을 지속적으로 유지하고 있음을 알 수 있다.

화자의 위치와 신호, 반향 환경 등을 다양하게 하며 위의 실험을 반복하여 입력 신호에 대한 학습 회수에 따른 방향 탐지 경우의 수를 <표 1>에 요약하였다. 실험은 다른 4명(남자 2명과 여자 2명)의 화자가 발음한 화자 신호를 <그림 3>에서 보는 바와 같이 화자의 위치가 30°, 45°, 60°, 75°일 때 실험적으로 생성한 채널 필터와 convolution한 결과를 이용하였다. 그 외 실험 환경 및 조건은 위의 실험에서 처럼 4.1 절을 따른다. 16가지 다른 마이크 입력 신호에 대해서 500번까지 매 100번의 학습 회수마다 실제 화자의 방향에 상응하는 시간지연과 이를 추정하기 위해 식 (6)의  $\tau_{12}$ 로부터 계산한 채널간의 최고값 차이를 비교하였다. <표 1>의 (a)는 추정값과 실제 화자의 방향에 상응하는 시간지연 차이 사이에 오차를 허용하





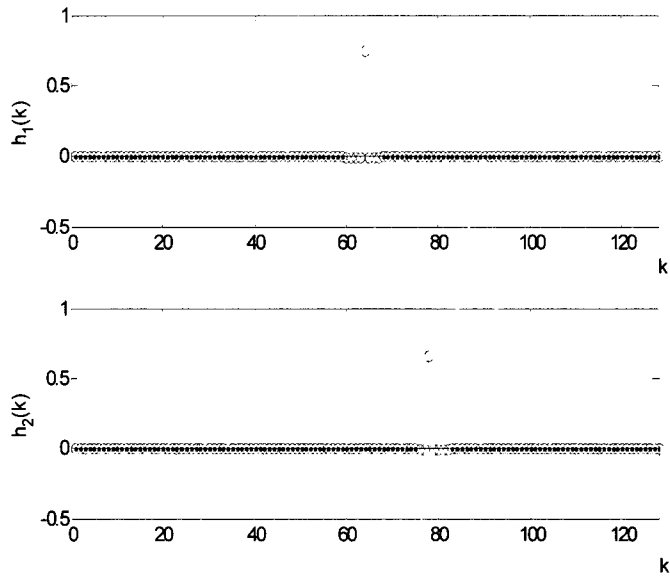
(a)



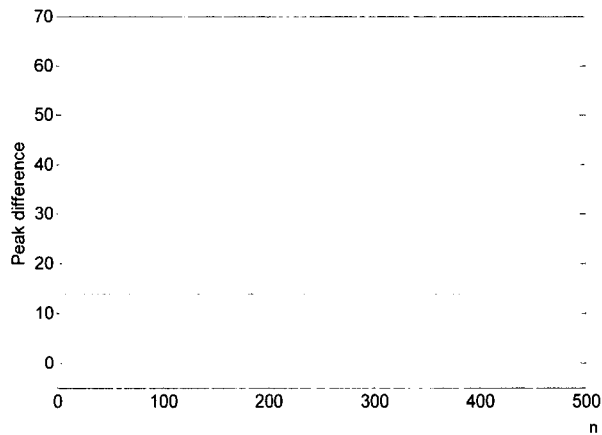
(b)

<그림 4> 기존 적용 고유값 분할방법의 채널 추정 결과; (a) 최종 추정 채널, (b) 입력 신호에 대한 학습 회수에 따른 추정 채널간의 최고값 차이의 변화

지 않았을 때, <표 1>의 (b)는  $\pm 1$  표본 주기만큼 차이나는 오차를 허용했을 때 탐지 경우의 수를 나타낸다. 반향 시간이 길어질수록 음향 채널 필터가 더욱 복잡해져서 추정 채널간의 최고값 차이를 정확하게 추정하기 어렵기 때문에 두 방법 모두 방향 탐지율이 감소하는 것을 확인할 수 있다. 그러나 기존 고유값 분할 방법의 경우에 학습이 진행되면서 추정 채널에서 직선 경로의 시간지연에 해당하는 최고값이 계속 변하기 때문에 추정 채널간의 최고값 차이도 변하여 화자 방향 탐



(a)



(b)

<그림 5> 음향 채널의 '성감' 특성을 이용한 제안한 방법의 채널 추정 결과; (a) 최종 추정 채널, (b) 입력 신호에 대한 학습 회수에 따른 추정 채널간의 최고값 차이의 변화

지울도 달라진다. 반면에 제안한 알고리즘은 실험한 모든 데이터에 대해 학습 회수에 영향을 받지 않고 추정한 채널에서 최고값을 쉽게 파악할 수 있었고, 추정 채널간 최고값 차이를 일정하게 유지하였다. 무엇보다도 반향 환경에서 화자 방향 탐지를 수행함에 있어서 추정 음향 채널의 '성감' 특성을 이용한 화자 방향 탐지 방법이 기존 적응 고유값 분할방법보다 일관되게 향상된 결과를 나타냄을 알 수

있다.

<표 1> 입력 신호에 대한 학습 회수에 따른 화자 방향 탐지 경우의 수; (a) 오차를 허용하지 않았을 때 탐지 경우의 수, (b)  $\pm 1$  표본 주기 오차를 허용하였을 때 탐지 경우의 수 (각 경우마다 16종류의 마이크 입력 신호에 대해 실험하였고, 화자 방향 탐지 백분율도 괄호 안에 표시하였다.)

(a)

학습 회수	$RT_{60} = 0.5$ 초		$RT_{60} = 0.7$ 초	
	고유값 분할	제안한 방법	고유값 분할	제안한 방법
100	10 (63%)	13 (81%)	8 (50%)	12 (75%)
200	10 (63%)	13 (81%)	4 (25%)	12 (75%)
300	9 (56%)	13 (81%)	5 (31%)	12 (75%)
400	8 (50%)	13 (81%)	5 (31%)	12 (75%)
500	6 (38%)	13 (81%)	2 (13%)	12 (75%)

(b)

학습 회수	$RT_{60} = 0.5$ 초		$RT_{60} = 0.7$ 초	
	고유값 분할	제안한 방법	고유값 분할	제안한 방법
100	10 (63%)	15 (94%)	8 (50%)	13 (81%)
200	10 (63%)	15 (94%)	4 (25%)	13 (81%)
300	9 (56%)	15 (94%)	5 (31%)	13 (81%)
400	8 (50%)	15 (94%)	5 (31%)	13 (81%)
500	6 (38%)	15 (94%)	2 (13%)	13 (81%)

## 5. 결론

본 논문은 반향 환경에서 화자 위치 탐지를 위한 시간지연 추정 알고리즘을 제안하였다. 이웃한 표본간에 높은 상관도를 가지고 있는 대부분의 자연 소리를 사용하여 화자에서 마이크까지의 음향 채널을 추정하게 되면 보통 백색화 현상이 일어나서 추정한 채널의 최대값과 그 외 큰 값의 차이가 크지 않아 학습이 진행됨에 따라서 직선 경로의 시간지연에 상응하는 최대값이 지속적으로 변한다. 이때, '성김' 특성을 음향 채널 추정 시에 반영하여 직선 경로의 시간지연에 상응하는 최대값을 손쉽게 파악할 수 있고, 추정 채널간 일정한 시간차를 얻을 수 있어서 화자의 위치 탐지 성능을 향상시킬 수 있다. 모의 실험결과는 제안한 방법이 반향 환경에서 화자 위치 탐지에 효과적임을 보여주었다.

## 참 고 문 헌

- [1] 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.
- [2] C. H. Knapp, G. C. Carter, "The generalized correlation method for estimation of time delay", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 24, No. 4, pp. 320-327, 1976.
- [3] S. Doclo, M. Moonen, "Robust adaptive time delay estimation for speaker localization in noisy and reverberant acoustic environments", *EURASIP Journal on Applied Signal Processing*, Vol. 2003, pp. 1110-1124, 2003.
- [4] Y.-I. Kim, R. M. Kil, "Estimation of interaural time differences based on zeros-crossings in noisy multisource environments", *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 15, No. 2, pp. 734-743, 2007.
- [5] J. Chen, J. Benesty, Y. Huang, "Time delay estimation in room acoustic environments: an overview", *EURASIP Journal on Applied Signal Processing*, Vol. 2006, pp. 1-19, 2006.
- [6] T. G. Dvorkind, S. Gannot, "Time difference of arrival estimation of speech source in a noisy and reverberant environment", *Signal Processing*, Vol. 85, No. 1, pp. 177-204, 2005.
- [7] Y. Huang, J. Benesty, "Adaptive multi-channel least mean square and Newton algorithms for blind channel identification", *Signal Processing*, Vol. 82, No. 8, pp. 1127-1138, 2002.
- [8] Y. Huang, J. Benesty, G. W. Elko, "Adaptive eigenvalue decomposition algorithm for real time acoustic source localization system", *Proc. ICASSP*, pp. 937-940, 1999.
- [9] 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.
- [10] A. Hyvarinen, J. Harhunen, E. Oja, *Independent Component Analysis*. John Wiley & Sons, 2001.
- [11] A. Hyvarinen, K. Raju, "Imposing sparsity on the mixing matrix in independent component analysis", *Neurocomputing*, Vol. 49, No. 1, pp. 151-162, 2002.
- [12] 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 22일

▶조지원(Ji-Won Cho)

주소: 121-742 서울특별시 마포구 신수동 1번지 서강대학교

소속: 서강대학교 전자공학과

전화: 02) 711-8916

E-mail: jiwonn85@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