

음향 채널 추정을 이용한 음질 향상

Speech Enhancement Using Acoustic Channel Estimation

최 영 근*, 김 기 만*, 박 규 식**
(Young-Keun Choi*, Ki-Man Kim*, Kyu-Sik Park**)

*한국해양대학교 전파공학과, **단국대학교 정보·컴퓨터학부 컴퓨터과학 전공
(접수일자: 2003년 3월 13일; 채택일자: 2003년 8월 13일)

최근 원격 회의 시스템에서 마이크로폰 어레이를 이용한 음원 위치 추정 및 음질 향상 기술이 연구되고 있다. 이러한 기술 가운데 대표적인 정합 필터 마이크로폰 어레이는 다수의 마이크로폰을 이용하여 음원의 위치를 추적하고 음원의 방향으로부터 입사되는 신호만을 수신할 수 있도록 한다. 그러나 이는 음향 공간의 기하학적인 형태 및 재질 등을 이용한 모델로부터 미리 계산된 음향 채널을 사용하기 때문에 실제 실내 환경과의 불일치로 인하여 성능이 저하된다. 본 논문에서는 수신된 음향 신호로부터 실내 음향 채널을 추정하고 이를 정합 필터 마이크로폰 어레이에 적용하였다. 연구된 방법은 실험을 통해 그 성능을 고찰하였다.

핵심용어: 음질 향상, 음향 채널 추정, 마이크로폰 어레이

투고분야: 음향 통신기술 분야 (6.4)

Recently, speaker localizing estimation technique has been rising in teleconference systems. In this paper, it was described to be able to enhance the speech quality through microphone array, and received the only signal of speaker. Unfortunately, as it using estimated the signal in advance, it is not matched in a real acoustic environment so it has poor performance. In this paper is proposed for Adaptive Matched Filter Microphone Array that estimated acoustic room environment from the received the signal and study of the efficiency through simulations.

Keywords: Speech enhancement, Acoustic channel estimation, Microphone array

ASK subject classification: Acoustic communication (6.4)

I. 서론

잡음이 섞인 음성을 이용하여 부호화기에 적용하게 되면 음성 모델에 기반을 둔 음성 부호화기의 성능을 떨어 뜨리고 음성 인식 시스템의 효율이 저하된다. 따라서 음성 처리에 있어서 잡음제거 또는 음질 향상은 음성 관련 시스템의 전처리 (pre-processing) 단으로써 반드시 필요하다. 최근 원격회의 시스템이나 지능형 차량 항법 시스템 등에 적용되는 멀티미디어 음향통신 시스템에서 고품질 음향 신호를 취득하기 위해 음원 위치 추정과 음질 향상 기술이 연구되고 있다. 이 기술에서는 다수의 마이크로폰 어레이를 이용, 음원의 위치를 파악하고 선택된 방향의 음질을 향상시킨다. 음질 향상을 위한 기존의 방법으로써 빔 형성 기법은 다수의 마이크로폰 입력 신호

들을 조합하여 빔 형성기 출력 단에서 최대 신호 대 잡음비를 갖도록 하는 것이다[1]. 그러나 이 방법은 마이크로폰의 수가 적은 경우 빔 폭이 커져서 원하지 않은 외부 소음도 수신할 수 있다. 최근 유색잡음이 첨가된 음성신호의 음질 향상을 위한 부공간 접근 방법에 관하여 활발하게 연구되어지고 있다[2]. 이 방법은 깨끗한 음성신호와 잡음 신호의 공분산 행렬들의 대각화 (diagonalization)에 기반하는 음질 향상 기법이다. 하지만 이 방법 또한 실제 적용시 계산량이 증가하는 문제점이 있다. 음향 채널 임펄스 응답을 이용하는 정합 필터 방법은 앞서 설명된 방법보다 향상된 성능을 나타내지만 미리 관심 영역의 모든 위치에서 임펄스 응답을 측정하거나 모델로부터 계산해야 하는 등의 문제점이 제기되고 있다[3]. 이외에도 스펙트럼 차감법 (Spectral subtraction)이나 위너 (Wiener) 필터 알고리즘 등이 연구되었다[4,5].

기존의 정합 필터 마이크로폰 어레이는 실내 공간으로부터 직접 측정 내지는 모델로부터 계산된 임펄스 응답

책임저자: 김기만 (kimkim@hanara.kmaritime.ac.kr)
506-791 부산광역시 영도구 동삼동 1번지
한국해양대학교 전파공학과
(전화: 051-410-4918; 팩스: 051-404-3986)

의 시간적 역변환 신호와 실제 마이크로폰에 수신된 신호 사이에 컨벌루션을 구하여 원 신호를 추정한다. 이는 이상적인 경우, 즉 실내 음향 모델이 정확한 경우에만 최적의 성능을 얻을 수 있다. 그러나 실제의 경우 실내 음향 환경은 시시각각 변화하므로 정합 필터 어레이의 성능은 저하되고, 매번 임펄스 응답을 따로 측정할 수도 없다. 이에 본 연구에서는 시시각각으로 변화하는 실내 음향 환경에 실시간으로 적응하기 위하여 마이크로폰으로부터 수신된 신호로부터 직접 실내 음향 임펄스 응답을 추정하고, 이를 이용하여 정합 필터 어레이를 구성하였다. 이때 임펄스 응답은 수신된 신호로부터 구성된 자기 상관 행렬의 잡음 부공간을 추정하여 계산된다. 또한 이 과정에는 많은 계산량이 요구되며 이를 극복하기 위하여 널리 알려진 적응 알고리즘을 이용하여 효율적으로 계산할 수 있다. 실험 결과에서 본 연구에서 제안된 방법의 효율성을 입증하였다.

II. 정합 필터 어레이

정합 필터 어레이는 일반 실내에서 잔향의 영향을 감소시키기 위한 방법으로 연구되었다. 실내 음향 환경을 살펴보면 먼저 그림 1과 같이 음원과 i 번째 마이크로폰 사이의 임펄스 응답을 $h_i(t)$ 라 할 때 원래의 음성 신호 $s(t)$ 는 $h_i(t)$ 와 컨벌루션되어 마이크로폰에 수신된다. 마이크로폰에 입력된 신호는 $h_i(t)$ 의 시간적 역필터 $h_i(-t)$ 를 통과하여 원래 음성 신호로 재현된다. 만약 마이크로폰의 수가 N 개라면 정합 필터 어레이의 출력 $y(t)$ 는 다음과 같다. 식 (1)에서 “*”는 컨벌루션을 의미한다.

$$y(t) = s(t) * \sum_{i=1}^N h_i(t) * h_i(-t) \quad (1)$$

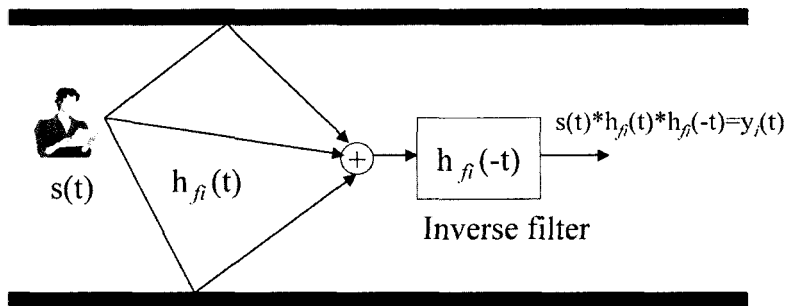


그림 1. 정합 필터 어레이의 개념도
Fig. 1. Configuration of matched filter array.

데이터의 샘플 개수가 많을 경우 이상적으로 임펄스 응답과 그 역필터 사이의 컨벌루션은 임펄스 형태가 된다. 결국 수신 신호와 역필터 사이의 컨벌루션, 정합 필터 어레이 출력은 음성 신호와 같아진다. 이러한 정합 필터 어레이는 빔 형성 기법에 비해 잔향에 강한 장점을 갖고 있으나 정확한 실제 실내 환경의 임펄스 응답을 알아야만 한다. 이를 위해서는 마이크로폰과 음원간의 임펄스 응답을 정확하게 알아야만 한다. 이상적으로 임펄스 응답의 길이는 무한하나 계산량의 증대가 문제가 되어 임펄스 응답의 길이를 조정해 주어야만 한다. 그 길이가 매우 길어지게 되면 역필터의 길이 또한 길어지게 되어 계산량이 크게 증가되는 단점을 갖고 있다. 이것을 보완하기 위하여 각 채널의 임펄스 응답의 길이를 조정해 주어야만 하는데 길이 조정에 따른 실제 임펄스 응답과의 오차로 인하여 시스템의 정확도에 영향을 미친다. 아울러 특정 위치에서의 고정 통신이 아닌 음원의 이동으로 인한 실내 통신 환경 변화에 능동적으로 대처하기 위하여 실내 음향 환경의 다양한 위치에서의 특성을 사전에 측정하여 데이터 베이스화하고 있어야 하는 단점이 있다. 그리고 정합 필터 어레이는 식 (1)에서 볼 수 있듯이 정합 필터링 과정에서 사용된 임펄스 응답의 길이만큼 시간 지연이 발생한다.

III. 음향 채널 추정을 이용한 음질 향상

Benesty는 TDOA (Time-Difference-of-Arrival) 추정을 위한 채널 임펄스 응답 추정 기법을 제안하였다[7]. 이 방법은 수신 센서 쌍 (pair)을 이용하여 임펄스 응답을 추정하는 것으로서 본 연구에서는 기존의 정합 필터 어레이의 문제점 가운데 하나인 시변 채널 문제를 해결하기 위해 [7]에서 연구된 채널 추정 알고리즘을 사용하고 나아가 적응 알고리즘을 적용하였다. 우선 i 번째 마이크로

른에 수신된 신호는 다음과 같이 표현할 수 있다.

$$r_i(t) = h_i(t) * s(t) + n_i(t) \quad (2)$$

$h_i(t)$ 는 i 번째 마이크로폰과 음원 신호간의 임펄스 응답이다. $n_i(t)$ 는 첨가 잡음 성분으로 여기서는 정규 분포를 갖는다고 가정한다. 수신된 신호로부터 채널 임펄스 응답을 계산하기 위하여 다음과 같은 과정을 거친다. 즉 각 채널에서 잡음 성분의 영향은 매우 작다고 가정하면, $x_i(t) = s(t) * h_i(t), (i=1, 2)$ 는 선형적인 특성에 의해 아래와 같아진다.

$$r_1(t) * h_2(t) = s(t) * h_1(t) * h_2(t) = x_2(t) * h_1(t) \quad (3)$$

윗 식 (3)을 벡터 연산으로 표현하면 다음과 같다.

$$X_1^T(t) H_2(t) = X_2^T(t) H_1(t) \quad (4)$$

여기서, $X_i(t)$ 는 아래와 같은 i 번째 채널의 수신 데이터 벡터이며, $H_i(t)$ 는 i 번째 채널의 임펄스 응답 벡터이다.

$$X_i(t) = [x_i(t) \ x_i(t-1) \ \dots \ x_i(t-M+1)]^T, i=1, 2 \quad (5)$$

$$H_i(t) = [h_{i,0} \ h_{i,1} \ \dots \ h_{i,M-1}]^T, i=1, 2 \quad (6)$$

이제 두 마이크로폰 출력 사이의 공분산 행렬은 다음과 같다.

$$R = \begin{bmatrix} R_{x_1 x_1} & R_{x_1 x_2} \\ R_{x_2 x_1} & R_{x_2 x_2} \end{bmatrix} \quad (7)$$

$$R_{x_i x_j} = E[x_i(t) \ x_j^T(t)], i, j=1, 2 \quad (8)$$

또한 임펄스 응답 $H(t)$ 로 구성된 벡터 u 를 다음과 같이 정의하자.

$$u = \begin{bmatrix} H_2(t) \\ -H_1(t) \end{bmatrix} \quad (9)$$

식 (4), (7), (9)로부터 $Ru=0$ 임을 알 수 있으며, 만약 잡음 성분을 갖는 채널이라면 벡터 u 의 norm을 1로 제한하면서 $u^T R u$ 를 최소화하는 해를 구하게 된다. 이 때 벡터 u 는 공분산 행렬 R 의 최소 고유치에 대응하는 고유 벡터가 된다. 이제 추정된 임펄스 응답 $\hat{h}_i(t)$ 를 이용하여 i 정합 필터 어레이를 다음과 같이 구성할 수 있다.

$$y(t) = \sum_{i=1}^2 \hat{h}_i(-t) * x_i(t) \quad (10)$$

시변 채널 임펄스 응답을 추정하기 위하여 2개의 채널 출력 데이터를 사용하였으나 정합 필터 어레이의 성능을 향상시키기 위해서는 채널 수가 많을수록 좋다. 따라서 마이크로폰을 쌍(pair)으로 배치하면 각 쌍으로부터 임펄스 응답을 추정하고 정합 필터를 구성하게 된다.

실제로 행렬 R 의 최소 고유치에 대응하는 고유 벡터 u 를 구하는 과정은 $O(2M)^3$ 에 비례하는 많은 계산량을 갖는다. 특히 실내 잔향이 긴 경우 짧은 임펄스 응답을 사용하면 그 성능이 크게 저하되는 것으로 실험적으로 증명되어 있다[8]. 따라서 실시간 처리 시스템으로 적용하기 위해서는 상대적으로 계산량이 적으면서도 효율적인 제한조건을 갖는 (LMS: Least Mean Square) 알고리즘을 사용하여 계산할 수 있다. 최적 값을 u_{opt} 라고 하고, 오차 신호 $e(t)$ 를 다음과 같이 정의한다.

$$e(t) = \frac{u(t)^T X(t)}{\|u(t)\|} \quad (11)$$

여기서 $x(t) = [x_1^T(t) \ x_2^T(t)]^T$ 이다. $e(t)$ 의 평균 제곱 값을 최소화하기 위해 $e(t)$ 의 미분은 다음과 같이 계산되어진다.

$$\nabla e(t) = \frac{1}{\|u(t)\|} \left[x(t) - e(t) \frac{u(t)}{\|u(t)\|} \right] \quad (12)$$

이제 제한조건을 갖는 LMS 알고리즘을 다음과 같이 얻을 수 있다.

$$u(t+1) = u(t) - \mu e(t) \nabla e(t) \quad (13)$$

여기서 μ 는 수렴 속도를 결정하는 임의의 상수 값이다. 식 (11)과 (12)를 (13)에 대입하면 다음과 같다.

$$u(t+1) = u(t) - \frac{\mu}{\|u(t)\|} \left[x(t) \ x(t)^T \frac{u(t)}{\|u(t)\|} - e(t)^2 \frac{u(t)}{\|u(t)\|} \right] \quad (14)$$

식 (14)를 이용하여 시변 채널 임펄스 응답을 계산할 수 있으며, 이로부터 적응 정합 필터를 구성한다. 그림 2에 제안된 방법의 흐름도를 나타내었다.

IV. 실험 및 결과 고찰

제안된 방법의 성능을 고찰하기 위하여 실험을 수행하였다. 실험에 사용된 마이크로폰은 오디오 테크니카사의 AT9500 모델로 핀 형태로써 무지향성을 가진다. 수신단은 마이크로폰 6개를 사용하여 어레이를 구성하였다. 마이크로폰 사이의 간격은 4 cm로 하여 그림 3과 같이 마이크로폰을 배열하였으며 음원은 마이크로폰 배열면에 수직되는 방향으로 3 m의 거리를 두어 배치하였다. 실험에는 무향실에서 취득한 남자의 음성신호를 사용하였다. 음원은 44.1 kHz의 샘플링 주파수를 가지며 스피커를 통하여 재생하였고 실내 공간을 통하여 마이크로폰으로부터 수신된 음성 신호로부터 음향 채널을 추정하고 정합 필터 어레이를 수행하였다. 음원과 마이크로폰 어레이는

지면으로부터 80 cm 떨어진 곳에 위치하였으며 실험이 수행된 실내 공간은 넓이 6.3 m, 폭 16.2 m, 높이 6 m이다.

실내 음향 환경 특성을 파악하고자 실내 임펄스 응답을 측정하였다. 음원으로는 의사랜덤 잡음 신호를 사용하였다. 그림 4(a)는 음원으로써 랜덤 잡음 신호를 사용하여 얻은 임펄스 응답으로 마이크로폰 1번에 해당하며, (b)는 앞의 경우와 같은 방법으로 마이크로폰 2번에 해당하는 신호이다. 두 마이크로폰간의 간격으로 인한 음원과의 거리 차로 인하여 초기 펄스 침투치 발생시점의 차이가 나타나며 침투치의 크기 또한 거리차에 의한 감쇄로 인하여 차이가 남을 알 수 있다.

잡음이 섞인 음성을 이용하면 음성 모델에 기반을 둔 음성 부호화기의 성능을 떨어뜨리고 음성 인식 시스템의 효율이 저하된다. 따라서 음성 처리에 있어서 잡음제거

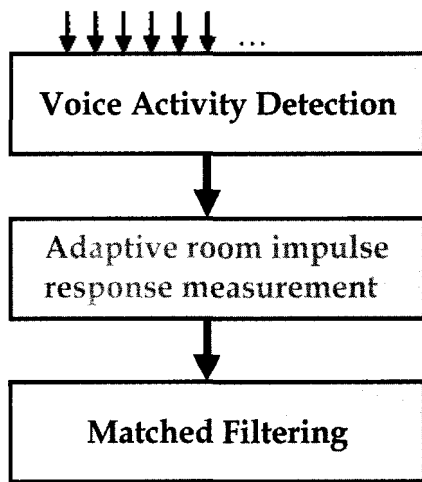


그림 2. 제안된 방법의 흐름도
Fig. 2. Block diagram of the proposed approach.

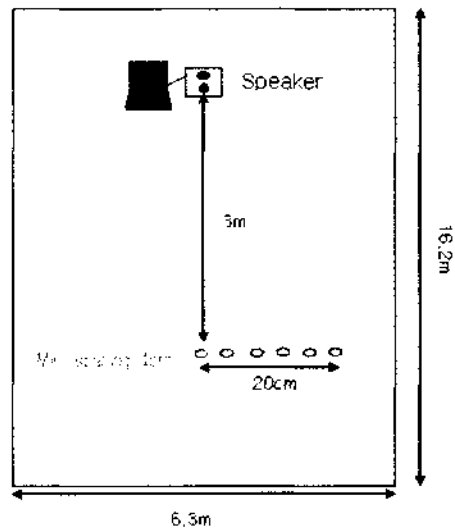
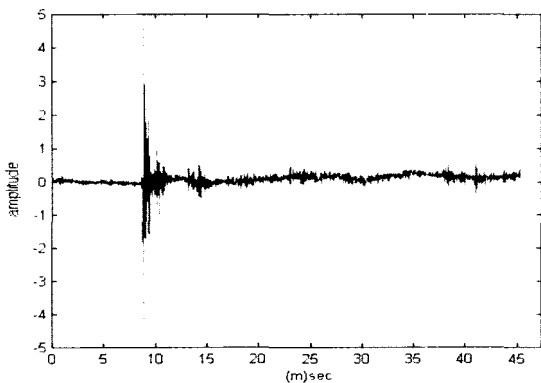
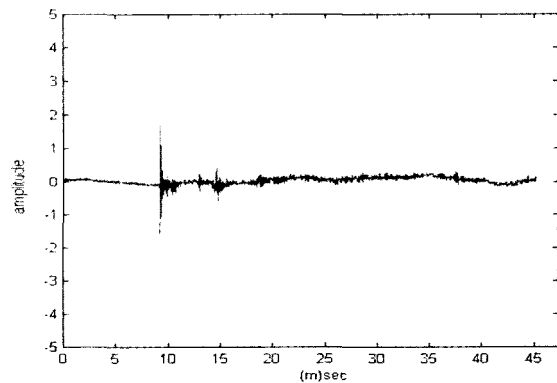


그림 3. 실험 배치도
Fig. 3. Setup for experiments.



(a) 채널 1
(a) Channel 1



(b) 채널 2
(b) Channel 2

그림 4. 측정된 채널 임펄스 응답
Fig. 4. Measured channel impulse response.

또는 음질 향상은 음성 관련 시스템의 전처리단으로써 반드시 필요하다. 그림 5에는 제안된 방법의 출력을 원래 음성 신호와 비교하였다. 그림 5(b)의 마이크론 출력

신호에서 초기 시간 지연이 발생하는데 이는 음원과 마이크론간의 거리차로 인한 도달 시간차에 기인한 것이다. 또한 적응 정합 필터를 처리한 결과에서 나타난 시간 지

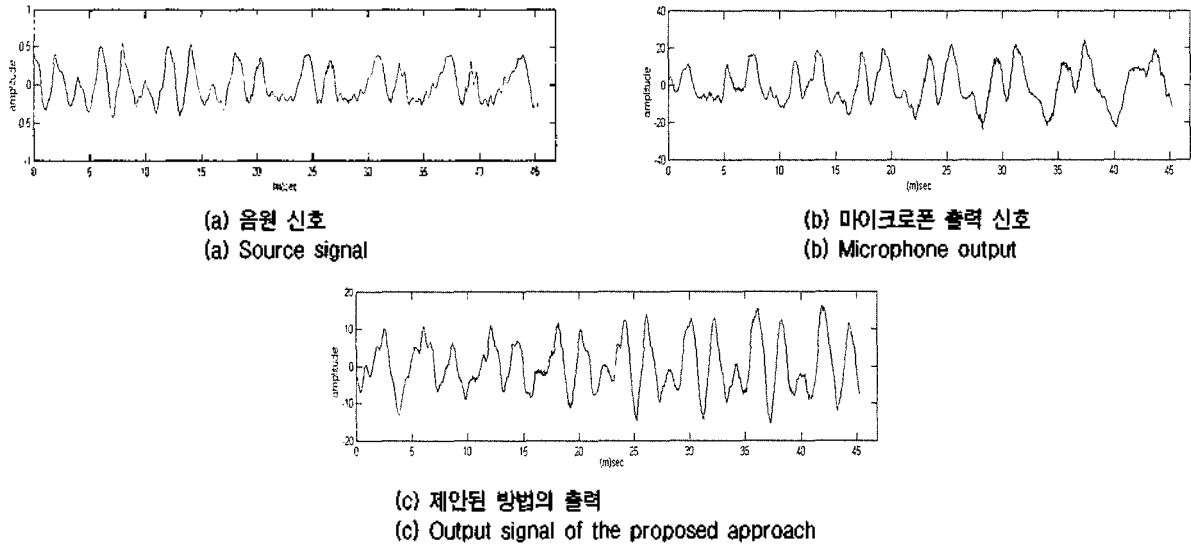


그림 5. 처리 결과
Fig. 5. Processing results.

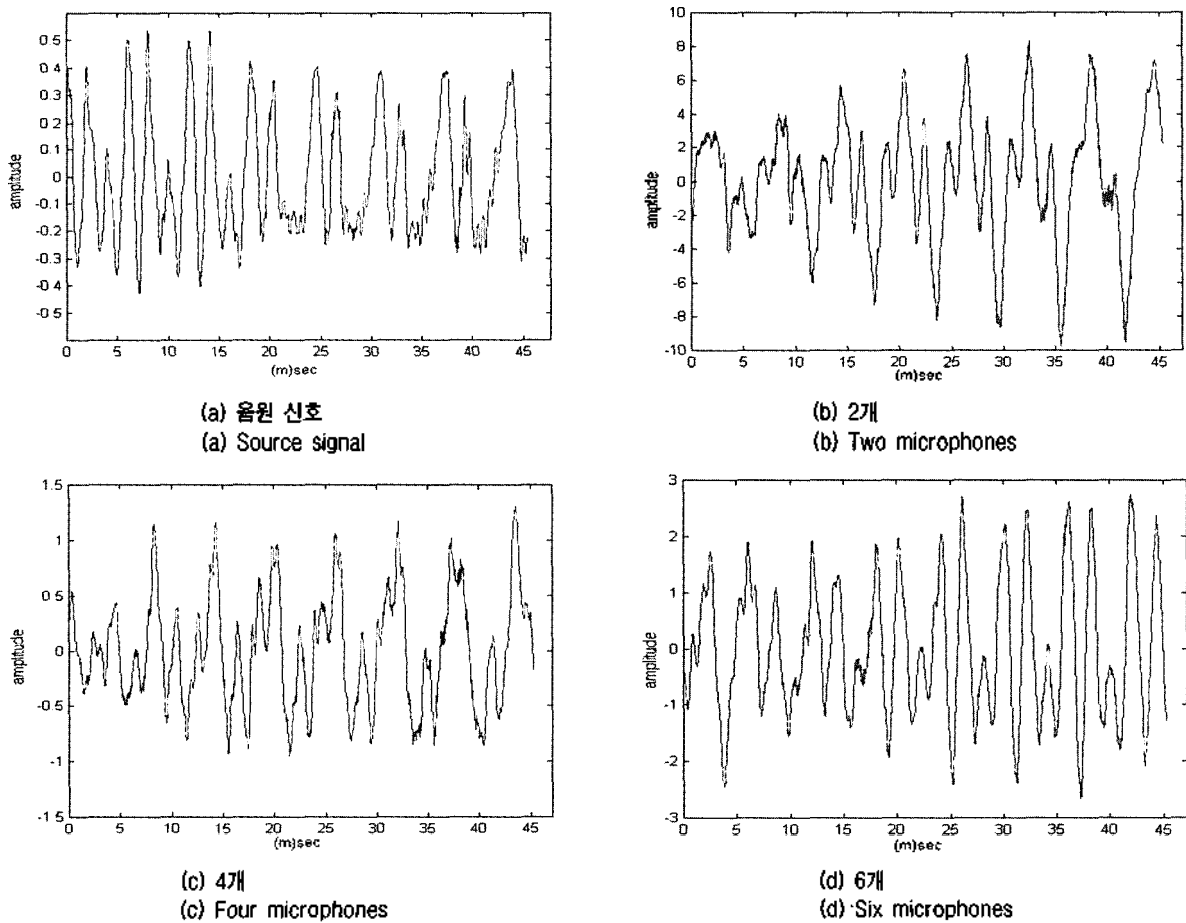


그림 6. 마이크론 개수에 따른 출력
Fig. 6. Output according to the number of the microphones.

연 현상은 정합 필터링 과정에서 나타나는 현상이다. 그림에서는 대략 24 msec 정도의 시간 지연이 나타났다. 또한 원 신호와 크기 차이가 나는데 이는 고유벡터를 구하는 과정에서 정규화하였기 때문에 발생한 현상이다.

그림 6은 마이크로폰 사용 개수에 따른 성능을 나타낸 그림이다. 마이크로폰의 개수가 늘어남에 따라 어레이에서 출력되는 신호가 원래 음성 신호에 가까워지는 것을 알 수 있다.

V. 결론

본 논문에서는 수신된 신호로부터 채널 임펄스 응답을 계산하고 정합 필터 어레이를 구성하였다. 이는 음향 신호와 두 개의 마이크로폰 사이의 임펄스 응답을 수신된 마이크의 신호의 공분산 행렬의 최소 고유치에 대응하는 고유벡터를 구함으로써 계산된다. 반향성분이 있는 실내 환경, 특히 시변 환경에서 효율적으로 적용될 수 있다.

추후 연구된 방법의 다양한 성능을 고찰해야 한다. 즉 실내 음향 특성, 이동 음원, 신호 대 잡음비 등과 같은 변수들에 따른 분석이 요구된다. 임펄스 응답의 길이 조정에 따른 오차 또한 연구되어야 하며 아울러 정합 필터 어레이의 또 다른 문제점인 역필터를 연산할 때 계산량을 줄이기 위한 방법들에 대해 연구되어야 한다. 또한 유색잡음이 포함된 음성신호의 부공간 해석에 대한 연구도 진행되어야 한다.

참고 문헌

1. P. Wauters, K. Eneman, K. Delaet and R. Lauwereins, "Adaptive speech beamforming using the TMS320C40 multi-DSP," *Texas Instrument Technical Note SPRA305*, Dec. 1995.
2. Y. Hu and P. C. Loizou, "A subspace approach enhancing speech corrupted by colored noise," *IEEE Signal Processing letters*, 9, 204-206, 2002.

3. Y. Kaneda and J. Ohga, "Adaptive microphone array system for noise reduction," *IEEE Trans. Acoust. Speech Signal Proc.*, 34 (6), 1391-1400, Dec. 1986.
4. Y. Grenier, "A microphone array for car environments," *Speech Communication*, 12 (1), 25-39, 1993.
5. R. J. Renomeron, D. V. Rabinikin, J. C. French and J. L. Flanagan, "Small-scale matched filter array processing for spatially selective sound capture," *134th Meeting of the Acoustical Society of America*, Dec. 1997.
6. E. E. Jan and J. Flanagan, "Sound capture from spatial volumes: matched filter processing of microphone arrays having randomly-distributed sensors," *Proc. ICASSP*, 917-920, 1996.
7. J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization," *J. Acoust. Soc. Am.*, 107 (1), January 2000.
8. D. V. Rabinikin, *Optimum Sensor Placement for Microphone Arrays*, Ph. D dissertation, Rutgers University, 1998.
9. J. Benesty and S. L. Gay, *Acoustic Signal Processing for Telecommunication*, Kluwer Academic Publishers, 2000.
10. M. Brandstein and D. Ward, *Microphone Arrays - Signal Processing Techniques and Applications*, Springer, 2001.
11. 장병건, 권태능, 변윤식, "마이크로폰 어레이를 위한 최적 패턴 형성," *한국음향학회지*, 16 (1), 47-53, 1997.
12. 이희원, "임펄스 응답을 이용한 실내음향 측정 시스템," *한국음향학회지*, 18 (5), 63-67, 1999.
13. 최희열, 최승진, "A time-delay correlation matching to source separation," 제19회 한국음향학회 음성통신 및 신호처리 학술대회 논문집, 221-224, 2002.

저자 약력

● 최 영 근 (Young-Keun Choi)

2003년 2월: 한국해양대학교 전파공학과 (공학사)
 2003년 3월 ~ 현재: 한국해양대학교 대학원 전파공학과 석사과정
 ※ 주관심분야: 음성 신호처리, 어레이 신호처리, DSP 하드웨어 구현



● 김 기 만 (Ki-Man Kim)

한국음향학회지 제22권 제5호 참조

● 박 규 식 (Kyu-Sik Park)

한국음향학회지 제20권 제1호 참조