적응 정합 필터 마이크로폰 어레이를 이용한 음질 향상

이외형, 최영근, 김기만, *박규식
한국해양대학교 전파공학과
* 단국대학교 정보컴퓨터학부 컴퓨터과학 전공

Speech Enhancement using Adaptive Matched Filter Microphone Array

Oe-Hyung Lee, Young-Keun Choi, Ki-Man Kim and *Kyu-Sik Park
Dept. of Radio Science & Eng., Korea Maritime University
* Dept. of Computer Science and Statistics, Dankook University
E-mail : kimkim@hanara.kmaritime.ac.kr

요 약

최근 영상 회의 시스템에서 화자 위치 추정 및 음질 향상 기술이 연구되고 있다. 이 시스템에서는 마이크로폰 어레이를 이용하여, 화자의 위치를 파악하여 화자의 방향으로 카메라를 자동으로 조정해 주고 그 방향으로부터 입사되는 신호만을 수신할 수 있도록 한다. 이를 위해 마이크로폰 어레이가 연구되어져 왔다. 본 연구에서는 시간에 따라 변화하는 음향 환경에 적응하는 적응 정합 필터 마이크로폰 어레이를 제안하고, 실험을 통해 그 성능을 고찰하였다.

I. 서 론

최근 영상 회의 시스템에서 화자 위치 추정과 음질 향상 기술이 부각되고 있다. 이 시스템에서는 마이크로폰 어레이를 이용, 화자의 위치를 파악하여 화자의 방향으로 카메라를 자동으로 조정하면서 음원을 선택하여 음질을 향상시킨다. 음질 향상 위한 기존의 방법으로써 적응 잡음 제거 기법, 능동 소음 제어 기법, 그리고 빔 형성 기법 등이 있다[1,2,3]. 적응 잡음 제거 기법은 적응 필터를 이용하여 주 채널에 입력되는 소음 신호를 보조 센서로 수신되는 소음 신호 사이의 상관도를 이용하여 제거한다. 적응 필터 구조로써 transversal, 격자형, Gram-Schmidt 및 혼합형 구조 등이 있으며, 적응 알고리즘으로 LMS (Least Mean Square), RLS(Recursive Least Square) 방법 등이 연구되었다. 이 방법의 성능은 주 채널로 입력되는 소음과 상관성이 높은 소음 신호를 보조 센서를 통해 획득하는데 있으며, 이는 다양한 차량 소음 환경을 고려할 때 많은 문제점을 갖고 있다. 능동 소음 제어 기법은 소음과 같은 크기를 가지면서 역 위상을 갖는 인공 소음을 만들어 스피커를 통해 방사함으로써 서로 상쇄 효과를 얻는 기법이다. 그러나 이 방법은 거의 단일 주파수 성분을 가지면서 저주파 대역 신호인 경우에만 적합하고 고주파 신호 성분이나 광대역 잡음인 경우 제어가 어렵다. 빔 형성 기법은 다수의 마이크로폰 입력 신호들을 조합하여 출력 단에서 최대 신호 대 잡음비를 갖도록 하는 것이다. 그러나 이 방법은 마이크로폰의 수가 적은 경우 빔 폭이 커져서 원하지 않은 외부 소음도 수신할 수 있다.

본 논문에서는 정합 필터 마이크로폰 어레이[4,5]를 이용하여 화자 위치 추적을 위한 방법을 연구하였다. 기존의 정합 필터 마이크로폰 어레이는 실내공간 모델로부터 계산된 임펄스 응답과 실제 마이크로폰에 수신된 신호 사이에 콘볼루션을 구한다. 이는 이상적인 경우, 즉 실내 음향 모델이 정확한 경우에만 원하는 성능을 얻을 수 있다. 그러나 실제의 경우 실내 음향 환경은 시시각각 변화하므로 정합필터 어레이의 성능은 저하된다. 이에 본 연

본 연구는 산업자원부의 산업기초기술연구개발사업의 연구비 지원으로 수행되었습니다.

구에서는 수신된 신호로부터 실내 음향 임펄스 응답을 추정하고, 이를 이용하여 정합 필터 어레이를 구성하였다. 이때 임펄스 응답은 수신된 신호로부터 구성한 자기 상관 행렬의 잡음 부공간을 추정함으로써 계산된다. 또한 이 과정은 이미 널리 알려진 적응 알고리듬을 이용하여 효율적으로 계산할 수 있다. 실험 결과에서 연구된 방법의 효용성을 입증하였다.

II. 정합 필터 어레이

정합 필터 어레이는 원래 영상 회의와 같은 일반 실내에서 잔향의 영향을 감소시키기 위한 방법으로써 연구되었다. 그림 1과 같이 화자와 i번째 마이크로폰 사이의 임펄스 응답을 $h_i(t)$라 할 때 원래의 음성 신호 s(t)는 $h_i(t)$와 콘볼류션 되어 마이크로폰에 입력된다. 마이크로폰에 입력된 신호는 $h_i(t)$의 역필터 $h_i(-t)$를 통과하여 원래 음성 신호로 재현된다. 만약 마이크로폰의 수가 N개라면 정합 필터 어레이의 출력 y(t)는 다음과 같다.

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

N이 충분히 크다면 임펄스 응답과 그 역필터 사이의 콘볼류션은 임펄스 형태가 되며 결국 정합 필터 어레이 출력은 음성 신호와 같아진다. 이러한 정합 필터 어레이는 빔 형성 기법에 비해 잔향에 강한 장점을 갖고 있으나 임펄스 응답의 길이가 긴 경우 역필터의 길이가 길어져 계산량이 크게 증가되는 단점을 갖고 있고, 아울러 대상 실내 음향 특성을 미리 알고 있어야만 한다.

III. 적응 정합 필터 마이크로폰 시스템

Benesty는 TDOA(Time-Difference-of-Arrival) 추정을 이용한 위치 추정에 적용하기 위하여 채널 임펄스 응답 추정 기법을 제안하였다[6]. 본 연구에서는 기존의 정합 필터 어레이의 문제점 가운데 하나인 시변 채널 임펄스 응답 문제를 해결하기 위한 적응 정합 필터 시스템을 제안하였으며, 이를 위해 [6]에서 연구된 적응 채널 추정 알고리듬을 적용하였다. 우선 i번째 마이크로폰에 수신된 신호는 다음과 같이 표현할 수 있다.

$$x_i(t) = h_i(t) * s(t) + n_i(t) \tag{2}$$

$h_i(t)$는 i번째 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.

$$x_1(t) * h_2(t) = s(t) * h_1(t) * h_2(t) = x_2(t) * h_1(t) \tag{3}$$

윗 식 (3)를 백터 연산으로 표현하면 다음과 같다.

$$X_1^T(t)\, H_2(t) = X_2^T(t)\, H_1(t) \tag{4}$$

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

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

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

이제 두 마이크로폰 사이의 공분산 행렬은 다음과 같다.

$$R = \begin{bmatrix} R_{x_1 x_1} & R_{x_1 x_2} \\ R_{x_2 x_1} & R_{x_2 x_2} \end{bmatrix} \tag{7}$$

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

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

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

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

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

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

실제로 행렬 R의 최소 고유치에 대응하는 고유 벡터 u을 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[7]. 따라서 실시간 처리 시스템으로

적용하기 위해서는 상대적으로 계산량이 적으면서도 효율적인 constrained Least Mean Square 알고리듬을 사용하여 반복해서 계산할 수 있다. 최적 값을 u_{opt} 라고 하고, 오차 신호 e(t)를 다음과 같이 정의한다.

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

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

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

이제 gradient-descent constrained LMS 알고리즘을 다음과 같이 얻을 수 있다.

$$u(t+1) = u(t) - \mu e(t)\nabla e(t) \tag{13}$$

여기서 μ는 수렴 속도를 결정하는 임의의 상수 값이다. 식 (11)과 (12)을 (13)에 대입하면 다음과 같다.

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

식 (14)을 이용하여 시변 채널 임펄스 응답을 계산할 수 있으며, 이로부터 적응 정합 필터를 구성한다. [6]에서는 식 (14) 보다 적은 계산량을 갖는 간소화된 알고리듬을 제안하기도 하였다. 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 된 형태로써 무지향성을 가지며, 44.1kHz의 샘플링 주파수를 갖도록 하였다. 마이크로폰 사이의 간격은 30cm 이고 음원은 그림 3과 같이 마이크로폰 배열의 중심으로부터 수직방향으로 전방 2m의 거리를 두었다. 실험이 수행된 실내 공간은 넓이 6.3m, 폭 16.2m 높이 6m 이다. 그림 4에 실험 결과를 나타내었다.

그림 4 (a)는 음원으로써 음성 신호 신호이며, (b)는 실내 전달 특성을 포함한 마이크로폰 출력 신호이다. 그림으로부터 실내 음향 특성으로 인해 많은 왜곡이 발생한 것을 볼 수 있다. 그림 4 (c)는 제안된 방법의 결과로 음원 신호와 거의 비슷한 형태를 갖는 것을 알 수 있으며, 그림에서 크기 차가 나는 것은 고유 벡터를 추정하는 과정에서 정규화(normalization) 과정이 포함되어 있기 때문이나 성능에는 영향을 미치지 않는다.

Ⅴ. 결 론

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 적용 될 수 있다.

추후 연구된 방법의 다양한 성능을 고찰해야 한다. 즉, 실내 음향 특성, 이동 음원, 신호 대 잡음비등과 같은 변수들에 따른 분석이 요구된다. 아울러 정합 필터 어레이의 또 다른 문제점인 역필터를 연산할 때 계산량을 줄이기 위한 방법들에 대해 연구되어져야 한다.

참고문헌

[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. Kaneda and J. Ohga, "Aadaptive microphone array system for noise reduction," *IEEE Trans. Acoust. Speech Signal Proc.*, vol.34, no.6, pp.1391-1400, Dec. 1986.

[3] Y. Grenier, "A microphone array for car environments," *Speech Communication*, vol.12, no.1, pp.25-39, 1993.

[4] R.J. Renomeron, D.V. Rabinkin, 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.

[5] E.E. Jan and J. Flanagan,"Sound capture from spatial volumes: matched filter processing of microphone arrays having randomly-distributed sensors," *Proc. ICASSP*, pp.917-920, 1996.

[6] J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization" *J. Acoust. Soc. Am.* 107 (1), January 2000.

[7] D.V. Rabinkin, *Optimum sensor placement for microphone arays*, Ph.D dissertation, Rutgers University, 1998.

[8] J. Benesty and S.L. Gay, *Acoustic signal processing for telecommunication*, Kluwer Academic Publishers, 2000.

[9] M. Brandstein and D. Ward, *Microphone Arrays - Signal Processing Techniques and Applications*, Springer, 2001.

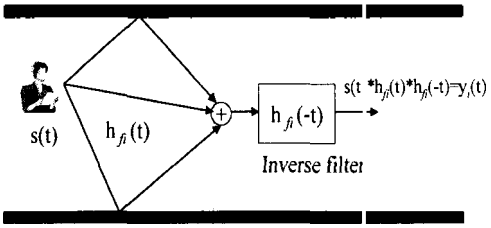

그림 1. 정합 필터 어레이의 개념도.

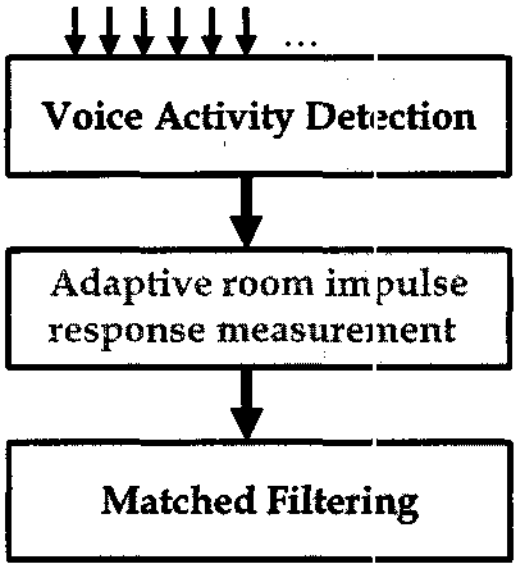

그림 2. 제안된 방법의 흐름도.

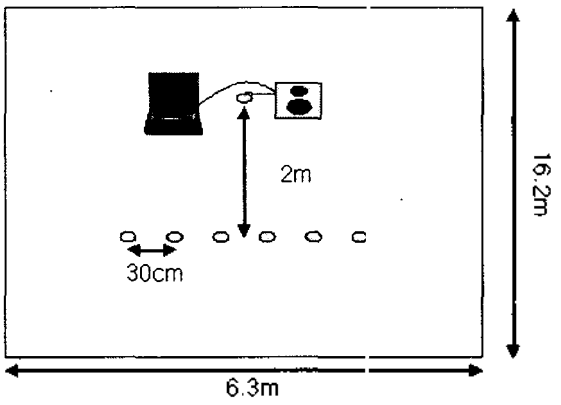

그림 3. 실험 배치도

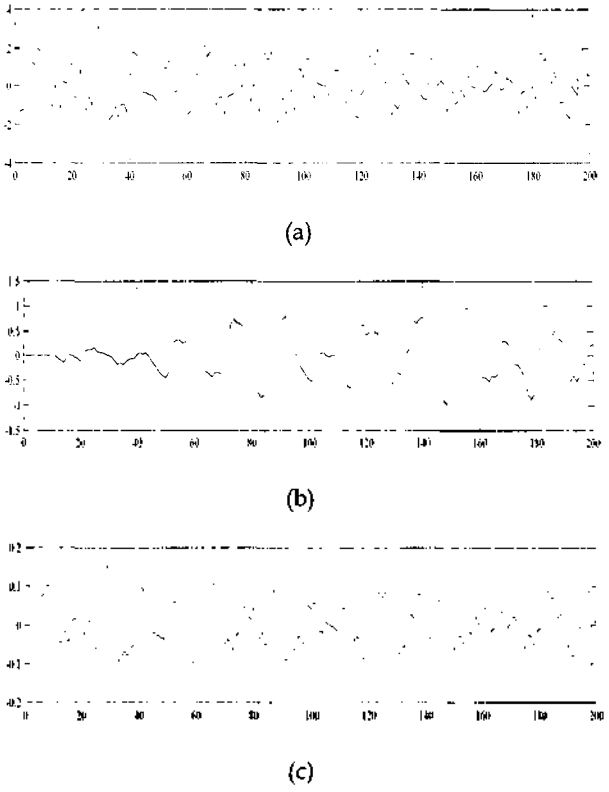

그림 4. (a) 음원 신호, (b) 마이크로폰 출력 신호, (c) 제안된 방법 출력.