상보적인 빔형성에 기반한 대수적 마이크로폰 배열을 이용한 음성개선

장병욱, 권홍석, 김시호, 배건성
경북대학교 전자·전기공학부

Speech Enhancement Using LLA Microphones Based on Complementary Beamforming

Byung Wook Jang, Hong Seok Kwon, Si Ho Kim, Keun Sung Bae
School of the Electronics & Electrical Engineering, Kyungpook National University
star@mir.knu.ac.kr

요 약

본 연구에서는 상보적인 빔형성에 기반한 대수적 마이크로폰 배열을 이용한 음성개선 시스템을 제안한다. 사람들이 많이 모여있는 회의실이나 사무실 환경에서는 백색잡음 보다 음성잡음, 즉, 다른 화자의 음성신호가 더 큰 영향력을 가질 수 있다. 따라서 대수적인 마이크로폰 배열을 사용함으로써 기존의 빔형성 기법에 비하여 저주파 영역에서의 성능을 향상시키고자 하였다. 모의실험 결과, 백색 가우시안 잡음에 대해서는 별다른 성능저하 없이 저주파 성분이 강한 음성잡음에 대해서는 우수한 성능을 가짐을 알 수 있었다.

Ⅰ. 서 론

여러 가지 잡음제거 기법 중에서 마이크로폰 배열을 이용한 방법은 가장 효과적인 방법 중의 하나인데, 지연합 빔형성 방법[1]과 스펙트럼 차감법[2], 적응 빔형성 방법[3] 등이 주로 사용된다. 지연합 빔형성은 신호가 배열로 전파되면, 적당히 지연시킨 후 더하여 다른 방향에서 전파되는 신호성분에 대하여 원하는 신호를 강화시키는 방법이다. 이 방법으로 높은 성능을 얻기 위하여 많은 수의 마이크로폰을 필요로 하며 신호의 도달각(Direction Of Arrival)을 미리 알아야 한다. 스펙트럼 차감법은 일반적으로 잡음이 섞인 입력신호의 스펙트럼에서 추정된 잡음신호의 스펙트럼을 차감하는 기법이다. 잡음신호의 스펙트럼은 묵음구간 검출을 통하여 이루어지므로 정확한 묵음구간의 검출이 필요하다. 만약 정확한 묵음구간의 검출이 이루어지지 않으면 잘못된 잡음을 추정하게 되어 성능저하를 일으킬 수 있으며 과추정으로 인하여 musical noise가 발생할 수 있다는 단점이 있다. 적응 빔형성을 이용한 방법은 알고리즘이 매우 광범위할 뿐만 아니라 널리 적용되는 기법으로 adaptive canceling 혹은 adaptive nulling 방법에 의하여 특정한 방향의 간섭을 줄이거나, 원하는 신호의 도달각을 미리 알지 못해도, 자동적으로 배열을 신호방향으로 조종하여 신호방향 이외의 방향에서 전파되는 잡음을 제거할 수 있다.

음성신호의 잡음제거를 위한 마이크로폰 배열 형태는 크게 선형 마이크로폰 배열과 비선형 마이크로폰 배열로 나눌 수 있는데, 구조적으로 간단한 선형 마이크로폰 배열이 주로 사용된다. 선형 마이크로폰은 다시 등간격 선형 배열과 대수적 선형 배열로 나눌 수 있으며, 대수적 선형 배열은 저주파 영역에서의 높은 상관성과 고주파 영역에서의 공간적 왜곡을 동시에 고려하기 위하여 사용되는 배열 방법이다[4]. 본 연구에서는 등간격 선형 배열에 비하여 저주파 영역에서의 성능 향상이 용이한 대수적 선형 배열을 이용한 음성개선 방법을 제안하고 시뮬레이션 하였다.

본 논문의 구성은 다음과 같다. 먼저 2장에서 상호보완적인 빔패턴을 갖는 등간격 선형 배열 빔형성기에 대하여 설명하고, 3장에서 제안한 대수적 선형 배열 빔형성기에 대해 설명한다. 그리고 4장에서 시뮬레이션을 통하여 제안한 방법의 성능을 제시하고, 마지막으로 5장에서 결론을 맺는다.

Ⅱ. 상호보완적인 빔패턴을 갖는 빔형성기

지연합 빔형성기는 모든 마이크로폰의 입력신호에 대하여 한 가지의 이득을 가진다. 하지만 상호보완적 빔형성 기술을 사용하면 각각의 마이크로폰에 대해 상호보완적인 두 가지의 이득을 사용함으로써 look direction 방향을 제외한 나머지 방향에 대해 빔형성기의 전체적인 이득을 낮출 수 있다[5]. 배열에 사용되는 마이크로폰의 갯수를 M이라고 하면 각 주파수에서 사용하는 상호보완 이득 벡터를 $\boldsymbol{g}=[g_1,\cdots,g_m,\cdots,g_M]$, $\boldsymbol{h}=[h_1,\cdots,h_m,\cdots,h_M]$로 표현할 수 있다. 상호보완적이라는 말은 어떤 주파수에서 임의의 방향에 대하여, 두 빔패턴이 look direction에 대해서는 똑같이 1의 이득을 가지고, 다른 방향으로 들어오는 신호에 대해서는 $|\boldsymbol{g}\boldsymbol{a}_d(f)| \gg |\boldsymbol{h}\boldsymbol{a}_d(f)|$, 또는 $|\boldsymbol{g}\boldsymbol{a}_d(f)| \ll |\boldsymbol{h}\boldsymbol{a}_d(f)|$의 조건을 만족하는 경우를 의미한다. 여기서, $\boldsymbol{a}_d(f)$는 각 배열요소에 도달각 θ_d로 입사하는 신호에 대한 위상차로서 steering vector라고 하며,

$$\boldsymbol{a}_d(f)=[a_{1,d}(f),\cdots,a_{m,d}(f),\cdots,a_{M,d}(f)]^T \quad (1)$$

$$a_{m,d}(f)=\exp(j2\pi f x_k\frac{\sin\theta_d}{c}) \quad (2)$$

로 표현할 수 있다. 여기서 c는 음속이고, x_k는 마이크로폰 배열의 중심에서 마이크로폰까지의 거리이다.

각각의 주파수에 대하여 마이크로폰에 입사하는 신호성분은 이득 벡터 $\boldsymbol{g}$와 $\boldsymbol{h}$에 의하여

$$S^g(f)=S_0(f)+\sum_{d\in\Omega}\boldsymbol{g}\boldsymbol{a}_d(f)\cdot N_d(f) \quad (3)$$

$$S^h(f)=S_0(f)+\sum_{d\in\Omega}\boldsymbol{h}\boldsymbol{a}_d(f)\cdot N_d(f) \quad (4)$$

가 된다. 여기서, $S_0(f)$는 look direction에서 입사하는 신호의 스펙트럼이고 $N_d(f)$는 look direction 이외의 방향에서 들어오는 잡음신호의 스펙트럼이며, Ω는 look direction에 해당하는 $\theta_d=0^\circ$ $(\equiv\theta_{d_0})$를 제외한 방향의 집합이다. 따라서 $S^g(f)$와 $S^h(f)$의 합과 차를 각각 $S^p(f)$, $S^r(f)$라고 하면

$$S^p(f)=2S_0(f)+\sum_{d\in\Omega}\{\boldsymbol{g}\boldsymbol{a}_d(f)+\boldsymbol{h}\boldsymbol{a}_d(f)\}\cdot N_d(f) \quad (5)$$

$$S^r(f)=\sum_{d\in\Omega}\{\boldsymbol{g}\boldsymbol{a}_d(f)-\boldsymbol{h}\boldsymbol{a}_d(f)\}\cdot N_d(f) \quad (6)$$

이 된다. 만약 $\boldsymbol{g}$와 $\boldsymbol{h}$에 의해 만들어지는 빔패턴 $|\boldsymbol{g}\boldsymbol{a}_d(f)|$와 $|\boldsymbol{h}\boldsymbol{a}_d(f)|$가 상호보완적이고 마이크로폰에 입사하는 잡음신호 사이에 상관성이 없다면, $S^p(f)$의 잡음성분에 대한 기대값을 $S^r(f)$의 기대값으로 근사화시킬 수 있다. 따라서 개선된 신호의 스펙트럼은

$$Y(f)=\frac{1}{2}\cdot||S^p(f)|^2-E[|S^r(f)|^2]|^{1/2}\cdot e^{j\phi(f)} \quad (7)$$

로 표현할 수 있다. 이 알고리즘을 이용하면 묵음구간의 검출 없이 스펙트럼 차감법을 이용하여 잡음을 제거할 수 있게되어 묵음구간 검출의 오류로 인한 시스템 성능의 저하를 막을 수 있다. 여기서, $Y(f)$는 개선된 음성신호의 스펙트럼을 나타내고, $\phi(f)$는 위상함수로서 지연합 빔형성기에 의해 얻을 수 있다. 이 시스템의 구조는 그림 1과 같다.

만약 잡음의 특성이 시간에 따라 일정하지 않을 때에는 $E[|S^r(f)|^2]$이 과추정되어 지나친 차감이 발생하여 시스템의 성능저하가 초래된다. 따라서 이러한 경우의 성능저하를 막기 위하여 식 (11)의 알고리즘을 사용한다.

$$|Y(f)|\equiv\begin{cases}\frac{1}{2}\cdot||S^p(f)|^2-|S^r(f)|^2|^{1/2}, \\ \quad (\text{if } \beta\cdot\langle|S^r(f)|^2\rangle > |S^r(f)|^2) \\ \frac{1}{2}\cdot||S^p(f)|^2-\langle|S^r(f)|^2\rangle|^{1/2}, \\ \quad (otherwise)\end{cases} \quad (8)$$

여기서, β의 값은 문턱값으로서 0.1을 사용하였다.

식 (8)을 만족시키는 상호보완적인 빔패턴 $|\boldsymbol{g}\boldsymbol{a}_d(f)|$와 $|\boldsymbol{h}\boldsymbol{a}_d(f)|$를 구하려면 $S^p(f)$의 기대값 중에서 잡음성분이 줄어들도록 $\boldsymbol{g}$와 $\boldsymbol{h}$의 빔패턴을 설계하여야 한다. 즉, look direction을 제외한 나머지 방향에서의 이득 $|\boldsymbol{g}\boldsymbol{a}_d(f)\cdot\boldsymbol{h}\boldsymbol{a}_d(f)|$를 작게 만들어야 한다. 이는

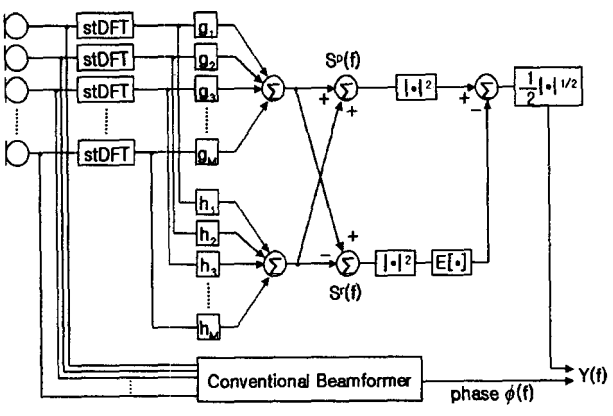


그림 1. 등간격 마이크로폰 배열을 갖는 상보적 빔형성

constrained LS problem으로 귀착되며, Gauss-Newton 방법[6]을 사용하여 g와 h를 구할 수 있다.

Ⅲ. 상호보완적인 빔패턴을 갖는 빔형성 시스템

3.1 등간격 선형 배열 시스템

마이크로폰을 등간격 선형으로 배열하고 입력신호는 10kHz로 표본화 할 경우, 모든 방향의 신호에 대해 spatial aliasing이 발생치 않으려면 마이크로폰 사이의 간격이 3.4 cm보다 작아야 하지만 look direction을 마이크로폰 배열에 수직방향으로만 고정하면 공간적인 왜곡이 발생하지 않는 마이크로폰 사이의 최대 간격은 6.8 cm가 된다. 이것은 3.4 cm의 간격을 갖는 선형 마이크로폰 배열보다 저주파 성분에서 주파수 선택성이 우수하다. 그러나 이 간격에 대하여 grating lobe가 발생하지 않는다는 것일 뿐, 고주파 성분에 대한 빔패턴은 좋지 않기 때문에 이것보다 작은 6 cm의 간격을 사용함으로써 고주파 성분에서의 빔패턴의 특성 저하를 최소화하였다.

3.2 대수적 선형 배열 시스템

사람의 귀는 고주파 성분보다 저주파 성분에 민감하기 때문에 음성개선 시스템은 고주파 성분에 대한 특성보다 저주파 성분에 대한 특성이 더 중요하다. 또한 사람들이 많이 모이는 실내공간에서는 주위의 음성신호가

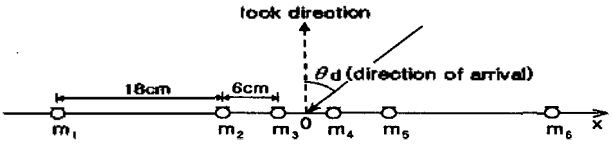


그림 2. 대수적 선형 마이크로폰 배열

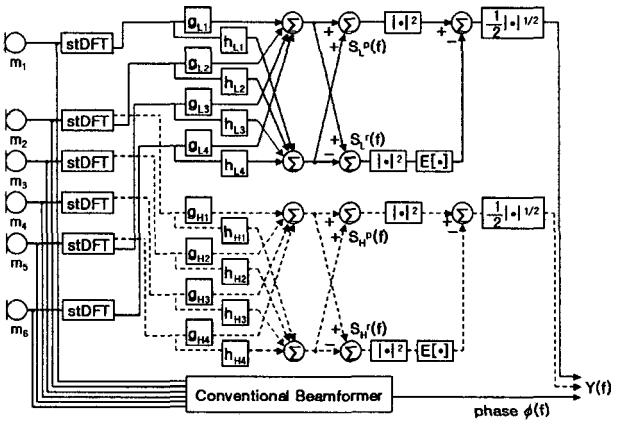


그림 3. 대수적 마이크로폰 배열을 갖는 상보적 빔형성기

잡음으로 많이 작용한다. 이런 음성신호는 저주파 성분에 높은 에너지를 가지기 때문에 음성개선 시스템의 저주파 성분에서의 특성이 강조된다. 따라서, 본 연구에서는 등간격 선형 마이크로폰 배열 대신에 대수적 선형 마이크로폰 배열을 사용하여 저주파 영역에서의 성능을 향상시킨 시스템을 제안하였다. 이에 따라 상호보완 이득벡터 g와 h도 저주파 대역과 고주파 대역에 대하여 다른 값을 사용하였다. 이렇게 함으로써 백색 가우시안 잡음의 경우에 대해서는 성능의 저하가 크게 없으면서도, 음성잡음의 경우에는 향상된 성능을 가지도록 하였다. 마이크로폰 사이의 간격은 그림 2처럼 1,667~5,000Hz의 고주파 성분에 대해서는 4개의 마이크로폰 m_2, m_3, m_4, m_5를 6 cm 간격으로 배열하였고, 0~1,666Hz의 저주파 성분에 대해서는 4개의 마이크로폰 m_1, m_2, m_5, m_6를 18 cm 간격으로 배열하여 사용하였다. 그림 3은 본 연구에서 제안한 시스템의 구조를 보인 것이다.

Ⅳ. 실험 및 검토

제안한 시스템의 성능을 측정하기 위하여 백색 가우시안 잡음, 음성신호 잡음의 두 가지를 사용하여 실험하였다. 백색 가우시안 잡음의 경우에는 SNR이 0dB인 경우와 10dB인 경우에 대하여 실험을 하였다. 본 논문에서 사용한 SNR의 정의는 식 (9)와 같다.

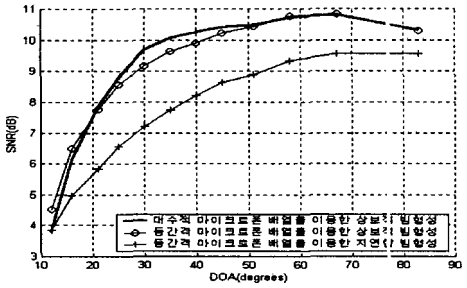
$$SNR(dB) = 10\log_{10}\left(\frac{\sum_i (s(i))^2}{\sum_i (n(i))^2}\right) \tag{9}$$

여기서, $s(i)$는 잡음이 섞이지 않은 음성신호의 샘플을 나타내고, $n(i)$는 잡음신호의 샘플이다.

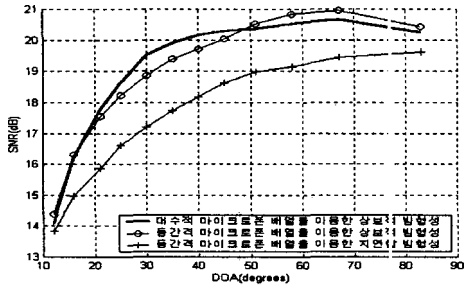
그림 4는 여러 방향(10°~90°)에서 들어오는 백색 가우시안 잡음에 대한 SNR 개선 정도를 나타낸 그림이다. 좌측의 그림은 SNR이 0dB인 경우의 결과이고, 우측의 그림은 SNR이 10dB인 잡음에 대한 결과이다. 굵은 실선은 6개의 대수적 마이크로폰 배열을 이용한 상보적 빔형성의 결과이고, 'o'선은 6개의 등간격 마이크로폰 배열을 이용한 상보적 빔형성의 결과이며, '+'선은 6개의 등간격 마이크로폰 배열을 사용한 지연합 빔형성의 결과이다. 그림에서 볼 수 있듯이 제안한 방법은 지연합 빔형성보다 0~2dB 정도 우수함을 볼 수 있고, 상보적 등간격 마이크로폰 배열과는 비슷한 성능을 보임을 알 수 있다.

그림 5는 여러 방향(10°~90°)에서 들어오는 음성잡음에 대한 개선 정도를 0dB와 10dB에 대하여 각각 나타낸 그림이다. 그림에서 볼 수 있듯이, 상보적 빔형성

기에서 모든 방향에 대하여 대수적 마이크로폰 배열이 등간격 마이크로폰 배열보다 크게는 2dB까지 SNR이 개선되었음을 볼 수 있으며, 지연합 빔형성에 대해서는 최고 5dB까지 향상되었음을 볼 수 있다.

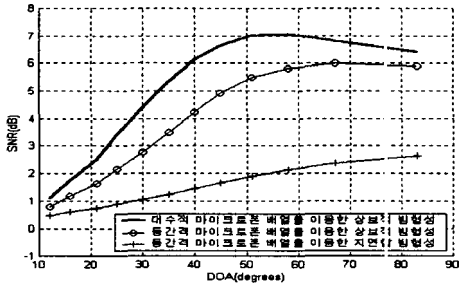


(a) 입력 SNR = 0dB

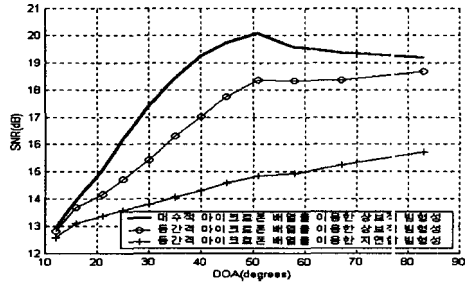


(b) 입력 SNR = 10dB

그림 4. 백색 가우시안 잡음에 대한 개선결과



(A) 입력 SNR = 0dB



(b) 입력 SNR = 10dB

그림 5. 음성잡음에 대한 개선결과

V. 결 론

상호보완적인 빔형성을 가지는 마이크로폰 배열은 지연합 빔형성기에 비하여 우수한 성능을 가진다. 마이크로폰을 등간격으로 배열하였을 경우에는 저주파 영역에서 빔패턴이 좋지 않기 때문에 에너지가 낮은 주파수에 많이 몰려있는 음성신호와 같은 잡음을 잘 제거하지 못하는 경향이 있다. 따라서 본 연구에서는 대수적인 마이크로폰 배열을 사용하여 낮은 주파수에서 좋은 특성을 가지도록 하여 에너지가 저주파 영역에 많이 분포되어 있는 음성신호와 같은 잡음 신호를 잘 제거하도록 하였다. 시뮬레이션 결과 백색 가우시안 잡음에 대해서는 제안한 방법이 지연합 빔형성에 비해서는 성능이 우수하였으며, 기존의 방법과는 비슷한 성능을 보였다. 그리고 음성잡음에 대해서는 지연합 빔형성에 비해서는 성능이 매우 우수하였으며, 기존의 방법보다도 우수한 결과를 얻을 수 있었다.

본 연구는 한국과학재단 목적기초연구(R01-1999-00233)지원으로 수행되었음.

참 고 문 헌

[1] J.L.Flanagan, J.D. Johnston, R. Zahn, and G.W. Elko, "Computer-steered Microphone Arrays for Sound Transduction in Large Rooms," J. Acoust. Soc. Am., vol.78, no.5, pp.1508-1518, 1985

[2] S.F. Boll, "Suppression of Acoustic Noise in Speech Enhancement in a Car Environment Using Wiener Filtering and Spectral Subtraction," IEEE Trans. Acoust., Speech & Signal Process., vol.ASSP-27, no.2, pp.113-120, 1979

[3] L.J. Griffiths and C.W. Jim, "An Alternative Approach to Linearly Constrained Adaptive Beamforming," IEEE Trans. Antennas & Propag., vol.AP-30, no.1, pp27-34, 1982

[4] Enrique Masgrau, Luis Aguilar and Eduardo Lleida, "Performance Comparison of Several Adaptive Schemes for Microphone Array Beamforming," ESCA, Eurospeech99, Budapest, Hungary, ISSN 1018-4074, pp.2615-2618, 1999

[5] Hiroshi Saruwatari, Shoji Kajita, Kazuya Takeda, Fumitada Itakura, "Speech Enhancement Using Nonlinear Microphone Array with Complementary Beamforming," IEEE, Acoustics, Speech and Signal Processing, vol.1, pp.69-72, 1999

[6] A.L. Peressini, F.E. Sullivan, J.J. Uhl, *The Mathematics of Nonlinear Programming*, Springer -Verlag, 1988