

마이크로폰 배열을 이용한 웨이브렛 도메인에서의 음성신호 개선

장 병 옥, 권 흥 석, 김 시 호, 배 건 성
경북대학교 전자·전기공학부

Microphone Array Processing in the Wavelet Domain for Speech Enhancement

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

요 약

마이크로폰을 배열을 이용한 음성개선 기법 중에서 저주파 영역에서의 높은 상관성과 고주파 영역에서의 spatial aliasing을 동시에 고려하기 위하여 대수적인 선형 마이크로폰 배열을 사용하고 웨이브렛 도메인에서의 Wiener 필터에 기반한 postfiltering을 수행하는 방법이 제안된 바 있는데[1], 본 논문에서는 이 방법의 문제점을 분석하고 해결방안을 제시하였다. 제안한 알고리즘을 사용하여 시뮬레이션한 결과, 마이크에 입력되는 음성신호의 SNR이 0dB일 때와 10dB일 때, 기존의 알고리즘에 비해 약 1.7dB와 2.5dB의 성능개선이 있었으며, 청취실험을 통해서도 음질의 향상을 확인할 수 있었다.

I. 서 론

여러 가지 잡음제거 기법 중에서 마이크로폰 배열을 이용한 방법은 가장 효과적인 방법 중의 하나이다. 그 중에서도 DS(Delay-and-Sum) beamforming[2]과 adaptive beamforming[3]이 주로 사용되는데 DS beamforming은 높은 성능을 얻기 위하여 많은 수의 마이크로폰을 필요로 하고 DOA(Direction of Arrival)를 정확하게 예측하지 못할 경우에는 시스템의 성능이 저하되는 단점이 있다. Adaptive beamforming은 주로 LMS 알고리즘에 기반하며 reverberant 환경에 적용되기 위해서는 많은 수의 필터계수를 필요로 하며 많은 계산량이 소요되어서 실시간 구현에 적합하지 않다. 그리고 마이크로폰 배열을 적당히 조합하여 비선형 음성 신호 처리를 하는 SS(Spectral Subtraction) 방법[4] 등

이 제안되었지만 묵음구간 검출의 비정확성과 잡음의 오추정으로 인하여 음질을 저하시키는 결과를 초래할 수도 있다.

한편 마이크로폰 배열을 사용하여 신호를 처리할 때 저주파 영역에서의 높은 상관성과 고주파 영역에서의 spatial aliasing을 동시에 고려하기 위하여 LLA(Logarithmic Linear Array)를 사용하고 웨이브렛 도메인에서 Wiener 필터에 기반한 postfiltering을 수행하는 방법이 제안된 바 있는데[1], 이것은 간축(decimation) 방법으로 웨이브렛 변환을 수행하므로 복원 후 왜곡으로 인하여 음질이 저하되는 문제점이 있다. 본 논문에서는 이러한 문제점을 제시, 분석하였고, 비간축(undecimation) 방법으로 웨이브렛 변환함으로써 이러한 문제점을 해결할 수 있음을 보이고 시뮬레이션을 통하여 성능이 향상됨을 확인하였다.

본 논문의 구성은 다음과 같다. 먼저 2장에서 웨이브렛 변환을 이용한 음성개선 방법과 문제점을 설명하고, 3장에서 해결 방안을 제시한다. 그리고 4장에서 제안한 방법에 의한 시뮬레이션 결과를 보이고 마지막으로 5장에서 결론을 맺는다.

II. 웨이브렛 변환을 이용한 음성개선

2.1 마이크로폰 배열

선형 마이크로폰 배열을 사용할 경우, 샘플링률이 8kHz일 때 이론적으로 spatial aliasing이 일어나지 않으려면 마이크로폰 사이의 간격이 4.25cm보다 작아야 한다. Nyquist 주파수에 맞게 마이크로폰 간격을 조절 한 ULA(Uniform Linear Array)의 경우에는 Nyquist 주파수에서는 좋은 특성을 가지지만 낮은 주파수 대역

에 대해서는 특성이 좋지 않다. 마이크로폰 사이의 간격이 4.25cm 이상인 ULA의 경우는 낮은 주파수 대역에서는 특성이 좋아지지만 높은 주파수 대역에서는 spatial aliasing이 발생하게 된다. 그러나 LLA의 경우는 주파수 대역을 대수적으로 나누어 spatial aliasing을 피하면서도 낮은 주파수 대역에서 우수한 성능을 가지게 할 수 있다[5]. 그림 1과 같이 6개의 마이크로폰을 대수적으로 배열하여 사용한다면 같은 수의 마이크로폰을 사용하는 ULA보다 낮은 주파수에서 우수한 spatial selectivity를 보인다.

2.2 웨이블릿 변환을 이용한 음성개선

웨이블릿 변환은 고주파 영역에서 우수한 시간 분해능을 가지고 저주파 영역에서 우수한 주파수 분해능을 가진다. 또한 주파수축으로 대수적 분석을 가능하게 하기 때문에 웨이블릿 변환을 통하여 원하는 주파수 대역으로 쉽게 분해할 수 있다(표 1). 일반적으로 간축 웨이블릿 변환은 그림 2와 같이 수행된다. 사용된 필터는 Daubechies I(Haar) 필터이며 PR-QMF(Perfect Reconstruction-Quadrature Mirror Filter)의 특성을 갖는다.

그림 3은 [1]에서 제안한 알고리즘의 구조도를 나타낸 것이다. 백색 잡음이 섞인 신호는 우선 conventional beamformer를 통과한다. Conventional beamforming은 원하는 신호의 DOA를 안다는 가정하에서 적당한 샘플의 지연을 통하여 수행된다. 이렇게 하여 일차적으로 잡음을 제거한 다음 각각의 신호를 웨이블릿 분석을 하여 각각 5개의 대역으로 나눈다. 그리고 각 대역마다 잡음을 제거할 필터계수를 추정하여 웨이블릿 변환 계수를 필터링한 후 웨이블릿 합성을 하면 잡음이 제거된 신호를 얻을 수 있다.

마이크로폰으로 들어오는 신호를 x 라고 하면, x 는 원하는 음성신호 s 와, s 와는 상관성이 없는 잡음신호 n 의 합으로 나타낼 수 있다. 이 신호를 웨이블릿 변환

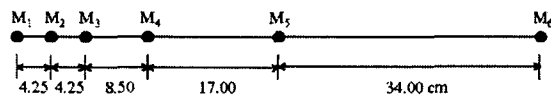


그림 1. 마이크로폰 배열

표 1. 주파수 분할을 위한 subarray

ϕ (부대역)	subarray	주파수 대역
1	M_1, M_5, M_6	0-250Hz
2	M_1, M_5, M_6	250-500Hz
3	M_1, M_4, M_5	500-1000Hz
4	M_1, M_3, M_4	1000-2000Hz
5	M_1, M_2, M_3	2000-4000Hz

행렬 A 를 통과시키면

$$X = Ax = As + An = S + N \quad (1)$$

으로 나타낼 수 있다. 여기서 S, N , 그리고 X 는 각각 s, n , 그리고 잡음이 섞인 신호의 웨이블릿 변환계수이다. 이것을 Wiener 필터, W 에 통과시켜서 잡음을 제거하면

$$\hat{S} = WX = W(S + N) \quad (2)$$

이 되고, 이것을 다시 역웨이블릿 변환을 하면

$$\hat{s} = A^{-1}\hat{S} \quad (3)$$

이 된다. 이 때, W 는 식 (4)와 같이 추정오차의 평균 자승값이 최소가 되도록 결정하며 그 해는 식 (5)와 같다.

$$\min E\{e^2\} = \min E\{[s - \hat{s}]^2\} \quad (4)$$

$$W = \frac{E\{S^2\}}{E\{X^2\}} \quad (5)$$

여기서, $E\{S^2\}$ 와 $E\{X^2\}$ 은 각각 원하는 신호와 잡음이 섞인 신호의 wavelet power spectrum이다. 만약 잡음 신호 사이에 상관성이 없다면 $E\{S^2\}$ 은 두 마이크로폰

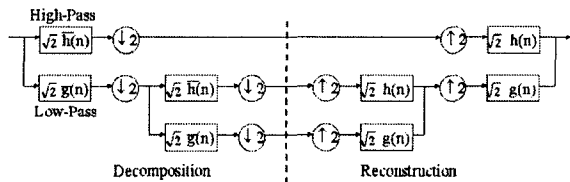


그림 2. 간축 웨이블릿 변환

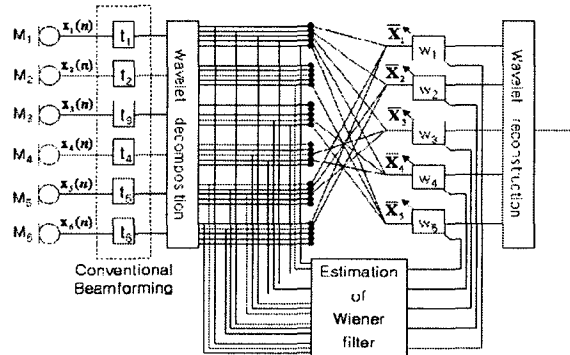


그림 3. 시스템 구조도

신호 $x_i, x_j (i \neq j)$ 의 웨이브렛 변환계수의 상호상관도로 표현될 수 있고, $E\{X^2\}$ 은 마이크로폰 신호 x_i 의 웨이브렛 변환계수의 자기상관도로 표현될 수 있으므로, 최종적으로 웨이브렛 도메인에서의 Wiener 필터계수는 식 (6)과 같다[6].

$$W_\phi = \frac{2 \sum_{i=1}^{M-1} \sum_{j=i+1}^M X_{i\phi}(n) \cdot X_{j\phi}(n)}{\frac{1}{M} \sum_{i=1}^M |X_{i\phi}(n)|^2} \quad (6)$$

여기서, ϕ 는 부대역을 나타내며 M 은 각 대역의 마이크로폰의 개수를 나타낸다. 그리고 $X_{i\phi}(n)$ 은 i 번째 마이크로폰, ϕ 번째 부대역의 웨이브렛 변환계수다. 이렇게 하여 대역마다 구한 값을 해당 대역의 Wiener 필터의 계수로 사용하고, 필터의 입력으로는 각 대역의 웨이브렛 변환계수의 평균값을 사용한다. 마지막으로 필터를 통과한 신호를 역웨이브렛 변환하면 잡음이 제거된 신호를 얻을 수 있다.

III. 문제점 및 개선방안

일반적으로 일정한 대역폭의 신호를 downsampling한 후 upsampling을 하면 왜곡이 발생한다. 그러나 웨이브렛 변환은 PR-QMF(Perfect Reconstruction-Quadrature Mirror Filter)를 사용하기 때문에 그림 4 처럼 역웨이브렛 변환을 한 후의 원하는 신호성분은 HPF를 통과한 신호와 LPF를 통과한 신호의 위상차가 없어서 남아있는 반면, 원하지 않는 신호성분은 위상차가 $\pm\pi$ 이고 크기는 동일하기 때문에 상쇄된다. 그림 4 에서 0, 또는 $\pm\pi$ 는 LPF를 통과한 신호에 대하여 HPF를 통과한 신호의 위상차를 나타낸다.

그런데 [1]에서 수행한 방법을 그대로 사용하면 웨이브렛 도메인에서 각 주파수대역별로 서로 다른 가중치를 두기 때문에 원하는 신호성분은 크기가 달라져도 위상차의 변화가 없어서 왜곡이 발생하지 않지만, 원하지 않는 신호성분의 경우는 위상차는 여전히 $\pm\pi$ 이면

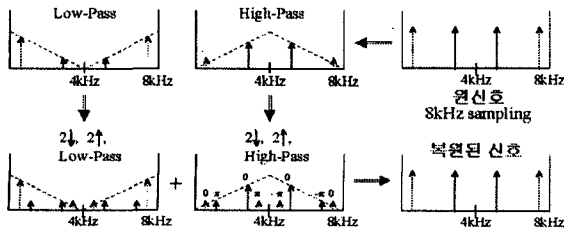


그림 4. 간축 웨이브렛 변환의 위상변화

서 LPF와 HPF를 통과한 신호의 크기가 서로 달라져서 서로 상쇄되지 않고 왜곡이 대칭적으로 발생한다.

이러한 현상을 없애기 위하여 본 연구에서는 그림 5 와 같이 비간축 웨이브렛 변환을 수행하였다. 비간축 웨이브렛 변환은 필터링된 신호의 길이를 입력신호의 길이와 동일하게 유지하면서, 다음 스케일의 신호를 얻기 위해 필터계수들 사이에 0을 삽입하여 신호와 convolution함으로써 간축과정을 거치지 않는다. 즉, 이 방법은 웨이브렛 변환과정에서 신호를 간축하는 것이 아니라 필터를 보간하기 때문에 이러한 왜곡의 영향을 피할 수 있다.

IV. 실험결과

실험에서 사용한 데이터는 8kHz로 표본화하고 16 비트로 양자화하였으며, 분석구간은 hanning 윈도우를 적용한 256샘플로 하였고 분석주기는 128샘플로 하였다. 실험은 SNR이 0dB, 10dB일 경우의 두 가지에 대하여 하였으며 그 중에서 SNR이 0dB일 경우에 대해서만 그림 6과 7에 나타내었다. 그림 6는 원음성, SNR이 0dB인 잡음이 섞인 음성신호, 간축 웨이브렛 변환으로 잡음이 제거된 음성신호, 그리고 비간축 방법으로 잡음이 제거된 음성신호를 나타내며 그림 7은 그림 6의 각각에 대한 스펙트로그램을 나타낸 것이다. 그림 6(c)와 (d)를 보면 간축 방법과 비간축 방법 둘 다, 잡음을 잘 제거함을 볼 수 있지만 그림 7(c)의 스펙트로그램을 보면 간축 방법을 사용하였을 때에는 대칭적인 왜곡이 나타났다. 그러나 비간축 방법을 사용하면 그림 7(d)에 나타나는 것처럼 왜곡이 없어진 것을 볼 수 있다. 그리고 청취실험을 통해 두 결과의 음질을 비교해 보아도 비간축 방법을 사용했을 경우가 간축 방법을 사용했을 경우보다 귀에 거슬리는 소리가 많이 줄어든 것을 확인할 수 있었다.

마이크에 입력되는 신호의 SNR이 0dB일 때 개선된 신호의 SNR은 간축 방법을 사용하면 약 9.3dB였고, 비간축 방법에 의해서 개선된 신호는 약 11.0dB였다. 그리고 입력신호의 SNR이 10dB일 때의 개선된 신호의 SNR은 간축 방법을 사용하면 16.6dB였고, 비간축 방법에 의해서 개선된 신호는 약 19.1dB였다. 따라서, 비

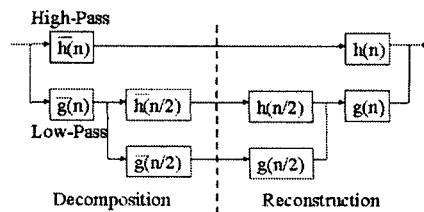


그림 5. 비간축 웨이브렛 변환

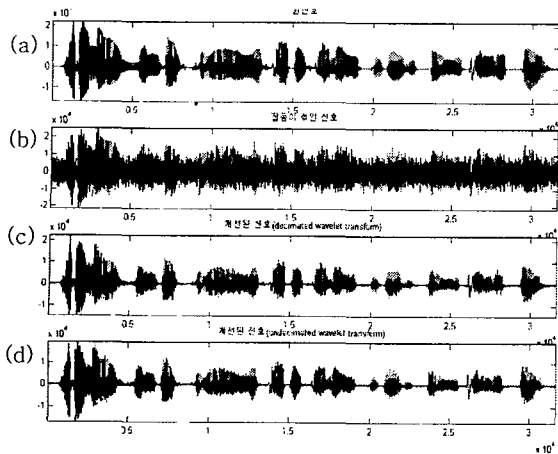


그림 6. 신호의 파형 (a)원신호, (b)잡음이 섞인 신호 (SNR=0dB), (c)개선된 신호(간축방법), (d)개선된 신호(비간축방법)

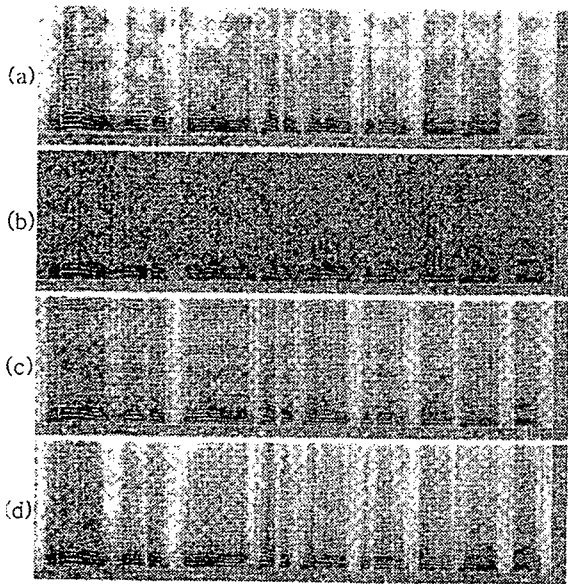


그림 7. 스펙트로그램 (a)원신호, (b)잡음이 섞인 신호 (SNR=0dB), (c)개선된 신호(간축방법), (d)개선된 신호(비간축방법)

간축 방법을 사용하면 SNR이 0dB, 10dB인 입력신호에 대해 비간축 방법보다 각각 약 1.7dB, 2.5dB 높은 SNR을 얻을 수 있었다.

V. 결론

대수적인 마이크로폰 배열을 사용하여 웨이브렛 도메인에서 Wiener 필터에 기반한 postfiltering을 수행하여 음성개선을 하는 방법이 제안된 바 있는데[1], 이것

은 간축 웨이브렛 변환을 수행하고 각 주파수대역마다 다른 계수값을 사용하기 때문에 복원 후 왜곡이 발생하여 음질이 저하되는 문제점이 있다. 본 논문에서는 이러한 문제점을 해결하고자 비간축 방법으로 웨이브렛 변환을 하였다. 그리고 비간축 웨이브렛 변환을 이용하여 시뮬레이션하였으며 기존방법의 문제점을 해결할 수 있음을 확인하였다. 실험에서 마이크로폰 입력신호의 SNR이 0dB일 때, 개선된 신호의 SNR은 간축 방법을 사용하면 약 9.3dB였고, 비간축 방법을 사용하면 약 11.0dB였다. 그리고 마이크 입력신호의 SNR이 10dB일 때, 개선된 신호의 SNR은 간축 방법을 사용하면 16.6dB였고, 비간축 방법을 사용하면 약 19.1dB였다. 따라서 SNR이 0dB와 10dB의 입력신호에 대하여 비간축 방법보다 각각 약 1.7dB, 2.5dB 높은 SNR을 얻을 수 있었고, 청취실험을 통해서도 음질의 향상을 확인할 수 있었다.

참고 문헌

- [1] Djamila Mahmoudi, "A Microphone Array for Speech Enhancement Using Multiresolution Wavelet Transform," ESCA, Eurospeech97, Rhodes, Greece, ISSN 1018-4074, pp.339-342
- [2] 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
- [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] 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
- [5] Masaaki Inoue, Satoshi Nakamura, Takeshi Yamada, Kiyohiro Shikano, "Microphone Array Design Measures for Hands-Free Speech Recognition," ESCA, Eurospeech97, Rhodes, Greece, ISSN 1018-4074, pp.331-334
- [6] Rainer Zelinski, "A Microphone Array with Adaptive Post-Filtering for Noise in Reverberation Rooms," Proc. of ICASSP 88, pp.2578-2581