

마이크로폰 배열에서 독립벡터분석 기법을 이용한 잡음음성의 음질 개선

Microphone Array Based Speech Enhancement Using Independent Vector Analysis

왕 썩 양¹⁾ · 전 성 일²⁾ · 배 건 성³⁾

Wang, Xinyang · Quan, Xingri · Bae, Keunsung

ABSTRACT

Speech enhancement aims to improve speech quality by removing background noise from noisy speech. Independent vector analysis is a type of frequency-domain independent component analysis method that is known to be free from the frequency bin permutation problem in the process of blind source separation from multi-channel inputs. This paper proposed a new method of microphone array based speech enhancement that combines independent vector analysis and beamforming techniques. Independent vector analysis is used to separate speech and noise components from multi-channel noisy speech, and delay-sum beamforming is used to determine the enhanced speech among the separated signals. To verify the effectiveness of the proposed method, experiments for computer simulated multi-channel noisy speech with various signal-to-noise ratios were carried out, and both PESQ and output signal-to-noise ratio were obtained as objective speech quality measures. Experimental results have shown that the proposed method is superior to the conventional microphone array based noise removal approach like GSC beamforming in the speech enhancement.

Keywords: speech enhancement, independent vector analysis, microphone array

1. 서론

다양한 형태의 배경잡음이 포함된 잡음음성에서 잡음성분을 제거하여 음질(speech quality)을 개선하는 기술을 음성개선(speech enhancement)이라고 하는데, 음성통신에서 음성의 명료도를 높이고 잡음으로 인한 청취자의 피로를 덜어주거나, 음성인식시스템의 인식성능을 향상시키기 위해 전처리 과정 등에서 사용되고 있는 기술로서 꾸준히 연구되고 있는 분야이다. 음성개선 방법으로는 크게 스펙트럼 차감법, MMSE-STSA 등과 같이 단일 채널 음성신호를 이용하는 방법[1]과 GSC(Generalized Sidelobe Canceller), DSB(Delay-Sum Beamforming) 등과 같이 마이크로폰 배열 기반의 다채널 음성

신호를 이용하는 방법[2]으로 나눌 수 있다.

독립성분분석(ICA: Independent Component Analysis) 기법은 다채널의 입력신호를 확률통계 관점에서 각각 독립적인 음원의 선형조합으로 가정하고 입력신호간의 독립성을 측정할 수 있는 다양한 척도(measure)를 이용하여 원래의 음원을 찾아내는 암묵적신호분리(blind source separation) 방법인데[3,4], 최근에 ICA를 음성개선에 적용하려는 연구도 발표되고 있다[5]. 일반적으로 시간영역에서 컨볼루션 형태로 모델링되는 입력 음성신호는 주파수영역에서 곱으로, 즉, 선형모델 형태로 표현되므로 마이크로폰 배열에서 얻어지는 다채널 음성신호에 ICA를 적용할 때 보통 주파수영역에서 많이 수행하게 된다. 그러나 주파수영역에서 ICA를 적용할 때 ICA 출력신호에서 주파수 빈(bin)이 이웃 신호의 주파수 빈과 뒤섞이는 문제가 발생하여 원래의 음원을 분리하기가 어렵게 되는데, 이 문제를 해결할 수 있는 방법으로서 독립벡터분석(IVA: Independent Vector Analysis) 기법이 제안되었다[6]. IVA는 주파수영역에서 ICA를 수행할 때, 각 주파수 빈 별로 독립적으로 ICA를 수행하는 대신에 각 신호의 주파수 빈 전체를 하나의 벡터로 간주하여 상호간의 의존성을 고려하면서 ICA를 수행하도록 함으로써 ICA에서 필연적으로 발생하는 주파수 빈의 뒤섞임

1) 경북대학교, wangxinyang123_0@hotmail.com

2) 경북대학교, comet263@hotmail.com

3) 경북대학교, ksbae@ee.knu.ac.kr, 교신저자

이 논문은 2012학년도 경북대학교 학술연구비에 의하여 연구되었음

접수일자: 2012년 8월 14일

수정일자: 2012년 11월 13일

게재결정: 2012년 11월 14일

문제를 해결한 방법이라 할 수 있다.

본 연구에서는 IVA 기법을 마이크로폰 배열에서 얻어지는 다채널 잡음음성 신호에 적용하여 잡음성분이 제거되고 향상된 음질의 음성신호를 얻는 음성개선 방법을 제안하였다. 제안한 방법에서는 먼저 다채널 잡음음성에 IVA 기법을 이용하여 잡음성분과 음성신호를 비롯한 음원신호로 분리하고, 마이크로폰 배열의 DSB(Delay-Sum Beamforming) 출력신호와 IVA로 분리된 출력신호 사이의 상관관계를 이용하여 IVA 출력신호 중에서 개선된 음성신호를 출력한다. 제안한 IVA 기반의 마이크로폰 배열 방식과 잡음제거 성능이 우수한 GSC 기법의 마이크로폰 배열 방식에 대해 컴퓨터 모의실험을 통해 음질개선 성능을 비교하였다. 실험결과 제안한 방법이 백색잡음(white noise) 및 배틀잡음(babble noise)에 대해 객관적인 음질 비교 척도인 PESQ (Perceptual Evaluation of Speech Quality)[7] 값에서 월등히 우수함을 보였다.

본 논문의 구성은 다음과 같다. 2장에서는 본 연구에서 제안한 IVA를 이용한 다채널 잡음음성의 음질 개선 방법에 대해 간략히 설명하고, 3장에서는 컴퓨터 모의실험 조건 및 과정, 음질개선 실험결과를 제시하며, 마지막으로 4장에서 결론을 맺는다.

2. IVA를 이용한 다채널 잡음음성의 음질개선

IVA는 주파수영역에서 ICA를 수행할 때 발생하는 주파수빈의 뒤섞임 문제를 해결해주지만 ICA와 마찬가지로 출력신호의 모호성 문제는 여전히 남게 된다. 출력신호의 모호성 문제란 <그림 1>에 보인 ICA를 이용한 혼합신호의 분리 과정에서 얻어진 출력신호 S1, S2 중에서 어느 것이 Source 1에 해당하는지를 알 수 없는 것을 의미하며, 본 연구에서는 Source 1 이외의 신호는 잡음에 해당되므로 마이크로폰 배열의 DSB 출력신호와 IVA를 이용하여 분리된 출력신호와의 상관관계를 조사하여 이러한 모호성 문제를 해결한다.

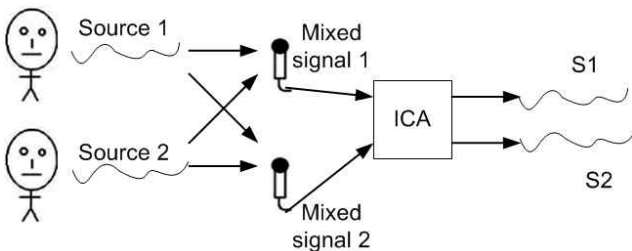


그림 1. ICA를 이용한 혼합신호의 분리 개념도
Figure 1. Blind source separation using ICA

IVA를 이용하여 잡음성분과 음성신호의 분리를 통한 제안한 음성개선 방법의 블록도는 <그림 2>와 같다.

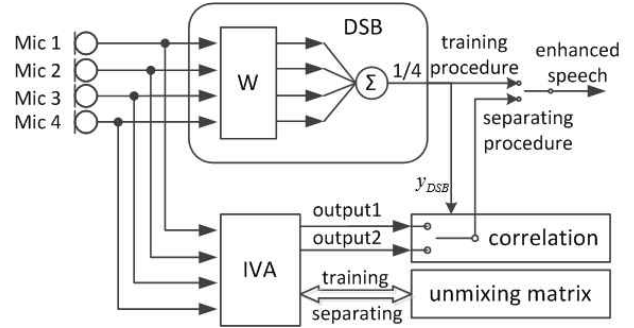


그림 2. IVA를 이용한 제안한 음성개선 방법의 블록도
Figure 2. Block diagram of proposed method using IVA

IVA를 실시간으로 동작하는 마이크로폰 배열에 적용하기 위해서 다음의 2가지 과정이 수행된다. 첫째는, 다채널 입력 잡음음성으로부터 음원이 분리된 출력신호를 얻기 위한 IVA 과정에서 분리행렬(unmixing matrix)을 구해야 하는데, 이 과정은 음성의 시작부분에서 1~2초 구간의 음성을 이용하여 수행하게 되며 그림 2에서 훈련/분리 과정으로 표시되어 있다. 이 구간에서는 DSB의 출력신호가 시스템의 출력신호가 된다. 둘째는, 구해진 분리행렬을 이용하여 IVA를 수행하면 다채널의 출력신호를 얻게 되는데, 잡음이 제거된 음성신호 이외에는 모두 잡음성분에 해당되므로 주요 신호 2개를 추출하여 DSB 출력신호와 IVA 출력신호와의 상관계수를 구하여 상관도가 높은 신호를 시스템의 출력신호로 한다. IVA에서 분리행렬을 구하는 과정과 이를 이용하여 출력신호를 얻는 독립성분분석 알고리즘의 자세한 과정은 [6]에 주어져 있다.

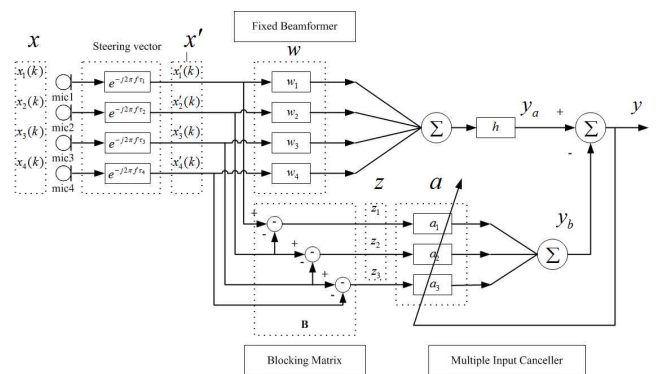


그림 3. 마이크로폰 배열 GSC 시스템의 블록도
Figure 3. Block diagram of GSC based microphone array system

제안한 방법과 음질개선 성능을 비교하기 위한 기존의 마이크로폰 배열 GSC 시스템의 블록도는 <그림 3>과 같다. GSC는 특정 방향의 잡음원에 의한 배경잡음 제거 성능을 향상시키기 위해 잡음신호를 추정하는 적응필터를 DSB에 적용한 것으로, 본 연구에서 적응필터의 차수는 32차, 필터계수의 갱신을 위한 알고리즘으로는 NLMS(Normalized Least Mean

Square)을 이용하였다.

음질개선의 성능평가를 위한 척도로는 일반적으로 많이 사용하는 객관적인 음질평가 척도인 PESQ와 신호대잡음비 (SNR: Signal-to-Noise Ratio)를 사용하였다. PESQ는 ITU-T P.862 권고안으로 표준화된 객관적인 음질평가 방법의 하나로 주관적인 평가결과와 상관도가 아주 높은 것으로 알려져 있으며, 공개된 자동음질 평가 프로그램을 이용할 수 있다. 출력신호의 SNR은 식 (1)로 정의되는데, 여기에서 $s(n)$ 은 원래의 깨끗한 음성신호이고, $s'(n)$ 은 IVA로 분리된 잡음이 제거된 음성이며, N 은 신호의 길이에 해당된다.

$$output\ SNR = \frac{\sum_{n=0}^{N-1} s(n)^2}{\sum_{n=0}^{N-1} [s(n) - s'(n)]^2} \quad (1)$$

3. 실험 및 검토

3.1 실험조건 및 데이터 생성

제안한 방법의 음질개선 성능을 평가하기 위하여 <그림 4>와 같은 마이크로폰 사이의 간격이 5cm인 4-채널 마이크로폰 배열을 가정하고, 음성신호는 마이크로폰 배열의 수직방향을 기준으로 0° 방향에서, 잡음신호는 40° 방향에서 입력되는 잡음음성을 생성하였다. 보통 음원의 위치가 마이크로폰에서 충분히 떨어지므로 평면파를 가정하면 각 채널에는 Mic1을 기준으로 식 (2)와 같은 시간지연을 가지는 신호가 수신되는데, 푸리에 변환을 이용하여 주파수영역에서 지연시간에 해당하는 위상변위를 줌으로써 각 채널의 잡음음성 신호를 생성할 수 있다.

$$\tau_n = (n-1)df_s \sin\theta / c \quad , \quad n = 1, 2, 3, 4 \quad (2)$$

여기서 n 은 채널 번호, f_s 는 샘플링주파수, c 는 공기 중에서의 음속으로 $340m/s$ 를 가정했다.

앞에서 설명한 방법에 따라 남녀 각각 3명씩의 6개 깨끗한 음성신호[8]에 대해 SNR이 -10dB, -5dB, 0dB, 5dB, 10dB가 되도록 가우시안 백색잡음을 섞은 잡음음성 30개와 배블잡음[9]을 섞은 잡음음성 30개를 4개의 각 채널에 대해 생성하였으며, 이때 신호의 샘플링주파수는 16kHz로 하였다. IVA 수행시에 STFT(Short-Time Fourier Transform)의 프레임 길이는 256 샘플로, 64 샘플씩 이동시키면서 분석/합성 하였다.

성능비교를 위한 마이크로폰 배열 기반의 GSC 구현 시에 위와 동일한 조건에서 최적의 결과를 얻을 수 있도록 예비실험을 통해 NLMS 알고리즘을 이용한 적응필터계수 갱신 과정에서의 step size 값을 설정하였다. <그림 5>는 가우시안 백색

잡음 및 배블잡음에 대해 GSC의 step size를 0.002 간격으로 바꾸어가면서 출력신호의 평균 SNR 값을 조사한 것이다. <그림 5(a)>의 백색잡음의 경우에는 대체로 step size 변화가 출력 SNR에 큰 영향을 보이지 않았지만, <그림 5(b)>의 배블잡음에서는 상대적으로 step size의 영향이 컸었다. 이러한 결과를 고려하여 본 연구에서는 step size 값을 0.01로 설정하여 GSC 관련 실험을 수행하였다.

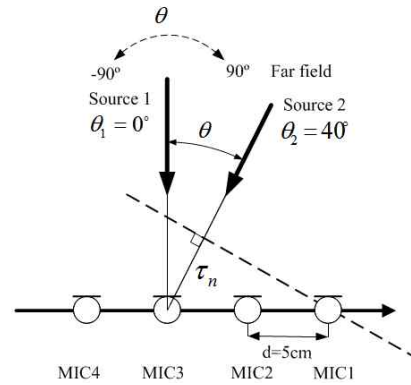
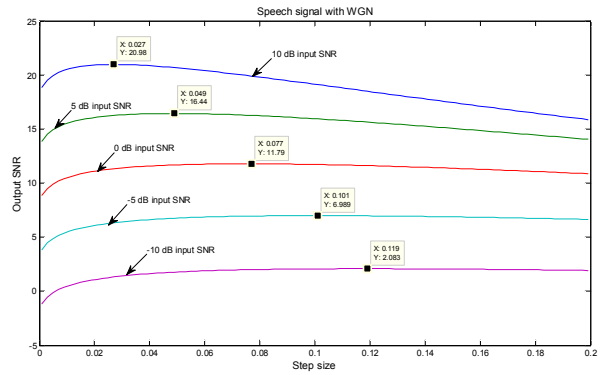
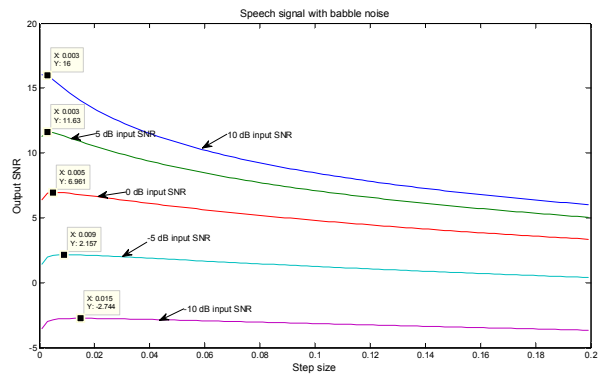


그림 4. 4-채널 선형 마이크로폰 배열
Figure 4. 4-channel linear microphone array



(a)



(b)

그림 5. NLMS 알고리즘의 step size에 따른 출력 SNR
(a) 가우시안 백색잡음의 경우 (b) 배블잡음의 경우
Figure 5. Output SNR respect to NLMS step size
(a) white Gaussian noise, (b) babble noise

3.2 실험결과

IVA의 분리행렬을 구하는 훈련과정에서 요구되는 적절한 데이터 길이를 조사하기 위하여 입력음성 신호의 시작 부분에서 훈련과정에 사용되는 프레임 수에 따라 잡음이 제거된 출력신호의 SNR을 조사하여 <그림 6>에 나타내었다. <그림 6(a)>는 가우시안 백색잡음이 포함된 잡음음성에 대한 결과이며, <그림 6(b)>는 배블잡음에 대한 결과이다. <그림 6>에서 보면 훈련과정에 사용되는 데이터의 길이가 60 프레임 이상이면 큰 변화를 보이지 않음을 볼 수 있는데, 이를 고려하여 본 실험에서는 약 1.28초 길이에 해당하는 80프레임을 분리행렬을 구하기 위한 훈련과정의 데이터로 사용하였다.

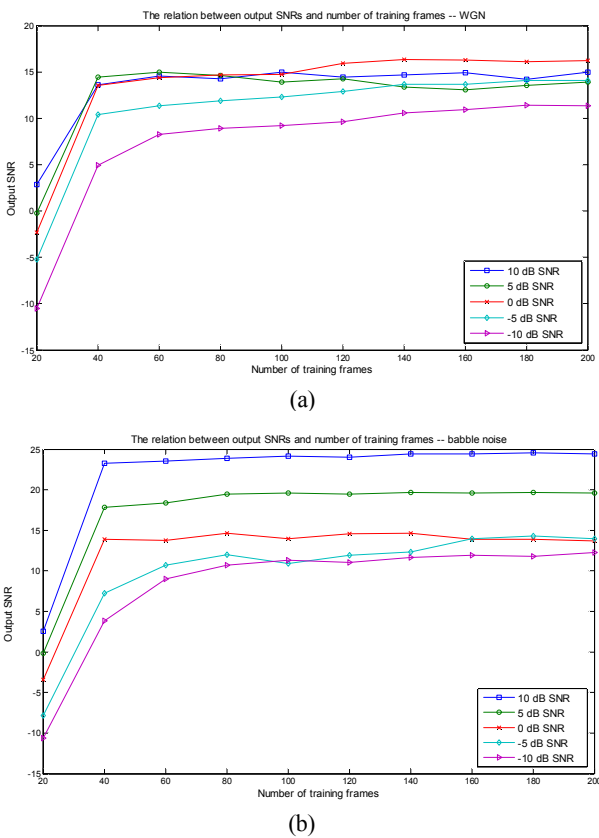


그림 6. IVA 훈련데이터의 프레임 수에 따른 출력신호의 SNR
 (a) 가우시안 백색잡음의 경우 (b) 배블잡음의 경우
 Figure 6. Output SNR respect to number of training frames
 (a) white Gaussian noise, (b) babble noise

기존의 마이크로폰 배열 기반의 GSC 방법과 IVA를 이용한 제안한 방법의 잡음제거 성능을 6명 화자에 대한 평균 PESQ 값으로 비교하여 <표 1>에 나타내었다. IVA를 이용한 제안한 음질개선 방법에서는 잡음음성에 포함된 잡음성분을 추정하여 차감하는 방식이 아니고 암묵적신호분리 기법으로 잡음성분과 음성신호 성분을 분리하는 방법이므로 사람이 청취시의 주관적인 음질평가에 가까운 PESQ 값에서 GSC에 비해 월등히 높은 값을 보임을 볼 수 있다. 특히 입력 잡음음성의 SNR

이 낮을수록, 즉, 잡음이 더 많을수록 기존의 잡음제거 기법에 비해 더 좋은 성능을 보임을 볼 수 있다.

<표 2>는 GSC와 제안한 방법에 대해 6명 화자에 대한 출력신호의 평균 SNR을 보인 것이다. <표 1>에서와 마찬가지로 전체적으로 제안한 방법이 GSC 방법에 비해 출력 SNR이 월등히 개선된 것을 볼 수 있다. 그런데 가우시안 백색 잡음의 경우 입력 잡음음성의 SNR이 상대적으로 높을 경우에는 제안한 방법이 PESQ 값이 훨씬 높음에도 불구하고 출력 SNR 값에서는 GSC에 비해 낮은 값을 얻어짐을 볼 수 있다.

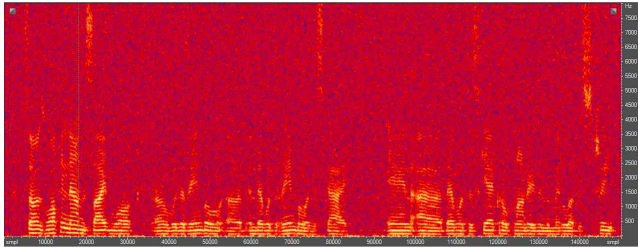
제안한 방법에서 입력 잡음음성의 SNR이 높을 경우에 출력 SNR의 개선 효과가 줄어드는 이유를 분석하기 위하여 출력신호의 스펙트로그램을 조사하였다. <그림 7>은 SNR이 10dB인 가우시안 백색잡음이 포함된 잡음음성과 제안한 방법으로 분리한 음성신호 및 잡음신호의 스펙트로그램을 보인 것이다. <그림 7(b)>와 <그림 7(c)>를 비교해보면 분리된 음성신호와 잡음신호의 고주파 쪽 대역의 스펙트로그램이 서로 바뀌어져 나타난 현상을 볼 수 있다. 다시 말해 주파수 빈의 뒤섞임 문제를 해결한 IVA를 잡음분리 문제에 적용할 때 잡음음성의 SNR이 높은 경우, 즉, 음성신호의 파워가 잡음신호의 파워보다 상대적으로 클 경우에는 고주파대역 및 저주파대역 전체가 뒤섞이는 문제가 발생하는 것으로 생각된다. 따라서 이 경우에 <표 2>에서와 같이 제안한 방법이 SNR 값에서는 GSC 방법보다 낮아지게 되지만 <그림 7(b)>에서 보듯이 음성신호 대부분의 에너지가 저주파 쪽 대역에 포함되어 있으므로 청취시의 음질 특성을 나타내는 PRSQ 값은 <표 1>에서 보듯이 제안한 방법이 훨씬 높은 값을 보이고 있다고 생각된다.

표 1. 백색잡음과 배블잡음에 대한 출력신호의 평균 PESQ
 Table 1. The comparison of average PESQ scores

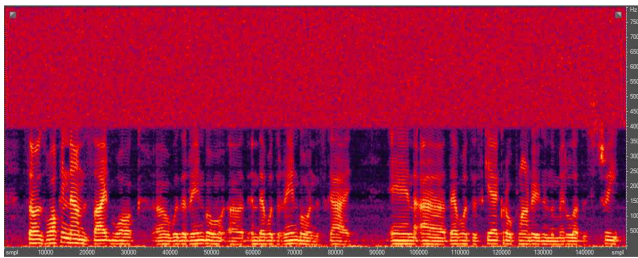
Input SNR (dB)	PESQ for white noise		PESQ for babble noise	
	GSC	Proposed method	GSC	Proposed method
10	2.51	3.30	2.50	3.41
5	2.23	3.22	2.31	3.37
0	1.91	3.34	2.05	3.25
-5	1.58	3.17	1.75	3.12
-10	1.28	3.02	1.43	2.98

표 2. 백색잡음과 배블잡음에 대한 출력신호의 평균 SNR
 Table 2. The comparison of average output SNR

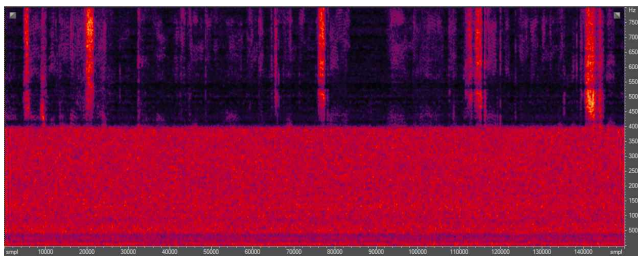
Input SNR (dB)	Output SNR for white noise (dB)		Output SNR for babble noise (dB)	
	GSC	Proposed method	GSC	Proposed method
10	20.14	14.27	14.78	23.91
5	15.34	14.60	11.20	19.49
0	10.40	14.70	6.86	14.67
-5	5.42	11.87	2.10	11.98
-10	0.43	8.94	-2.82	10.71



(a)



(b)



(c)

그림 7. 백색잡음이 첨가된 10dB 잡음음성 및 IVA 출력신호

(a) 10dB 잡음음성 신호의 스펙트로그램

(b) IVA로 분리한 개선된 음성신호의 스펙트로그램

(c) IVA로 분리한 잡음신호의 스펙트로그램

Figure 7. Spectrograms of (a) noisy signal with 10dB WGN,

(b) speech signal seperated by IVA,

(c) noise signal seperated by IVA

4. 결론

본 논문에서는 IVA를 이용한 마이크로폰 배열 기반의 음성 개선 방법을 제안하였다. 입력되는 다채널 잡음음성 신호에서 처음 1~2초 길이의 데이터로 IVA의 분리행렬을 구하고, 구해진 분리행렬을 이용하여 다채널 잡음음성에서 음성신호를 비롯하여 잡음신호들을 분리해 낸다. IVA의 출력신호의 모호성 문제, 즉, 분리된 신호 중에서 잡음이 제거된 음성신호를 찾아 내는 문제는 마이크로폰 배열의 DSB 출력과 IVA 출력신호와 상관함수를 이용하여 해결하였다. 컴퓨터 모의실험을 통하여 IVA를 이용한 제안한 방법이 기존의 마이크로폰 배열 기반의 GSC 방법에 비해 음질개선 면에서 월등히 우수함을 보였다.

입력 잡음음성의 SNR이 상대적으로 높은 경우에는 백색잡음의 경우 IVA로 분리된 신호의 고주파 쪽 대역이 서로 뒤섞

이는 현상이 일부 관찰되었지만, 음성신호의 에너지 대부분이 저주파 대역에 존재하므로 이 경우에도 PESQ로 측정되는 음질 면에서는 GSC 보다 우수한 결과를 보였다. 하지만 추후 고주파 대역에서 발생하는 스펙트럼 빈의 뒤섞임 문제에 대한 연구 및 해결이 요구된다.

참고문헌

[1] Ephraim, Y. & Malah, D. (1984). Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator, *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 32, 1109-1121.

[2] Benesty, J. et al. (2008). *Microphone array signal processing*. Berlin: Springer Verlag.

[3] Hyvarinen, A., Karhunen, J. & Oja, E. (2001). *Independent component analysis*. New York: Wiley.

[4] Haykin, S. (2009). *Neural networks and learning machines*. New Jersey: Pearson Prentice Hall.

[5] Asano, F., Ikeda, S., Ogawa, M., Asoh, H. & Kitawaki, N. (2001). A combined approach of array processing and independent component analysis for blind separation of acoustic signals, in *Proc.IEEE Int. Conf. Acoust., Speech Signal Process.*, 2729-2732.

[6] Kim, T., Attias, H., Lee, S. & Lee, T. (2007). Blind source separation exploiting higher-order frequency dependencies, *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, 70-79.

[7] PESQ, <http://www.itu.int/rec/T-REC-P.862/en>

[8] Donohue, K. D. Systems Array Processing Toolbox. <http://www.engr.uky.edu/~donohue/>.

[9] NoiseX-92, <http://www.speech.cs.cmu.edu/comp.speech/Section1/Data/noisex.html>

• 왕성양 (Wang, Xingyang)

Email: wangxingyang123_0@hotmail.com

관심분야: 음성신호처리, 음성인식

2011~현재 中國 鄭州中原遙感 科學技術有限會社 연구원

• 전성일 (Quan, Xingri)

경북대학교 전자전기컴퓨터학부

대구 북구 산격동 1370번지

Tel: 053-940-8627

Email: comet263@hotmail.com

관심분야: 음성신호처리, ICA, 심음(PCG)신호 분석 및 응용
현재 대학원 전자전기컴퓨터학부 박사과정

• **배건성 (Bae, Keunsung), 교신저자**

경북대학교 IT대학 전자공학부

대구 북구 산격동 1370번지

Tel: 053-950-5527

Email: ksbae@mir.knu.ac.kr

관심분야: 음성신호처리, 디지털신호처리, 수중음향신호처리

1979~현재 전자공학부 교수