
차량용 음성인식을 위한 주변잡음에 강건한 블라인드 음원분리

Robust Blind Source Separation to Noisy Environment For Speech Recognition in Car

김현태*, 박장식**
동의대학교 멀티미디어공학과*, 동의과학대학 디지털정보전자과**

Hyun-Tae Kim(htaekim@deu.ac.kr)*, Jang-Sik Park(jsipark@dit.ac.kr)**

요약

독립성분분석을 사용한 암묵신호분리의 성능은 잔향이 존재하는 환경에서 잔류 누설 성분(cross-talk) 때문에 현저히 저하된다. 본 논문에서는 잔류 누설 성분을 제거하기 위한 후처리 방법을 제안한다. 제안하는 방법은 주파수 영역에서의 변형된 NLMS(normalized least mean square) 필터를 사용하며 필터의 역할은 잔류 누설 성분을 유발하는 누설 경로를 추정하는 데 있다. 특정 채널에서 잔류하는 누설 성분은 상대 채널의 직접 성분에 해당되므로 관측되는 상대 채널의 입력신호를 이용하여 누설 경로를 추정할 수 있다. 변형된 NLMS 필터는 필터 입력 신호의 전력과 추정 오차 신호의 전력을 함께 고려하여 정규화한다. 특정 채널의 직접 신호 성분은 적응 필터에서 잡음처럼 동작하여 결국 적응필터가 오조정되기 때문에 제안하는 방법을 통해 적응필터의 오조정을 방지할 수 있다. 음성 신호를 사용한 컴퓨터 시뮬레이션 결과를 통해 제안하는 방법이 후처리를 사용하지 않은 경우에 비해 잡음 제거 성능(NRR)이 약 3dB 정도 개선되는 것을 확인 할 수 있다.

■ 중심어 : □주파수영역독립 성분분 석□후처리 □변형된 NLMS□누설(cross-talk) 제거□잡음제거율□

Abstract

The performance of blind source separation(BSS) using independent component analysis(ICA) declines significantly in a reverberant environment. A post-processing method proposed in this paper was designed to remove the residual component precisely. The proposed method used modified NLMS(normalized least mean square) filter in frequency domain, to estimate cross-talk path that causes residual cross-talk components. Residual cross-talk components in one channel is correspond to direct components in another channel. Therefore, we can estimate cross-talk path using another channel input signals from adaptive filter. Step size is normalized by input signal power in conventional NLMS filter, but it is normalized by sum of input signal power and error signal power in modified NLMS filter. By using this method, we can prevent misadjustment of filter weights. The estimated residual cross-talk components are subtracted by non-stationary spectral subtraction. The computer simulation results using speech signals show that the proposed method improves the noise reduction ratio(NRR) by approximately 3dB on conventional FDICA.

■ keyword : □FDICA□Post-Processing□Modified NLMS□Cross-Talk Cancellation□Noise Reduction Ratio□

* 본 논문은 2004학년도 동의대학교 교내연구비에 의하여 연구되었습니다. 2004AA167

접수번호 : #081113-005

심사완료일 : 2008년 12월 18일

접수일자 : 2008년 11월 13일

교신저자 : 김현태, e-mail : htaekim@deu.ac.kr

I. 서론

암묵 신호 분리 기법은 관측된 혼합신호를 이용하여 본래의 독립된 음원 신호를 추정하는 기술이다. 독립성 분석(ICA, Independent component analysis)은 대표적인 암묵 신호 분리 기법으로 특히 신호들이 선형으로 중첩되어 섞인 선형 혼합(instantaneous mixture)인 경우에 효과적이다[1-3]. 그러나 잔향이 존재하는 환경에서는 독립 성분 분석을 이용한 암묵 신호 분리 기법의 성능은 현저히 저하된다[4-6]. 특히, 임펄스 응답을 사용하여 독립성분분석으로 분리한 시스템은 상대 채널 신호(jammer signal)의 직접음(direct sound) 성분을 제거할 수 있어도 잔향을 제거할 수 없다는 최근 연구도 있다 [7]. 이처럼 대부분의 실제 환경은 잔향이 존재하는 나선형 혼합(convolutive mixture)이며 이러한 환경에서는 독립성분분석만으로는 만족할 만한 성능을 기대할 수 없다. 이에 대한 대안으로 후처리 기법이 제안되었다. Mikai 는 주파수영역 독립성분분석을 이용한 분리 시스템의 출력에 NLMS 필터를 이용하여 잔류 누설 경로 성분을 추정하는 후처리 방법을 제안하였다[8]. 이 방법은 누설 경로 성분을 추정하는 필터 입력 신호에 직접 경로 성분이 포함되어있어 이 성분이 잡음과 같은 역할을 하기 때문에 필터의 적응 상수를 직접 경로 성분의 전력값으로 정규화시켜 성능 저하를 방지한다. 그러나 필터 계수의 갱신 과정에 포함되는 오차신호에는 상대방 직접 경로 성분과 주변 잡음 성분 등이 포함되며 이 성분이 진동하게 되면 필터 계수값도 함께 진동하여 최종적으로는 성능 저하를 가져오는 단점을 지닌다. 특히 달리는 자동차 실내와 같이 주변 잡음 성분이 강한 환경에서는 더욱 성능이 저하된다.

본 논문에서는 오차 신호의 전력값을 입력 신호 전력값과 더한 값으로 적응 상수를 정규화하여 주변 잡음과 상대방 직접 경로 성분에 강건한 특성을 보이는 새로운 방법을 제안한다. 제안한 후처리 알고리즘을 적용함으로써 기존의 NLMS 적응필터에서 발생하는 직접성분에 의한 계수 오조정을 감소시킴으로써 잔류 성분 제거 성능을 향상시킬 수 있다. 제안하는 방법이 주변 잡음과 상대방 경로의 직접 성분에 대해 강건한 특성을 보이는 것

을 컴퓨터 시뮬레이션을 통해 확인한다.

II. 주파수영역 독립성분분석에서의 후처리

1. 주파수영역 독립성분분석을 이용한 암묵신호분리

서로 다른 음원 신호가 \leq 알 때, 개의 마이크로부터 관측되는 신호는 \leq 었다. 그리고 주파수영역 독립성분분석으로 분리된 신호를 \leq 라고 할 때 암묵신호분리 모델에서는 다음 식으로 들 수 있다.

$$(1)$$

$$(2)$$

식 (1)에서 는 음원 에서 마이크 까지의 임펄스 응답이며, 마이크 에서 관측되는 신호 는 서로 다른 위치의 음원 신호들과 마이크까지의 경로에 해당하는 임펄스 응답 간의 나선형 혼합으로 주어지는 신호들의 선형 합이다. 이 처럼 서로 다른 음원 신호로부터 마이크 관측 신호까지 이르는 과정을 혼합 시스템(mixing system)이라 일컫는다. 그리고 식 (2)에서 는 마이크 에서 관측된 혼합 신호로부터 다시 음원 성분을 분리해내는 필터의 계수를 나타내며 이 때 필터는 보통 FIR 필터로 주어진다. 이 처럼 혼합 신호로부터 원 음원 신호를 분리해내는 과정을 분리 시스템(unmixing system)이라 일컫는다. 그리고 식 (1) 과 (2)에서 사용한 * 기호는 콘볼루션(convolution) 연산을 의미한다.

상기 식 (1) 과 (2)에서 주어진 것과 같이 시간 영역의 나선형 혼합은 주파수 영역에서는 선형 혼합에 대응된다. 따라서 잔향이 존재하는 환경에서의 암묵신호분리 문제 해결을 위해 시간 영역 독립성분분석 알고리즘을 주파수 영역으로 옮길 필요가 있다. 식 (1)~(2)에 단구간 이산 푸리에 변환을 적용시켜 식 (3)~(4)를 얻는다.

여기서

는 주파수 영역의 원 음원 신호에 대한 추정된 신호 벡터이며, 는 분리 행렬이다 때 번째 프레임에 대해 주파수 영역에서 관측된 신호 는 각각 독립적으로 존재하는 음원 신호들의 선형 혼합으로 가정할 수 있어 주어진 관측 신호를 이용하여 와 가상 신호 독립이 되게 하는 를 결정하는 방법이 주파수 영역 독립성분분석이다[9].

혼합 시스템과 분리 시스템을 연이은 시스템을 라 두면, 이며 암묵신호분리를 통해 분리된 신호 는 식 (5)와 같다.

(3) 스템의 출력에 NLMS 필터를 이용하여 잔류 누설 경로

(4) 이 방법은 누설 경로 성분을 추정하는 필터 입력 신호에 직접 경로 성분이 포함되어있어 이 성분이 일종의 잡음과 같은 역할을 하기 때문에 필터의 적응 상수를 직접 경로 성분의 전력값으로 정규화시켜 성능 저하를 방지한다.

(9)

(10)

(11)

여기서,

(5)

는 필터 탭 입력 벡터이고 는 추정 오차이다. 그러나 필터 계수의 갱신과정에 포함되는 오차신호에는 상대방 직접 경로 성분과 주변 잡음 성분 등이 포함되며 이 성분이 진동하게 되면 필터 계수 값도 함께 진동하여 최종적으로는 성능 저하를 가져올 수 있다. 특히 달리는 자동차 실내와 같이 주변 잡음 성분이 강한 환경에서는 더욱 문제가 된다.

여기서 를 원 음원 부터 유래된 성분, 즉 직계 성분과 를 제외한 다른 음원으로부터 유래된 성분, 즉 누설(cross-talk) 성분으로 나누면 식 (6)과 같다.

(6)

식 (6)과 같이 나눌 때 식 (5)에서 는 직계 경로에 해당하고 는 누설 경로에 해당한다.

(7)

(8)

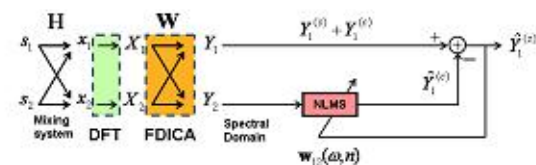


그림 1. NLMS를 이용한 후처리 방법

III. 제안하는 주변 잡음에 강건한 후처리 방법

2. NLMS 필터를 이용한 주파수영역 후처리

Mikai 는 주파수영역 독립성분분석을 이용한 분리 시

2장에서 언급한 바와 같이 특히 달리는 자동차 실내와 같은 환경에서는 독립성분분석을 사용한 암묵신호분리의 성능은 잔향 환경에서 심하게 저하된다. 이 문제를 해

결하기 위해 NLMS 필터계수의 갱신식을 변형한 새로운 알고리즘을 제안한다. 제안하는 알고리즘에서 적응상수는 필터 입력신호의 전력과 추정 오차 전력의 합으로 정규화된다. 이를 통해 추정 오차에 포함되는 상대 채널 신호의 직접음 성분이 큰 경우에도 적응 필터 계수의 오조정을 줄일 수 있으며 또한 주변 잡음이 큰 환경에서도 필터 계수의 오조정을 줄일 수 있다. 제안하는 알고리즘의 적응 필터 계수의 갱신 절차는 식 (12) 및 식(13)과 같다.

$$\begin{aligned} & \parallel \\ & \parallel \\ & \parallel \end{aligned} \tag{12}$$

$$\tag{13}$$

$$\tag{14}$$

여기서,

는 필터 입력신호이고 는 추정 오차이다. 그리고 는 적응 상수이며 가 매우 작은 값을 가질 때 분모가 불안정해지는 것을 막기 위해 미소 값 를 둔다. 누설 경로 추정 필터를 적용하여 잔류 누설 성분의 추정치를 식 (14)와 같이 얻는다.

이 값을 이용하여 스펙트럼 차감법을 통해 잔류 누설을 제거하고 직계 성분을 획득한다. [그림 2]와 [그림 3]은 일 때 제안하는 방법에 대한 블록도이다.

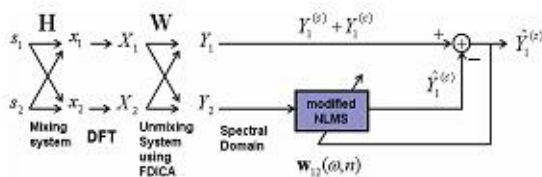


그림 2. 제안하는 후처리 방법

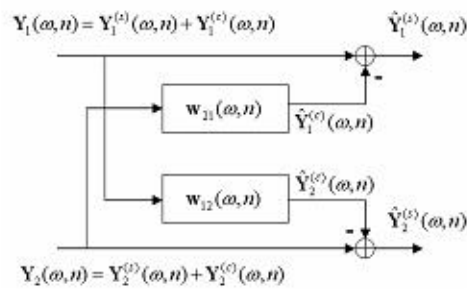


그림 3. 제안하는 후처리 방법 상세도

IV. 실험 결과 및 검토

주변 잡음에 유입되는 달리는 자동차 실내를 대상으로 인 경우에 대해 실험하였다. 마이크 센서의 위치는 운전석 위의 선바이저 중간에 한 개를 배치하고 나머지 한 개는 룸 미러 중간에 배치하여 운전석과 조수석으로부터 각각의 마이크 센서에 이르는 임펄스 응답을 측정하였다(그림 4)(그림 5). 유입되는 잡음 측정 환경은 중형 자동차 실내에서 에어컨 off, 창문 닫은 상태에서 아스팔트 노면을 평균 시속 50 km 시내 주행 시에 측정하였다.

사용한 음원으로는 남성과 여성 각 1명이 서로 다른 문장을 발음하여 한 조의 원 음성 데이터로 사용하며 전체적으로는 3명의 남성과 3명의 여성이 서로 조를 달리하고 3가지 조합의 문장을 발음하여 총 12 조의 원 음성 데이터를 이용하였다. 각 음성은 16비트 양자화수와 16kHz의 샘플링 주파수를 가지며 단구간 이산 푸리에 변환의 프레임 길이는 1024 탭으로 두었고, 프레임 중첩은 64 탭, 창함수는 해밍창을 사용한다. 필터 갱신 때 사용한 적응상수는 0.2로 두었다. 누설 경로 추정을 위한 필터의 길이는 NLMS 및 제안하는 방법 모두 16 탭으로 두었다. 성능 비교에 사용한 척도는 잡음 제거 성능(NRR)를 사용하였으며 출력 신호대 잡음비(SNR)에서 입력 신호대 잡음비의 차를 dB로 나타낸 값으로 식 (15)와 같다[3].

$$\tag{15}$$

- (16) 내며 제안하는 방법에 의해 누설 성분이 줄어든 것을 확인할 수 있다.

- (17) 그림 7은 자동차 소음 환경에서 12가지 음성 신호 조합의 전체 주파수 성분에 대하여 잡음제거비(NRR)를 구한 결과이다. 제안한 방법의 경우 후처리를 적용하지 않은 FDICA와 비교했을 때 약 3~5 dB이 개선됨을 보여준다. 12가지 음성 신호 조합 중에서 기존의 NLMS 알고리즘으로 후처리한 결과가 다소 성능이 좋게 나타나는 경우는 제안하는 알고리즘이 NLMS 알고리즘에 비하여 다소 수렴속도가 느리기 때문에 성능이 저하된다. 그러나 전반적으로는 제안하는 알고리즘의 성능이 우수하며 기존의 NLMS 후처리 알고리즘과 비교 했을 때 평균적으로 1~2 dB 향상을 나타낸다.

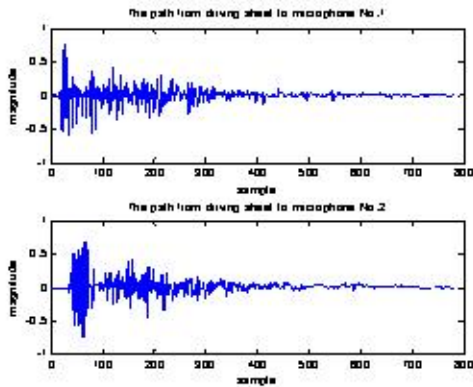


그림 4. 실험에 사용된 임펄스 응답(운전석)

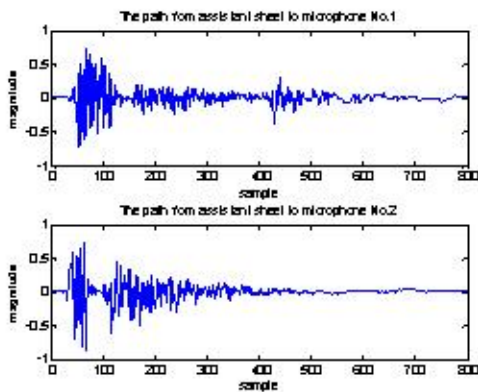


그림 5. 실험에 사용된 임펄스 응답(조수석)

[그림 6]은 암묵 신호분리 시스템에서 협대역 스펙트럼(1,024개의 주파수 bin(bin)에서 144번째 bin(bin)의 샘플에 따른 변화)의 한 예로써 가로축은 샘플 수를 나타낸다. [그림 6]의 (a)와 (b)는 주파수영역 독립성분분석으로 분리한 두 신호의 스펙트럼이다. 제안한 후처리 방법을 적용한 결과는 각각 [그림 6]의 (c)와 (d)에 보였다. [그림 6]에서 원들은 서로 상대 채널에 누설 성분을 나타

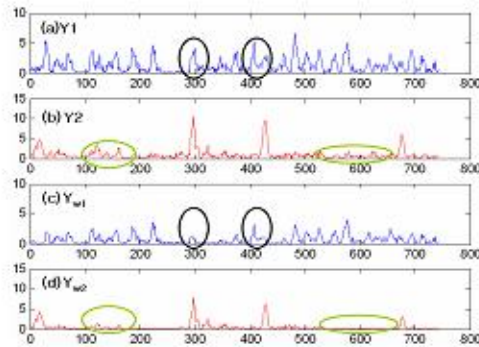


그림 6. 특정 주파수 bin(bin)에서 관찰한 제안하는 후처리 방법 적용 전후의 진폭 스펙트럼 (a) 2채널의 잔류 누설성분이 포함된 1 채널 출력 (b) 1채널의 잔류 누설성분이 포함된 2 채널 출력 (c) 1채널 출력에 대하여 후처리한 결과 (d) 2채널 출력에 대하여 후처리한 결과

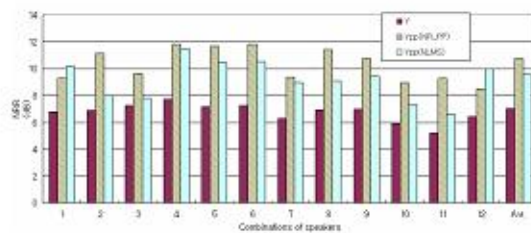


그림 7. 12가지 발음 조합에 대한 NRR 비교(Y: 후처리하지 않은 경우, NR_PP: 제안하는 후처리 방법, NLMS: 기존의 후처리 방법)

V. 결론

독립성분분석을 사용한 암묵신호분리의 성능은 잔향이 존재하는 환경에서 현저히 저하된다. 이러한 성능 저하는 잔향에 포함되어 있는 잔류 누설 성분에 주로 기인한다. 잔류 누설 성분을 제거하기 위하여 기존의 NLMS 필터를 이용한 방법은 추정오차 신호 즉, 후처리 신호 성분에 의하여 NLMS 필터의 계수가 오조정되어 잔류 성분 제거 성능이 떨어진다.

본 논문에서는 잔류 성분을 보다 정교하게 제거하기 위한 후처리 방법을 제안하였다. 제안하는 방법은 누설 경로 추정을 위해 필터의 적응 상수를 필터 입력 신호의 전력과 추정 오차 신호의 전력을 더한 값으로 정규화한다. 제안하는 후처리 방법은 적응필터의 계수 오조정 감소되어 주변 잡음에 강건한 특성을 지닌다. 자동차 실내 처럼 주변 잡음이 유입되는 환경에서 다양한 음성 발음 조합의 신호를 사용한 컴퓨터 시뮬레이션 결과를 통해 제안하는 방법이 후처리를 사용하지 않은 경우에 비해 잡음 제거 성능(NRR)이 약 3dB 정도 개선되었다.

참고 문헌

- [1] A. Hyvarinen, J. Karhnen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.
- [2] S. Haykin, *Unsupervised Adaptive Filtering*, John Wiley & Sons, 2000.
- [3] S. Mäkelä, T. Jung, A. J. Bell, D. Ggahremani, and T. J. Sejnowski, "Blind separation of auditory event-related brain response into independent components," *Proceedings on National Academic Science, USA*, pp.10979-10984, 1997.
- [4] S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, "Fundamental limitation of frequency domain blind separation for convolutive mixture of speech," *Proceedings on ICASSP 2001*, pp.2737-2740, 2001.

- [5] M. Z. Ikram and D. R. Morgan, "Exploring permutation inconsistency in blind separation of speech signals in a reverberant environment," *Proceedings on ICASSP 2000*, pp.1041-1044, 2000.
- [6] S. Araki, A. Blin, R. Mukai, S. Sawada, and S. Makino, "Underdetermined blind separation for speech in real environments with sparseness and ICA," *Proceedings on ICASSP 2004, Vol.3*, pp.17-21, pp.881-884, 2004.
- [7] R. Mukai, S. Araki, and S. Makino, "Separation and dereverberation performance of frequency domain blind source separation," *Proceedings on Int. Conf. on Independent Component Analysis and Blind Signal Separation(ICA2001)*, pp.230-235, 2001.
- [8] R. Mukai, S. Araki, H. Sawada, and S. Makino, "Removal of residual cross-talk components in blind source separation using LMS filters," *Proceedings on ICASSP 2002*, pp.435-444, 2002.
- [9] S. I. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," *Advances in Neural Information Processing Systems 8*, MIT Press, Cambridge MA, 1996.

저자 소개

김 현 태 (Hyun-Tae Kim)

정회원



- 1989년 2월 : 부산대학교 전자공학과(공학사)
- 1995년 2월 : 부산대학교 전자공학과(공학석사)
- 2000년 2월 : 부산대학교 전자공학과(공학박사)

• 2002년 3월 ~ 현재 : 동의대학교 멀티미디어공학과 조교수

<관심분야> : 음성 및 음향신호처리, 입체음향

박 장 식(Jang-Sik Park)

정회원



- 1992년 2월 : 부산대학교 전자공학
학과(공학사)
- 1994년 2월 : 부산대학교 전자공학
학과(공학석사)
- 1999년 2월 : 부산대학교 전자공학
학과(공학박사)

•1997년 3월 ~ 현재 : 동의공업대학 영상정보과 부
교수

<관심분야> : 음성 및 음향신호처리, 멀티미디어통신,
입체음향