

반향환경에서의 가변 적응 상수를 이용한 FDICA와 여러 BSS 알고리즘과의 비교

Comparison for the variable step-size FDICA with
BSS algorithm in reverberant condition

박근수, 박장식*, 손경식
부산대학교, 동의과학대학*

Park Keun-Soo, Park Jang-Sik*, Son Kyung-Sik
Pusan National Univ.,
Donggeui Institute of Technology*

요약

본 논문은 기존의 주파수 영역 독립 성분 분석 (FDICA : frequency domain independent component analysis) 방법에 가변 적응 상수를 이용한 알고리즘을 제안한다. 여러 반향 환경에 대한 모의실험으로 기존의 temporal structure(TA)알고리즘, FDICA알고리즘과 비교하여 그 분리 성능을 평가하고 비교한다. 실제 녹음한 혼합 음원에 대한 분리정도를 비교 분석한다.

Abstract

This paper proposes a variable step size parameter method in frequency domain ICA (FDICA). The FDICA and the temporal analysis (TA) algorithm are experimented for blind source separation (BSS). This paper will compare the separation qualities of these two algorithms in various reverberation environments. Furthermore, it is shown that the proposed technique has the better separation performance than those of two methods especially in recorded data.

I. 서론

디지털 신호처리 기술의 발전으로 음성통신 및 음성인식 시스템을 활용한 다양한 음성 서비스들이 제공되고 있지만 보다 효과적으로 시스템에 적용하기 위해서는 성능 개선이 필요하다. 음성통신에 있어 음질을 향상시키고 음성인식률을 개선하기 위해서는 원하는 신호와 원하지 않는 주변잡음 등의 신호를 분리하는 기술이 필수적으로 요구된다. 최근 이러한 통신 시스템, 생물의학계 그리고 각종 A/V 시스템에서는 다중 센서에서 관찰된 수신 신호만으로 일반적으로 알 수 없는 음원/신호원을 추정하는 블라인드 음원 분리 (blind source separation : BSS)에 관한 음

용 및 연구가 활발하다[1].

BSS를 위해 많이 이용하는 방법 중 하나는 N 개의 통계적으로 독립인 입력들이 각각 독립적인 채널을 거쳐 M 개의 다른 출력으로 혼합된 신호를 음원에 대한 분포도나 사전정보 없이 음원을 추정하는 것이다. 이 방법을 독립 성분 분석 (independent component analysis : ICA)이라 한다. 이러한 ICA에는 비선형(non-linear) 또는 시간 지연된(time-delayed) 비상관관계(decorrelation) 등을 이용한 방법들이 있다[2].

비선형 함수를 이용한 고차 통계 특성을 사용하는 시간 영역에서의 ICA 처리 기법은 선형 혼합

(instantaneous mixture)에 대한 분리 성능은 우수하지만, 일반적인 방(room)에서의 음원의 콘볼루션 혼합(convolutive mixture)에 관한 BSS를 수행할 때, 그 분리 성능이 저하된다. 이러한 문제를 해결하기 위해 주파수 영역에서의 ICA (frequency domain ICA : FDICA) 기법이 제안되었다. 일정한 시간 단위의 블록마다 DFT를 이용한 주파수영역에서의 처리는 각 주파수 빈(bin)에서의 선형적인 혼합으로 가정할 수 있다. 이 방법은 그 알고리즘이 간단하고 각 주파수마다 독립적인 분리를 수행하는 장점이 있다. 그러나 각 주파수마다의 독립적인 분리 수행은 크기 (scaling) 문제와 순열(permutation) 문제를 발생시키므로 이를 해결하기 위한 여러 알고리즘들이 제안되었다[3].

시간 지연된 비상관관계를 이용한 독립 요소 분석 방법은 앞선 비선형 함수를 이용한 고차 통계 특성을 사용하는 것이 아니라, 입력 신호의 시간적인 구조(temporal structure)를 사용한다. 이 알고리즘을 Temporal Analysis 기법이라 정의한다. 음성 신호는 주로 10~20msec의 시간을 기준으로 긴 시간동안은 비정적(nonstationary)이지만, 짧은 시간동안은 정적(stationary)인 시간 구조를 가진다고 가정할 수 있다. 이러한 특성을 사용하여 상관 함수와 같은 이차 통계 특성만을 이용한 BSS가 가능하며, Murata와 Ikeda는 상관 함수를 이용한 BSS를 시간-주파수 영역에서 해석하여 우수한 결과를 나타냈다 [5].

본 논문에서는 FDICA와 Murata와 Ikeda[9]에 의해 제시된 Temporal Analysis 기법의 BSS 방법을 비교 분석한다. 그리고 FDICA의 적용이 각 주파수마다 동등한 조건으로 수립할 수 있도록, 입력 신호의 전력에 따른 정규화[5] 처리 기법을 제안하고 비교, 분석한다. 모의 데이터와 실제 녹음된 데이터를 이용한 실험 결과, 제안하는 알고리즘의 기존의 방법에 비해 우수한 분리 성능을 나타내고 있음을 보여준다.

II. 주파수 영역 독립 성분 분석 (FDICA)

N개의 통계적으로 독립인 음원들이 각각 독립적인 채널을 거쳐 M 개의 다른 마이크로 혼합 입력된 신호의 주파수 영역에서의 모델링을 다음과 같이 한다.

$$\mathbf{X}(w, \tau) = \mathbf{H}(w) \mathbf{S}(w, \tau), \quad (1)$$

여기서 w 는 각 주파수 빈 성분들, τ 는 프레임을 나타낸다.

그리고, $\mathbf{S}(w, \tau) = [S_1(w, \tau), \dots, S_N(w, \tau)]^T$ 는 각 주파수에서의 음원 신호를 나타내며, $\mathbf{X}(w, \tau) = [X_1(w, \tau), \dots, X_M(w, \tau)]^T$ 는 마이크의 수신 신호를, $\mathbf{H}(w)$ 는 혼합 행렬을 각각 나타낸다. 각 주파수에서의 분리 과정은 다음과 같다.

$$\mathbf{Y}(w, \tau) = \mathbf{W}(w) \mathbf{X}(w, \tau), \quad (2)$$

여기서 $\mathbf{Y}(w, \tau) = [Y_1(w, \tau), \dots, Y_L(w, \tau)]^T$ 는 추정된 신호를 표현하며, $\mathbf{W}(w)$ 는 $Y_i(w, \tau)$ 의 독립성을 이용하여 얻은 분리 행렬 (또는 필터)를 나타낸다. 앞으로 수식적 간략화를 위해 w 와 τ 를 생략해 표현한다. 다음은 주파수 영역에서의 분리 행렬을 구할 때 가장 널리 이용되는 Amari의 갱신 알고리즘이다 [2].

$$\begin{aligned} \Delta \mathbf{W} &= \mu [\mathbf{I} - \Phi(\mathbf{Y}) \mathbf{Y}^H] \mathbf{W} \\ \Phi(Y_i) &= \tanh [re(Y_i)] + j \tanh [im(Y_i)] \end{aligned} \quad (3)$$

여기서, \mathbf{Y}^H 는 \mathbf{Y} 의 공액 전치 행렬을, 그리고 $re[Y_i]$ 와 $im[Y_i]$ 는 각각 \mathbf{Y} 의 실수 부분과 허수 부분을 나타낸다. 비선형 함수 $\Phi(Y_i)$, 즉 $\tanh(\cdot)$ 는 각 주파수 빈마다 독립적으로 적용되어 동작한다. 행렬 \mathbf{I} 는 단위행렬을 표시하며, 상수 μ 는 적용 상수

로서 알고리즘의 수렴속도와 안정성을 결정하는 주요한 파라미터이다.

적용 상수는 일정한 값으로 적용되므로, 각 주파수마다 다른 크기로 입력되는 신호에 따라 그 적용 속도가 일정하지 않게 되므로 적용상수를 입력의 전력에 따라 정규화 하여 갱신시키는 방법을 제안한다.

III. Temporal Analysis 알고리즘

시간 지연된 비상관관계를 이용하여 음원을 분리하는 것은 신호의 정적인 구간을 해치지 않는 범위에서의 지연 값의 수만큼의 상관함수를 동시 직각화(simultaneous diagonalization)하여 혼합된 신호들 사이의 상관관계를 없애 주는 것이다. 이러한 방식을 시간 주파수 영역에서 시행하는 것은 시간 영역에서 가지게 되는 초기값 문제에 영향을 받지 않기 위함이다. 시간-주파수 영역에서의 입력 신호의 자기 상관함수를 나타내면 식(4)과 같다.

$$V = \langle X_i(w, \tau) X_j(w, \tau)^T \rangle, \quad (4)$$

$$i, j = 1, 2, \dots, r.$$

여기서 i, j 는 프레임 지표이며, r 는 동시 직각화를 수행할 프레임 개수이다. 이 상관 함수를 이용한 미지 신호 분리는 백색화(whitening or sphering)와 회전화(rotation) 두 가지 과정으로 구성된다. 자기 상관함수를 나타내는 행렬의 고유값 분리에서 얻는 고유벡터로 구성된 고유 행렬 V 를 이용하여 입력 신호를 백색화(sphering)한다. 이런 백색화 과정 이후에도 아직 회전에 관한 애매성(ambiguity)이 남게 된다. 회전의 애매성을 해결하기 위해서는 여러 시간차의 상관 함수 행렬의 비직각(off-diagonal) 성분이 최소값을 가지도록 하는 C 를 구한다.

$$\langle YY^H \rangle = CV \langle XX^H \rangle (CV)^H. \quad (5)$$

기본적으로 이러한 방식은 음원이 WSS(Wide

Sense Stationary)일 때에 사용이 가능하지만 비정적이라고 하더라도, 그 정도가 크지 않다면 상관함수의 적절한 평균값을 취하면 신호의 분리가 가능하다. 이 방식은 고차 통계를 사용하지 않아도 되고 비선형 함수를 최적화하기 위해서 반복적인 계산(iterative calculation)을 해야 하는 비선형 비상관관계를 이용한 FDICA보다 연산이 간단하며 실제 하드웨어 구현의 측면에서도 이점을 지닌다. 그러나, 주파수 영역으로서 변환시 사용하는 윈도우의 길이가 룸 임펄스 응답(room impulse response)의 길이와 밀접한 관계를 가지기 때문에, 방의 환경에 따라 분리의 완성도가 달라진다는 문제점을 가지고 있다.

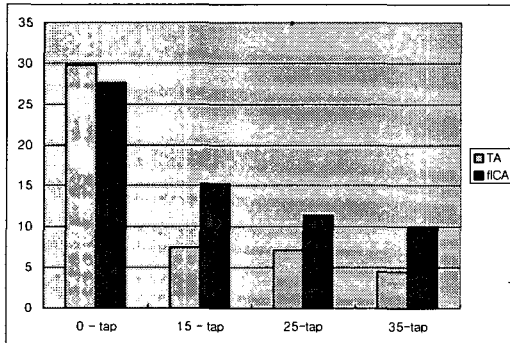
IV. 알고리즘의 비교 분석

모의실험을 양질의 음성 데이터를 임의로 구성한 필터를 통과한 경우와 실제 방에서 녹음한 경우에 대해서 분리 성능을 측정한다. 이 모의 실험은 입력 신호원이 2개, 센서가 2개로 가정한다. 16kHz 샘플링 주파수와 16bit 양자화 된 총 길이는 160,000의 여성, 남성 음성을 사용한다. 실측된 데이터는 일반적인 사무실 환경에서 두 화자간의 거리와 화자와 마이크의 거리는 모두 60cm로 하여 측정된 것이다. 프레임 처리를 위해 256 샘플 길이(16ms)의 Hamming를 사용한다. TA 알고리즘의 동시 직각화를 시행할 행렬의 개수는 30개로 선택했고 윈도우의 이동 시간은 1.25ms로 한다. 혼합 필터의 길이에 따른 신호대 잡음비(Signal-to-Interference ratio : SIR)로서 결과 신호에 대한 성능을 평가한다.

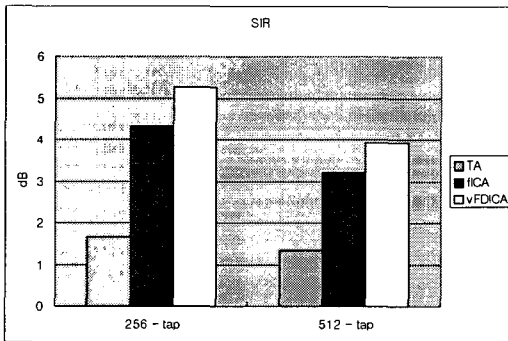
표 1은 FA알고리즘과 FDICA알고리즘의 혼합 필터의 길이에 따른 SIR을 보여 주고 있다. 선형 혼합의 경우(혼합 필터 길이=0), FA 알고리즘은 FDICA 방법보다 그 분리 감도가 우수하며, 계산속도가 빠르다. 그러나 혼합 필터의 길이가 길어짐에 따라, 즉 콘볼루션 혼합에 대해서는 그 분리 감도가 FDICA기 FA보다 우수한 결과를 나타내고 있다. 특히, 음향 잔

향의 환경이 악화될수록 FDICA의 성능이 TA보다 우수한 분리 결과를 나타냄을 표 2에서 볼 수 있다.

[표 1] 채널 필터 길이에 따른 SIR(혼합 필터의 길이: 0,15,25,35탭인 경우)



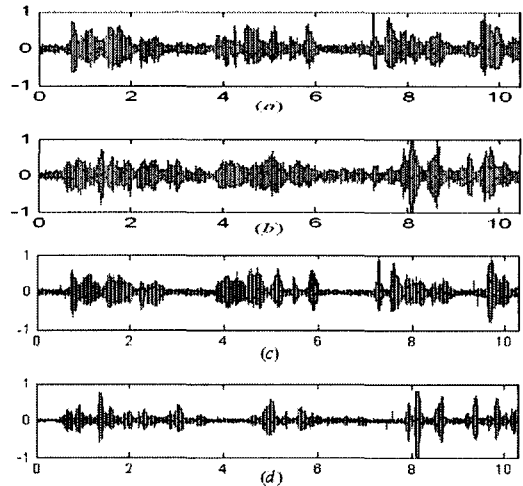
[표 2] 채널 필터 길이에 따른 SIR(혼합 필터의 길이: 256,512탭인 경우)



그리고, SIR 제안하는 가변 적응 상수를 가진 FDICA의 성능이 가장 우수함을 알 수 있다.

그림 1은 일반 사무실에서 실제 녹음한 혼합 신호를 분리한 결과를 나타낸 것이다. 이 실험은 원음에 대한 정보를 얻을 수 없기 때문에 SIR의 수치적 계량화를 보여 줄 수 없기 때문에 그 파형을 나타낸다. (a),(b)는 TA 알고리즘의 추정된 두 원음을, (c),(d)는 FDICA의 두 추정 음원을 나타낸다. FDICA가 잡음에 대한 영향도 적고, 청취 평가에서도 우수함을 보여주었다. 역시 FDICA의 방법이 우수함을 알 수 있다. 이것은 보통 사무실의 환경은 선형혼합이 아닌 콘볼루션 혼

합으로 가정되기 때문이다.



▶▶ 그림 1. (a),(b) TA알고리즘의 분리 결과
(c),(d) FDICA 알고리즘의 분리 결과

V. 결론

본 논문은 FDICA와 FA 알고리즘의 분리 성능을 비교분석하였으며, 제안하는 가변 적응 상수 알고리즘을 FDICA에 적용하여 성능 개선을 보여주고 있다. FA 방법은 그 처리 속도가 빠르고, 선형 혼합의 경우에 FDICA 방법보다 우수한 성능을 나타내지만, 콘볼루션 혼합의 경우에 분리 성능이 FDICA가 더 유리한 것을 알 수 있다. 그리고 각 주파수 성분마다 수렴 속도를 일정하게 해주는 제안하는 가변 적응 상수 알고리즘은 급격한 계산량의 증가 없이도 음원을 분리하는데 기존의 두 방법보다 우수한 성능을 나타내고 있음을 모의 실험과 실제 측정된 데이터에 대한 분리 결과로 보여주고 있다. 그러나, 음향 잔향이 긴 환경에서의 블라인드 기법에 대한 연구가 여전히 요구되고 있다.

■ 참고 문헌 ■

- [1] Bell A. J., Sejnowski T.J., "An Information

- Maximization Approach to Blind Separation and Blind Deconvolution," *Neural Computation*, 7, pp.1129-1159, 1995.
- [2] Amari, S-I Chichoki and HH. Yang, "A New Learning Algorithm for Blind Signal Separation," In *Advances in Neural Information Processing Systems*, 8, MIT Press, Cambridge, MA, 1997.
- [3] Shoji Makino, "Blind Source Separation of Convulsive Mixture of Speech" *Adaptive Signal Processing*, : pp.195-207, August, 2002.
- [4] Noburu Murata, Shiro Ikeda, "An approach to blind source separation based on temporal structure of speech signals" *Neurocomputing* 41 . pp.1-24, 2001.
- [5] Haykin, S., *Adaptive filter theory*, Third edition, Prentice-Hall, 1996.