음성 신호의 주파수 대역별 자기 공분산 기울기 분포

김선일*

Distribution of the Slopes of Autocovariances of Speech Signals in Frequency Bands

Seonil Kim*

요 약

자기 공분산 기울기를 이용하여 음성 신호와 배경 잡음 신호를 구분할 때 구분 가능성을 높이기 위해 주파수 영역에서 음성 신호의 자기 공분산 기울기를 최대화하는 주파수 대역을 찾아내었다. 디지털 샘플링 된 음성 신호를 일정한 개수의 신호로 이루어진 블록으로 나눈 후 각 블록에 고속푸리에변환(Fast Fourier Transform, FFT)을 하여 주파수 영역으로 변환한 다음 임의의 주파수 대역에서 각 블록에서의 공분산을 구하고 이 공분산 값들을 연결하는 직선 근사를 한 후에 이 직선의 기울기를 자기 공분산 기울기로 사용하는데 이 값은 음성 신호의 특성 상 주파수 대역별로 차이가 있다. 따라서 어느 주파수 대역에서 자기 공분산 기울기가 크게 나타나는지 200개의 남성 음성 파일을 이용하여 주파수 대역별로 비교 분석하였다.

ABSTRACT

The frequency bands were discovered which maximize the slopes of autocovariances of speech signals in frequency domain to increase the possibility of segregation between speech signals and background noise signal. A speech signal is divided into blocks which include multiples of sampled data, then those blocks are transformed to frequency domain using Fast Fourier Transform(FFT). To find linear equation by Linear Regression, the coefficients of autocovariance within blocks of some frequency band are used. The slope of the linear equation which is called the slope of autocovariance is varied from band to band according to the characteristics of the speech signal. Using speech signals of a man which consist of 200 files, the coefficients of the slopes of autocovariances are analyzed and compared from band to band.

키워드

자기 공분산, 기울기, ICA, 음성 신호, 직선 근사

Key word

Autocovariance, Slope, ICA, Speech Signal, Linear Regression

* 정회원 : 거제대학교(교신저자, seonil@koje.ac.kr)

접수일자 : 2013. 02. 01
심사완료일자 : 2013. 03. 05

Open Access http://dx.doi.org/10.6109/jkiice.2013.17.5.1076

Ⅰ. 서 론

실시간 음성 인식에는 메인 프레임 컴퓨터 급의 계산 및 처리 능력이 필요하고 이러한 컴퓨터를 휴대하거나 작은 공간에 설치한다는 것이 공간 문제나, 비용 문제로 불가능해서 실시간 활용이 어려웠으나 최근 통신 및 유무선 네트워크 기술의 발전으로 음성 신호를 압축하여 메인 프레임 컴퓨터에 전송하고 메인 프레임 컴퓨터가 방대한 데이터베이스를 활용하여 음성을 인식한 후 그 결과를 전송하는 방식으로 많이 활용되고 있다. 특히 스마트폰 기술이 발전하고 막대한 양이 보급됨에 따라 소비자들의 손끝에서 음성 인식의 열매를 맛보는 시대가 되었다. 하지만 음성 인식은 아직도 잡음에 취약한 것이 사실이다. 음성이 원래 잡음에 취약하므로 사람의 경우에도 배경잡음이 존재하는 상황에서 상대방의 대화 내용을 다 알아 듣기가 쉽지 않다. 하지만 컴퓨터를 사용하면 잡음을 제거하는 기술을 적용하여 사람보다 더 잡음에 강한 인식기를 만들 수도 있다.

음성을 하나의 신호원으로 보고 배경 잡음을 또 다른 신호원으로 본다면 BSS(Blind Source Separation) 기술을 이용할 수 있다[1]. BSS에서는 신호들이 서로 독립적이라는 전제하에 최대한 독립적인 방향으로 분리해 내는 ICA(Independent Component Analysis) [2-4], 신호들의 상관관계가 최소가 되도록 변환하는 CCA (Canonical Correlation Analysis)[5,6] 등이 있다. 하지만 음성 신호와 배경 잡음을 각각 독립된 하나의 신호원으로 보고 앞에서 언급한 방법으로 각 신호를 성공적으로 분리해 내었을 경우에 둘 중의 어느 것이 음성 신호인지 단순히 구분하기가 쉽지 않다. 최근 음성 인식 기능을 자동차에 구현하려는 경향이 증가하면서 자동차 배기음이 배경 잡음으로 존재하는 경우에 대한 연구가 이루어지고 있으며[7-8] 이 경우, 자동차 소음과 음성이 확연히 다른 특성을 보이고 있음을 알 수 있다[9-10].

분리된 두 신호 중 음성 신호를 구분해 내기 위해서는 신호를 일정한 크기의 블록(Block)로 나누고 각각의 블록에 FFT를 적용하여 주파수 성분을 분석한 후 각 주파수 대역별로 전체 신호에 대한 자기 공분산을 구하고 이 공분산 값들을 그래프로 나타내면 각 값들을 연결하는 직선을 그릴 수 있는데 이 직선의 기울기를 이용하여 음성 신호와 배경 잡음을 구분해 낼 수 있다[11].

Ⅱ. 주파수 영역에서의 자기 공분산 기울기

분리된 시간 영역 음성 신호 S 를 FFT가 가능한 $n=2^p$(p는 임의의 자연수)개의 데이터로 구성된 블록으로 나누어 준다. 신호 S 를 행렬로 나타내면 식(1)과 같다. 식(1)과 같은 행렬에서 각 열, 즉 $S_1\ S_2 \cdots\ S_m$이 이 블록에 해당된다. 물론 마지막 열의 데이터 수가 2^p 개에 못 미칠 때에는 0을 넣어서 맞추어 준다. 따라서 총 m 개의 블록이 생기게 된다.

$$S=\begin{bmatrix} s_{1,1} & s_{1,2} \cdots & s_{1,m} \\ s_{2,1} & s_{2,2} \cdots & s_{2,m} \\ \vdots & \vdots & \cdots \ \vdots \\ s_{n,1} & s_{n,2} \cdots & s_{n,m} \end{bmatrix}=[S_1\ S_2 \cdots\ S_m] \tag{1}$$

식(1)에서 i 번 째 행벡터 S_i 는

$$S_i=\left[s_{1,i}\ s_{2,i} \cdots\ s_{n,i}\right]^T \tag{2}$$

이고 i 는 1부터 m 이다.

각 S_i에 대해 FFT를 구하면 식(3)과 같은 결과를 얻을 수 있다. S 는 시간 영역의 신호이지만 F 는 주파수 영역에서 각 블록의 주파수 성분을 나타낸다.

$$F=\begin{bmatrix} f_{1,1} & f_{1,2} \cdots & f_{1,m} \\ f_{2,1} & f_{2,2} \cdots & f_{2,m} \\ \vdots & \vdots & \cdots \ \vdots \\ f_{n,1} & f_{n,2} \cdots & f_{n,m} \end{bmatrix}=[F_1\ F_2 \cdots\ F_m] \tag{3}$$

$$F_i=\left[f_{1,i}\ f_{2,i} \cdots\ f_{n,i}\right]^T \tag{4}$$

이고 i 는 1부터 m이다.

식(3)에서

$$F^j=\left[f_{j,1}\ f_{j,2} \cdots\ f_{j,m}\right] \tag{5}$$

이고 j 는 1부터 n일 때 F는

$$F=\begin{bmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,m} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,m} \\ \vdots & \vdots & \cdots & \vdots \\ f_{n,1} & f_{n,2} & \cdots & f_{n,m} \end{bmatrix}=[F^1\ F^2 \cdots F^m]^T \quad (6)$$

식(6)과 같이 나타낼 수 있는데 F^j는 각 블록의 특정 주파수 대역을 나타낸다.

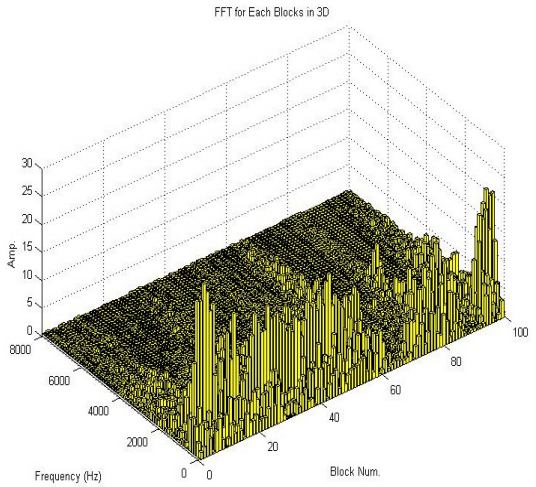


그림 1. F 행렬의 3차원 도표
Fig. 1 Three Dimensional Graph for F Matrix

16㎑로 샘플링한 신호에 대해 3차원 그래프로 F를 그리면 그림 1과 같다.

그림 1의 x축의 Block Num.은 식(3)의 m에 해당된다. y축의 Frequency는 식(3)의 n에 해당되는데 이를 0에서 8㎑의 주파수로 변환하여 나타내었다. z 축은 FFT의 절대값이다. 주로 저주파 쪽에 큰 값이 나타나는 것을 관찰할 수 있다.

행벡터 F^j 각각에 대해 자기 공분산을 구하면 식(7)과 같이 나타낼 수 있다.

$$C=\begin{bmatrix} c_{1,-\tau} & c_{1,-\tau+1} & \cdots & c_{1,0} & \cdots & c_{1,\tau-1} & c_{1,\tau} \\ c_{2,-\tau} & c_{2,-\tau+1} & \cdots & c_{2,0} & \cdots & c_{2,\tau-1} & c_{2,\tau} \\ \vdots & \vdots & & \vdots & & \vdots & \vdots \\ c_{n,-\tau} & c_{n,-\tau+1} & \cdots & c_{n,0} & \cdots & c_{n,\tau-1} & c_{n,\tau} \end{bmatrix} \quad (7)$$

$$=\left[C_1\ C_2\ \cdots\ C_n\right]^T$$

여기서 C_i 는

$$C_i=\left[c_{i,-\tau}\ c_{i,-\tau+1} \cdots c_{i,0} \cdots c_{i,\tau-1}\ c_{i,\tau}\right] \quad (8)$$

이고 i 는 1부터 n까지의 정수이다. 여기서

$$c_{i,\tau}=\sum_{p=1}^{m-|\tau|}(f_{i,p+|\tau|}-\mu)(f_{i,p}-\mu) \quad (9)$$

$$\mu=\frac{1}{m}\sum_{q=1}^{m}f_{i,q} \quad (10)$$

이며 τ=0 일 때 1 이 되도록 정규화 하였다.

τ 는 자기 공분산을 구할 때 데이터가 어긋나는 정도이다. 신호가 주기성이 있다든지 하는 음성일 경우에 τ가 커지면 공분산값이 작아지게 되나 자동차 배기소음과 같은 신호는 τ에 상관없이 일정한 값을 보이는 경향을 갖는다[9][12]. 이 값은 좌우 대칭으로 나타나고 정규화하면 $c_{i,0}$ 는 항상 1이므로 아무런 정보를 주지 못한다. 그래서 식(8)에서 $c_{i,0}$와 그 오른쪽 값들을 제외하면 식(11)과 같이 표현할 수 있다.

$$C'=\begin{bmatrix} c_{1,-\tau} & c_{1,-\tau+1} & \cdots & c_{1,-1} \\ c_{2,-\tau} & c_{2,-\tau+1} & \cdots & c_{2,-1} \\ \vdots & \vdots & \cdots & \vdots \\ c_{n,-\tau} & c_{n,-\tau+1} & \cdots & c_{n,-1} \end{bmatrix}=\left[C_1'\ C_2' \cdots\ C_n'\right]^T \quad (11)$$

여기서 C_i' 는 식(12)와 같다.

$$C_i'=\left[c_{i,-\tau}\ c_{i,-\tau+1} \cdots c_{i,-1}\right] \quad (12)$$

C_i'에 Linear Regression[13]을 적용하면 식(13)과 같은 직선의 방정식에서 기울기 a와 절편 b를 구할 수 있다.

$$y=ax+b \quad (13)$$

그림 2에서 $n=128$, $\tau=10$ 일 때 C_i의 그래프와 C_i'을 이용해 Linear Regression으로 구한 직선의 방정식을 볼 수 있다.

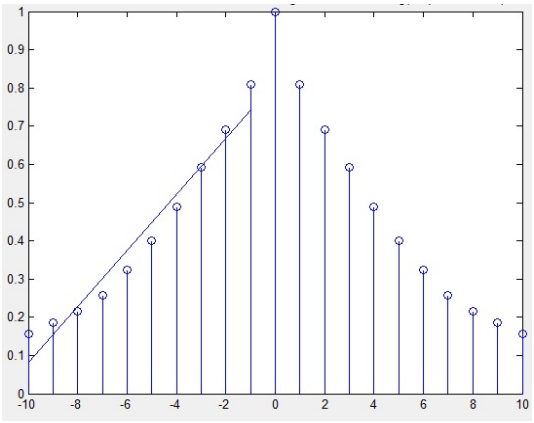


그림 2. 임의 대역에서 자기 공분산 C_i, 그리고 $C_i^{'}$을 이용한 직선 근사, $n=128$
Fig. 2 Autocovariances C_i and Linearly Regressed Line using $C_i^{'}$ at a band, $n=128$

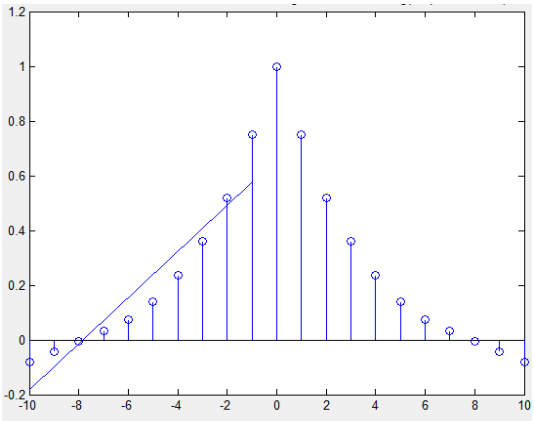


그림 3. 임의 대역에서 자기 공분산 C_i, 그리고 $C_i^{'}$을 이용한 직선 근사, $n=256$
Fig. 3 Autocovariances C_i and Linearly Regressed Line using $C_i^{'}$ at a band, $n=256$

이렇게 구한 기울기 a를 자기 공분산 기울기라고 하자. $n=128$일 경우 음성 신호에서는 τ의 증가에 따라 자기 공분산 계수가 서서히 감소하는데 비해서 자동차 배기음과 같은 잡음에서는 τ의 변화에 따른 자기 공분산 계수의 변화가 미미하다. 따라서 두 종류의 신호 사이에서 자기 공분산 기울기가 중요한 정보를 제공하고 있다.

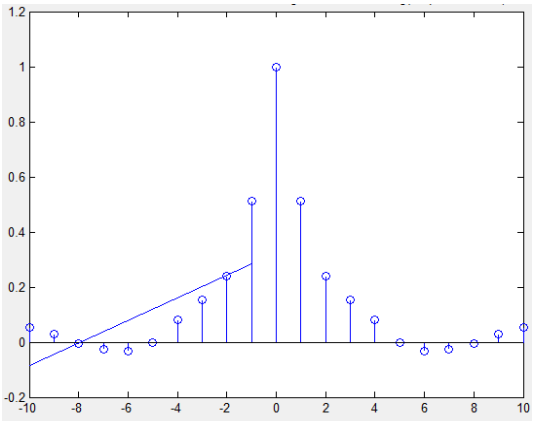


그림 4. 임의 대역에서 자기 공분산 C_i, 그리고 $C_i^{'}$을 이용한 직선 근사, $n=512$
Fig. 4 Autocovariances C_i and Linearly Regressed Line using $C_i^{'}$ at a band, $n=512$

하지만 $n=256$이거나 $n=512$ 일 때는 그림 3과 그림 4처럼 음성 신호의 자기 공분산 분포 형태가 비선형적인 모양을 더 강하게 띠게 된다. 이는 n의 크기가 커질수록, 즉 블록의 크기가 커질수록 블록과 블록 사이의 거리가 멀어지는 셈이므로 블록간의 주파수 특성이 더 많이 달라지기 때문으로 해석할 수 있다. 그래서 자기 공분산 값들의 분포에서 비선형적인 특성이 더 두드러지게 되어 직선 근사를 통한 기울기 값을 유용한 특성으로 사용하기 어렵다. 음성 신호 내의 가까운 데이터끼리의 공분산이 자동차 소음등과 같은 배경 잡음과 큰 차이를 갖는 특성을 이용하고자 할 때 n 이 커지는 것은 분명히 큰 장애물이 됨을 알 수 있다.

Ⅲ. 주파수 대역별 자기 공분산 기울기 분포

음성 신호에서 기대할 수 있는 정보는 낮은 주파수 영역은 성도에 대한 정보를 가지고 있다는 것이다. 따라서 주파수가 낮은 영역에서는 자기 공분산 기울기가 크고 높은 주파수 영역에서는 이 값이 상대적으로 작을 것이라는 예측이 가능하다. 그렇다면 자기 공분산 기울기가 큰 주파수 대역을 찾아 이 대역의 자기 공분산 기울기를 이용해 배경 잡음과 구분하면 성공 가능성이 더 높아질

것이다. 따라서 $n=128$ 이고 $\tau=10$ 일 때 각 주파수 대역에서의 자기 공분산 기울기를 계산하였다. $\tau=10$ 이상으로 해도 계산량만 늘어나고 성능의 개선은 기대할 수 없어 $\tau=10$으로 정하였다[9].

사용된 데이터는 남성 뉴스 앵커의 뉴스 음성신호를 16㎑로 샘플링한 것이다. 총 200개의 음성 파일 중 각 100개씩에 대해 자기 공분산 기울기의 분포를 구하여 그래프로 나타내었다.

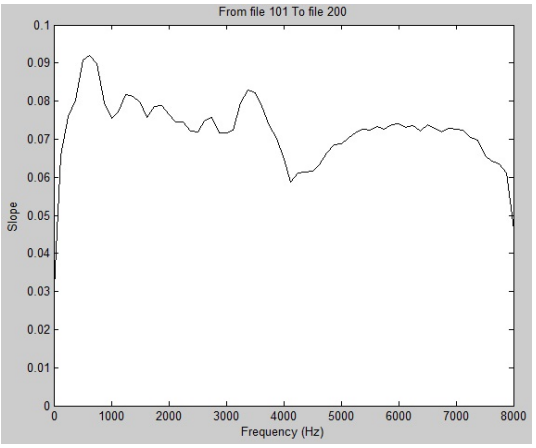


그림 5. 자기 공분산 기울기 분포(그룹 1)
Fig. 5 The Distribution of the slopes of Autocovariances(Group 1)

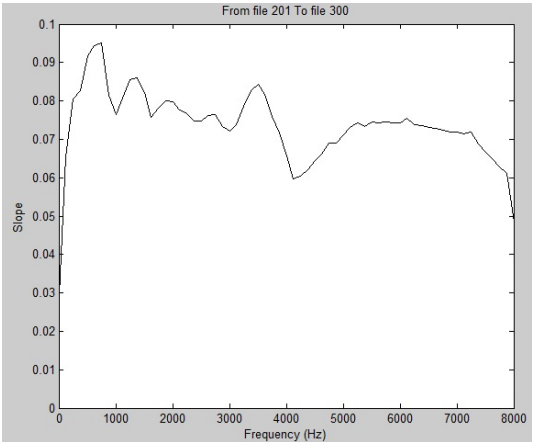


그림 6. 자기 공분산 기울기 분포(그룹 2)
Fig. 6 The Distribution of the slopes of Autocovariances(Group 2)

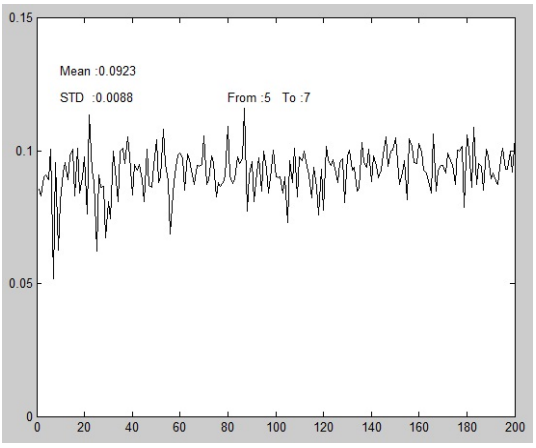


그림 7. 200개 음성 파일에 대한 자기 공분산 기울기(600Hz～900Hz)
Fig. 7 The Slopes of Autocovariances for 200 Speech Signal Files(600Hz～900Hz)

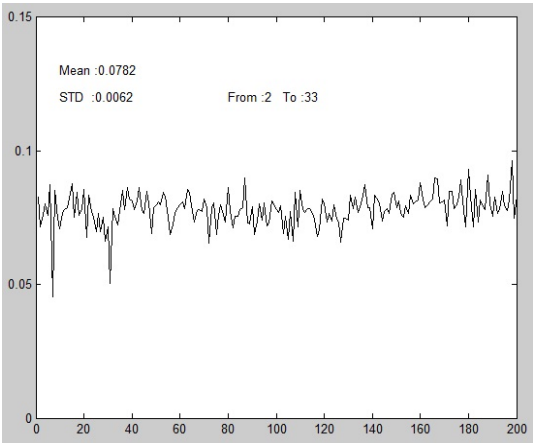


그림 8. 200개 음성 파일에 대한 자기 공분산 기울기(～4㎑)
Fig. 8 The Slopes of Autocovariances for 200 Speech Signal Files(～4㎑)

100개를 한 그룹으로 나타내면 그림 5는 총 두 그룹 중 첫 번째 그룹, 그림 6은 두 번째 그룹에 대한 분포도이다. 두 그룹 다 4㎑ 이내의 대역에서 평균적으로 큰 값을 나타내고 있으며 그 중에서도 600Hz에서 900Hz 대 사이에서 큰 값들이 관찰된다. 이를 확인하기 위해 개별 음성 신호의 각 주파수 대역별 자기 공분산 기울기를 구하고

총 200개의 음성 파일에 대해서 그래프로 도시하였다.

그림 7은 자기 공분산 기울기가 가장 큰 대역인 600 Hz에서 900Hz 사이의 자기 공분산 기울기의 평균을 구한 것이다. 총 200개의 음성 파일을 대상으로 하고 200개 파일의 자기 공분산 기울기의 평균과 표준 편차도 구하였다.

자기 공분산 기울기를 구하는 주파수 구간을 600Hz에서 900Hz로 했을 때는 그림 7과 같고 200개 음성 파일의 평균 자기 공분산 기울기가 0.0923이며 표준 편차는 0.0088이다. 주파수 구간 4㎑ 이내인 그림 8에서는 평균값이 0.0782이고 표준편차가 0.0062이다. 전체 주파수 구간에서는 평균값이 0.0735이고 표준 편차는 0.0051이다. 예상한대로 600Hz에서 900Hz 대역의 자기 공분산 기울기의 평균값이 가장 높고 그 다음이 4㎑ 까지, 전체 구간의 순서이다. 표준 편차는 대상 데이터 수가 많아질수록 줄어드는 경향이 있다. 그림 7의 경우를 보면 자기 공분산 기울기가 가장 작은 값도 0.05보다 큰 값을 유지하지만 그림 8을 보면 나머지 그렇지 못한 것을 관찰할 수 있다.

Ⅳ. 결 론

음성 신호와 배경 잡음을 구분할 때 자기 공분산 기울기를 구하는 주파수 대역에 대한 정확한 분석이 이루어야 어느 대역의 자기 공분산 기울기를 사용하는 것이 타당한지 그 근거를 가질 수 있다. 하지만 이러한 분석 없이 저주파 대역이 기울기가 클 것이라는 가정 하에 대역을 설정해서 사용하였다[12]. 이번 연구를 통해 자기 공분산 기울기가 비교적 큰 영역을 찾아 배경 잡음과의 여유를 더 확보할 수 있게 되어 배경 잡음과 음성 신호를 분리해 낼 때 더 높은 신뢰성을 확보할 수 있게 되었다.

그러나 이번 연구에서 사용된 음성 데이터는 데이터의 양은 많지만 한 사람의 것이다. 사람에 따라 성도의 모양도 다르고 따라서 성도에 대한 정보를 제공하는 저주파 영역도 약간 씩 다를 수 있다. 또한 여자와 남자는 그 구조가 많이 다를 수 있으므로 앞으로 여성을 포함한 좀 더 다양한 음성 데이터를 확보해서 연구해야 할 과제가 남아있다. 또 배경 잡음은 음성 신호보다 더 다양한 형태를 띠게 되므로 다양한 배경 잡음을 가정하고 그에 해당하는 데이터를 확보해서 음성 신호와 배경 잡음 신호를 구분하는데 자기 공분산 기울기가 유용한 수단임을 확고히 증명할 필요가 있을 것이다.

참고문헌

[1] J. F. Cardoso, "Blind signal separation: statistical principles," Proc. IEEE, vol. 9, no. 10, pp. 2009-2-25, Oct., 1988.

[2] A. Hyvarinen and E. Oja, "Independent component analysis: algorithms and applications," Neural Networks, vol. 13, no. 4/5, pp. 411-430, 2000.

[3] A. Hyvarinen, "Fast and Robust Fixed-Point Algorithms for Independent Component Analysis," IEEE Trans. On Neural Networks, vol. 10, no. 2, pp. 626-634, May, 1999.

[4] Pl. Conon, "Independent component analysis, A new concept?," Signal Processing, vol. 36, pp. 287-314, 1994.

[5] W. Liu, D. Mandic, and A. Cichocki, " Analysis and Online Realization of CCA Approach for Blind Source Separation," IEEE Transaction on Neural Networks, Vol. 18, No. 3, September 2007.

[6] 김선일 "정준 상관 분석을 이용한 잡음 섞인 음성 신호의 분리," 한국정보통신학회 종합학술대회논문집, 춘계16권, 1호, pp. 164-167, 동명대학교, 2012.

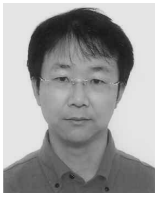
[7] H. Saruwatari, K. Sawai, T. Nishikawa, A. Lee, K. Shikano, A. Kaminuma, M. Sakata and D. Saitoh. "Speech Enhancement Based on Blind Source Separation in Car Environments," Proc. 21st International Conference on Data Engineering." pp. 1205, 05-08 April, 2005.

[8] J. Lee, H, Jung, T. Lee and S. Lee, "SPEECH CODING AND NOISE REDUCTION USING ICA-BASED SPEECH FEATURES," International Workshop on independent component analysis and blind signal separation, pp. 417-422, 19-22 June, 2000, Helsinki, Finland.

[9] 김선일, "주파수 영역 자기 공분산 기울기를 이용한 음성과 자동차 소음 신호의 구분," 한국해양정

보통신학회 논문지, 제15권, 10호, 10월, 2011.
[10] 김선일, “ICA로 분리한 신호의 분류,” 대한전자공학회 논문지, 제47권, IE-4호, 12월, 2010.
[11] 김선일, “음성 및 음성 관련 신호의 주파수 및 Quefrency 영역에서의 자기공분산 변화,” 해양정보통신 종합학술대회논문집, 춘계15권, 1호, pp. 340-343, 대구 EXCO, 2011.
[12] 김선일, 양성룡 “배경 잡음을 제거하는 음성 신호 잡음 제거기의 구현”, 대한전자공학회 논문지, 제49권, IE-2호, pp. 24-29, 6월, 2012.
[13] R. Johnson, K. Tsui, *Statistical Reasoning and Methods,* John Wiley & Sons, Inc. 1998.

저자소개



김선일(Seonil Kim)

1983년 아주대학교 전자공학과 공학사
1985년 아주대학교 전자공학과 공학석사
1996년 아주대학교 전자공학과 공학박사
1985년～1990년 한국기계연구원 선임연구원
1990년～현재 거제대학교 교수
1997년 Visiting Professor, CAIP Center, Rutgers Univ., Piscataway, N.J., USA
2007년 Visiting Professor, Dept. of ECE, Georgia Institute of Technology, Atlanta, G.A., USA
※관심분야 : 신호 처리, 음성 인식