

음성 신호의 주파수 영역에서의 주파수 대역별 공분산 행렬의 고유값 분석

김선일*

*거제대학교

Analysis of Eigenvalues of Covariance Matrices of Speech Signals in Frequency Domain for Various Bands

Seonil Kim*

*Koje College

E-mail : seonil@koje.ac.kr

요 약

음성 신호는 자음 신호와 모음 신호의 결합으로 이루어져 있지만 그 특성상 자음보다는 모음 신호의 지속시간이 길다. 따라서 전체적으로 음성 신호 블록들 사이의 상관관계가 상당히 크다고 간주할 수 있다. 하지만 같은 음성 신호 내에서도 주파수 대역별로 그 상관관계가 다르게 나타난다. 음성신호를 128개의 데이터를 갖는 블록들로 나눈 후 각 블록의 FFT를 구한다. 여러 주파수 대역별 FFT 값으로 부터 이웃 블록들과의 공분산 행렬을 구하고 이 행렬로부터 고유값을 계산해 낸다. 이중 첫 번째 고유값은 주성분과 관련이 있다. 다양한 주파수 대역별로 주성분을 구한 후 이 주성분의 값들이 대역별로 어떻게 나타나는지 그 분포를 알아보고 어떤 대역의 공분산 행렬의 고유값을 선택해야 더 안정적인 결과를 얻을 수 있을지 분석한다.

ABSTRACT

Speech Signals consist of signals of consonants and vowels, but the lasting time of vowels is much longer than that of consonants. It can be assumed that the correlations between signal blocks in speech signal is very high. But the correlations between signal blocks in various frequency bands can be quite different. Each speech signal is divided into blocks which have 128 speech data. FFT is applied to each block. Various frequency areas of the results of FFT are taken and Covariance matrix between blocks in a speech signal is extracted and finally eigenvalues of those matrix are obtained. It is studied that in the eigenvalues of various frequency bands which band can be used to get more reliable result.

키워드

음성 신호, 상관관계, 고유값, 공분산, 주성분

I. 서 론

음성 인식 기술의 진보는 이제 일상 생활에도 파고 들어서 많은 사람들이 사용하고 있는 손 안의 컴퓨터, 스마트폰에서도 음성 인식 기능은 기본 기능으로 자리 잡았다.

음성 인식은 기계가 아닌 인간이어도 주변 환경에 많은 영향을 받는다. 주변에 잡음이 많으면 그 만큼 인식에 어려움을 겪고 사람의 경우에는

뭐라고 했는지 다시 물어봐야 한다.

하지만 기계에서는 다시 물어보기가 쉬운 일이 아니다. 따라서 주변 잡음이 존재할 때 이를 음성 신호와 구별해 내는 전처리 과정이 꼭 필요하다.

음성 이외의 모든 소리나 자기 음성 이외의 다른 음성을 모두 잡음으로 간주할 수 있으나 잡음의 범위를 자동차 소음이나 자동차 엔진 소음으로 단순화하게 되면 음성 신호를 자동차 엔진 신호와 구분하는 문제가 된다.

자동차 엔진 소음과 섞인 음성 신호는

ICA[1,2,3]와 같은 통계적 방법을 이용해 분리해 낼 수 있다. 하지만 분리해 낸 신호 중 어느 것이 음성 신호이고 어느 것이 자동차 소음인지 가려내어야 한다.

분리된 신호로부터 음성 신호를 가려내기 위해서는 특정 신호와의 상관계수를 이용하는 방법 [4], 각 신호를 블록으로 나누어 각 블록의 자기 공분산 값들을 연결하는 직선을 구해 이 직선의 기울기 값으로 구별하는 방법[5,6,7]과 주성분 분석[8]을 통해 구별하는 방법[9,10]이 제시되어 있다.

주성분 분석을 이용한 방법[9,10]에서 신호의 저주파 부분만 사용하여 주성분 분석을 하였으나 이에 대한 근거를 제시하지 않고 음성 신호의 저주파 부분이 모음 성분이 몰려 있는 주파수 대역이라 주성분 값이 클 것으로 예상하고 주파수가 낮은 구간의 주성분만 사용한 것에 대한 근거를 제시하였다.

따라서 주파수 대역을 크게 6개 구간으로 나누고 이 구간별로 음성 신호의 각 블록간 공분산 행렬의 첫 번째 고유값(신호의 첫 번째 주성분)의 변화를 연구하였다.

FFT 변환하는 방법과 주성분 분석에 대한 이론은 참고문헌 [5,6]과 동일하게 사용하였고 제안 및 검증에서 주파수 대역별 주성분 변화를 제시하였다.

II. 신호의 FFT 방법

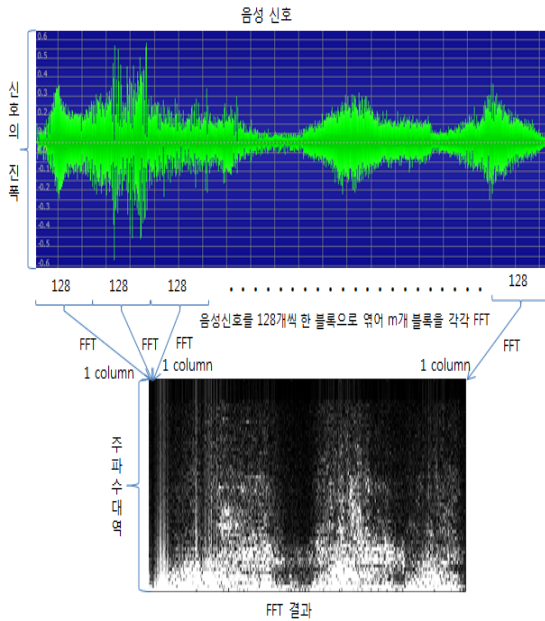


그림 1. 음성신호의 각 블록을 FFT 방법

그림 1과 같이 시간 영역에서 신호 S 를 FFT(Fast Fourier Transform)가 가능한 $n = 2^p$ (p 는

임의의 자연수)개의 데이터로 구성된 m 개의 블록으로 나누어 준 후 각 블록에 대해 FFT를 계산한다.[3]

III. 주성분 분석

가중 행렬

$$W = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,m} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,m} \\ \vdots & \vdots & \cdots & \vdots \\ w_{n,1} & w_{n,2} & \cdots & w_{n,m} \end{bmatrix} = [W_1 \ W_2 \ \cdots \ W_m] \quad (1)$$

에서 $W_i = [w_{1,i} \ w_{2,i} \ \cdots \ w_{n,i}]^T$ 이고 이를 가중 벡터라 한다.

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,m} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,m} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,m} \end{bmatrix} = [X_1 \ X_2 \ \cdots \ X_m] \quad (2)$$

에서 $X_i = [x_{1,i} \ x_{2,i} \ \cdots \ x_{n,i}]^T$

일 때

$$Y_1 = W_1^T X \quad (3)$$

j 는 1부터 m 까지 이다.

그런데

$$Y = \begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,m} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,m} \\ \vdots & \vdots & \cdots & \vdots \\ y_{n,1} & y_{n,2} & \cdots & y_{n,m} \end{bmatrix} = [Y_1 \ Y_2 \ \cdots \ Y_n]^T \quad (4)$$

에서 $Y_i = [y_{i,1} \ y_{i,2} \ \cdots \ y_{i,m}]$

Y_1 의 공분산이 최대이면 이 Y_1 을 X 의 첫 번째 주성분이라고 한다.

공분산은 가중치 벡터 W_1 의 norm과 방향에 따라 달라지고 W_1 의 norm이 커짐에 따라 무한정 커지게 된다. 따라서 W_1 의 norm을 일정하게 해야 할 필요가 있는데 일반적으로 norm을 1로 한다. 여기서 norm은 다음과 같이 Euclidean 거리로 정의한다.

$$\|W_1\| = (W_1^T W_1)^{\frac{1}{2}} = \left[\sum_{k=1}^n w_{k,1}^2 \right]^{\frac{1}{2}} \quad (5)$$

이런 조건을 만족시키는 가중치 벡터 W_1 을 구하면

$$E\{Y_1^2\} = E\{(W_1^T X)^2\} = W_1^T E\{XX^T\} W_1 \quad (6)$$

$$= W_1^T C_X W_1$$

여기서 $\|W_1\| = 1$

이고

X 의 평균이 0 이라면

$$C_X = E\{XX^T\} \quad (7)$$

이다.

이 식의 주성분 분석 해는 행렬식 C_X 의 단위 길이 고유벡터 e_1, \dots, e_n 이라는 사실은 잘 알려져 있다. 따라서 고유벡터의 고유값들이

$$d_1 \geq d_2 \geq \dots \geq d_n \text{ 이라면}$$

$W_1 = e_1$ 이다.

따라서

$$E\{Y_1^2\} = E\{(e_1^T X)^2\} = e_1^T E\{XX^T\} e_1 \quad (8)$$

$$= e_1^T C_X e_1 = e_1^T d_1 e_1 = d_1$$

이다.

즉 Y 의 분산은 X 의 분산의 고유값이다.

IV. 제안 및 검증

16kHz로 샘플링한 음성 신호를 이용하여 128개의 신호를 하나의 블록으로 잡아 FFT를 수행하였다. 따라서 신호의 지속 시간에 따라 블록 전체의 수가 달라진다. 자동차 엔진 소음 신호는 급가속 소음, 일반 가속 소음, 정속 주행 엔진 소음 신호를 사용하였다. 음성 신호로는 총 81개 방송 뉴스를 사용하였다.

음성신호를 서로 겹치지 않는 블록으로 나누고 각 블록을 FFT 신호로 바꾸었다. FFT 크기는 블록 내의 데이터 개수와 같은 128로 하였다.

음성신호는 각 블록간의 상관관계가 높아서 주성분 값이 크게 나온다. 따라서 이를 이용하면 자동차 소음과 구분이 가능하다.

그림 2는 주파수가 약 1500Hz 이하인 구간을 사용하여 첫 번째 주성분 값을 구한 결과이다. 최대 약 250과 최저 약 50 사이에 값들이 존재한다. 하지만 그림 3과 같이 다음 1500에서 2600Hz 대역에서는 최대 약 33에서 최저 약 4 사이의 값으로 현저히 떨어지는 것을 발견할 수 있다. 나머지 그림을 살펴 보면 주파수 대역을 높일수록 점점 값이 작아지는 것을 관찰할 수 있다. 따라서 음성

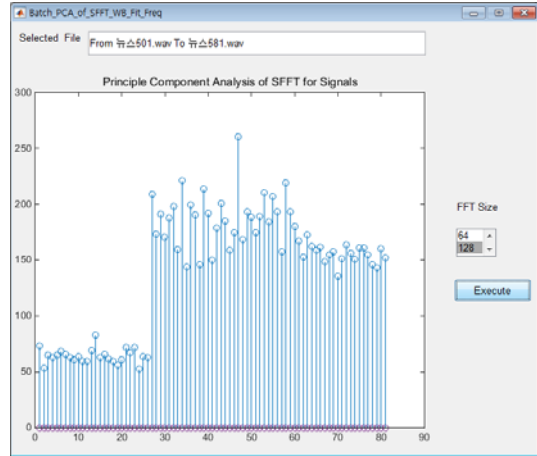


그림 2. 81개 음성 신호의 첫 번째 주성분 값 분포(약 1500Hz 이하 대역)

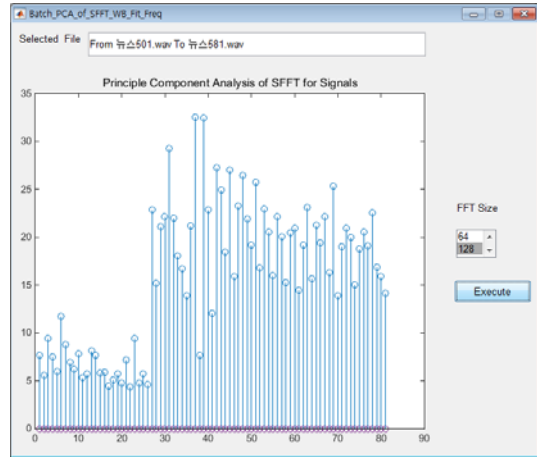


그림 3. 81개 음성 신호의 첫 번째 주성분 값 분포(약 1500-2600Hz 대역)

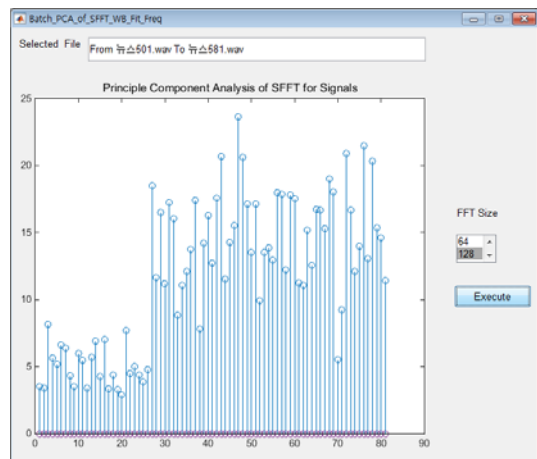


그림 4. 81개 음성 신호의 첫 번째 주성분 값 분포(약 2800-3900Hz 대역)

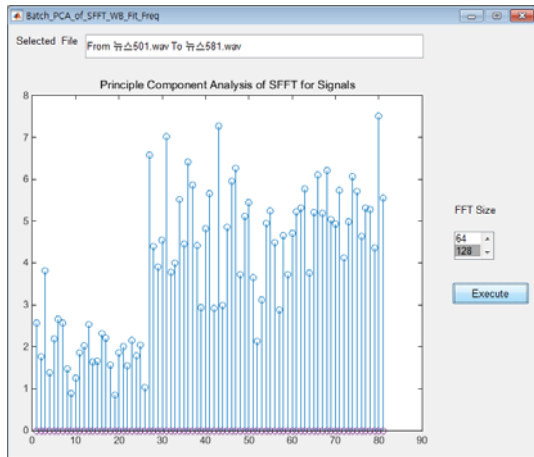


그림 5. 81개 음성 신호의 첫 번째 주성분 값 분포(약 4000-5100Hz 대역)

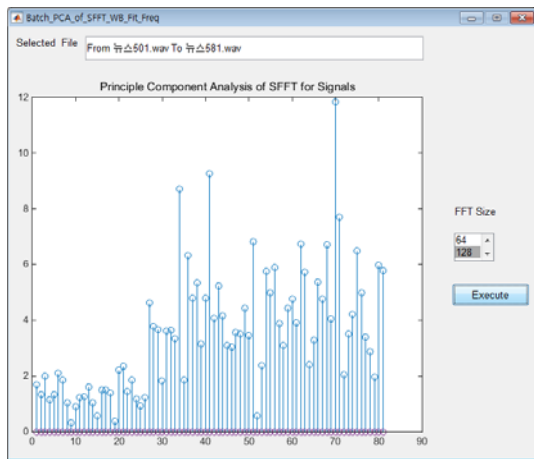


그림 6. 81개 음성 신호의 첫 번째 주성분 값 분포(약 5200-6400Hz 대역)

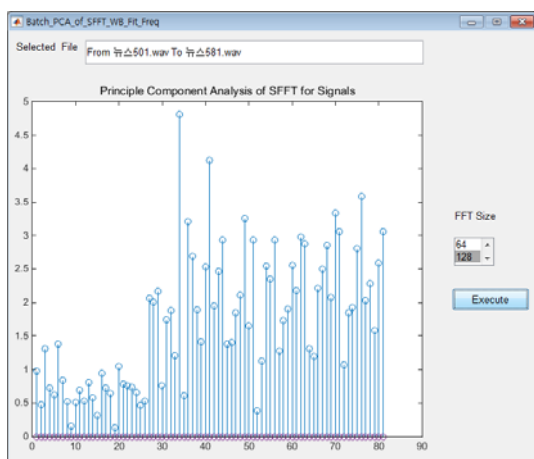


그림 7. 81개 음성 신호의 첫 번째 주성분 값 분포(약 6500-7600Hz 대역)

신호의 주성분 분석을 사용할 때는 저주파 구간에서 주성분 분석을 하고 이를 사용하여야 큰 값의 주성분 값, 즉 의미있는 주성분 값을 얻을 수 있다.

참고문헌

- [1] A. Hyvarinen and E. Oja, "Independent component analysis: algorithms and applications," Neural Networks, vol. 13, no. 4/5, pp. 411-430, 2000.
- [2] 김선일, "자동차 소음 환경에서 ICA를 이용한 한국어 인식", 한국정보처리학회논문지, 기술교육 제 2권, 제3호, pp. 203-207, 2007.
- [3] 김선일, "잡음 섞인 한국어 인식을 위한 ICA 비교 연구," 대한전자공학회논문지, 제45권 1E편, 제 3호, pp. 20-26, 2008.
- [4] 김선일, "ICA로 분리한 신호의 분류," 대한전자공학회논문지, 제 47권 1E편, 제 4호, pp. 10-17, 2010.
- [5] 김선일, "주파수 영역 자기 공분산 기술을 이용한 음성과 자동차 소음 신호의 구분," 한국해양정보통신학회 논문지, 제15권, 10호, 10월, 2011.
- [6] 김선일, 양성룡, "배경 잡음을 제거하는 음성 신호 잡음 제거기의 구현," 대한전자공학회논문지, 제 49권 1E편, 제 2호, pp. 24-29, 2012.
- [7] 김선일, "블록 크기 변화에 따른 자기 공분산 기술기 변화," 한국정보기술학회 하계학술대회는문집, 5월, 2013..
- [8] R. Johnson, K. Tsui, Statistical Reasoning and Methods, John Wiley & Sons, Inc. 1998.
- [9] 김선일, "주성분 분석을 이용한 자동차 소음과 음성의 구분," 한국정보기술학회 하계종합학술대회논문집, 5월, 2014
- [10] 김선일, "음성 신호의 주파수 영역에서의 공분산행렬의 고유값 분석", 한국정보통신학회 종합학술대회논문집, 2015 춘계 19권 1호