음성 신호의 주파수 영역에서의 공분산행렬의 고유값 분석

김선일*

*거제대학교

Analysis of Eigenvalues of Covariance Matrices of Speech Signals in Frequency Domain

Seonil Kim*

*Koje College

E-mail : seonil@koje.ac.kr

요 약

음성 신호는 자음 신호과 모음 신호의 결합으로 이루어져 있지만 그 특성상 자음보다는 모음 신호의 지속시간이 길다. 따라서 전체적으로 음성 신호 블록들 사이의 상관관계가 상당히 크다고 간주할 수 있다. 음성신호를 128개의 데이터를 갖는 블록들로 나눈 후 각 블록의 FFT를 구한다. 이 중에서 모음의 에너지가 집중되어 있는 저주파수 부분만 취하여 이웃 블록들과의 공분산 행렬을 구하고 이 행렬로부터 고유값을 계산해 낸다. 이 중 첫 번 째 고유값은 주성분과 관련이 있다. 다양한 음성 파일들을 이용하여 비교적 값이 큰 첫 번째, 두 번째, 세 번째 고유값과 이들을 합한 고유값이 각 음성 파일에서 어떻게 나타나는지 그 분포를 알아보고 이것들이 음성신호가 아닌 자동차 소음 신호와 같은 잡음 신호의 고유값 분포와 어떻게 다른지 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. Each speech signal is divided into blocks which have 128 speech data. FFT is applied to each block. Low frequency areas of the results of FFT is taken and Covariance matrix between blocks in a speech signal is extracted and finally eigenvalues of those matrix are obtained. It is studied that what the distribution of eigenvalues of various speech files is. The differences between speech signals and noise signals from cars are also studied.

키워드

음성 신호, 상관관계, 고유값, 공분산

Ⅰ. 서 론

음성 인식 기술은 이미 우리 주변을 파고들어서 휴대전화에서 음성을 인식하여 각종 서비스를 실시해주는 앱이 등장하고 있고 어느 정도 실효성을 보여주고 있다.

음성을 인식하는 능력은 음성의 품질과 상당부분 관련되어 있고 조용한 사무실 환경에서 작동하도록 설계된 프로그램은 시끄러운 바깥 환경에서 사용할 때 인식률 면에서 기대 이하의 결과를 나타내게 된다. 이런 현상은 굳이 기계가 아닌 사람이 대화를 나눌 때도 나타나는 현상이어서 배경잡음을 인식하고자 하는 음성 신호와 분리해 주는 작업은 음성 인식 이전에 거쳐야 할 중요한 단계이다.

수 많은 소음을 음성 신호와 분리하려면 소음에 대한 분석부터 해야하지만 이를 자동차로 제한하면 문제가 조금 단순해진다. 자동차가 움직일 때 발생하는 소음은 가속할 때 많이 발생하고 정속 주행 중일 때도 주로 엔진소리나 배기음 소리가 배경 잡음으로 들어가게 된다.

따라서 음성 인식 기능을 자동차에 적용하려고 하면 이런 자동차 소음과 음성 신호의 특징을 찾아내고 이를 이용해 두 신호를 분간해 내는 것이

중요하다.

음성 신호와 자동차 배경 잡음은 ICA[1]와 같은 통계적 방법을 이용해 분리해 낼 수 있다. 하지만 분리해 낸 신호 중 어느 것이 배경 잡음이고 어느 것이 음성 신호인지 분간하려면 다른 기술이 필요하다.

이 기술에는 각 신호를 블록으로 나누어 각 블록의 자기공분산 값을 계산한 다음에 각 블록의 자기공분산 값들을 연결하는 직선을 구해 이 직선의 기울기 값으로 분간하는 방법과[2,3] 주성분 분석을 통해 분간하는 방법이 제안되어 있다.[4]

제안된 주성분 분석 방법은 주 성분 하나만 이용하여 분리된 두 신호의 주성분을 구하고 그 값의 크기가 큰 것은 음성신호, 작은 것은 자동차 배경잡음 신호로 분간하는 것이었다. 따라서 좀 더 안정적으로 신호를 분간할 수 있는 방법이 필요하고 여기서 이를 제시하고자 한다.

먼저 FFT 변환하는 방법에[5] 대해 살펴보고 주성분 분석에 대한 이론을[5] 고찰한 다음 이를 응용한 결과를 제시하였다.

II. FFT 변환

시간 영역에서 신호 S를 식(1)과 같이 FFT(Fast Fourier Transform)가 가능한 $n=2^p$(p는 임의의 자연수)개의 데이터로 구성된 m 개의 블록으로 나누어 준 후 각 블록에 대해 FFT를 계산한다.[3]

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

여기서

$$S_j=[s_{1,j}\ s_{2,j}\ \cdots\ s_{n,j}]^T \tag{2}$$

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

그런데

$$F_j=[f_{1,j}\ f_{2,j}\ \cdots\ f_{n,j}]^T \tag{4}$$

여기서 j는 1부터 m이고 m은 신호 전체를 n으로 나누어서 얻을 수 있는 수이다.

식 (3)에서

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

이고 i는 1부터 n일 때 F는

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

식 (6)에서 F^i는 각 블록의 특정 주파수 대역이다.

III. 주성분 분석

가중 행렬

$$W=\begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,m} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,m} \\ \vdots & \vdots & \cdots & \vdots \\ w_{n,1} & w_{n,2} & \cdots & w_{n,m} \end{bmatrix}=[W_1\ W_2\ \cdots\ W_m] \tag{7}$$

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

$$X=\begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,m} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,m} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,m} \end{bmatrix}=[X_1\ X_2\ \cdots\ X_m] \tag{8}$$

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

일 때

$$Y_1=W_1^TX \tag{9}$$

j는 1부터 m까지 이다.

그런데

$$Y=\begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,m} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,m} \\ \vdots & \vdots & \cdots & \vdots \\ y_{n,1} & y_{n,2} & \cdots & y_{n,m} \end{bmatrix}=[Y_1\ Y_2\ \cdots\ Y_n]^T \tag{10}$$

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

Y_1 의 공분산이 최대이면 이 Y_1을 X의 첫 번째 주성분이라고 한다.

공분산은 가중치 벡터 W_1의 norm과 방향에 따라 달라지고 W_1의 norm이 커짐에 따라 무한정 커지게 된다. 따라서 W_1의 norm을 일정하게 해야 할 필요가 있는데 일반적으로 norm을 1로 한다. 여기서 norm은 다음과 같이 Euclidean 거리로 정의한다.

$$\| W_1 \| = (W_1^T W_1)^{\frac{1}{2}} = \left[\sum_{k=1}^{n} w_{k,1}^2 \right]^{\frac{1}{2}} \quad (11)$$

이런 조건을 만족시키는 가중치 벡터 W_1을 구하면

$$E\{ Y_1^2 \} = E\{ (W_1^T X)^2 \} = W_1^T E\{ XX^T \} W_1 \quad (12)$$
$$= W_1^T C_X W_1$$

여기서 $\| W_1 \| = 1$

이고

X 의 평균이 0 이라면

$$C_X = E\{ XX^T \} \quad (13)$$

이다.

이 식의 주성분 분석 해는 행렬식 C_X의 단위 길이 고유벡터 $e_1, \cdots, e_n$ 이라는 사실은 잘 알려져 있다. 따라서 고유벡터의 고유값들이

$d_1 \geq d_2 \geq \cdots \geq d_n$ 이라면

$W_1 = e_1$ 이다.
따라서

$$E\{ Y_1^2 \} = E\{ (e_1^T X)^2 \} = e_1^T E\{ XX^T \} e_1 \quad (14)$$
$$= e_1^T C_X e_1 = e_1^T d_1 e_1 = d_1$$

이다.

즉 Y의 분산은 X의 분산의 고유값이다.

Ⅳ. 제안 및 검증

16㎑로 샘플링한 음성 신호를 이용하여 128개의 신호를 하나의 블록으로 잡아 FFT를 수행하였다. 따라서 신호의 지속 시간에 따라 블록 전체의 수가 달라진다.

음성신호를 서로 겹치지 않는 블록으로 나누고 각 블록을 FFT 신호로 바꾸어 그 중 저주파 성분에 해당하는 성분만 사용하여 주성분을 구한다. 음성신호는 각 블록간의 상관관계가 높아서 주성분 값이 크게 나온다. 하지만 자동차 배경 잡음도 음성 신호의 주성분과 별로 차이가 나지 않는 경우가 발생한다. 따라서 주성분만으로 두 신호를 분간하기가 어려울 수도 있다.

각 신호의 구분 가능성을 높이기 위하여 주성분 뿐 만 아니라 그 다음 성분도 이용하는 방법을 제안한다. 즉 첫 번째 주성분과 두 번째, 세

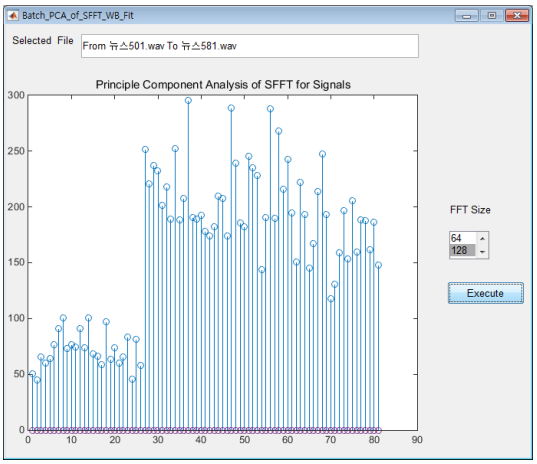

그림 1. 81개 음성 파일에 대한 고유값 d_1

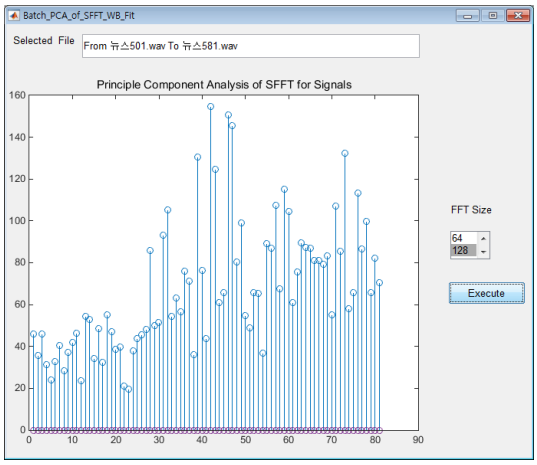

그림 2. 81개 음성 파일에 대한 고유값 d_2

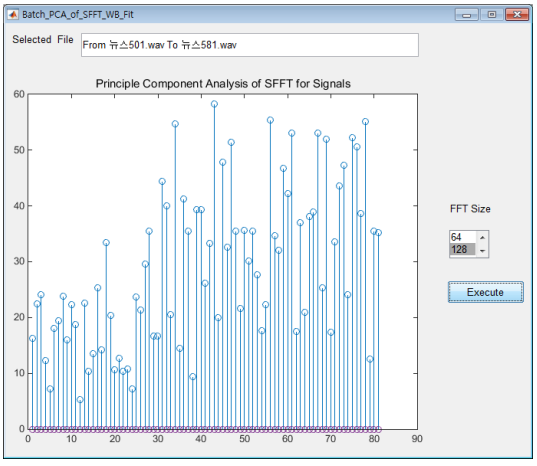

그림 3. 81개 음성 파일에 대한 고유값 d_3

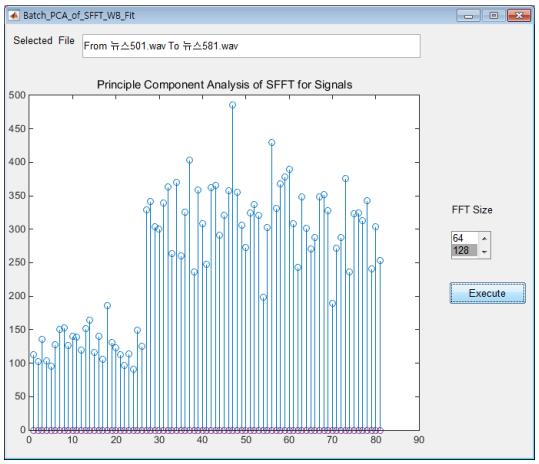

그림 4. 81개 음성 파일에 대한 고유값 $d_1+d_2+d_3$

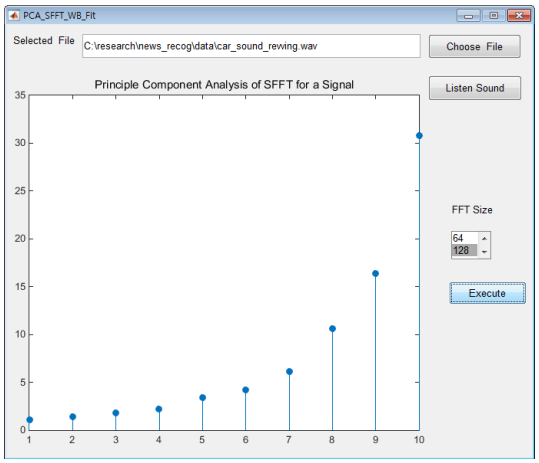

그림 5. 자동차 소음1의 고유값 분포(10개)

번째 주성분을 더하여 사용하면 두 신호를 구분하는 가능성을 더 높일 수 있다. 일반적으로 음성 신호의 두 번째 주성분과 세 번째 주 성분이 자동차 배경 잡음의 그것들보다 값들이 현저하게 크게 나타나는 것을 인지한다면 각 주성분들을 더하여 사용하는 것이 각 신호를 구분할 가능성을 더욱 높여준다고 생각하는 것이 무리가 아닐 것이다.

그림 1부터 그림 7을 살펴보면 음성 신호의 경우에 d_2나 d_3도 비교적 큰 값을 가진다. 반면에 자동차 소음은 d_1이나 d_2나 d_3가 비교적 작은 값을 가지고 있으므로 자동차 소음과 음성 신호를 구별하기 위해 d_1 만 사용하는 것보다 그림 4에 나타나 있는 것처럼 d_1+d_2+d_3를 사용하는 것이 더 유리한 결과를 보여줄 가능성이 높다.

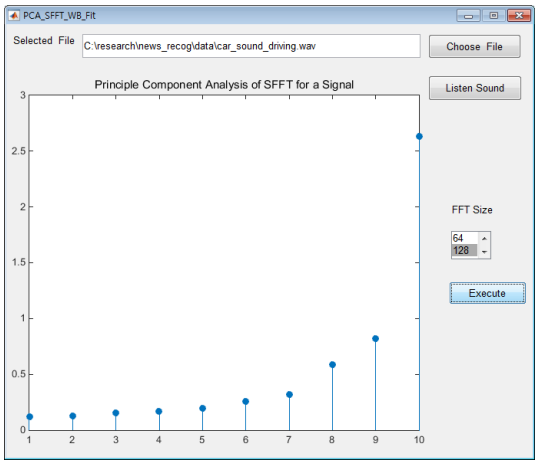

그림 6. 자동차 소음2의 고유값 분포(10개)

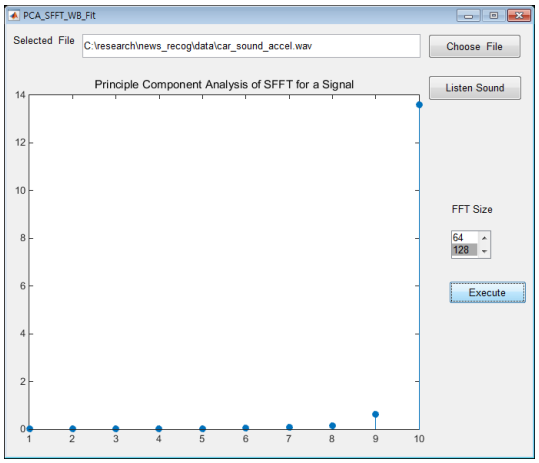

그림 7. 자동차 소음3의 고유값 분포(10개)

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, "주파수 영역 자기 공분산 기울기를 이용한 음성과 자동차 소음 신호의 구분," 한국해양정보통신학회 논문지, 제15권, 10호, 10월, 2011..

[3] 김선일, "블록 크기 변화에 따른 자기 공분산 기울기 변화," 한국정보기술학회 하계학술대회논문집, 5월, 2013..

[4] R. Johnson, K. Tsui, Statistical Reasoning and Methods, John Wiley & Sons, Inc. 1998.

[5] 김선일, "주성분 분석을 이용한 자동차 소음과 음성의 구분," 한국정보기술학회 하계종합학술대회논문집, 5월, 2014