

독립성분분석을 이용한 강인한 음성인식 (Robust Speech Recognition Using Independent Component Analysis)

임형규(Hyung-Kyu Im)¹⁾ 이창기(Chang-Ki Lee)²⁾

요 약

기존 음성 인식의 실세계 적용에서 큰 문제점은 잡음이다. 본 논문에서는 잡음이 섞인 음성 신호로부터 잡음 성분을 분리해 내는 방법을 제안한다. 이 방법은 잡음이 섞인 음성 신호에 독립성분분석(ICA:Independent Component Analysis)을 사용한 암묵 신호 분리(blind source separation)를 적용하여 잡음 성분을 제거하게 된다. 잡음이 혼합된 음성 신호에 독립성분분석을 전처리(preprocessing) 과정에 이용함으로써 인식 성능을 향상시킬 수 있다. 깨끗한 음성 신호에 음악과 거리잡음을 섞었을 경우 인식률이 잡음 없는 음성의 인식률보다 각각 최대 14.98%, 13.78%까지 저하되었다. 그러나 독립성분분석으로 복원된 음성의 경우 잡음 없는 음성의 인식률 수준(각각 97.39%, 96.49%)으로 나타났으며, 독립성분분석을 이용한 음성의 잡음 제거가 인식률 향상에 좋은 결과를 가져옴을 확인 할 수 있다.

ABSTRACT

Noisy speech recognition is one of most important problems in speech recognition. In this paper, a method which efficiently removes the mixed noise with speech, is proposed. The proposed method is based on the ICA to separate the mixed noise. ICA(Independent component analysis) is a signal processing technique, whose goal is to express a set of random variables as linear combinations of components that are statistically as independent from each other as possible.

1) 정회원 : 서남대학교

2) 정회원 : 서남대학교

I. 서 론

음성 인식에서 잡음(noise)은 음성 인식률을 저하시키므로, 이러한 잡음들을 처리하여 음성 인식률 향상시킬 수 있도록 하여야 한다.

본 논문에서는 독립 성분 분석 (ICA: Independent Component Analysis)을 이용하여 잡음이 섞인 음성 신호로부터 잡음 성분을 분리 및 제거하였다. 독립성분분석은 블라인드 신호 분리(blind source separation)문제를 해결하는 방법으로 가장 널리 사용되고 있다. 블라인드 신호분리란 아무런 사전 지식 없이 선형 혼합된 신호로부터 알려지지 않은 소스원이나 혼합 행렬을 구해내는 것이 목적인 통계적인 방법이다. 이중 가장 널리 쓰이고 있는 방법인 독립성분분석은 오직 소스원들이 서로 통계적으로 독립이라는 사실만 가지고 데이터의 고차원 통계 성분을 이용하여 혼합된 신호들로부터 독립된 음원 신호등을 분리해 해 내는 방법이다.

본 논문에서는 독립성분분석을 이용한 음성 신호의 잡음 제거의 성능 평가를 위해서 SNR(Signal to Noise Ratio)과 상관관계수(correlation coefficients)를 비교하였다. 또한 독립성분분석 적용에 따른 인식 성능을 평가 하였다.

II. 독립성분분석을 이용한 잡음 제거

2.1. 독립성분분석

독립성분분석은 통계적인 기법으로, 독립된 정보와 다른 정보들 사이의 상관관계를 변환하는 기술이다[1, 2, 3]. 즉, 특징이 상이한 둘 이상의 신호들이 선형적으로 혼합되어 구성된 확률 변수(random variable)들을 통계적인 방법에 따라서 독립적인 신호들로 분리하는 것을 통칭한다.

독립성분분석은 임의의 신호들을 몇 개의 신호의 가중치가 곱해진 혼합으로 가정되고, 정보이론에 기반하여 혼합된 여러 신호들 간의 통계적인 의존성을 정의한다. 그리고 신경회로망에서

사용되는 학습 방법에 따라 신호들 간의 의존성을 최소로 하는 가중치를 추정하고 이를 곱하면서 독립적인 신호들을 얻을 수 있다.

먼저 q 개의 변수 x_1, x_2, \dots, x_q 가 p 개의 확률적으로 독립적인 변수 s_1, s_2, \dots, s_p 의 선형결합으로 이루어졌다고 가정한다. q 는 p 보다 작지 않으며 평균은 0인 확률 변수들이다. 이 확률 변수들은 두 개의 확률 벡터 $x = [x_1, x_2, \dots, x_q]^T$ 와

$s = [s_1, s_2, \dots, s_p]^T$ 을 구성하며, x 는 s 와 혼합 행렬 A 의 곱으로 표현된다.

$$x = As \tag{1}$$

독립성분분석의 목적은 원래의 벡터 s 와 선형 혼합 행렬(linear mixing matrix) A 의 추정치를 구하는 것이다. y 를 s 의 추정치, W 를 A^{-1} 의 추정치로 표현하면 식 (1)은 식 (2)와 같이 바뀌어 쓸 수 있다. 이러한 방법을 통해 혼합 행렬(mixing matrix)에 의해 x 로 변환되어 관측되는 원 신호(origin source) s 를 구할 수 있다.

$$s \cong y = Wx \tag{2}$$

$$W = W + (r * (I - \Phi(y) * y) * W) \tag{3}$$

식 3에서 r 은 학습 비율을 말한다.

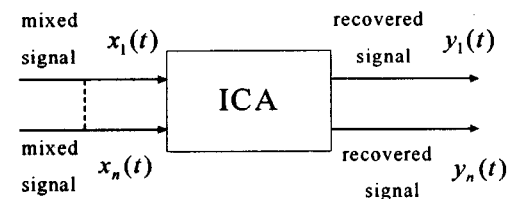


그림 1. 독립성분분석을 이용한 혼합신호분리 블록도

Fig. 1. Block diagram of mixed source separation based on Independent Component Analysis.

혼합 행렬 A 대신 관찰 신호를 독립신호로 변환하는 W 를 구함으로써 독립성분분석의 해(solution) 즉, 원 신호원(original source) s 를 구할 수 있다[4, 5]. 즉 독립성분분석은 독립 신호 s 에 가까운 q 차 신호 y 와 행렬 W 를 찾아내는 과정이라 할 수 있다.

그림 1은 독립성분분석을 이용하여 혼합된 신호 원으로부터 독립된 신호원을 구하는 간단한 블록도이다.

2.2. 독립성분분석에 의한 음성의 잡음 제거

본 논문에서는 음성의 잡음 제거에 있어서 독립성분분석을 이용하는 것이 얼마나 효과적인 것인가를 나타내기 위하여 음성 신호에 음악 신호와 거리 잡음을 첨가한 후에 독립성분분석을 이용하여 다시 이 신호들을 분리해 내었다.

그림 2는 실험에 사용된 원(origin) 음성 신호, 음악 및 거리 잡음 신호이다.

이 데이터들은 음성과 음악 잡음 신호, 음성과 거리 잡음 신호가 식 (1)에서와 같이 2 by 2 선형 혼합 행렬 A 에 의해 서로 선형적으로 혼합되고, 잡음이 섞인 음성 신호가 된다. 이 잡음이 혼합된 신호들에 독립성분분석을 적용하여 독립된 음원 신호들로 분리해 내고, 그 결과 잡음이 섞인 음성 신호로부터 잡음을 제거할 수 있게 된다.

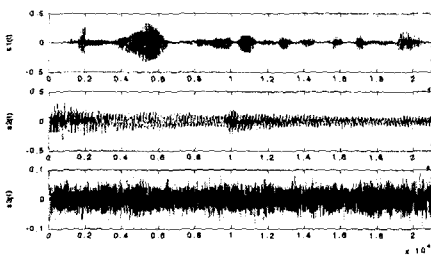


그림 2. 원(original) 음성, 음악 및 거리 잡음 신호

Fig. 2. Original speech, music and street noise signals.

그림 3은 이 전체적인 과정을 나타낸 블록도이다. 그림 3의 과정에 따라 음성 신호에 각각 음악 신호와 거리 잡음을 혼합한 신호를 각각 그림 4와 5에 나타냈다. 이 신호들은 잡음이 섞인 음성 신호들로 그림 3에서의 $x_1(t)$ 와 $x_2(t)$ 의 입력 신호가 되고 이 잡음이 혼합된 음성 신호들에 독립성분분석을 적용하여 $y_1(t)$ 와 $y_2(t)$ 의 신호로 분리해 낼 수 있다. 따라서 잡음이 분리 및 제거되어 잡음이 섞이기 전 깨끗한 음성 신호를 얻어 낼 수 있다.

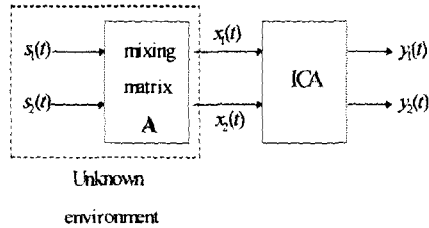


그림 3. 혼합된 신호원으로부터 독립된 신호원을 구하는 블록도

Fig. 3. Block diagram of independent source separation using Independent Component Analysis.

잡음 제거 실험에 사용된 음성 데이터는 11,025Hz, 16bits로 양자화 되었으며 약 21,000 sample를 갖는 “3주에 스무 대 씩 있을겁니다”를 발성한 데이터이다. 잡음 데이터는 음악과 거리 잡음을 사용하였다. 실험은 Matlab을 이용하였다.

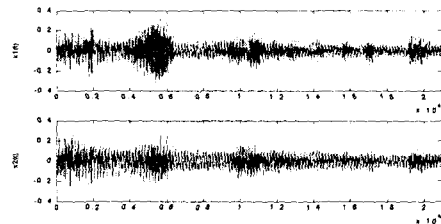


그림 4. 음성 신호와 음악 잡음이 섞인 신호
Fig. 4. Speech-music noise mixtures.

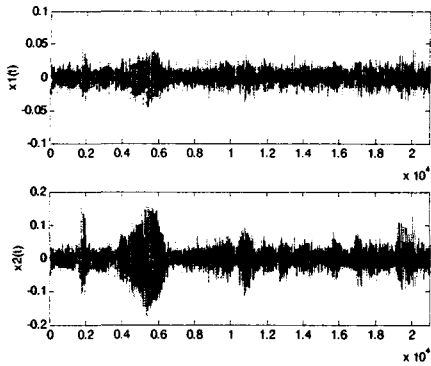


그림 5. 음성 신호와 거리 잡음이 섞인 신호
Fig. 5. Speech-street noise mixtures.

III. 결 과

본 논문에서는 음성 신호와 잡음이 혼합되어 있는 경우 잡음을 제거하고 원래의 음성 신호를 복원해내는 목적으로 독립성분분석을 이용하였다.

그림 6과 7은 잡음이 혼합된 음성 신호(그림 4와 5)에 독립성분분석을 이용하여 잡음 섞인 음성 신호로부터 잡음을 분리 및 제거하고 복원된 원 잡음 신호와 음성 신호를 나타낸다.

표 1과 2는 독립성분분석을 하기 전의 음성과 잡음이 혼합된 신호와 독립성분분석을 한 후 얻어진 잡음이 제거된 음성 신호의 SNR을 나타냈다. 독립성분분석을 행하기 전의 잡음이 혼합된 음성 신호와 독립성분분석을 행한 후 얻어진 음성 신호의 SNR를 서로 비교해 보면 독립성분분석 후 얻은 음성 신호의 SNR이 높음을 알 수 있다. 이는 독립성분분석을 통해 음성 신호의 잡음을 효과적으로 제거하였음을 나타낸다. 또한, 원(origin) 음성 신호와 잡음이 혼합된 음성 신호 및 독립성분분석을 행한 후 잡음이 제거된 음성 신호의 상관관계 계수(correlation coefficients)를 나타낸다. 상관관계 계수의 정의는 식 (4)와 같다.

$$\rho_{ij} = \frac{Cov(X_i, Y_j)}{\sigma(X_i)\sigma(Y_j)} \quad (4)$$

여기서, ρ_{ij} 는 X_i 와 Y_j 의 상관관계 계수이다.

3.1. 음성 신호와 음악 잡음 신호의 분리

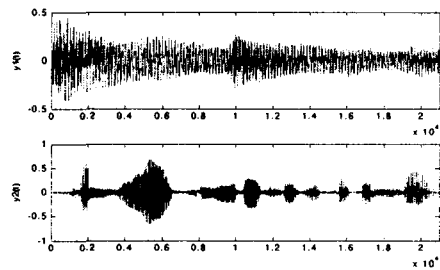


그림 6. 독립성분분석에 의해 분리된 음악 신호(위)와 음성 신호(아래)
Fig. 6. Recovered music(top) and speech(bottom) signals

표 1. 음성-음악 잡음 혼합 신호와 독립성분분석으로 잡음 분리된 신호의 SNR과 상관관계 계수

Table 1. SNR and correlation coefficients of speech-music noise mixtures and recovered speech

신호 구분	SNR(db)	상관관계 계수
음성-음악 혼합 신호 1 ($x_1(t)$)	-5.3147	0.3624
음성-음악 혼합 신호 2 ($x_2(t)$)	-7.5749	0.2681
ICA 후 음성 신호 ($y_2(t)$)	24.1322	0.9981

3.2. 음성 신호와 거리 잡음의 분리

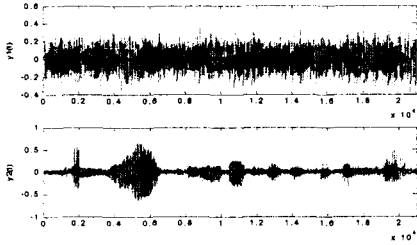


그림 7. 독립성분분석에 의해 분리된 거리 잡음(위)과 음성 신호(아래)

Fig. 7. Recovered street noise(top) and speech signals(bottom).

표 2. 음성-거리 잡음 혼합 신호와 독립성분분석으로 잡음 분리된 신호의 SNR과 상관관계 계수

Table 2. SNR and correlation coefficients of speech- street noise mixtures and recovered speech.

신호 구분	SNR(db)	상관관계 계수
음성-음악 혼합 신호 1 ($x_1(t)$)	-3.1267	0.3157
음성-음악 혼합 신호 2 ($x_2(t)$)	5.8429	0.4236
ICA 후 음성 신호 ($y_2(t)$)	22.3152	0.9817

표 1, 2에서 독립성분분석을 통해 잡음을 분리한 음성 신호의 상관관계 계수가 1에 근접함을 확인할 수 있는데 이는 잡음의 분리가 잘됐음을 나타내고, 독립성분분석을 이용하여 잡음이 섞인 음성으로부터 효과적으로 잡음과 음성을 분리해낼 수 있음을 확인할 수 있었다.

3.3. 인식 실험 결과

음성 인식 실험은 잡음이 혼합된 음성 신호에 독립성분 분석을 적용하여 잡음을 분리해주기 전후의 인식 성능을 비교하였다.

인식 실험에 사용된 음성 DB는 PBW(Phonetically Balanced Word)이다. PBW 452단어에 각각 음악과 거리 잡음을 혼합

한 후 학습 데이터 350, 테스트 데이터 102로 분류하여 mono-phone 단계에서 인식 실험하였다. 또한 잡음 데이터는 음악과 거리 잡음 신호를 사용하였다. 실험에는 Matlab을 이용하였고, 객관적인 인식 실험 결과를 나타내기 위해서 HTK 2.2를 사용하였으며, 특징 파라미터로는 MFCC 39차를 사용하였다.

음악, 거리 잡음이 섞인 음성 신호에 독립성분 분석을 적용하기 전후의 인식률을 표 3, 4에 나타냈다.

표 3. 독립성분분석 적용 후 인식률(음악 잡음)

Table 3. Recognition rate according to ICA application. (music noise)

	원(origin) 음성신호	음성-음악 혼합 신호 1 ($x_1(t)$)	음성-음악 혼합 신호 2 ($x_2(t)$)	ICA 후 음성 신호 ($y_2(t)$)
인식률 (%)	98.04	83.06	85.99	97.39

표 4. 독립성분분석 적용 후 인식률(거리 잡음)

Table 4. Recognition rate according to ICA application. (street noise)

	원(origin) 음성신호	음성-거리 혼합 신호 1 ($x_1(t)$)	음성-거리 혼합 신호 2 ($x_2(t)$)	ICA 후 음성 신호 ($y_2(t)$)
인식률 (%)	98.04	84.26	86.93	96.49

표 3, 4의 결과를 보면, 음악과 거리잡음을 섞었을 경우 인식률이 잡음 없는 음성의 인식률보다 각각 최대 14.98%, 13.78%까지 저하되었으나, 독립성분분석으로 복원된 음성의 경우 잡음 없는 음성의 인식률 수준으로 나타났으며, 독립성분분석을 이용한 음성의 잡음 제거가 인식률 향상에 좋은 결과를 가져옴을 확인할 수 있다.

VI. 결론

본 논문에서는 음성 신호로부터 인식에 사용하는 음성 정보를 얻어내는 과정에서 섞여 있는

잡음 성분을 효과적으로 제거하여 잡음 환경에서 인식 성능을 향상시키는 방법을 제안하였다. 제안한 방법은 독립성분분석을 사용하여 음성 인식에 있어 가장 큰 장애 요소인 잡음 성분을 제거하였다.

본 논문에서 사용한 실험 방법은 깨끗한 음성에 각각 음악, 거리 잡음을 섞은 후 이 잡음 섞인 음성 신호에 독립성분분석을 적용하여, 독립성분분석 전·후의 SNR (Signal to Noise Ratio), 상관관계계수(correlation coefficient) 및 인식을 서로 비교 분석하였다.

그 결과, SNR은 최대 31.7db, 25.4db가 증가하였고, 독립성분분석 후 상관관계계수는 각각 0.9981, 0.9817로 원 음성신호에 근접하였다. 또한 독립성분분석 후 인식 성능은 97.39%, 96.49%로 잡음 섞은 음성 신호보다 최고 14.33%, 12.23% 증가하였다.

향후 다양한 환경의 잡음 섞인 음성 인식에 독립성분분석을 통해 잡음을 제거한다면 좋은 성능을 보일 수 있을 것이다.

11(2): 609-633, 1996.

[5] K. Torkola, "Blind Separation of Convolved sources based on Information Maximization," In IEEE Workshops on Neural Networks for Signal Processing, pp. 423-432, 1996.

참 고 문 헌

- [1] P. Comon, "Independent component analysis-a new concept?," Signal Processing, 36, pp. 2940-2943, 1996.
- [2] J.-F. Cardoso and B. Hvam Laheld, "Equivariant adaptive source separation," IEEE Trans. on Signal Processing, vol. 44 no. 12, pp. 3017-3030, 1996.
- [3] N. Delfosse and P. Loubaton, "Adaptive blind separation of convolutive mixtures," In Proc. ICASSP96, pp. 2940-2943, 1996.
- [4] T-W. Lee, M. Girolami and T. J. Sejnowski. "Independent Component Analysis using an Extended Infomax Algorithm for Mixed Sub-Gaussian and Super-Gaussian Sources," Neural Computation, vol.