

논문 2010-47IE-4-2

ICA로 분리한 신호의 분류

(Classification of Signals Segregated using ICA)

김 선 일*

(Seonil Kim)

요 약

ICA(Independent Component Analysis)를 이용하여 신호를 분리했을 때 그 중 어느 것이 원하는 신호인지 알아낼 수 있는 일반적인 방법이 없다. 본 논문에서는 자동차 배기음에 오염된 음성 신호를 가정하고 이를 ICA를 이용해 분리했을 때 분리된 신호에서 어느 것이 배기음이고 음성 신호인지 구별할 수 있는 방법을 제시하였다. 음성 신호는 음성 신호와의 상관계수가 가장 크게 나타날 것으로 예상되므로 오염된 음성 신호와 같은 동일인의 단모음 '아', '오', '우' 신호와 타인의 단모음 신호를 이용하여 분리된 각 신호와의 상관계수를 구하되 일괄 방식, 최대값 방식, 평균값 방식 등 세가지 방식으로 구하고 각 방식마다 '아', '오', '우'와의 상관 계수로 구분하고 이외에 투표 방법, 합산 방법을 추가한 다섯 가지 방법을 시도하여 가장 좋은 분류율을 나타내는 방식 및 방법을 제시하였다.

Abstract

There is no general method to find out from signals of the channel outputs of ICA(Independent Component Analysis) which is what you want. Assuming speech signals contaminated with the sound from the muffler of a car, this paper presents the method which shows what you want. It is anticipated that speech signals will show larger correlation coefficients for speech signals than others. Batch, maximum and average method were proposed using 'ah', 'oh', 'woo' vowels whose signals were spoken by the same person who spoke the speech signals and using the same vowels whose signals are by another person. With the correlation coefficients which were calculated for each vowel, voting and summation methods were added. This paper shows what the best is among several methods tried.

Keywords : ICA, BSS, 상관계수, 분류, 음성 신호, 배기음

I. 서 론

일반적으로 기계가 음성 신호를 인식하고자 할 때 가장 걸림돌이 되는 신호가 배경 잡음이다. 인식하고자 하는 음성 신호의 배경 잡음을 얼마나 제거하느냐 하는 것은 음성 인식의 인식률을 높일 수 있는 관건이다. 오늘날 음성 인식 기술의 발달로 배경 잡음이 없는 환경에서 녹음한 음성이나 사무실 환경에서 녹음된 음성의 인식률은 현저히 개선되었고 상당한 수준에 올라 와 있지만 배경 잡음이 존재하는 일상생활에서 기계를 이용

해 음성을 인식하는 데는 상당한 한계가 있는 것이 사실이다. 음성 인식 기술이 사회 전반적으로 다양하게 응용되지 못하고 있는 데는 이런 현실적인 이유가 가장 크다.

일반적으로 배경 잡음이라 하면 일상생활에 존재하는 다양한 음원으로 부터 나온 소리를 일컫는데 그것이 다른 사람의 음성일 수도 있고 자동차의 배기음일 수도 있으며 기타 사무기기에서 나오는 소리일 수도 있다. 각테일파티 효과라고 해서 여러 사람들이 음성거리는 소리 가운데서 본인이 주의를 집중해서 듣고자 하는 사람의 소리만 듣게 되는 능력이 인간에게 있다고 알려져 있다. 기계적으로 배경 잡음을 음성 신호와 분리해 내는 방법은 일반적으로 필터링을 사용하나 이 방법은 주

* 정회원, 거제대학 조선정보기술학부
(Dept. of Information Technology for Shipbuilding,
Koje College)

접수일자: 2010년9월7일, 수정완료일: 2010년12월7일

파수 영역이 서로 겹치지 않을 경우에 신호의 손상 없이 잡음을 제거할 수 있지만 주파수가 중복될 경우에 손실을 감수해야만 한다. 특히 배경 잡음이 다른 사람의 음성일 때는 필터링이 불가능하다.

최근 자동차에 텔레매틱스 기술이 적용되면서 궁극적으로 사람의 음성으로 자동차의 각종 기기를 제어하거나 자동차를 조종하려는 목표를 실현시키기 위해 각종 기술을 개발하고 있다. 주행 중에 자동차에서 나오는 주 배경 잡음은 자동차 배기음으로 특히 가속할 때 상당히 큰 배기음을 경험할 수 있다.

여러 신호원에서 나온 관찰되지 않은 신호가 섞인 여러 관찰된 신호로부터 관찰되지 않은 신호를 분리해 내는 것을 Blind Source Separation(BSS)^[1]이라고 하는데 관찰된 신호에서 음원을 분리해 내는 것을 일반적으로 일컬으며 BSS를 하기 위한 방법으로 가장 많이 쓰이는 것이 Independent Component Analysis(ICA)^[2-3]이다.

ICA는 BSS에서 쓰이는 여러 가지 방법 중 하나이며 음원 신호를 분리해 내기 위해 Kurtosis^[4]나 Negentropy^[5]등을 사용하는데 이는 Nongaussianity를 최대화해서 관찰한 신호에서 관찰되지 않은 서로 독립적인 신호를 분리해 내는 방법이다. 일반적으로 여러 가지 신호가 서로 섞일수록 밀도 함수가 가우스 분포화하는 경향이 있다고 알려져 있으므로 그와 반대로 밀도 함수가 가우스 분포에서 멀어지는 신호를 찾아내는 방법이다.

하지만 ICA를 이용해서 음원을 분리해 내더라도 어느 음원이 원하는 신호인지 구별해 내야 하는 문제점이 있다. 현재 제시된 ICA에서는 음원이 두 개 일 때 2개 이상의 마이크로폰을 이용하여 신호를 입력한 후 두 개의 독립적인 신호를 분리해서 출력하게 되는데 신호를 분리하기만 하지 이 신호가 어느 음원인지 제시해 주지 못한다.

본 연구에서는 자동차 배기음과 섞인 음성 신호에서 두 신호를 분리해 내고 분리된 신호에서 어느 것이 음성 신호인지 분리하는 방법을 제시하고자 한다. 자동차 배기음은 음성 신호와 확연히 구별되지만 두 음성을 가려내는 방법은 마땅치 않다. 하지만 음성 신호는 유성음을 가지고 있고 이는 자동차 신호와 구별될 수 있는 특징이라고 판단하고 음성 신호에서 유성음 신호를 채취한 후 이 신호와 ICA로 분리해 낸 각 신호와의 상관계수를 계산하고 이 값의 크기를 비교하여 두 신호를

구별해 내는 방법을 제안한다.

본 논문은 먼저 BSS가 무엇인지 알아보고 이어서 ICA와 상관계수를 언급한 후에 실험 결과에 대해 논의하는 순서로 서술되어 있다.

II. Blind Source Separation

여러 음원에서 나온 신호가 마이크로폰에 도착한 시점에 서로 중첩되어 마이크로폰에 입력되는데 각 음원의 신호들에 대한 정보가 전혀 없는 상태에서 중첩되전의 신호를 분리해 내는 것을 BSS라고 일컫는다.

m 개의 음원에서 나온 신호가 서로 중첩되어 n 개의 중첩된 신호를 형성했다고 하자. 즉 m 개의 서로 다른 신호들이 n 개의 마이크로폰에 입력되었을 경우에 음원 신호는

$$\mathbf{s} = [s_1 \ s_2 \ s_3 \dots s_m]^T \tag{1}$$

이 음원 신호는 관측되지 아니한 신호이며 서로 독립적이라고 가정한다.

이 신호들이 마이크로폰에 도달하여 서로 중첩되었을 때 신호는

$$\mathbf{x} = [x_1 \ x_2 \ x_3 \dots x_n]^T \tag{2}$$

$x_1 \ x_2 \ x_3 \dots x_n$ 은 각 마이크로폰의 출력 신호이며 관찰 가능한 신호이다.

각 출력 신호가 마이크로폰에서 합성될 때 각 마이크로폰의 특성 및 신호와의 거리 등에 따라 합성 행렬 A에 의해

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{3}$$

와 같이 합성된다.

음원 신호가 3개이고 마이크로폰이 3개이면 A는

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{32} & a_{12} & a_{23} \end{pmatrix} \tag{4}$$

와 같이 표현되는 매트릭스이고 이 때 합성 신호는

$$x_1 = a_{11}s_1 + a_{12}s_2 + a_{13}s_3 \tag{5}$$

$$x_2 = a_{21}s_1 + a_{22}s_2 + a_{23}s_3 \tag{6}$$

$$x_3 = a_{31}s_1 + a_{32}s_2 + a_{33}s_3 \tag{7}$$

합성 행렬 A 를 알면 다음 식과 같이

$$s = A^{-1}x = Wx \quad (8)$$

$$W = A^{-1} \quad (9)$$

A 의 역행렬 W 를 구하여 음원 신호 s 를 구할 수 있다.

여기서 우리가 관측할 수 있는 신호는 x 뿐이다. 일반적으로 이 식에서 합성 행렬 A 를 구하기 위해서는 x 뿐만 아니라 s 도 알고 있어야 하지만 현실은 그렇지 못하다. 따라서 s 의 각 신호 $s_1, s_2, s_3, \dots, s_m$ 이 서로 독립적이라고 가정하고 각 신호의 독립성을 최대한 높이는 방향으로 행렬 W 를 추정하여 관측된 신호 x 로부터 음원 신호 s 를 구하게 되는데 이 방법을 Independent Component Analysis라고 한다.

III. Independent Component Analysis

서로 독립적인 여러 신호가 섞인 상태에서 독립적인 신호를 추정해 내는 ICA를 구현하는 방법은 여러 가지가 있다.

Central Limit Theorem^[6]에 의하면 둘 이상의 신호를 더해서 얻어지는 신호의 분포는 원래 신호의 분포보다 더 정규화 되는 경향이 있다. 즉 데이터의 양이 많을수록 그 데이터의 분포는 Gauss 함수에 가까워진다. 따라서 마이크로폰을 통해 합성된 신호는 합성되기 전의 신호보다 더 정규 분포하게 된다. 이런 특성을 이용해서 합성된 신호로부터 분리되는 신호들의 Nongaussianity가 큰 방향으로 분리되도록 W 를 구하면 음원 신호에 대한 정보 없이도 섞이기 전의 음원 신호들을 추정해낼 수 있다.

Nongaussianity를 최대화하는 방법으로는 4차 모멘트를 이용하는 Kurtosis^[4]를 구하는 방법이 있다. 임의의 신호 y 의 Kurtosis^[2]는

$$kurt(y) = E(y^4) - 3(E(y^2))^2 \quad (10)$$

와 같이 구할 수 있다. 백색처리를 통해 $E(y^2) = 1$ 이 되게 하면 10번 식의 오른쪽 끝항이 -3으로 바뀌게 된다. 하지만 수식의 4승 항목 때문에 큰 신호 몇몇이 전체 값을 대표하는 경우가 생기는 단점이 있다. 따라서 이런 문제점을 개선한 것이 Negentropy이다. Negentropy도 Nongaussianity를 측정하는 한 방법으로 이용 되는데 이것은 엔트로피에 근거를 둔 것으로 확률

이 $p_y(x)$ 인 랜덤 벡터의 차등 엔트로피는

$$H(y) = - \int p_y(x) \log p_y(x) dx \quad (11)$$

로 나타낼 수 있다. 그런데 가우스 변수는 가장 큰 엔트로피를 나타낸다. 따라서 다음 식 12와 같이 가우스 랜덤 변수인 y_{gauss} 의 엔트로피에서 y 의 엔트로피를 빼면

$$J(y) = H(y_{gauss}) - H(y) \quad (12)$$

y 의 Nongaussianity를 알 수 있으며 이 J 를 Negentropy^[2]라고 한다.

Kurtosis나 Negentropy가 Nongaussianity를 최대화하는 방법이라면 Maximum Likelihood를 최대화 하는 방법도 있고 이와 유사하면서 출력 엔트로피를 최대화 하는 Infomax^[7]도 있으며 출력 사이의 Mutual Information을 최소화하는 방법^[8]도 있다. 각 방법을 구현하기 위해서는 Gradient 알고리즘^[9]이나 Fast Fixed-Point 알고리즘^[5](FastICA)을 주로 사용한다. 본 연구에서는 Infomax를 이용한 Bell의 알고리즘^[9]에 Amaral^[10,11]의 제안에 따른 Natural Gradient 개념을 적용하여 역행렬 계산을 피하고 수렴 속도를 증가시켰다.

IV. 상관계수

상관관계란 둘 또는 그 이상의 랜덤 변수나 관찰 데이터 사이의 통계적 연관성을 말한다. 상관관계를 나타내는 계수를 상관계수라고 하는데 일반적으로 Pearson의 표본상관계수를 지칭하며 두 랜덤 변수 사이의 상관계수는 다음과 같이 정의된다^[13].

$$r(x, y) = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{d_1} \right) \left(\frac{y_i - \bar{y}}{d_2} \right) \quad (13)$$

여기서 평균값 \bar{x} 및 \bar{y} 는

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j \quad (14)$$

$$\bar{y} = \frac{1}{n} \sum_{j=1}^n y_j \quad (15)$$

이고 표준편차 d_1, d_2

$$d_1 = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2} \quad (16)$$

$$d_2 = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (y_k - \bar{y})^2} \quad (17)$$

이다.

수식에서 추정할 수 있듯이 표본상관계수는 그 크기가 1을 초과할 수 없고 -1과 1 사이의 값을 가진다. 1이나 -1인 경우 강한 상관관계를 가지고 0에 가까울수록 상관도가 떨어진다.

V. 음원 분리과 분류

남성 앵커의 뉴스 음성 신호와 자동차의 배기음 신호를 사용하였다. 실제 방송 뉴스에서 81개의 길거나 짧은 문장으로 이루어진 음성을 녹취하여 사용하였으며 배기음으로는 가속시의 배기음이 포함된 신호를 사용하였다.

실험에 사용된 배기음 신호는 그림 1^[12]의 좌측과 같다. 음성 신호와 자동차 배기음이 두 개의 마이크에 동시에 입력되는 상황을 가정해서 두 마이크로폰에 입력되는 배기음 대 음성 신호의 평균 전력비가 2dB 정도 차이 나게 조정하여 한 쪽 마이크로폰이 0dB이면 다른 쪽은 2dB이 되도록 섞어서 입력하였다. 그림 1의 오른쪽에 배기음과 혼합된 여러 음성 신호들 중의 하나가 나타나 있다.

그림 2^[12]를 보면 분리된 음성 신호가 그림 왼쪽에, 배기음은 오른쪽에 나타나 있다. 경우에 따라 오른쪽에 음성 신호가, 왼쪽에 배기음이 나타나기도 한다. 특정 출력에 어느 신호가 나타날지는 알 수가 없다. [12]에서는 자동차 배기음과 혼합된 음성 신호를 ICA로 분리해 내기만 하였고 분리된 신호 중 어느 것이 음성이고 어느 신호가 배기음인지 알 수 없었다. 본 연구에서는 분리된 신호가 음성 신호인지 배기음 신호인지 구분해 내기 위해 분리된 각 신호에 대하여 주어진 특정 신호와의 상관계수를 구하여 음성과 배기음을 구분해 내고자

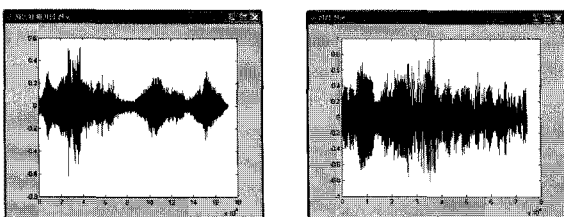


그림 1. 자동차 배기음 신호(왼쪽) 및 혼합 신호
Fig. 1. Muffler noise signal(left) and mixed signal.

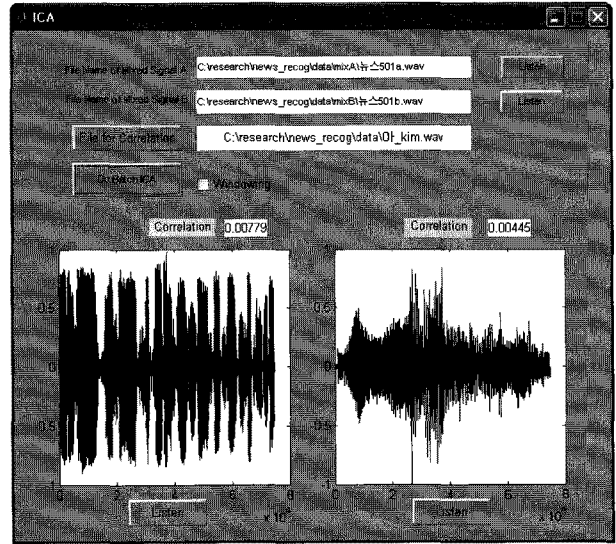


그림 2. 분리된 음성 신호와 자동차 배기음
Fig. 2. Separated speech signal and muffler noise.

한다.

상관계수를 구하는데 사용한 신호로는 자동차 배기음과 음성 신호의 조음 원리가 다른 점을 이용하여 사람이 말할 때 포함될 수밖에 없는 유성음 중 복모음을 제외하고 단 모음 중 ‘아, 오, 우’ 를 이용하였다. 실험에 사용된 남성 앵커의 음성 신호 중에서 ‘아, 오, 우’ 를 무작위 채취하고 이 앵커가 아닌 다른 한 남자의 ‘아, 오, 우’ 도 녹취하여 실험에 사용하였다.

분리된 신호는 분류에 사용된 모음 신호보다 상당히 지속 시간이 길기 때문에 음성 신호와 모음을 복제하여 지속 시간을 맞춰준 후 신호 전체에 대한 상관계수를 한 번에 구하는 일괄 방식과 모음 신호를 주어진 구간 만큼씩 이동시키면서 계수를 구하는 이동 방식을 사용하였고 이동 방식은 또한 최대값 방식과 평균값 방식으로 구분하여 총 세 가지 방식을 시도하였다.

1) 일괄 방식

배기음과 음성 신호가 섞인 혼합 신호의 길이와, 상관계수를 구하는데 사용될 ‘아, 오, 우’ 각 모음 신호의 길이가 서로 다르므로 두 신호의 길이를 맞추어 주기 위해 신호 길이가 짧은 모음 신호를 복사해서 혼합 신호의 길이와 맞추어 주는 작업이 선행되어야 한다. 신호를 분리하면 크기가 T_1 개인 분리 신호 \hat{s}_1, \hat{s}_2 가 얻어진다.

$$\hat{s}_1 = [\hat{s}_1(1) \hat{s}_1(2) \hat{s}_1(3) \dots \hat{s}_1(T_1)] \quad (18)$$

표 1. 일괄 방식으로 분류한 결과

Table 1. Results of classification by batch processing.

구분	방법	인식문장수(개)	분류율(%)
동일인	아	67	82.7
	오	67	82.7
	우	44	54.3
	투표	69	85.2
	합산	73	90.1
타인	아	45	55.6
	오	70	86.4
	우	58	71.6
	투표	65	80.2
	합산	72	88.9

$$\hat{s}_2 = [\hat{s}_2(1) \hat{s}_2(2) \hat{s}_2(3) \dots \hat{s}_2(T_1)] \quad (19)$$

그리고 크기가 T_2 인 모음 신호는

$$v = [v(1) v(2) v(3) \dots v(T_2)] \quad (20)$$

대부분의 경우에 $T_1 > T_2$ 이므로

$$f = \text{fix}\left(\frac{T_1}{T_2}\right) \quad (21)$$

로 배수 f 를 얻을 수 있다. fix 는 실수의 소수점을 모두 버리고 정수만을 되돌려 주는 함수이다.

T_1 에서 T_2 의 f 배를 빼면 다음과 같이 차이 d 를 얻을 수 있다.

$$d = T_1 - f * T_2 \quad (22)$$

따라서

$$v_d = [v(1) v(2) v(3) \dots v(d)] \quad (23)$$

일 때, 크기가 T_1 과 같아지도록 늘린 모음 신호 v_f 는

$$v_f = \underbrace{[v \ v \ \dots \ v]}_{f\text{개}} \ v_d \quad (24)$$

이다. 식 13을 이용해 v_f 과 \hat{s}_i 와의 상관 계수

$$r_i = r(v_f, \hat{s}_i) \quad (25)$$

를 구하고 둘을 다음과 같이 비교하여 최대값의 상관 계수를 보이는 신호를 음성 신호로 판정한다. 표 1에서 결과를 확인할 수 있다.

표 2. 최대값 방식으로 분류한 결과

Table 2. Results of classification by peak value.

구분	방법	분류율(%)				
		100	200	500	1000	1600
동일인	아	100.0	100.0	96.3	87.7	75.3
	오	96.3	95.1	93.8	84.0	86.4
	우	100.0	100.0	100.0	98.8	97.5
	투표	100.0	100.0	100.0	98.8	96.3
	합산	100.0	100.0	100.0	100.0	98.8
타인	아	24.7	72.8	54.3	37.0	92.6
	오	96.3	95.1	98.8	91.4	88.9
	우	64.2	70.4	61.7	44.5	60.5
	투표	71.6	88.9	85.2	56.8	91.4
	합산	79.0	91.4	87.7	64.2	92.6

$$\begin{cases} m = \underset{i}{\operatorname{argmax}}(r_i) \\ \text{Speech Signal} = \hat{s}_m \end{cases} \quad (26)$$

2) 최대값 방식

각 모음에 대해 분리된 신호들과의 상관계수를 StepSize 개수 만큼의 간격으로 이동하면서 상관계수를 구하여 그 중에서 제일 큰 값을 대표 상관계수로 정하고 분리된 두 신호와의 대표 상관계수 중에 값이 더 큰 것을 음성 신호로 분류하였다. StepSize는 100개, 200개, 500개, 1000개, 1600개로 정하였다.

$$\hat{s}_{ik} = [\hat{s}_i(k+1) \hat{s}_i(k+2) \dots \hat{s}_i(k+T_2)] \quad (27)$$

일 때 분리된 각 신호에 대해 다음 알고리즘을 적용한다.

```

for i = 1 to m
  r_i = 0
  for j = 1 to f
    k = (j-1)*StepSize
    r_i = max(r_i, r(v, \hat{s}_{ik}))
  end
end

```

$\max()$ 는 두 신호 중에 최대값을 돌려주는 함수이다.

이렇게 구해진 r_i 에 대해 식 26을 적용해서 음성신호를 구별한다.

100개 이하의 간격으로 이동하면 분리된 신호 전 구

간에 대해 상관계수를 구하는 횟수가 증가하여 계산 시간이 상당히 많이 걸려 100개 샘플 간격 이하는 시도하지 않았다. 표 2에 그 결과를 나타내었다.

3) 평균값 방식

최대값 방식과 마찬가지로 정해진 이동 간격으로 이동하면서 상관계수를 구하되 다음 알고리즘과 같이 신호 전 구간에서 걸쳐 상관계수를 구한 후 이들의 평균값을 대표 상관계수로 사용하였다.

```

for i = 1 to m
    rs = 0
    for j = 1 to f
        k = (j - 1) * StepSize
        ri = r(v, sik)
        rs = rs + ri
    end
    ri = ri / f
end
    
```

이렇게 구해진 r_i에 대해 식 26처럼 상관계수가 더 큰 신호를 음성신호로 판정하였고 표 3에 결과가 나타나 있다.

표에서 '동일인'은 상관계수를 구하는데 쓰인 모음이 배기음과 섞이게 되는 음성 신호와 같은 사람의 모음을 나타내고 '타인'은 다른 사람임을 나타낸다. '아'는 상관계수를 구하는데 '아' 음성 신호를 사용하였으며 '오'는 '오' 음성 신호를, '우'는 '우' 음성 신호를 사용하였음을 나타낸다. '투표'의 경우에는 '아', '오', '우'와의 대표 상관계수를 각각 구하여 어느 출력이 음성 신호인지 판정한 후 각 판정 결과를 모아서 셋 중에 둘 이상의 판정 결과를 인정하는 방식이다. 예를 들어 ICA로 분리한 A, B 두 출력 중에서 각 모음을 이용한 분류 결과 '아'의 경우에 A를 음성 신호로, '오'의 경우에도 A를 음성 신호로, '우'의 경우에는 B를 음성 신호로 판정했을 경우에 A로 판정한 결과가 B로 판정한 결과보다 많으므로 A가 음성 신호라고 분류한다. '합산'의 경우에는 '아', '오', '우'와의 대표 상관계수를 각각 구한 후 이를 다 더하여 그 결과를 다시 대표 상관값으로 선정하여 이 값이 더 큰 값을 음성 신호로 분류하는 방식이다. 앞에서 제시한 방법을 이용하여 분류한 결과 평균값

표 3. 평균값 방식으로 분류한 결과
Table 3. Results of classification by average.

구분	간격 방법	분류율(%)				
		100	200	500	1000	1600
동 일 인	아	40.7	33.3	58.0	53.1	37.0
	오	100.0	100.0	100.0	100.0	100.0
	우	98.8	98.8	100.0	98.8	100.0
	투표	100.0	98.8	100.0	98.8	100.0
	합산	100.0	100.0	100.0	100.0	100.0
타 인	아	51.9	46.9	45.7	35.8	64.2
	오	97.5	92.6	88.9	81.5	97.5
	우	87.7	90.1	69.1	60.5	87.7
	투표	93.8	88.9	70.4	59.3	95.1
	합산	96.3	95.1	87.7	71.6	96.3

방식의 합산 방법을 사용하고 분류 대상의 음성과 같은 사람 즉 동일인의 모음을 사용했을 때 가장 좋은 결과를 얻을 수 있었다(표 3). 그리고 이것은 이동 간격과 무관하게 나타났다. 동일한 방법으로 타인의 모음을 사용했을 때는 이동 간격이 100일 때 분류율 96.3%로 가장 좋은 결과를 얻고 이동 간격이 증가할 때 마다 분류율이 감소했지만 이동 간격이 1600 일 때 분류율이 96.3%로 갑자기 증가하였다. 하지만 흐름을 보면 이동 간격이 100일 때 가장 좋은 결과를 얻을 수 있을 것으로 기대할 수 있다. 분류 대상 음성과의 상관계수를 구할 때 사용하는 모음이 분류 대상 음성과 같은 화자로부터 나온 것이라면 당연히 상관계수가 더 높으리라고 기대할 수 있지만 일괄 방식의 표 1에서 보듯 '오'와 '우'에서 그렇지 않은 결과가 나왔고 평균값 방식의 표

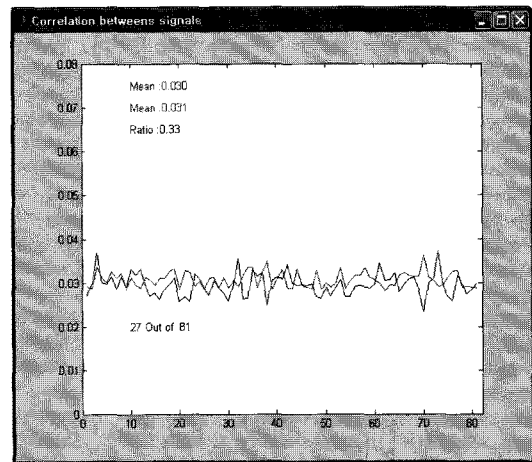


그림 3. 81개 문장 전체에 대한 상관계수 그래프 ('아', 평균값 방식)
Fig. 3. Correlation graph for 81 sentences. ('ah', by average)

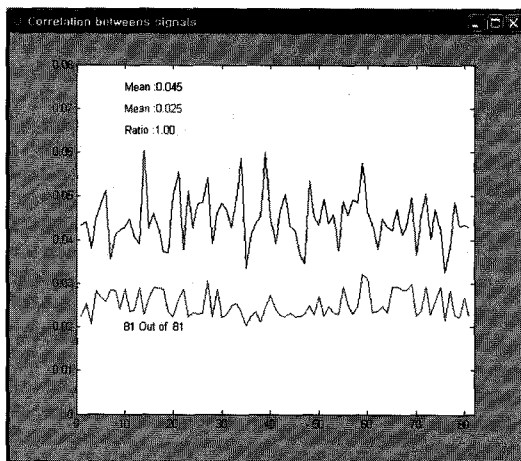


그림 4. 81개 문장 전체에 대한 상관계수 그래프 ('오', 평균값 방식)

Fig. 4. Correlation graph for 81 sentences. ('oh', by average)

3의 '아'에서도 기대에 반하는 결과를 볼 수 있다. 이는 동일인의 음성이라도 파형이 일정하지 않을 수 있고 또 분리된 음성에 섞여서 남아 있는 자동차 배기음의 영향일 수도 있으며 특정 모음의 파형이 배기음과 유사할 수도 있기 때문이다.

그림 3은 전체 81개 문장에 대해 평균값 방식으로 구하여 각 문장의 상관계수를 그래프로 그린 상관계수 그래프다. 분류율이 제일 좋지 않은 동일인 '아'에 대한 그래프로 200개 씩 이동하면서 상관계수를 구하여 평균값으로 분류한 결과이다. 총 81개의 음성 신호에 대해 상관계수를 구하여 그래프로 나타내었다. 그림 4는 같은 경우의 '오'에 대한 결과다. '아'에 대한 결과는 분류율이 33.3%인데 전체 81개 문장의 음성에 대한 상관계수의 평균값도 음성신호에 대해서는 0.030, 배기음에 대해서는 0.031로 거의 같은 값을 보여주고 그래프 상에서도 비슷한 영역에 겹쳐서 나타나므로 서로 구분하기 어려워 분류율이 상당히 낮다. 반면에 '오'의 경우에는 음성 신호와 배기음의 81개 문장에 대한 상관계수의 평균값이 0.045와 0.025로 상당히 많이 차이 나고 그래프 상으로도 확연하게 구분됨을 알 수 있다.

IV. 결 론

음성과 자동차의 배기음이 섞여 있을 때 이를 ICA로 분리한 후 분리된 신호에서 배기음과 음성 신호를 분류해 내는 방법을 제시하였다. 음성 신호는 음성 신호와의 상관계수가 클 것이라는 가정하에 잡음과 유사한 특

성을 지니고 있는 자음을 배제하고 유성음 중에서 둘 이상의 모음으로 구성된 복모음도 배제하면 단모음만 남는다. 단모음 중에서 '아', '오', '우'의 음성을 녹음하여 이 모음 신호들과 분리된 신호와의 상관계수를 구하여 계수가 더 큰 신호를 음성 신호로 분류하였다. 사용된 모음의 음성 신호로는 배기음과 섞인 음성 신호와 동일한 사람의 음성 신호, 다른 사람의 음성 신호를 모두 사용하였다.

상관계수를 계산하여 음성 신호와 배기음을 분류하기 위해 세가지 방식을 사용하였다. 계산에 사용된 모음은 분리된 음성 신호 및 배기음보다 지속 시간이 짧으므로 모음을 복제하여 음성 신호 및 배기음과 지속 시간이 같도록 만들어 상관계수를 구한 일괄 방식과 음성 신호 및 배기음 신호 내에서 주어진 간격으로 이동하면서 이동된 지점에서 값을 구하는 이동 방식으로 접근했다. 후자의 경우에 이동 간격에 따른 수만큼의 상관계수가 구해지는데 그 중에서 최대값을 갖는 상관계수를 대푯값으로 취하는 최대값 방식과 전체를 평균해서 대푯값으로 취하는 평균값 방식을 시도하였다. 각 방식에서 각 모음 '아', '오', '우'의 대푯값을 이용한 분류 후 그 결과를 이용하여 셋 중에 둘 이상의 판정 결과를 인정하는 투표 방법, 각 대푯값 세 개를 다 더한 후 그 값을 새로운 대푯값으로 사용하여 분류하는 방법 등 다섯 가지 방법을 적용하였다. 실험 결과 일괄 방식보다는 이동 방식이 더 좋은 결과를 보여주었으며 이동 방식의 경우에 합산 방법이 동일인 음성 및 타인의 음성에 대해서도 더 좋은 결과를 안겨주었다. 이동 간격이 좁을수록 좋은 결과를 보여 주었지만 간격이 좁으면 계산 시간이 많이 걸리므로 적절한 타협이 필요하다.

참 고 문 헌

- [1] J. F. Cardoso, "Blind signal separation: statistical principles," *Prod. IEEE*, vol. 9, no. 10, pp. 2009-2025, Oct., 1998.
- [2] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley and Sons, 2000.
- [3] P. Comon, "Independent component analysis, A new concept?," *Signal Processing*, vol. 36, pp. 287-314, 1994.
- [4] J. P. LeBlanc and P. L. De Leon, "Speech Separation by Kurtosis Maximization," *Proc. ICASSP*, vol. 2, pp. 1029-1032, 1998.

- [5] A. Hyvarinen, "Fast and Robust Fixed-Point Algorithms for Independent Component Analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626-634, May, 1999.
- [6] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 1991.
- [7] D. Obradovic, G. Deco, "Information Maximization and Independent Component Analysis: Is There a Difference?" *Neural Computing*, vol. 10, no. 8, pp. 2085-2101, Nov., 1998.
- [8] A. Hyvarinen, "Survey on independent component analysis," *Neural Computing Surveys* 2, pp. 94-128, 1999.
- [9] A. J. Bell and Terrence J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129-1159, 1995.
- [10] S. Amari, A. Cichocki, H. H. Yang, "A New Learning Algorithm for Blind Signal Separation", In *Advances in Neural Information Processing System 8*. Cambridge, MA:MIT Press, pp. 757-763, 1996.
- [11] S. Amari, "Natural Gradient Works Efficiently in Learning," *Neural Computation*, vol. 10, no. 2, pp 251-276, Feb., 1998.
- [12] 김선일, "잡음 섞인 한국어 인식을 위한 ICA 비교 연구", *대한전자공학회 논문지*, 제45권 1E편, 제3호, 2008년 9월.
- [13] 김우철 외 8인, *현대통계학*, 영지문화사, 1988.

— 저 자 소 개 —



김 선 일(정회원)
 1983년 아주대학교 전자공학과
 학사
 1985년 아주대학교 전자공학과
 석사
 1996년 아주대학교 전자공학과
 박사

1985년~1990년 한국기계연구원 선임연구원

1990년~현재 거제대학 교수

1997년 Visiting Professor, CAIP Center,
 Rutgers University, N.J., USA

2007년 Visiting Professor, Department of ECE,
 Georgia Institute of Technology, G.A.,
 USA

<주관심분야 : 신호처리, 음성 인식>