한국음향학회지 제32권 제1호 pp. 79~85 (2013)
The Journal of the Acoustical Society of Korea Vol.32, No.1 (2013)
http://dx.doi.org/10.7776/ASK.2013.32.1.079

pISSN : 1225-4428
eISSN : 2287-3775

CASA 시스템의 비모수적 상관 특징 추출을 이용한 목적 음성 분리

Target Speech Segregation Using Non-parametric Correlation Feature Extraction in CASA System

최태웅†, 김순협

(Tae-woong Choi† and Soon-Hyub Kim)

광운대학교 컴퓨터공학과

(접수일자: 2012년 9월 20일; 수정일자: 2012년 10월 25일; 채택일자: 2012년 11월 19일)

초 록: CASA 시스템의 특징 추출은 시간의 연속성과 채널 간 유사성을 이용하여 청각 요소의 상관지도를 구성하여 사용한다. 채널 간 유사성을 교차 상관 계수를 이용하여 특징 추출 할 경우 상관성을 정량적으로 나타내기 위해 계산량이 많은 단점이 있다. 따라서 본 논문에서는 특징 추출 시 계산 량을 줄이기 위한 방법으로 비모수적 상관 계수를 이용한 특징 추출 방법을 제안하고 이를 CASA 시스템을 통하여 목적 음성을 분리하는 실험을 수행하였다. 목적 음성의 분리 성능을 평가하기 위하여 신호 대 잡음비를 측정한 결과, 제안 방식이 기존 방식에 비해 평균 0.14 dB의 미세한 성능 개선을 보였다.

핵심용어: CASA 시스템, 특징 추출, 스피어만 상관계수, 목적 음성 분리, 음성 인식

ABSTRACT: Feature extraction of CASA system uses time continuity and channel similarity and makes correlogram of auditory elements for the use. In case of using feature extraction with cross correlation coefficient for channel similarity, it has much computational complexity in order to display correlation quantitatively. Therefore, this paper suggests feature extraction method using non-parametric correlation coefficient in order to reduce computational complexity when extracting the feature and tests to segregate target speech by CASA system. As a result of measuring SNR (Signal to Noise Ratio) for the performance evaluation of target speech segregation, the proposed method shows a slight improvement of 0.14 dB on average over the conventional method.

Key words: CASA system, Feature extraction, Spearman correlation Coefficient, Target speech segregation, Speech recognition
PACS numbers: 43.72. -p

I. 서 론

인간만이 가지는 능력 중 여러 음성이 혼합하여 잡음처럼 들리더라도 목적하는 소리만 집중하여 청취할 수 있는 인지 능력을 자동 음성 인식 시스템이나 음성 통신 시스템에 접목하기 위해 다양한 연구가 진행되고 있으며 목적하는 소리를 자동으로 분리하거나 배경적 잡음의 간섭을 억제하는 시스템을 개발하기 위한 연구가 계속되고 있다.[1]

목적 음성 분리를 위해서는 음성 신호와 간섭 신호의 본질 속성이 고려되어야 하며 목적 음성의 본질 속성을 분석하기 위해 인간의 청각 기관 인지 특성을 반영한 청각 장면 분석(ASA: Auditory Scene Analysis)[2]이 제기되었다. 청각 장면 분석은 음향 심리학적인 용어로 음향 환경에 대한 인지과정을 시각 정보의 인지 과정으로 청각 신호 인지 장면(scene)을 시각화하였다.

CASA(Computational Auditory Scene Analysis)[3] 시스템은 목적 음성을 분리하기 위하여 청각신경 분

†**Corresponding author:** Tae-woong Choi(dami73@kw.ac.kr)
Department of computer engineering, 447-1, Wolgye-Dong, Nowon-GU, Seoul, 139-701, Republic of Korea
(Tel: 82-2-940-5123, Fax: 82-2-940-8919)

석, 특징 추출, 세그먼테이션, 그룹화, 재합성 과정을 수행한다. CASA 시스템 과정 중 특징 추출 과정은 시간의 연속성과 채널 간 유사성을 이용한 청각 요소의 상관지도를 구성하여 사용한다. 채널 간 유사성을 교차 상관 계수를 이용하여 특징 추출을 하는 CASA 시스템은 상관성을 정량적으로 나타내기 위해 계산 량이 많은 단점이 있다.

따라서 본 논문에서는 특징 추출 시 계산 량을 줄이기 위한 방법으로 비모수적 상관 계수인 스피어만(spearman) 상관 계수를 이용한 특징 추출 방법을 제안하였으며 이를 CASA 시스템을 통하여 목적 음성 분리 실험을 수행하였다. 목적 음성의 분리 성능 평가를 위해 OHIO 주립대학 PNL에서 채집한 비음성 소리(non-speech sounds) 환경 잡음을 사용하여 깨끗한 음성과 혼합한 후 목적 음성 신호를 분리하여 신호 대 잡음비(SNR)를 측정한 결과 기존방법에 비해 평균 0.14 dB 증가로 제안한 방법이 기존방법에 비해 미세하나마 우수함을 보였다.

논문의 구성은 2장에서 CASA 시스템에 대해 설명하고 3장에서 비모수적 상관 계수인 스피어만 상관 계수를 이용한 특징 추출 방법을 설명하였다. 4장에서는 제안한 방법을 검증하기 위한 실험 환경과 실험 결과에 대해 기술하였으며 5장에서 결론 및 향후 계획에 대하여 이야기한다.

II. CASA 시스템

CASA 시스템은 음향심리학(psychoacoustic)과 생물학(biology)적 발견을 컴퓨터에서 구현하여 인간의 청각 처리 구조를 이해하고 인간의 청각 시스템과 동일하게 동작하는 자동 기계를 구현하는데 목적이 있으며 이를 실현하기 위하여 이상적인 마스크(ideal binary mask)를 설계하려는 목표를 가지고 있다.

CASA 시스템은 추상화 단계에서 인간의 청각 기관 기능을 모방하고 음향 신호의 이해를 장면 분석으로 해석하는 방법이며 Fig. 1과 같이 처리 절차를 나타낸다.[4]

청각 신경 분석(peripheral analysis)은 청각 기관을 모델링하는 과정이며 대역 통과 필터링, 프레임별 창 함수(windowing), 시간-주파수(T-F) 단위 표현을 거쳐 청각 장면으로 분해한다. 모델링에 이용되는 청각 기관은 외이(outer ear), 중이(middle ear), 내이(inner ear)로 세 부분으로 구분한다. 외이는 소리를 전달하는 관의 모양을 가지고 있으며 이를 통해 얻어진 음향 신호는 중이의 고막(tympanic membrane)을 진동한다. 이 신호는 세 개의 뼈(malleus, incus, stapes)를 통해 증폭 과정을 거쳐 내이의 달팽이관(cochlea) 내에 있는 유모 세포(hair cell)로 전달된 후 청각 신경(auditory nerve)을 통해 뇌로 전달된다.

신호와 시스템의 관점에서 외이와 중이는 신호의 강도가 거의 선형성을 보이고 있고 이곳에서 일어나는 공명(resonances)은 간단한 선형 필터로 모델링이 가능하며 선형 필터의 형태인 전 강조(pre- emphasis) 고역 통과 필터(high-pass filter)를 식(1)과 같이 표현하였다.

$$y(t) = x(t) - 0.95x(t-1). \tag{1}$$

$x(t)$는 입력 신호를 나타내며 t는 시간의 인덱스를 나타나고 $y(t)$는 출력을 나타낸다.

외이와 중이를 거친 소리는 달팽이관에 입력되어 달팽이관 내에 존재하는 유모 세포를 통해 특정 구간별로 주파수 응답 특성을 표현한다. CASA 시스템에서는 이를 모델링하기 위하여 감마톤 필터(gamma tone filter)[5]채널을 가진 ERB 필터 뱅크(Equivalent Rectangular Bandwidth filter bank)[6]를 사용한다. 감마톤 필터는 시간 영역에서 식(2) 와 같이 임펄스 응답 $h(t)$로 정의된다.

$$h(t) = kt^{n-1}\exp(-2\pi Bt)\cos(2\pi f_c t + \phi). \tag{2}$$

k는 출력 이득을 나타내고 B는 필터 대역폭을 나타내며 n은 필터의 차수를 나타낸다. f_c는 중심 주파

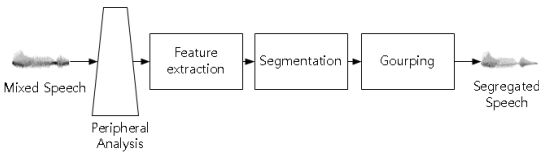

Fig. 1. Schematic diagram of a typical CASA system.

수를 나타내며 ϕ는 위상을 나타낸다. 감마톤 필터는 대역폭이 1.0109일 때와 필터 차수가 4일 때 청각 특성에 가장 잘 부합된다.

출력 이득 k는 중심 주파수에서 1로 설정하기 위하여 식(3)과 같이 변형하여 계산한다.

$$k = \frac{1}{\sum_{l=1}^{N} t^{n-1}\exp(-2\pi b ERB(f_c)t)}. \quad (3)$$

N은 임펄스응답의 길이를 나타내고 감마톤 필터 대역폭은 ERB 대역폭에 의해 결정된다. 필터의 주파수 응답$|H(f)|$와 최대 이득 $|H(f_{\max})|$이 주어졌을 때 ERB는 식(3)과 같이 나타낸다.

$$ERB = \frac{\int |H(f)|^2 df}{|H(f_{\max})|^2}. \quad (4)$$

ERB는 일정한 이득$|H(f_{\max})|$과 에너지를 갖는 사각(rectangular)필터의 대역폭과 감마톤 필터의 에너지가 같아지는 대역폭이다. ERB는 Glasberg와 Moore[7]의 제안을 따르고 있다.

$$E(f_0) = 24.7log_{10}(4.37f_0+1). \quad (5)$$

$E(f_0)$는 ERB 필터뱅크 개수를 나타내고 f_0는 필터의 중심 주파수를 나타내며 각 중심주파수는 최저 주파수와 최고 주파수 사이에 필터뱅크 채널수에 따라 분포한다. 감마톤의 ERB 필터뱅크를 거친 신호는 달팽이관 내에 존재하는 유모 세포 신호를 모델링한다.

외이, 중이, 내이의 청각 기관을 모델링하여 T-F 단위 계수를 하나의 장면으로 나타낸 달팽이관지도(cochleargram)를 Fig. 2와 같이 표현한다.

특징 추출 과정은 청각 기관을 모델링하여 얻은 달팽이관지도를 바탕으로 필터뱅크에 대한 주파수 응답 계수의 시간 축 상의 주기성(periodicity)과 주파수 축 상의 채널 간 주파수 유사성(similarity)을 얻는 과정이다.

세그먼테이션 과정은 특징 추출에서 얻은 상관지도를 바탕으로 T-F 단위를 세분화하는 과정이다. 세그먼트들은 신호들의 고유 속성을 표시(mark)하는 역할을 하고 이의 조합들은 신호를 마스킹(masking)하는 역할을 수행한다.

이후 동일 성격의 세그먼트들을 그룹화 과정과 재합성 과정을 거쳐 분리된 목적 음성을 얻는다.

III. 비모수적 상관 계수를 이용한 특징 추출

특징 추출은 필터뱅크에 대한 시간 축 상의 주기성과 채널 간 주파수 유사성을 얻기 위한 두 가지 과정으로 분리된다.

첫 번째 과정은 128채널에 대한 주파수 응답 계수의 시간 축 상의 주기성 특징을 찾는 방법으로 자기상관 계수(autocorrelation)[8] 함수(ACF)를 이용하며 식(6)과 같이 나타낸다.

$$ACF(t,f,\tau) = \sum_{i=0}^{\infty} r(t-T,f)r(t-T-\tau,f)w(T). \quad (6)$$

식(7)은 식(6)에 대한 정규화된 응답을 나타낸다.

$$a_n(t,f,\tau) = \frac{ACF(t,f,\tau)}{ACF(t,f,0)}. \quad (7)$$

채널 주파수 f에서 지연 τ에 사각 윈도우 함수 $\omega(T)$를 곱하여 연산하며 윈도우 크기는 20 ms를 10 ms 간격으로 계산하여 주기성을 표현하는 완성된

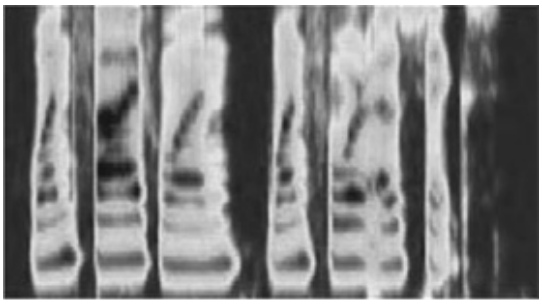

Fig. 2. Cochleagram.

장면의 상관지도로 표현된다.

Fig. 3은 자기 상관 계수를 이용한 자기상관지도 구조와 장면을 나타낸다.

두 번째 과정은 주파수 축 상의 채널 간 주파수 유사성 판단을 위해 교차 상관(cross correlation)[8] 계수를 이용하며 자기 상관의 입력으로 채널 간의 상관 정도를 계산하여 채널 간에 주기 패턴의 유사성을 판단한다. 식(8)은 채널의 중심 주파수 f_1, f_2의 상관도를 나타낸다.

$$CCF(f_1,f_2,t) = \frac{2\sum_{\tau} a_n(t,f_1,\tau)a_n(t,f_2,\tau)}{\sum_{\tau} a_n(t,f_1,\tau)^2 + \sum_{\tau} a_n(t,f_2,\tau)^2} \quad (8)$$

$a_n(t,f,\tau)$는 자기 상관 계수를 나타내고 계산된 계수는 정규화를 통해 0~1사이의 값을 갖는다. 채널의 주기 패턴에 대한 상관이 높으면 1에 가까운 값을 나타내고 상관이 낮으면 0에 가까운 값을 나타낸다.

교차 상관 계수를 이용하여 특징 추출을 하면 채널과 채널을 교차하여 비교하므로 비교 횟수에 따라 계산량이 결정되고 비교 횟수의 증가는 계산량의 증가로 나타난다. 따라서 비모수적 상관관계인 스피어만 상관 계수를 이용하여 채널별로 순위를 정하여 비교 횟수를 줄이는 방법을 사용하였다.

비모수적 상관인 스피어만의 로우ρ(rho)[9]는 측정치 변수나 순서 형 변수들의 상관관계를 자료의 순위 값에 의하여 계산하는 방법으로 순서 형 변수들의 상관관계를 계산한다. 데이터의 값 대신 순위를 이용하는 상관 계수이며 데이터를 작은 것부터 차례로 순위를 정하여 서열 순서로 바꾸어 상관 계수를 구하는 방법이다. 두 채널간의 연관 관계가 있는지 없는지를 밝혀주며 데이터에 이상점이 있거나 표본 크기가 작을 때 유용하게 사용된다. 자기 상관 지도를 바탕으로 식(9)와 같이 유사도를 측정한다.

$$r = 1 - \frac{6\sum_{i=0}^{n}(R(x_i)-R(y_i))^2}{n(n^2-1)}. \quad (9)$$

정렬을 통하여 순위를 얻은 채널 변수 x의 i번째 자기 상관 계수의 순위를 $R(x_i)$로 나타내고 관측한 채널 변수 y의 i번째 자기 상관 계수의 순위를 $R(y_i)$로 나타내며 전체 서열 순위는 r에 의해 구해진 순위 값으로 결정된다. n은 연산을 위한 전체 채널을 나타낸다. 두 채널 사이의 상관 계수 연산은 프레임별 각 채널의 자기 상관의 지연을 대상으로 하고 음성의 기본 주파수($F0$)가 존재하는 구간인 500 Hz 범위까지 수행한다.

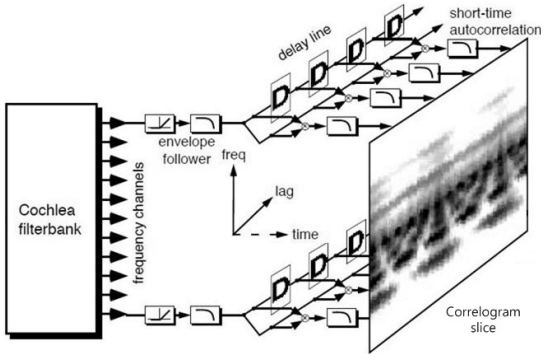

Fig. 3. Structure of autocorrelation correlogram and scene.

IV. 실험 및 고찰

교차 상관 과 비모수상관의 특징 추출의 음성 분리성능 평가를 위하여 OHIO대학 PNL(Perception and Neurodynamics Laboratory)의 비음성 소리(non-speech sounds)[10] 와 깨끗한 음성인 ETRI 445 PBW 음성데이터베이스의 "가운데", "과일", "교양", "금융", "우주" 발성을 각각 혼합하였으며 혼합된 잡음은 신호 대 잡음비를 5dB로 고정하여 목적 음성 분리 실험을 수행하였다. Table 1에는 OHIO대학 PNL 잡음 유형중 실험에 사용한 잡음 유형을 나타내었다.

교차상관과 비모수상관 특징추출에 대해 세그먼테이션을 수행하여 얻은 세그먼트들의 지도를 통하여 세그먼트들의 분리된 차이를 Fig. 4에 나타내었다.

음성 신호 분리 성능 실험을 위하여 그룹화 과정을 수행하고 재합성을 통하여 얻어진 목적 음성 분리를 수행하여 Fig. 5에 나타내었다. Fig. 5에 표현된 신호는 목적음성 분리 후 "가운데" 발성의 교차상관과 비모수상관 특징 추출에 대한 비교를 위한 그림이다.

Table 1. Types of noise in PNL noises database.

No	Type	Description
1	Crown	N1-N17
2	Machine	N18-N29
3	Alarm and siren	N30-N43
4	Traffic and car	N44-N46
5	Animal	N47-N55
6	Water	N56-N69
7	Wind	N70-N78
8	Bell	N79-N82
9	Cough	N83-N85
10	Clap	N86
11	Snore	N87
12	Click	N88
13	Laugh	N89-N90
14	Yawn	N91-N92
15	Cry	N93
16	Shower	N94
17	Tooth brushing	N95
18	Footsteps	N96-N97
19	Door moving	N98
20	Phone dialing	N99-N100

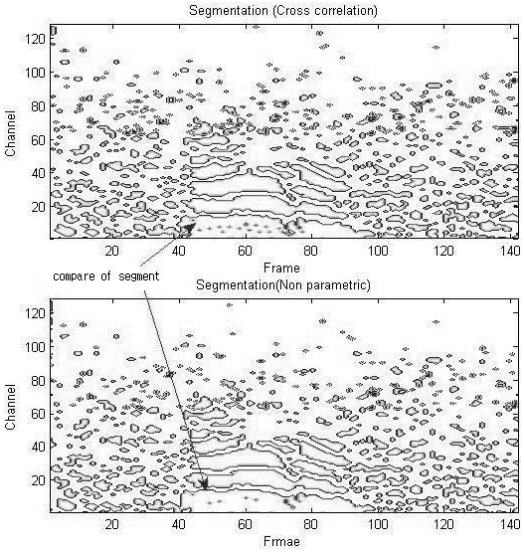

Fig. 4. Comparison of cross correlation feature and non-parametric correlation feature in segmentation.

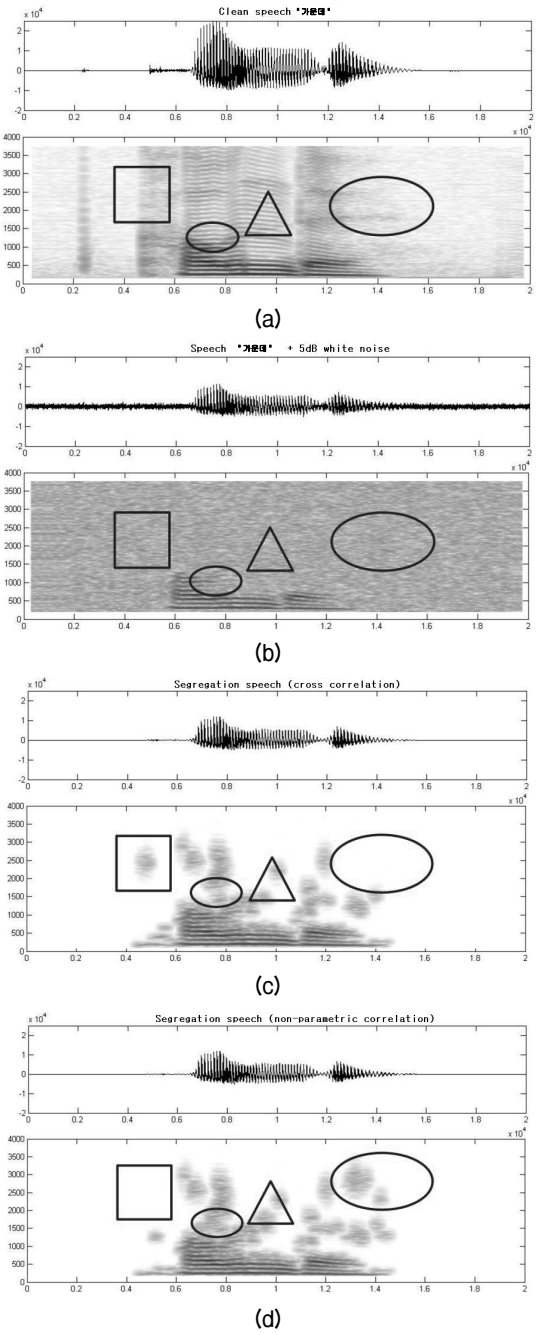

Fig. 5. Comparison of cross correlation feature and non-parametric feature in speech segregation (a) clean speech (b) noisy speech (c) segregated speech using cross correlation (d) segregated speech using non-parametric correlation.

Fig. 5에서 표시되어진 4부분으로 구분하여 교차상관 특징 추출 방법과 비모수상관 특징 추출 방법을 시각과 청각을 이용하여 단순 비교한 결과 'ㅏ', 'ㅜ'에 해당하는 작은 원으로 표시된 부분에서는 교차상관 특징 추출 방법이 비모수상관 특징 추출 방법에 비해 분리 성능이 좀 더 우수하게 나타났으며 "가"에서 'ㄱ'에 해당하는 사각형으로 표시된 부분에서는 비모수상관 방법이 좀 더 우수한 분리 성능

을 나타내었다. “가운데” ‘ㄴ’에 해당하는 삼각형으로 표시된 부분에서는 교차상관과 비모수적 상관 특징 추출 모두 유사한 분리 성능을 나타내었으며 큰 원으로 표시된 부분에서는 교차상관 특징 추출 방법이 좀 더 우수한 분리 성능을 나타내었다.

단순 비교에 대한 신뢰성을 검증하기 위해 분리 성능 평가의 정량적 측정인 신호 대 잡음비[11]를 이용하여 성능 평가를 수행하였다.

$$SNR(dB) = 10\log_{10}\frac{\sum_{n=0}^{N-1} x^2(n)}{\sum_{n=0}^{N-1}[x(n)-\hat{x}(n)]^2}. \qquad (10)$$

정량적 측정을 위한 신호 대 잡음비를 식(10)에 나타내었으며 $x(n)$은 잡음과 혼합되기 전 깨끗한 음성을 나타내고 $\hat{x}(n)$은 깨끗한 음성과 잡음이 혼합된 신호를 입력으로 얻어진 분리 음성을 나타낸다.

Table 2 와 Table 3은 목적음성 분리 후 “가운데”, “금융” 발성의 특징 추출에 따른 신호 대 잡음비의 성능 평가를 나타낸다.

Table 2, Table 3에서와 같은 방법으로 “과일”, “교양”, “우주” 발성의 특징 추출에 따른 신호 대 잡음비의 성능을 평가하기 위한 실험을 수행하였으며 그 결과를 Table 4에 나타내었다.

목적음성 분리 전 5 dB의 신호를 이용하여 분리 실험을 수행한 후 교차 상관의 신호 대 잡음비 평균이 14.72 dB 로 9.72 dB 향상됨 을 보였고, 제안한 방법의 신호 대 잡음비 평균이 14.86 dB로 9.86 dB 향상됨을 보였다. 목적음성 분리 수행 후 두 방법 모두 신호 대 잡음비가 향상되었고 교차상관 과 제안한 방법의 특징추출을 비교한 결과 제안한 비모수상관을 이용한 특징 추출 방법이 0.14 dB 향상을 보여 미

Table 2. SNR comparison of the conventional method and the proposed method for the utterance “가운데”.

Noise NO	cross correlation (a)	proposed correlation (b)	difference (b)-(a)
	SNR(dB)	SNR(dB)	SNR(dB)
1	13.82	14.27	0.45
2	11.55	11.95	0.40
3	15.95	15.78	-0.17
4	11.54	11.84	0.29
5	11.67	11.68	0.01
6	13.76	13.74	-0.02
7	11.33	11.35	0.02
8	20.77	20.20	-0.57
9	13.74	12.12	-1.62
10	15.76	16.12	0.35
11	10.96	11.89	0.92
12	15.01	16.50	1.49
13	17.42	16.51	-0.91
14	14.33	14.79	0.47
15	20.05	20.09	0.04
16	19.11	19.28	0.17
17	15.19	15.44	0.25
18	18.31	18.40	0.10
19	11.51	11.44	-0.07
20	19.31	19.73	0.41
average	15.05	15.15	0.10

Table 3. SNR comparison of the conventional method and the proposed method for the utterance “금융”.

Noise NO	cross correlation (a)	proposed correlation (b)	difference (b)-(a)
	SNR(dB)	SNR(dB)	SNR(dB)
1	15.39	15.19	-0.20
2	12.79	12.79	0.00
3	15.80	16.05	0.25
4	12.51	12.73	0.22
5	15.86	15.89	0.04
6	14.08	14.06	-0.03
7	11.35	11.63	0.27
8	12.86	13.04	0.18
9	16.70	17.14	0.44
10	17.09	17.56	0.47
11	12.50	12.51	0.01
12	16.10	16.95	0.85
13	15.02	14.99	-0.03
14	18.11	18.46	0.35
15	19.46	19.37	-0.08
16	17.20	17.90	0.70
17	14.08	14.13	0.05
18	15.52	15.69	0.17
19	11.06	11.04	-0.02
20	9.45	9.45	0.00
average	14.65	14.83	0.18

Table 4. Overall SNR comparison of the conventional method and the proposed method.

utterance	cross correlation (a)	proposed correlation (b)	difference (b)-(a)
	SNR(dB)	SNR(dB)	SNR(dB)
가운데	15.05	15.15	0.10
과일	15.15	15.26	0.11
교양	14.72	14.88	0.16
금융	14.65	14.83	0.18
우주	14.04	14.19	0.15
average	14.72	14.86	0.14

세하나마 기존방식에 비해 우수함을 보였다.

V. 결론 및 향후계획

본 논문은 채널 간 유사성을 교차 상관 계수를 이용하여 특징 추출을 하는 CASA 시스템에서 계산 량을 줄이기 위한 방법으로 비모수적 상관 계수인 스피어만 상관 계수를 이용한 특징 추출 방법을 제안하여 이를 CASA 시스템을 통해 목적 음성 분리 실험을 수행하였다. 목적 음성의 분리 성능 평가를 위해 OHIO 주립대학 PNL 100 non-speech sounds 환경 잡음과 깨끗한 음성을 혼합한 후 목적 음성 신호를 분리하여 신호 대 잡음비를 측정 하였다. 실험 결과 제안한 비모수상관 특징 추출이 교차상관 특징 추출보다 신호 대 잡음비가 0.14 dB 증가하여 제안한 방법을 통해 얻은 특징 추출이 목적 음성 분리 성능에서 미세하나마 더 우수함을 확인할 수 있었다. 향후 다양한 환경잡음과 음성, 다수화자 혼합 음성을 대상으로 제안하는 방법을 평가할 계획이다.

참고문헌

1. S. M. Naqvi, M. Yu and J. A. Chamber, "A multimodal approach to blind source separation of moving sources," IEEE Trans. Signal Process. **4**, 895-910 (2010).
2. A. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound* (Cambridge, MIT Press, USA: MA, 1990).
3. Y. Shao, S. Srinivasan, Z. Jin, and D. L. Wang, "A computational auditory scene analysis system for robust speech recognition," Comput. Speech Lang. **24**, 77-93 (2010).
4. P. Li, Y. Guan, B. Xu, and W. Liu "Monaural speech separation based on computational auditory analysis and objective quality assessment of speech," IEEE Trans. Audio Speech Lang. Process. **14**, 2014-2022 (2006).
5. A. P. Klapuri, "Multipitch analysis of polyphonic music and speech signals using an auditory model," IEEE Trans. Audio Speech Lang. Process. **16**, 255-266 (2008).
6. L. Lin and E. Ambikairajah, "Auditory filterbank inversion," in Proc. ISCAS 2001, 537-540 (2001).
7. B. R. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," Hearing Research **47**, 103-138 (1990).
8. G. Hu, and D. L. Wang, "A tandem algorithm for pitch estimation and voiced speech segregation," IEEE Trans. Audio Speech Lang. Process. **18**, 2067- 2079 (2010).
9. S. Y. Cho, D. M Sun and Z. D. Qiu, "A spearman correlation coefficient ranking for matching-score fusion on speaker recognition," in Proc. TENCON, 736-741 (2011).
10. G. Hu, Perception and Neurodynamics Laboratory, http://www.cse.ohio-state.edu/pnl/corpus/, 2010.
11. D. L. Wang and G. J. Brown, "Separation of speech from interfering sounds based on oscillatory correlation," IEEE Trans. Neural Networks **10**, 684-697 (1999).

저자 약력

▸ 최 태 웅(Tae-woong Choi)

2001년 2월 : 호원대학교 컴퓨터공학과 학사
2003년 2월 : 광운대학교 컴퓨터공학과 석사
2003년 2월 ~ 현재 : 광운대학교 컴퓨터 공학과 박사과정
〈관심분야〉 음성 인식, 음성 신호 분리

▸ 김 순 협(Soon-Hyub Kim)

1974년: 울산대학교 전자공학과 학사
1976년: 연세대학교 대학원 전자공학과 석사
1983년: 연세대학교 대학원 전자공학과 박사
1979년 ~ 현재: 광운대학교 컴퓨터공학과 교수
〈관심 분야〉 음성인식