

논문 2014-51-5-21

계산적 청각 장면 분석 시스템에서 가중치 상호상관계수를 이용한 음성 분리

(Speech Segmentation using
Weighted Cross-correlation in CASA System)

김 정 호*, 강 철 호**

(JungHo Kim[©] and ChulHo Kang)

요 약

계산적 청각 장면 분석 시스템의 특징 추출은 시간 연속성과 주파수 채널간에 유사성을 이용하여 청각 요소의 상관지도를 구성한다. 세그멘테이션은 상호상관계수 함수를 이용하여 2진 마스크를 구성하고, 마스크 성분 1(음성)은 동일한 주기성과 동기를 가진다. 그러나 채널간에 비슷한 주기성을 갖지만 지연이 있는 경우에 음성으로 잘못 결정되는 문제가 있다. 본 논문에서는 세그멘테이션에서 가중치 상호상관계수를 이용해 채널간에 유사성의 변별력을 높이는 방법을 제안한다. 계산적 청각 장면 분석 시스템의 음성 분리 성능을 평가하기 위하여 배경 잡음(사이렌, 기계, 백색, 자동차, 군중) 환경에서 신호 대 잡음비(5dB, 0dB)의 변화에 따라 실험을 수행하였다. 본 논문에서는 기존의 방법과 제안한 방법과 비교한 결과, 제안한 방법이 기존의 방법에 비하여 각각 신호 대 잡음비 5dB에서 2.75dB 그리고 0dB에서 4.84dB 향상되었다.

Abstract

The feature extraction mechanism of the CASA(Computational Auditory Scene Analysis) system uses time continuity and frequency channel similarity to compose a correlogram of auditory elements. In segmentation, we compose a binary mask by using cross-correlation function, mask 1(speech) has the same periodicity and synchronization. However, when there is delay between autocorrelation signals with the same periodicity, it is determined as a speech, which is considered to be a drawback. In this paper, we proposed an algorithm to improve discrimination of channel similarity using Weighted Cross-correlation in segmentation. We conducted experiments to evaluate the speech segregation performance of the CASA system in background noise(siren, machine, white, car, crowd) environments by changing SNR 5dB and 0dB. In this paper, we compared the proposed algorithm to the conventional algorithm. The performance of the proposed algorithm has been improved as following: improvement of 2.75dB at SNR 5dB and 4.84dB at SNR 0dB for background noise environment.

Keywords : CASA, Binary Mask, VAD

* 정회원, 광운대학교 전자통신공학과
(Department of Electronics and Communication
Engineering, Kwangwoon University)

© Corresponding Author(E-mail: anpaul@kw.ac.kr)

※ 이 논문은 2012년도 광운대학교 교내 학술연구비에
의하여 연구되었음.

접수일자: 2014년2월19일, 수정일자: 2014년4월22일

수정완료: 2014년4월30일

I. 서 론

최근 편리한 정보 시스템의 인터페이스 수단으로 음성 인식 시스템의 많은 발전이 이루어지고 있다. 그러나 음성 인식 시스템 연구의 가장 큰 어려움은 잡음으로 인식률이 감소된다. 음성 인식 성능 향상을 위해 잡

음 제거와 음성 향상 알고리즘을 이용한다. 예로 적응 필터(adaptive filter), 독립 성분 분석(ICA: Independent Component Analysis)과 계산적 청각 장면 분석(CASA: Computational Auditory Scene Analysis)등 다양한 방법들이 연구되고 있다.

인간의 청각 시스템은 청각 장면 분석(Auditory Scene Analysis)^[1]을 통해 배경 잡음(Background noise)이나 여러 사람들이 동시에 말하는 상황에서도 특정 목적을 가지는 음성 신호를 청취할 수 있는데 이러한 현상을 칵테일 파티 효과(Cocktail party effect)라고 한다.

음성 인식 시스템의 성능 향상을 위하여 배경 잡음 또는 칵테일 파티 문제로부터 음성 분리(Speech Segregation)는 중요한 애플리케이션으로 사용된다. 적응 필터^[2]는 잡음 억압으로 특정 음성 신호를 청취할 수 없고, 통계적인 방법을 이용한 독립 성분 분석^[3]는 두 개 이상의 마이크로 환경적인 영향을 받는다. 하지만 계산적 청각 장면 분석 시스템은 단일 마이크(monaural microphone) 환경에서 청각 장면 분석 원리를 기반으로 음성 분리를 한다. 음성 분리를 수행하려면 음성 신호의 본질적 속성을 분석해야 한다. 음향심리학(Psychoacoustic)과 생물학적(Biology) 발견에 기반을 둔 계산적 청각 장면 분석 시스템은 다양한 잡음에 대한 사전적 정보 및 학습 없이 잡음이 혼합된 음성을 분리한다^[4].

계산적 청각 장면 분석 시스템은 혼합된 음성 신호를 시간-주파수(Time-Frequency) 단위로 분해한다. 세그먼테이션(Segmentation)에서 2진 마스크(Binary Mask)로 변환하고, 그룹핑(Grouping)에서 같은 소스에 있는 세그먼트 그룹을 동일한 스트림(stream)으로 구성한다. 계산적 청각 장면 분석 시스템의 목적은 이상적인 2진 마스크(Ideal Binary Mask) 구성이다^[5]. 2진 마스크는 시간-주파수 영역에서 상호상관계수와 임계값을 비교해 마스크 성분 0(잡음 또는 간섭신호)과 1(고조파 또는 음성 신호)을 결정한다.

본 논문에서는 세그먼테이션 과정에서 2진 마스크를 구성할 때 가중치 상호상관계수를 이용하여 채널간 유사성의 변별력을 높이는 방법을 제안한다. 제안한 방법의 성능 평가는 백색잡음과 100 non-speech sounds를 이용하여 신호 대 잡음비(SNR 0~5dB)의 변화에 따라 위너 필터(Wiener filter), 기존의 계산적 청각 장면 분

석 방법(Hu and Wang)^[6], 기존의 계산적 청각 장면 분석 시스템의 청각 장면과 PAR를 이용한 음성 검출 방법^[7]과 제안한 방법의 성능을 비교하였다.

논문의 구성은 다음과 같다. II장에서 계산적 청각 장면 분석 알고리즘 대해 알아보고 III장에서는 제안한 알고리즘에 대해 소개한다. IV장은 실험 결과에 대해 기술하며, V장에서 결론을 내린다.

II. 계산적 청각 장면 분석 알고리즘

2.1 청각 신경의 신호 분석

외이와 중이를 통해 증폭된 신호는 달팽이관 안에 있는 세포막의 유모 세포(Hair Cell)에서 가청 주파수 영역별로 인식한다.

사람의 청각 기관(외이, 중이, 달팽이관)을 모델링하기 위해 계산적 청각 장면 분석 시스템은 감마톤 필터(Gammatone filter) 채널을 가진 ERB 필터뱅크(Equivalent Rectangular Bandwidth filterbank)를 사용한다. 입력 신호는 감마톤을 갖는 128 채널로 분해되고, ERB 필터뱅크의 중심 주파수는 50~5000Hz 대역을 가진다. ERB 필터뱅크는 Glasberg와 Moore 알고리즘을 사용한다^[8].

$$g(t, f_c) = t^{n-1} e^{-2\pi Bt} \cos(2\pi f_c t + \phi) \quad (1)$$

식(1)에서 B는 필터 대역폭, n은 필터 차수이고 n=4를 사용하였다. f_c 는 중심 주파수를 나타낸다.

청각 분석(Auditory Analysis)에서 사람의 청각 모델인 ERB 필터뱅크를 이용해 각 채널 별로 주파수 분석을 할 수 있다. 각 필터 채널 출력의 프레임은 20ms이고 10ms 이동한다. 시간-주파수(T-F)로 표현된 감마톤 ERB 필터뱅크 응답을 코클리어그램(Cochleagram)이라고 한다.

2.2 특징 추출

특징 추출(Feature Extraction)은 청각 신경 분석 후 얻어진 청각 신호로부터 고차원의 표현을 통해 중간단계 정보로 나타내고, 청각 신호의 특징 정보를 매핑(Mapping)하는 과정이다. 청각 신경 신호 분석 데이터를 바탕으로 시간축의 주기성과 주파수축의 유사성을 얻기 위한 과정으로 자기상관 함수(Autocorrelation

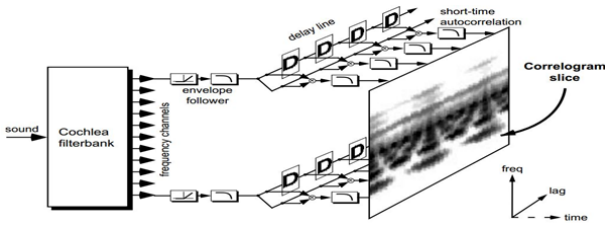


그림 1. 자기상관지도 구성과 장면
Fig. 1. Structure of autocorrelation correlogram and scene.

function)와 상호상관 함수(Cross-correlation function)를 이용한다.

첫째 각 채널의 프레임에서 시간 지연(time delay 또는 lag time)의 일정 기간내에 주기성(periodicity)을 찾는 방법으로 자기상관 함수를 계산한다.

$$A(c, m, \tau) = \frac{\sum_{n=0}^{N-1} h(c, mT+n)h(c, mT+n+\tau)}{\sqrt{\sum_{n=0}^{N-1} h^2(c, mT+n) \sum_n h^2(c, mT+n+\tau)}} \quad (2)$$

$A(c, m, \tau)$ 는 자기상관 함수이다. n 은 이산시간, c 는 채널, m 은 프레임이다. $h(c, t)$ 는 유모 세포(Hair cell) 응답이다. $T=10ms$ 는 프레임 이동(frame shift)이고, N 은 T 의 2배인 $20ms$ 이다. τ 는 시간 지연이고, τ 영역은 $[0ms \sim 12.5ms]$ 이다.

그림 1은 청각 신경의 발화 패턴에 대한 자기상관인 상관지도(Correlogram)로 표현한다. 상관지도는 피치 현상을 잘 모델링 할 수 있는 방법이다.

둘째 상호상관 함수는 주파수축 상에서 채널 간의 동일한 음성(또는 주기성) 성분에 대한 유사성을 측정한다.

$$C(c, m) = \frac{\sum_{\tau} [A(c, m, \tau) - \overline{A(c, m)}] [A(c+1, m, \tau) - \overline{A(c+1, m)}]}{\sqrt{\sum_{\tau} [A(c, m, \tau) - \overline{A(c, m)}]^2 \sum_{\tau} [A(c+1, m, \tau) - \overline{A(c+1, m)}]^2}} \quad (3)$$

$C(c, m)$ 는 상호상관 함수이다. $A(c, m, \tau)$ 는 자기상관 함수이고, $\overline{A(c, m)}$ 는 자기상관 함수의 평균이다. τ 영역은 $[0ms \sim 12.5ms]$ 이다.

2.3 계산적 청각 장면 분석 시스템

계산적 청각 장면 분석 시스템은 음향심리학과 생물학적 발견을 기반으로 인간 청각 시스템의 기능을 모방

하고, 음향 신호의 이해를 장면 분석으로 해석한다. 계산적 청각 장면 분석 시스템의 처리 과정은 크게 세그먼테이션(Segmentation)와 그룹핑(Grouping)로 구분한다. 세그먼테이션 과정은 음성 신호를 시간-주파수 영역으로 분해하고, 분해된 청각 특징 정보를 2진 마스크로 변환한다. 그룹핑은 같은 소스에 있는 세그먼트들을 하나의 스트림으로 구성한 후 음성 세그먼트 그룹과 잡음 세그먼트 그룹으로 분리한다. 잡음이 분리된 음성 신호 정보를 시간영역의 파형으로 재합성한다.

III. 제안한 알고리즘

계산적 청각 장면 분석 시스템의 목적은 이상적인 2진 마스크를 구성하는 것이다. 세그먼테이션에서 시간 연속성(time continuity)과 상호상관 함수를 이용해 2진 마스크를 구성한다. 마스크 성분 1은 resolved(음성 또는 고조파)와 0은 unresolved(잡음 또는 간섭)로 표현한다. 그러나 실제 환경에서 음성은 잡음에 영향을 받기 때문에 이상적인 2진 마스크 구성에 어려움이 있다. 또한 배경 잡음 환경에서 세그먼테이션을 구성할 때 문제점이 있다. 첫 번째, 프레임은 10ms 씩 이동하면서 일부 음성과 잡음은 중복(overlap)된다. 두 번째, 잡음이 음성과 같은 주기성을 가지면 음성으로 결정된다. 그림

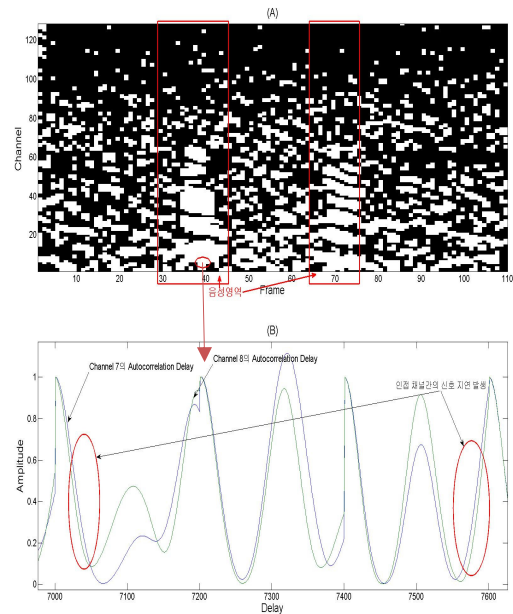


그림 2. 세그먼테이션에서 자기상관 함수 지연 비교
Fig. 2. Autocorrelation delay comparison in segmentation.

2(B)와 같이 채널간에 비슷한 주기성을 가지면서 지연 시간(Delay time) 있는 경우에 음성으로 결정된다. 이러한 문제로 계산적 청각 장면 분석 시스템은 성능 감소로 이어진다.

그림 2(A)는 백색잡음이 혼합된 음성 신호(SNR 0dB)의 2진 마스크이다. 흰색은 음성이고, 검정색은 잡음이다. 음성 영역은 직접 레이블(Label) 하였다. 그림 2(B)는 그림 2(A)의 음성 영역 35~37 프레임과 채널 7과 8을 확대한 자기상관 신호의 지연 비교이다. 그림 2(B)는 두 신호가 비슷한 주기성을 갖지만 신호간에 지연이 있다. 2진 마스크는 상호상관 계수와 임계값(0.985)을 비교해 음성과 잡음을 결정하는데 그림 2와 같이 채널간에 신호 지연이 발생한 경우에 잡음을 음성으로 잘못 결정되는 문제점이 있다.

3.1 제안한 알고리즘

세그먼테이션은 상호상관계수와 임계값을 비교해 2진 마스크를 구성한다. 마스크 성분 1인 음성은 채널간에 두 신호의 지연이 없고(동기), 같은 주기성을 가진다. 그러나 그림 2(B)와 같이 비슷한 주기성을 가지면서 두 신호간에 지연이 있는 경우에 잡음(또는 간섭)이 음성으로 잘 못 결정되는 문제가 있다. 본 논문에서는 세그먼테이션 과정에서 상호상관 함수의 유사성에 대한 변별력을 높이는 방법을 제안한다.

그림 3은 제안한 방법을 혼합한 계산적 청각 장면 분석 시스템의 음성 분리 구조이다.

그림 4는 제안한 가중치가 적용된 상호상관 함수 알고리즘 흐름도이다. 채널간 두 신호가 비슷한 주기성을 가지면서 지연이 있는 경우에 마스크 성분은 잡음이 아닌 음성으로 결정되는 문제를 개선하기 위해 채널간 지연 시간(d: delay time) 정보에 따라 가중치가 적용된

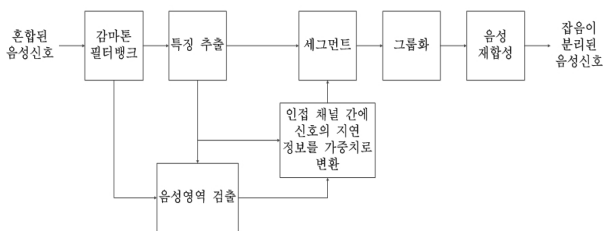


그림 3. 제안한 계산적 청각 장면 분석 시스템의 음성 분리 구조

Fig. 3. Speech segregation schematic of proposed CASA system.

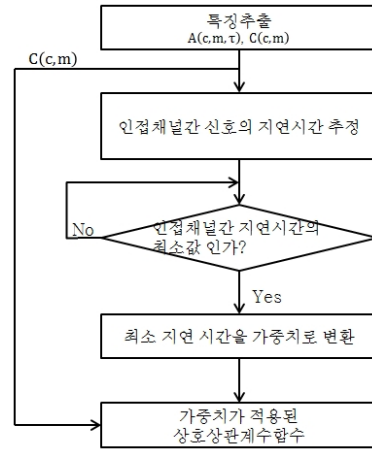


그림 4. 제안한 가중치가 적용된 상호상관 함수 알고리즘 흐름도

Fig. 4. Flow chart of proposed Weighted Cross-correlation algorithm.

상호상관 함수를 제안한다.

다 채널 마이크로폰 어레이(Microphone array)는 다양한 환경에서 음향 소스(Acoustic source) 추적(Tracking)과 위치(Location)를 알기 위해 시간 지연 추정(Time Delay Estimation) 알고리즘을 사용한다. 채널간에 신호의 지연 정보를 추정하기 위해 시간 지연 추정 알고리즘을 이용하여 지연에 따라 가중치로 변환하여 상호상관 함수에 적용하였다.

본 논문에서는 다양한 시간 지연 추정 알고리즘 중에서 계산 복잡성, 정밀도와 정확도를 고려해 AMDF (Average Magnitude Difference Function)와 ASDF (Average Square Difference Function) 알고리즘을 이용해 채널간의 시간 지연 정보를 추정한다^[9]. 최소 에러(Minimum error)를 기반으로 하는 AMDF와 ASDF 알고리즘은 잡음이 없는 경우에 두 신호 사이 지연 정보를 정확하게 찾을 수 있다. 본 논문에서는 AMDF 알고리즘을 이용한다.

$$R_{AMDF}(c, m, d) = \frac{1}{T-d} \sum_{\tau=d}^T |A(c, m, \tau) - A(c+1, m, \tau+d)| \quad (4)$$

$$D_{AMDF}(c, m) = \underset{d}{\operatorname{argmin}} R_{AMDF}(c, m, d) \quad (5)$$

식(4)은 자기상관 함수로부터 채널간에 지연에 따라 값을 계산한다. τ 는 자기상관 함수의 지연이고, d 는 채널간 두 신호의 지연 시간이다. 식(1)에서 자기상관 함수의 지연(τ)과 식(4)에서 지연(d)으로 구분한다. d 영

역은 [0ms,1ms]이고, T는 자기상관 함수의 최대 지연 (12.5ms)이다.

식(5)는 $R_{AMDF}(c,m,d)$ 을 이용하여 각각의 채널과 프레임에서 최소 지연 값을 가지는 위치를 찾는다.

$$w_{AMDF}(c,m) = e^{-\frac{1}{2} \left(\frac{D_{AMDF}(c,m) - \mu}{\sigma} \right)^2} \quad (6)$$

식(6)은 채널간 지연 시간 정보에 따라 가우시안 분포 함수를 이용해 가중치를 계산한다. 여기서 가우시안 분포 함수의 평균($\mu=0$) 이고, 표준편차($\sigma=4$)이다. 가우시안 분포 함수는 지연($D_{AMDF}(c,m)=0$)이 없는 경우에 최대 1 값을 가진다. 지연이 있는 경우에는 시간에 따라 최대 1 보다 작은 값을 가지게 된다. 그러나 전체 채널과 프레임에서 제안한 방법을 계산하려면 많은 계산량이 필요하다. 그림 3에서 제안한 방법을 음성 영역만 적용하면 계산량을 감소 할 수 있다. 청각 장면과 PAR를 이용한 음성 검출 방법^[7]을 제안한 방법에 적용하였다. 청각 장면과 PAR를 이용한 음성 검출 방법은 코클리어그램의 주기 성분과 비주기 성분으로 분해 후 통계적 모델을 이용한 음성 영역을 검출 방법으로 신호 대 잡음비 변화에도 민감하지 않고 우수한 음성 영역 검출 성능을 보여주고 있다.

$$w_{AMDF}(c,m) = \begin{cases} e^{-\frac{1}{2} \left(\frac{D_{AMDF}(c,m) - \mu}{\sigma} \right)^2} & \text{음성영역} \\ 0.5 & \text{others} \end{cases} \quad (7)$$

식(6)은 식(7)와 같이 음성 영역에서만 채널간 지연 시간 정보에 따라 가중치를 계산한다.

$$\hat{C}(c,m) = C(c,m) \times w_{AMDF}(c,m) \quad (8)$$

식(8)은 가중치가 적용된 상호상관 함수 이다.

IV. 실험 결과

본 논문에서 실험에 사용된 데이터는 ETRI의 PBW 음성 데이터베이스 445개의 단어 음성, 백색잡음과 OHIO 대학 PNL의 100 non-speech sounds^[10]을 사용하였다. 남자 5명과 여자 3명의 깨끗한 음성과 배경 잡음(일정한 잡음(Stationary noise)과 일정하지 않는 잡음(Non-stationary noise))이 혼합된 음성에 대해 신호 대 잡음비를 비교하였다. 5개 배경 잡음은 일정한 잡음(기계(N28), 사이렌(N35))과 일정하지 않는 잡음(백색,

자동차(N45)과 군중(N2))이다.

$$SNR = 10 \log_{10} \left(\frac{\sum_t s^2(t)}{\sum_t (s(t) - s_v(t))^2} \right) \quad (9)$$

식(9)에서 $s(t)$ 는 깨끗한 음성이고, $s_v(t)$ 는 계산적 청각 장면 분석 시스템으로부터 잡음이 분리된 음성 또는 필터 응답의 음성이다.

그림 5는 제안한 방법을 이용해 그림 2(A)에서 채널 간에 비슷한 주기성 가지면서 지연 있는 경우에 잡음이

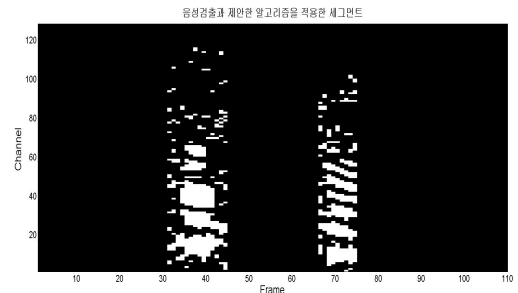


그림 5. 제안한 방법을 적용한 세그멘테이션
Fig. 5. Segmentation using Proposed algorithm.

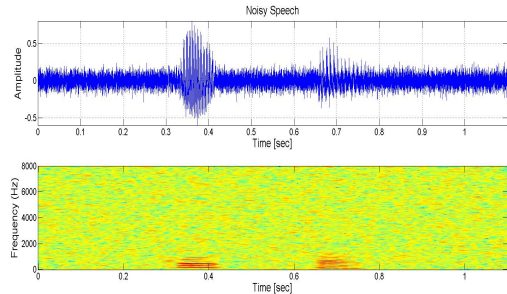


그림 6. 백색잡음이 혼합된 음성신호(SNR 0dB)
Fig. 6. Noisy speech created by additive white noise at (SNR 0dB).

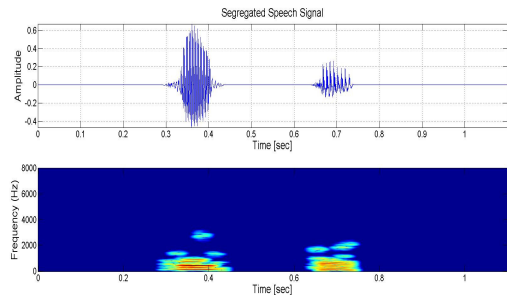


그림 7. 제안한 방법을 이용해 잡음이 분리된 음성신호
Fig. 7. Segregated speech signal using proposed algorithm.

음성으로 결정되는 문제를 개선하였다. 제안한 방법은 향상된 음성 세그먼트를 구성할 수 있다.

그림 6은 백색 잡음이 혼합된 신호 대 잡음비가 0dB 인 남성 음성 (“꽃밭”)이다.

그림 7은 음성 분리 성능 실험을 위하여 그룹화 과정을 수행하고 재합성을 통해 얻어진 잡음이 분리된 음성이다.

그림 8과 그림 9는 일정한 잡음(사이렌, 기계)과 변하는 잡음(백색, 자동차, 군중) 환경에서 기존 방법과 제안한 방법의 신호 대 잡음비의 변화(SNR 5dB, 0dB)에 대한 실험 결과를 보여주고 있다.

표 1은 신호 대 잡음비 5dB와 0dB에서 기존 방법과 제안한 방법의 신호 대 잡음비 성능을 평균으로 비교하였다. 제안한 방법은 기존 방법 보다 5dB에서 2.75dB와 0dB에서 4.84dB 향상됨을 확인할 수 있었다.

위너 필터는 일정한 잡음에서 우수한 잡음 제거 성능

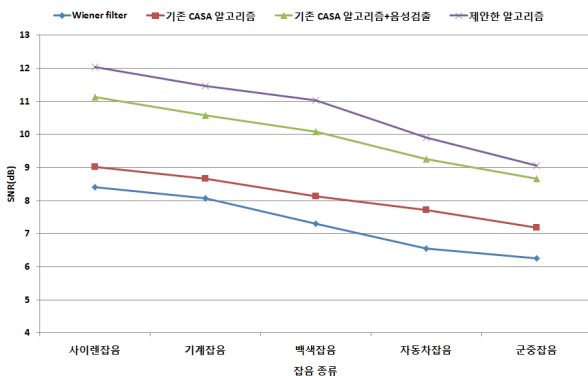


그림 8. 신호 대 잡음비 5dB에서 음성 분리 성능 평가
Fig. 8. Segregation speech performance evaluation in SNR 5dB.

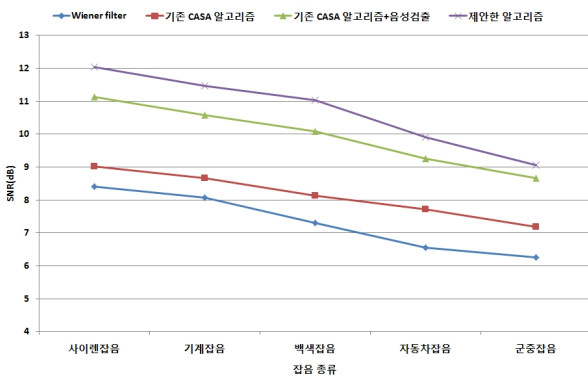


그림 9. 신호 대 잡음비 0dB에서 음성 분리 성능 평가
Fig. 9. Segregation speech performance evaluation in SNR 0dB.

표 1. 기존 방법과 제안한 방법의 신호 대 잡음비 비교

Table 1. SNR comparison of the conventional algorithm and the proposed algorithm.

알고리즘	SNR	5dB	0dB
		평균	평균
위너 필터		7.32dB	4.16dB
기존의 CASA		8.14dB	6.85dB
기존의 CASA+ 음성검출		9.94dB	8.3dB
제안한 알고리즘		10.7dB	9.00dB

을 보여주고 있지만 일정하지 않는 잡음에서는 잡음 제거에 어려움이 있다. 이와 반대로 계산적 청각 장면 분석 방법은 일정하지 않는 잡음 환경에서도 향상된 신호 대 잡음비를 보여주고 있다. 기존의 계산적 청각 장면 분석 방법에도 세그먼테이션에서 배경 잡음을 포함한 2진 마스크 구성으로 분리된 음성 신호에는 잔여 잡음이 있고, 음성 검출 문제^[11]로 음성 분리의 성능 감소 발생한다. 제안한 방법과 기존 계산적 청각 장면 분석 방법을 비교한 결과 제안한 방법의 신호 대 잡음비가 향상됨을 보여주었다.

V. 결 론

본 논문은 채널간 유사성을 상호상관 계수를 이용하여 세그먼테이션을 구성하는 CASA 시스템에서 채널간 유사성의 변별을 높이기 위한 방법으로 가중치 상호상관계수를 이용한 세그먼테이션 방법을 제안하여 이를 계산적 청각 장면 분석 시스템을 통해 음성 분리 실험을 수행하였다. 음성의 분리 성분을 평가하기 위해서 깨끗한 음성과 배경 잡음을 혼합한 후 음성 신호를 분리하여 신호 대 잡음비에 대한 성능 평가를 수행하였다. 실험 결과 제안한 방법이 기존 방법보다 신호 대 잡음비가 5dB에서 2.75dB와 0dB에서 4.84dB 증가하였고, 제안한 방법이 잔여 잡음 제거와 음성 분리 성능이 향상됨을 확인할 수 있었다.

REFERENCES

- [1] A. S. Bregman, "Auditory Scene Analysis: The Perceptual Organization of Sound," MIT Press, 1994.
- [2] Loizou and Philipos C., "Speech Enhancement: Theory and Practice," Crc Press, 2007.

- [3] A. Hyvärinen, J. Karhunen and K. Oja, "Independent Component Analysis," Wiley-Interscience, 2001.
- [4] D. L. Wang and G. J. Brown, "Computational Auditory Scene Analysis," Wiley-IEEE Press, 2006.
- [5] D. L. Wang, "On Ideal Binary Mask As the Computational Goal of Auditory Scene Analysis," Speech Separation by Humans and Machines, pp. 181-197, Kluwer Academic, Norwell MA, 2005.
- [6] G. Hu and D. L. Wang, "Monaural speech segregation based on pitch tracking and amplitude modulation," IEEE Trans. on Neural Networks, vol. 15, no. 5, pp. 1135-1150, September 2004
- [7] Jung-Ho Kim, Hyung-Hwa Ko, Chul-Ho Kang, "A Study on Voice Activity Detection Using Auditory Scene and Periodic to Aperiodic Component Ratio in CASA System," Journal of The Institute of Electronics Engineers of Korea, vol. 50, no. 10, pp. 181-187, October 2013.
- [8] B. R. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," Hearing Research, vol. 47, no. 2, pp. 103-138, August 1990.
- [9] G. Jacovitti and G. Scarano, "Discrete Time Techniques for Time Delay Estimation," IEEE Trans. on Signal Processing, vol. 41, no. 2, pp. 525-533, February 1993.
- [10] G. Hu and PNL, "100 Nonspeech Sounds," <http://www.cse.ohio-state.edu/pnl/corpus>
- [11] G. Hu and D. L. Wang, "Auditory Segmentation Based on Onset and Offset Analysis," IEEE Tran. on Audio, Speech, and Language Processing, vol. 15, no. 2, pp. 396-405, February 2007.

 저 자 소 개



김 정 호(정회원)

2000년 대전대학교 통신공학과

학사 졸업

2002년 광운대학교 전자통신

공학과 석사 졸업

2006년~현재 광운대학교

전자공학과 박사 수료

<주관심분야 : 음성신호처리, 음성분리, 잡음처리>



강 철 호(정회원)

1975년 한양대학교 전자공학과

학사 졸업.

1979년 서울대학교 전자공학과

석사 졸업.

1988년 서울대학교 전자공학과

박사 졸업.

1977년~1981년 국방과학연구소

1983년~현재 광운대학교 전자통신공학과 교수

<주관심분야 : 음성신호처리, 통신신호처리>