

반향제거를 위한 음성특징 기반의 동시통화 검출 기법

Speech Feature based Double-talk Detector for Acoustic Echo Cancellation

박준은*, 이윤재*, 김기현*, 고한석**

Jun-Eun Park*, Yoon-Jae Lee*, Ki-Hyeon Kim*, Han-Seok Ko**

Abstract

In this paper, a speech feature based double-talk detector method is proposed for an acoustic echo cancellation in hands-free communication system. The double-talk detector is an important element, since it controls the update of the adaptive filter for an acoustic echo cancellation. In previous research, the double talk detector is considered in the signal processing stage without taking the speech characteristics into account. However, in the proposed method, speech features which are used for the speech recognition is used for the discriminative features between the far-end and near-end speech. We obtained a substantial improvement over the previous double-talk detector methods using the only signal in time domain.

요 약

본 논문에서는 핸즈프리 통신에서의 반향제거를 위한 음성 특징 기반의 동시통화 검출 기법을 제안한다. 동시통화 검출은 반향제거를 위한 적응 필터의 적응을 제어하는 역할을 하기 때문에 매우 중요한 분야이다. 이전까지의 연구에서는 동시통화 검출을 음성의 특징에 대한 고려 없이 단순히 신호처리 영역에서만 이루어졌다. 하지만 제안한 기법에서는 음성인식으로 사용되는 음성 특징을 핸즈프리 통신상에서의 근단 화자와 원단화자 사이의 차별성을 가지는 특징으로 사용하였다. 제안한 방식이 시간 축에서의 과형만을 이용하여 판단하는 동시통화검출기보다 우수한 성능을 나타내는 것을 실험을 통하여 입증하였다.

Key words : 핸즈프리 반향제거, 적응필터, 동시통화 검출, MFCC, LPCCC 기반의 음성특징

1. 서론

자동차 환경에서 핸즈프리는 운전자의 안전과 관련된 중요한 시스템으로 그 사용이 점점 확대되고 있다. 운전자의 안전과 편의를 위하여 사용되는 핸즈프리는 음성 입출력을 위하여 마이크와 스피커가 사용

된다. 스피커로부터 나오는 상대방(원단화자)의 목소리가 다시 마이크로 들어가 원단화자가 자신의 목소리를 다시 듣게 되는 소리를 반향, 즉 에코라고 한다. 이러한 반향으로 인해 원단화자는 전화통화상에서 불편함을 느끼게 되는 단점이 있으며 이를 해결하기 위해 다양한 방법으로 반향제거를 하는 연구가 진행되어 왔다.[1]-[3] 반향제거는 크게 반향을 제거하기 위한 적응필터, 동시통화검출기로 구성되어 있으며 잔여반향제거나 잡음제거 기법을 추가할 수도 있다. 본 논문에서는 이러한 구성 요소 중 동시통화검출 분야에서 새로운 기법을 제안한다.

동시통화란 자동차내의 핸즈프리 사용자(근단 화자)와 원단화자가 동시에 발생하는 상황을 말하지만 본 논문에서는 더욱 일반적인 상황으로 원단화자의

* 高麗大學校 電子 컴퓨터 工學科

(Department of Electronics and Computer Engineering, Korea University)

★ 교신저자 (Corresponding author): 고한석

※ 감사의 글 (Acknowledgment) : 본 연구는 서울시 산학연 협력사업의 연구결과로 수행되었습니다.

接受日:2009年 5月 24日, 修正完了日: 2009年 6月 9日

발성 유무와 상관없이 근단 화자가 발성하였을 경우를 동시통화로 가정한다. 반향제거를 위한 적응 필터는 원단화자의 목소리가 다시 마이크로 들어온 것을 원단화자에게 전송이 되지 않도록 최소화 시켜주는 방향으로 적응된다. 하지만 원단화자와 근단 화자가 동시에 발성하였을 경우 적응 필터는 두 화자의 음성을 줄여주는 방향으로 적응되기 때문에 잘못된 방향으로 필터가 적응되게 된다. 또는 원단화자가 발성하지 않고 근단 화자만 발성하였을 경우에도 필터 적응이 잘못된 방향으로 적응되게 된다. 따라서 이러한 경우에는 반향제거가 효과적으로 되지 못하며 잔여 반향이 많이 생기거나 마이크 입력신호가 발산하는 현상이 생기게 된다. 이러한 현상을 방지하기 위하여 동시통화 검출기가 필요하며 동시통화가 검출되었을 때, 필터의 적응을 중지시켜서 잘못된 방향으로 필터가 적응되는 것을 방지하여야 한다.

기존의 연구에서 동시통화 검출은 주로 상호상관(cross correlation)을 이용한 기법 [4][5], Geigel 기법 [6], 적응필터의 변화기반의 기법 [7] 등이 주로 사용된다. 하지만 이러한 기법들은 단순한 시간상에서의 신호과형에 대한 분석에만 의존한 기법들이다. 이러한 기법들은 배경 잡음이 들어가거나 핸드프리 스피커의 음량이 변하거나 원단 또는 근단 화자의 목소리 크기가 변할 경우 자동적으로 문턱치 구하는 것이 어렵게 되며 문턱치와의 비교를 위한 특징들이 상당히 불안정하게 되어 안정적인 동시통화 검출이 어렵게 된다.

따라서 본 논문에서는 안정적인 동시통화 검출을 위한 기법으로 음성인식에 사용되는 특징들을 동시통화 검출을 위한 특징으로 사용한다. 원단화자가 혼자 이야기 하고 있을 경우 스피커로 재생되기 전의 원단화자 신호와 반향신호는 같은 단어의 유사한 음성으로 들리기 때문에 음성인식을 위한 특징들이 유사한 값을 가지게 될 것이며 이때 근단 화자가 발성하였을 경우 원단화자의 음성 특징과 상관성이 떨어지는 특징이 추출되게 된다. 따라서 이러한 특징의 유사성을 기준으로 동시통화를 판단하게 된다.

본 논문에서는 동시통화 검출을 위한 음성 특징으로 Mel-Frequency Cepstral Coefficients (MFCC)와 Linear Predictive Coding Cepstral Coefficients (LPCCC)의 두 특징을 사용하였다.

II. 본론

1. 일반적인 반향제거 시스템

그림 1. 은 일반적인 반향제거 시스템의 구성도를

나타낸다. 원단화자(far-end speaker)가 k 시간에 발성한 소리, $x(k)$ 는 자동차내의 핸드프리 스피커를 통하여 근단 화자(near-end speaker)에 들리게 된다.

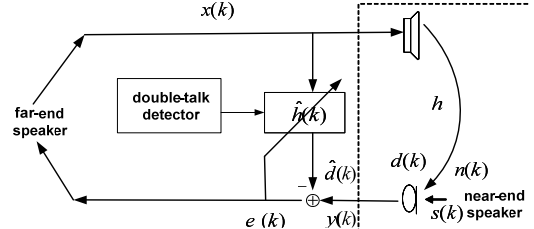


Fig. 1. Structure of the acoustic echo cancellation system
그림 1. 핸드프리 반향제거의 시스템 구성도

하지만 이 신호는 자동차내 핸드프리의 마이크로 다시 들어가게 되는데 이때 자동차 실내의 h 의 주파수 응답(impulse response)을 가지는 공간 경로를 통해 컨벌루션(convolution)되며 $d(k)$ 라는 반향신호가 되어 원단화자에게 전송이 된다. 근단 화자의 신호 $s(k)$ 및 잡음 $n(k)$ 가 포함된 마이크 입력 신호 $y(k)$ 가 되며 식 (1)과 같이 표현된다.

$$y(k) = d(k) + s(k) + n(k) \quad (1)$$

반향이 제거된 신호 $e(k)$ 는 식 (2)와 같이 나타낸다.

$$e(k) = y(k) - \hat{d}(k) \quad (2)$$

$$\hat{d}(k) = \hat{\mathbf{h}}(k) \otimes \mathbf{x}(k)$$

\otimes 는 컨벌루션 연산을 나타내며 $\hat{d}(k)$ 는 추정된 반향 신호이며 마이크 입력 신호에서 추정된 반향신호를 차감하여 반향을 제거해주게 된다. 추정된 주파수 응답 $\hat{\mathbf{h}}(k)$ 와 $\mathbf{x}(k)$ 는 각각 다음과 같이 나타내며,

$$\hat{\mathbf{h}}(k) = [\hat{h}_0(k), \hat{h}_1(k), \dots, \hat{h}_{N-1}(k)],$$

$$\mathbf{x}(k) = [x(k), x(k-1), \dots, x(k-N+1)], \quad N$$

은 모델링한 주파수응답의 필터 개수를 나타낸다.

추정된 적응필터 $\hat{\mathbf{h}}(k)$ 는 Normalized Least Mean Square (NLMS) 알고리즘을 통하여 다음식과 같이 매 샘플마다 재추정된다.[1]

$$\hat{\mathbf{h}}(k+1) = \hat{\mathbf{h}}(k) + \frac{\mu}{\mathbf{x}^T(k)\mathbf{x}(k)} e(k)\mathbf{x}(k) \quad (3)$$

μ 는 step size 로, 재추정시 어느 정도의 크기로 보정 해주는지에 대한 상수이다. 서론에서 언급하였듯 이, 식 (3)의 필터 적용은 근단 화자가 없을 경우에만 적용되어야 하며 이를 위하여 동시통화 검출기가 필요하다.

2. 기존의 동시통화 검출 알고리즘

기존의 동시통화 검출 알고리즘들은 크게 상호 상관관계를 이용하거나 에너지를 이용한 알고리즘들이 있다. 그중 상관관계를 이용한 알고리즘 중에 대표적인 것이 short-term normalized correlation (STNC)기반의 동시 통화 검출기이며 식 (4)와 같이 특징을 구한다.[5]

$$d_S(k) = \max_l \frac{\sum_{i=0}^{M-1} x(k-l-i)y(k-i)}{\sum_{i=0}^{M-1} |x(k-l-i)y(k-i)|} \quad (4)$$

M 은 상관관계를 계산하는 윈도우 길이가 되며 l 은 지연시간과 관계된 샘플 값이다. 상호 상관관계를 이용하는 것은 스피커 출력되기 전의 원단 신호와 마이크 입력의 반향 신호는 서로 상관관계가 상대적으로 크며, 이때 근단 화자 신호가 더해지면 상관관계가 떨어진다는 점을 고려한 것이다.

에너지 기반의 알고리즘은 대표적으로 Geigel algorithm 이 있으며 식 (5)와 같이 정의된다.[6]

$$d_G(k) = \frac{|y(k)|}{\max_{x(k)} x(k), \dots, x(k-n+1)} \quad (5)$$

에너지 기반의 Geigel 알고리즘은 반향신호에 근단 화자 신호가 더해지면 상대적으로 에너지가 더 커진다는 점을 고려한 알고리즘이다.

위 두 식에서 나온 값을 각각의 정해진 문턱치와 비교하여 식(4)의 경우는 문턱치보다 작으면 동시통화 발생으로 판단하며 식(5)의 경우는 문턱치보다 클 경우 동시통화 발생으로 판단한다.

2. 제안된 동시통화 검출 알고리즘

음성인식에 사용되는 음성특징은 다양한 사람이 같은 단어를 발성하더라도 그 특징이 유사하게 된다. 반향 신호 $d(k)$ 가 원단화자의 신호 $x(k)$ 의 공간 경

로를 통과한 신호이기는 하지만 음향학적 특성상 같은 단어가 들리게 된다. 이와 같은 사실은 $x(k)$ 로부터 추출한 음성 특징이 $d(k)$ 로부터 추출한 음성 특징과 상당한 유사성을 가진다는 것을 의미한다. 만약 근단 화자가 원단화자와 정확히 동일한 시간에 같은 단어를 발성하지 않는다고 가정하면, $x(k)$ 로부터 추출한 특징과 반향신호와 근단 화자 신호가 합해진 $d(k) + s(k)$ 로부터 추출한 특징 사이에는 유사성이 떨어지게 된다. 또한 제안된 기법은 단어 또는 음소에 기반을 둔 특징이므로 신호 자체의 크기에 크게 상관없이 기존에 시간축 신호영역 기반의 특징보다 강건한 특성을 지니게 된다. 따라서 이러한 특성에 기반을 두어 본 논문에서는 음성 인식에 사용되는 음성 특징 기반의 동시통화 검출 기법을 제안한다. 다양한 음성 특징 중에서 본 논문에서는 널리 알려진 MFCC 와 LPCCC 두 가지 특징을 사용하여 비교한다.

가. MFCC 기반의 특징

MFCC 는 Mel-Frequency Cepstral Coefficients의 약어로 현대 음성인식 시스템에서 가장 널리 쓰이는 음성 특징이다. 가장 큰 특성은 인간 청각의 특성을 반영한 것이라 볼 수 있으며 MFCC 특징 추출 과정은 다음과 같다.[8]

(1) 프레임(Framing)

음성 신호가 짧은 구간에서는 정상적(stationary) 상태에 있다고 가정하기 때문에 짧은 구간(약 25ms 정도)으로 나누어 프레임 처리를 한다.

(2) 프리-엠퍼시스(Pre-emphasis)

유성음의 경우 발생 시스템 구조에 의해 20dB/decade 의 신호 감쇠가 일어나기 때문에 이를 보상해주기 위해 프리-엠퍼시스를 한다. pre-emphasis 는 다음 식과 같이 1차 Finite Impulse Response (FIR) 로 구현된 필터이다.

$$x_{pre}(k) = x(k) - 0.97x(k-1) \quad (6)$$

(3) 윈도우 및 푸리에 변환

프리-엠퍼시스를 통과한 신호를 푸리에 변환(Fourier Transform)을 하기 위해 윈도우를 씌우게 된다. 윈도우는 일반적으로 해밍(hamming) 윈도우를 사용한다. 윈도우가 씌어진 신호는 푸리에 변환을 통해 주파수 영역으로 변환 된다.

(4) 멜-필터뱅크(Mel-filter bank)

주파수 영역으로 변환된 신호는 멜 필터뱅크(Mel-filter bank)를 통과하게 되는데 멜 필터뱅크는 인간의 청각 특성을 반영한 필터뱅크이다. 그림 2와 같이 멜 필터뱅크는 저주파 대역에서는 필터 대역(band)이 선형적인 간격을 가지며 고주파 대역으로 갈수록 필터의 대역이 증가하는 것을 알 수 있다. 이는 인간의 청각 특성이 고주파로 갈수록 신호 주파수 변화에 덜 민감해지는 특성을 반영한 것이다. 필터뱅크를 통과한 신호의 에너지가 각 필터뱅크의 출력이 되며 필터뱅크의 개수만큼의 특징이 나온다. 이 에너지 특징에 로그를 취한다.

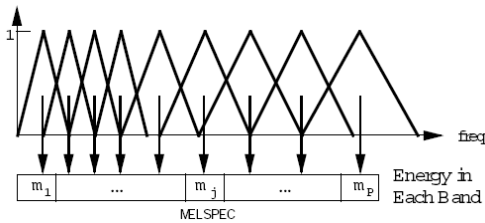


Fig. 2. Mel-filter bank
그림 2. 멜 필터뱅크 그림

(5) 이산 코사인변환 (discrete cosine transform)

마지막으로 멜 필터뱅크를 통과한 에너지 값들을 이산 코사인변환을 하며, 출력 값의 첫 12개의 계수를 최종적인 특징으로 사용한다.

나. LPCCC 기반의 특징

LPCCC 특징은 LPC 계수를 이용하여 cepstral coefficient를 구한 것이다. 사람이 다른 음소들을 발음하는 것은 성대와 입술 사이의 성도 모양이 다르기 때문이며 음소가 다름에 따라서 달라지는 것이 신호의 주파수 대역에서의 스펙트럼의 포락선(envelop)이다. 이 스펙트럼의 포락선은 LPC 계수로 모델링이 될 수 있기 때문에 LPC 계수가 중요한 특징으로 사용될 수 있다. LPC 계수 역시 프레임마다 구해지며 MFCC와 유사하게 프리-엠펜시스를 거치고 윈도우를 씌운 뒤에 자기상관 기법을 통하여 구할 수 있다.[9]

LPCCC는 LPC 계수 $a_m, m=1, \dots, p$ 를 이용하여 구해지며 여기서 p 는 LPC 계수의 개수이다. p 개의 LPCCC 계수 c_m 은 식 (7)과 같이 회귀 기법을 이용해 구해진다.[9]

$$c_m = a_m + \sum_{k=1}^{m-1} \frac{k}{m} c_k a_{m-k} \quad (7)$$

$$c_0 = \ln \sigma^2$$

σ^2 은 LPC 모델의 에너지 이득(gain)을 의미한다. LPCCC는 LPC 계수보다 더욱 안정적이고 강건한 특징으로 MFCC보다 널리 쓰이지는 않으나 음성인식을 위한 특징으로 사용될 수 있다.

다. 동시통화 검출 기법

동시통화 검출을 하기 위해 프레임 단위로 음성 특징을 추출한다. 스피커 재생 직전의 원단화자의 음성 신호와 반향이 되어 마이크 입력으로 들어오는 반향 신호 사이의 시간지연은 같은 시간의 프레임단위로 특징을 추출하여 비교할 수 있을 정도의 미미한 정도로 가정하며 실제 실험에서도 이를 확인하였다.

먼저 다음의 l 번째 프레임의 원단화자 및 마이크 입력 신호

$$\mathbf{x}(l) = [x_l(k), x_l(k+1), \dots, x_l(k+L-1)],$$

$$\mathbf{y}(l) = [y_l(k), y_l(k+1), \dots, y_l(k+L-1)]$$

로부터 음성특징 추출을 한다. L 은 한 프레임 샘플 개수를 의미한다. 근단 화자가 없을 경우 두 특징은 유사성을 가지게 된다. 유사성을 판단하기 위해 두 특징간의 거리 제곱의 합을 새로운 특징($feat_l$)으로 이용하여 문턱치를 경계로 식 (8)의 조건을 만족하면 동시통화 상황으로 판단한다.

$$feat_l = \sum_{i=1}^m (f_f(i) - f_n(i))^2 \geq \eta \quad (8)$$

$f_f(i), f_n(i), m, \eta$ 은 각각 원단화자 신호의 특징, 근단 화자 신호의 특징, 특징의 차원수, 문턱치 값을 나타낸다.

3. 실험 및 실험결과

실험에 사용한 데이터는 8kHz, 16bit PCM 음성으로 원거리 화자 음성(far-end-speech) 데이터 $x(k)$ 는 약 4-5초의 녹음된 데이터이다. 반향 데이터는 핸드프리 상황을 설정하여 녹음하였으며 근단 화자의 정확한 위치과약을 위하여 본 논문에서는 반향신호에 따로 녹음된 근단 화자 신호를 인공적으로 부가하여 실험하였다. 본 논문에서는 식 (1)에서의 잡음이 없다

고 가정하고 원거리 화자의 음성 데이터 $x(k)$ 를 조용한 환경의 사무실에서 스피커와 마이크를 통해 재생과 녹음을 동시에 진행하였다. 또한 근거리 화자의 음성을 따로 녹음한 뒤 Near-end speech to Far-end speech Ratio(NFR)에 변화를 주어 인공적으로 반향 신호에 근단 화자 신호를 부가하여 마이크 입력 데이터 $y(k)$ 를 얻었다. NFR은 아래 식과 같이 근단 화자와 원단화자 신호의 파워비로 정의된다.

$$NFR = 10 \log_{10} \left(\frac{E[s(k)^2]}{E[x(k)^2]} \right) \quad (9)$$

모든 실험은 25ms의 프레임 단위로 특징을 추출하였으며 식 (8)의 문턱치는 실험적으로 구하였다. 두 특징 모두 12차 특징을 이용하였으며 MFCC 특징의 경우 550, LPCCC 특징의 경우 40으로 문턱치를 정하였다. 문턱치의 경우 그 값에 따라 원하는 성능 사이의 교환관계 (trade-off)가 성립한다. 즉, 문턱치를 크게 하면 동시통화가 아닌 구간은 확실히 찾을 확률이 높아지나 동시통화구간이지만 아니라고 판단할 확률도 높아진다. 반대로 문턱치를 작게 잡으면 동시통화구간은 확실히 검출할 확률이 높아지나 동시통화구간이 아닌 구간도 동시통화로 검출할 확률 역시 높아지게 된다.

그림 3은 원단화자 신호와 녹음된 반향 신호의 파형 그림이다.

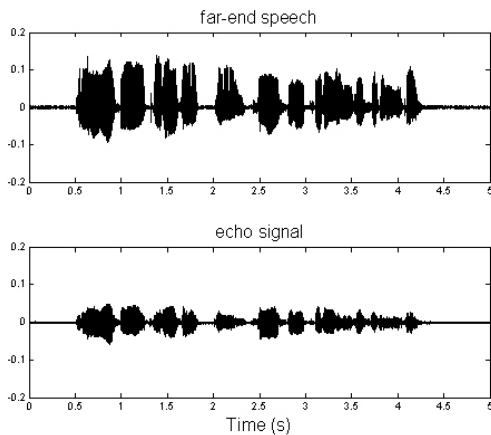


Fig.3. signal of the far-end speech and echo signal

그림 3. 원단화자 및 반향 신호

실험은 각각의 특징에 따라 두 가지 다른 NFR을 가지는 반향을 부가하였다. NFR이 클수록 근단 화

자의 음성 신호의 파워가 커지기 때문에 보다 더 동시통화 검출이 쉬울 것이라는 것을 직감할 수 있다.

그림 4는 NFR이 -4.3dB 일 경우의 MFCC 기반의 실험결과이며 그림 5는 NFR이 1.6dB 일 경우의 MFCC 기반의 실험결과이다.

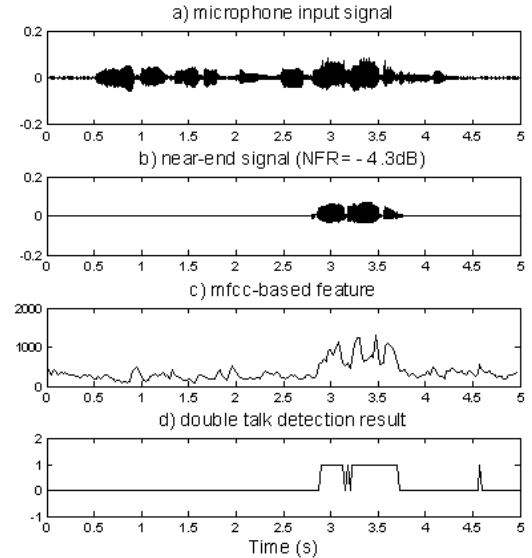


Fig. 4. Results : a) microphone input signal, b) near-end signal (NFR= -4.3dB), c) MFCC-based feature, d) double talk detection result

그림 4. 실험결과 : a) 근단 화자 신호와 반향신호가 더해진 마이크 입력 신호, b) NFR = -4.3dB 일 경우의 근단 화자신호, c) MFCC 기반의 특징, d) 동시통화 검출 결과

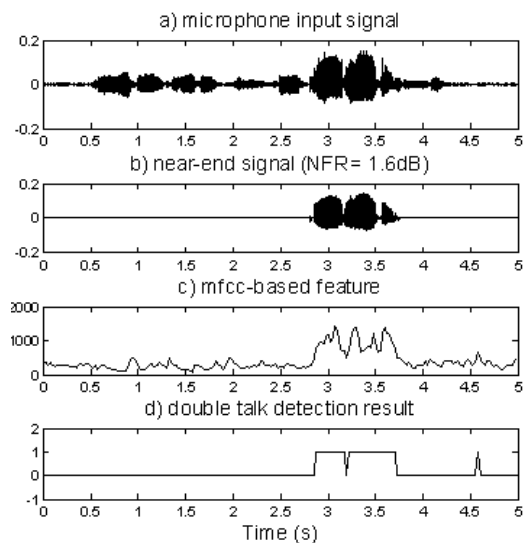


Fig. 5. MFCC based Results : (NFR= 1.6 dB)
 그림 5. NFR = 1.6dB 일 경우의 MFCC 기반의 동시통화
 검출 결과

그림 4의 경우는 근단 화자 신호와 반향신호의 파워비가 유사하여 a)만 봤을 경우 어느 구간이 근단 화자 신호가 존재하는지 눈으로 정확한 판단이 힘들다는 것을 알 수 있다. 하지만 MFCC 기반의 특징을 뽑았을 경우 c) 에서부터 동시통화 구간과 아닌 구간이 구별이 된다는 것을 확인할 수 있다. 이를 기반으로 문턱치를 적용하여 동시통화 검출을 하였을 경우 문턱치를 550으로 주었을 때 그림 4. d) 와 같은 결과를 얻을 수 있었다. 동시통화 구간이 아닌 구간에서 동시통화라고 판단한 경우가 이 실험에서는 하나의 구간이 나왔으며 동시통화구간이지만 근단 화자의 발성사이의 짧은 휴지 구간에서 동시통화가 아니라고 판단한 경우가 나왔다. 하지만 이 실험결과는 충분히 만족스러운 실험결과라고 판단되며 NFR 이 더 커진 신호의 경우는 그림 5에서 더욱 안정적인 실험 결과가 나온 것을 확인할 수 있다.

그림 6과 7은 LPCCC 기반의 특징으로 동시통화 검출을 한 경우이다. LPCCC 기반의 특징 역시 MFCC 특징과 같은 패턴을 가지며 안정적으로 동시통화 구간과 아닌 구간이 분별된다는 것을 그림 6, 7의 c) 그림을 통하여 알 수 있으며 NFR 이 커진 그림 7에서 더욱 안정적으로 검출된다는 것을 확인할 수 있다.

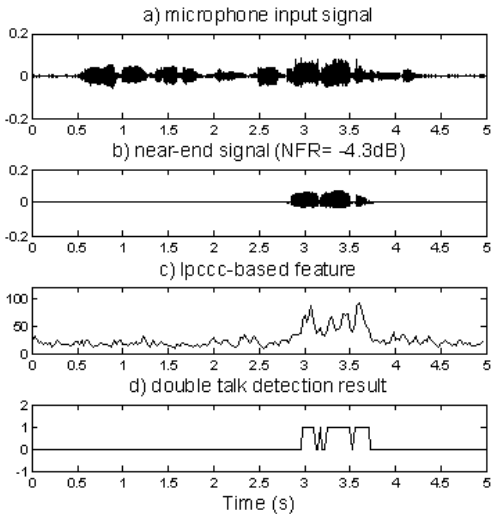


Fig. 6. LPCCC based Results : (NFR= -4.3dB)
 그림 6. NFR = -4.3dB 일 경우의 LPCCC 기반의 동시통화
 검출 결과

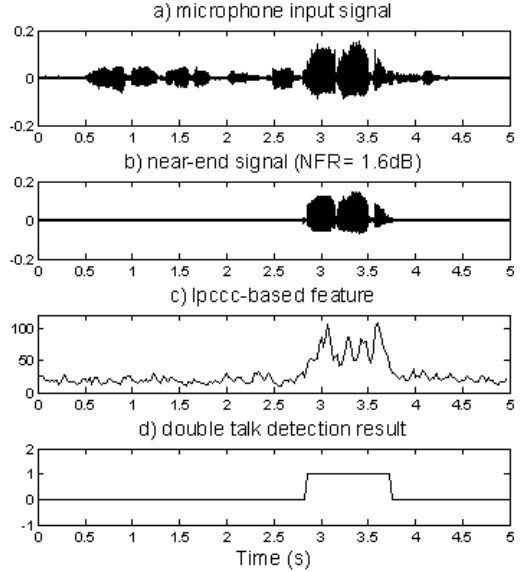


Fig. 7. LPCCC based Results : (NFR= 1.6dB)
 그림 7. NFR = 1.6dB 일 경우의 LPCCC 기반의 동시통화
 검출 결과

그림 8은 기존의 STNC 와 Geigel 알고리즘 기반의 특징을 뽑은 그림이며 NFR=1.6dB 일 경우에 대하여 실험한 것이다.

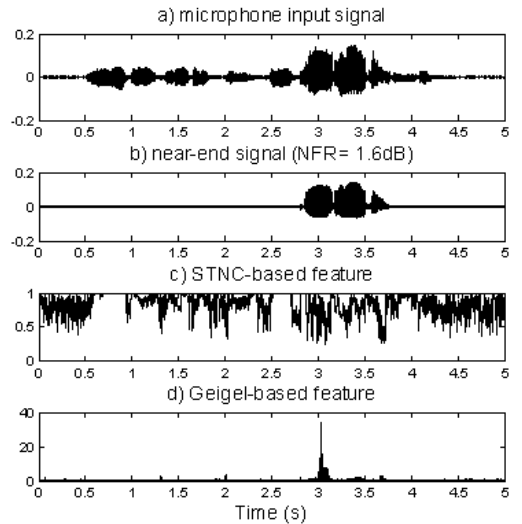


Fig. 8. STNC and Geigel-based Results : (NFR= 1.6dB)
 그림 8. NFR = 1.6dB 일 경우의 STNC 와 Geigel 기반의
 동시통화 검출 결과

식 4에서의 l 은 70, M 은 200 으로 설정하였으며 . 식 5

에서의 n 은 50으로 설정하였다. STNC 기반의 특징은 동시통화구간에서 상관관계가 줄어들기 때문에 그 값이 작아져야한다. 그림 8. c)에서 다른 구간에 비해 동시통화구간에서 그 값이 분별성있게 작아지는 것을 확인하기 힘들며 신뢰성 있는 문턱치를 구하는 것도 불가능해 보인다. d) 에서의 Geigel 알고리즘은 에너지 기반으로 하기 때문에 STNC 보다는 안정적으로 나오는 것을 볼 수 있으나 이 특징 역시 동시통화 구간에서 대부분의 값들이 동시통화 구간이 아닌 구간에서 가지는 값보다 분별성이 뛰어난 것은 아니라고 판단된다. 게다가 근단 화자의 음성 에너지가 작아질 경우에는 더욱 분별성이 떨어지는 특징이 될 것이라고 예상할 수 있다. 제안한 알고리즘들의 특징은 프레임 단위로 특징 추출이 이루어지며 기존알고리즘은 시간축의 샘플 단위로 매 샘플이 바뀔 때 마다 연산이 이루어진다. 전체 데이터에 대한 연산량을 비교한 결과 MFCC 기반의 특징추출은 STNC 와 Geigel 알고리즘보다 약 1.3배, 1.8배 의 연산속도를 가지며 LPCCC 기반의 특징추출은 STNC 와 Geigel 알고리즘보다 약 1.2배, 1.7배 의 연산속도를 가진다. 제안한 기법들이 기존의 알고리즘보다 연산이 많은 단점이 있으나 프레임단위의 실시간 처리에는 문제가 없으며 보다 안정적인 성능을 보이는 장점을 가진다.

III 결론

본 논문에서는 핸드프리 통신에서의 반향제거를 위한 음성특징 기반의 동시통화 검출기를 제안하였다. 사용된 음성특징은 음성인식에 사용되는 특징으로 MFCC 와 LPCCC 이다. 이러한 특징은 원단화자와 근단 화자 다른 발성에 대한 특성을 반영한 것이며 일반적인 시간축에서의 파형으로만 동시통화검출을 하는 것보다 강한 특징이 될 수 있음을 본 실험을 통하여 확인하였다. 본 논문에서 제안한 동시통화검출기를 가지는 반향제거기가 음성인식에 사용된다면 음성특징을 추출하는 과정에서 자연스럽게 동시통화 검출도 할 수 있는 장점을 가지게 될 것이다.

참고문헌

- [1] Bernard Widrow and Samuel D. Stearns, *Adaptive Signal Processing*, Prentice Hall, 1985
- [2] Donald L.Duttwiler, "Proportionate Normalized Least-Mean-Squares Adaptation in Echo Cancelers", *IEEE Transactions on speech and audio processing*, Vol. 8, No.5, pp 508-518, 2000.
- [3] John F. Doherty, Rajiv Porayath, "A Robust Echo Canceler for Acoustic Environments", *IEEE Transactions on circuits and systems*, Vol.44, No.5,

pp 389-396, 1997

- [4] Jacob Benesty , Dennis R. Morgan and Jun H. Cho, "A New Class of Doubletalk Detectors Based on Cross-Correlation", *IEEE Transactions on Speech and Audio Processing*, Vol.8, No2, pp168-172, 2000
- [5] C. Breining, P. Dreiseitel, E. Hänslar, A. Mader, B. Nitsch, H. Puder, T.Schertler, G. Schmidt, and J.Tilp, "Acoustic echo control - An application of very high-order adaptive filters," *IEEE Signal Processing Magazine*, Vol. 16, No. 4, pp. 42-69, July 1999
- [6] Donald L. Duttweiler, "A twelve-channel digital echo canceler," *IEEE Transactions on Communications*, Vol. 26, No. 5, pp. 647-653, May. 1978
- [7] Per Ahgren, *On system identification and acoustic echo cancellation*, Ph.D. thesis, Uppsala University, 2004
- [8] ETSI standard document, *Speech Processing, Transmission and Quality aspects (STQ), 2000 ; Distributed speech recognition; Front-end feature extraction algorithm; Compression algorithms*, ETSI ES 201 108 v1.1.3 (2000-04)
- [9] Lawrence Rabiner, Biing-Hwang Juang, *Fundamentals of Speech Recognition*, Prentice Hall PTR, 1993

저 자 소 개

박 준 은 (학생회원)



1992년 : 경북대학교 전자공학과 졸업 (공학사)
 1997년~현재 : 고려대학교 대학원 전자전기공학과 (공학석사)
 <주관심분야> 신호처리, 음성인식, 반향제거

이 윤 재 (정회원)

2003년 : 고려대학교 전기전자전
파공학과 졸업 (공학사)
2003년~현재 : 고려대학교 대학원
전자컴퓨터공학과 (석박사통합과
정)
<주관심분야> 신호처리, 음성인식,
반향제거

김 기 현 (비회원)

2003년 : 고려대학교 전기전자전
파공학과 졸업 (공학사)
2003년~현재 : 고려대학교 대학원
전자컴퓨터공학과 (석박사통합과
정)
<주관심분야> 신호처리,
음성인식, 반향제거

고 한 석 (비회원)

1982년 : 카네기멜론대학교
전기공학과 졸업 (공학사)
1986년 : 메릴랜드대학교 대학원
시스템공학과 (공학석사)
1988년 : 존스홉킨스대학교
대학원 전기공학과 (공학석사)
1992년 : 미국 카톨릭대학교

대학원 전기공학과 (공학박사)

1995년 3월~현재 : 고려대학교 전자전기공학과
교수

<주관심분야> 영상 및 음성 신호처리, 패턴인식,
데이터 융합