음성인식기 구현을 위한 잡음에 강인한 음성구간 검출기법

Robust Speech Segmentation Method in Noise Environment for Speech Recognizer

김창근, 박정원, 권호민, 허강인
Chang-Keun Kim, Jeong-Won Park, Ho-Min Kwon, Kang-In Hur

동아대학교 전자공학과
Dept. of Electronic Engineering, Dong-A University
e-mail : kihur@daunet.donga.ac.kr

요 약

실시간 음성 인식기의 구현에 있어서 선행되어야 할 과제는 신뢰성 있는 음성구간 검출과 적절한 음성특징벡터를 구하는 것이다. 그러나, 주변 잡음이 인가되는 환경에서는 신뢰성 있는 음성구간 검출이 어렵게 되어 적절한 음성특징벡터를 구할 수 없게 되어 최종적으로 인식기의 성능 저하를 초래하게 된다. 이러한 문제점을 보완하기 위하여 본 논문에서는 일반적으로 사용되어지는 단구간 파워 스펙토럼 외에 잡음에 강인한 특성을 가질 수 있도록 하는 새로운 특징 파라메터로써 스펙트럼 밀도비교척도와 선형회귀를 이용한 선형결정함수를 사용하였다. 이러한 두 가지 파라메터를 추가하여 주변 잡음의 크기에 따라 각각의 파라메터를 적절한 가중치로 조합하여 음성구간 결정을 수행한 다음 DTW를 사용하여 인식실험을 한 결과 주변 잡음이 존재하는 환경에서도 강인한 특성을 가짐을 확인할 수 있었다.

Abstract

One of the most important subjects in the implementation of real time speech recognizer is to design both reliable VAD(Voice Activity Detection) and suitable speech feature vector. But, because it is difficult to calculate reliable VAD in the environment having surrounding noise, designed suitable speech feature vector may not be obtained. Solving this problem, in this paper, we implement not only short time power spectrum which is generally used but also two additive parameters, the comparison measure of spectrum density having robust property in noise and linear discriminant function using linear regression, then perform VAD by using the combination of each parameter having apt weight in other magnitudes of surrounding noise and confirm that proposed parameters show a robust characteristic in circumstances having surrounding noise by using DTW(Dynamic Time Warping) in recognition experiment

Key words : VAD, Detection Parameter, Real-Time Speech Recognizer, DTW, SNR

Ⅰ. 서 론

급격한 발전을 거듭하는 정보통신분야에서 음성인식기술은 점차 중요한 분야로 자리잡고 있으며 차세대 핵심기술로 각광받고 있다. 하지만, 실생활에서 다양하게 첨가되는 주변 잡음의 영향은 음성인식기술의 전 분야에서 많은 어려운 문제들로 남아있다. 그 중 음성구간검출 분야에서도 예외는 아니다.

일반적인 음성구간검출 방법으로는 단구간 파워 스펙트럼과 영 교차율에 의한 방법이 있다. 이 방법은 간단한 수학적 계산에 의해 빠르게 음성구간검출을 할 수 있다는 장점이 있으나 입력음성에 잡음이 첨가되면 활용의 폭이 좁아지는 단점이 있다.[1]-[4]

실시간으로 음성구간 검출을 수행할 경우 주변잡음에 의한 원음성정보의 왜곡현상을 보상하기 위하여 전처리 과정으로 잡음을 제거하는 방법을 사용할 수도 있다.[5][6] 그러나, 실시간으로 적용하기 위해 전용 하드웨어로 구현할 경우 입력음성의 일정 구간을 버퍼에 보관

하여 잡음을 제거하고 잡음이 제거된 음성에 대한 음성 구간검출을 시행하는 두 번의 처리과정으로 인한 비용의 증가, 반응시간 지연, 고성능의 하드웨어 사용 등의 문제가 발생한다. 따라서, 실시간 음성인식 장치를 구현할 경우 음성구간 검출을 위한 고려사항으로는 적은 계산량과 효과적인 알고리즘의 제안이 선행되어야 할 것이다.[7]

본 논문에서는 실시간으로 음성의 구간만을 자동으로 검출하는 방법을 구현함에 있어 계산의 복잡도를 줄이고 주변의 잡음이 부가되어 있는 환경에서도 일정수준 이상의 인식성능을 얻을 수 있는 방법을 제안한다. 기존에 주로 사용되는 입력음성의 단구간 파워를 사용하는 방법 외에도 잡음의 영향을 측정하여 부가잡음의 효과를 감소할 수 있는 스펙트럼 밀도비교척도와 선형회귀를 이용한 선형결정함수를 추가하여 3종류의 파라미터에 주변잡음의 크기에 따른 각각의 가중치를 곱하여 얻어진 정보를 혼합한 판정함수로 음성구간의 유무를 판정하는 알고리즘을 제안하고 인식방법으로는 DTW(dynamic time warping)를 사용하여 인식실험을 수행하였다. 제안한 알고리즘에 의한 음성인식기를 구현하여 각기 다른 신호대 잡음비(SNR)로 구성되어 있는 음성데이터에서 실험을 수행하여 상당한 주변잡음이 부가된 환경에서도, SNR이 우수한 음성데이터에서 보여주던 인식성능을 유지함을 확인하였으며 본 논문에서 제안한 자동단어분할기법이 실시간 음성인식기에 적용하기에 효과적임을 확인할 수 있었다.

본 논문의 구성은 다음과 같다. 2장에서는 음성구간검출을 위한 파라메터의 선택, 각 파라메터에 대한 간단한 설명과 특성에 대해 기술하고 3장에서는 제안된 파라메터에 의한 음성구간 검출방법과 각기 다른 SNR에서의 음성구간 검출결과에 대해 기술하였다. 마지막으로, 4장에는 검출된 음성구간으로 인식실험을 한 결과에 대해 논의 할 것이다.

II. 음성구간검출 파라메터

실시간 동작하는 연속음성인식기를 구현함에 있어서 선행되어야 할 과정은 음성의 구간검출과 검출한 음성의 실시간 자동 분할이다. 또한, 실시간으로 동작하는 환경에 대하여 고려되어야 할 사항으로는 첫째로는 주변잡음에 의한 음성정보의 왜곡을 어떻게 보상을 하느냐는 문제가 있으며 두 번째의 문제는 계산량이 적어야 한다는 것이다.

이러한 문제점들을 해결하기 위하여 사용되는 파라메터로는 단구간 에너지, 영교차율, 모음의 주기성, 스펙트럼 분포, 켑스트럼 정보와 스펙트럼 비교법 등 많은 파라메터를 사용하고 있다. 그러나, 이러한 많은 분석방법을 단독으로만 사용하여서는 적용 가능한 신뢰성 있는 정보를 얻기란 참으로 어려운 형편이다.[8]

본 논문에서는 각각의 파라메터들이 가지는 정보를 상호 보완할 수 있도록 적용하여 잡음의 영향과 계산량의 감소라는 두 가지의 문제점을 해결할 수 있도록 3종류의 파라메터를 입력음성에 적용하여 얻어지는 각각의 파라메터 정보를 혼합하여 얻어진 결정함수를 사용하여 실시간으로 음성구간만을 검출하는 신뢰성 있는 방법을 제안한다.

1. 단구간 프레임 에너지

첫 번째로 적용한 파라메터로는 식(1)로 정의되는 각 프레임의 파워스펙트럼을 모두 합한 에너지를 사용하였다. 여기서 N은 프레임 길이, n는 프레임 번호, $S_n(k)$는 현 프레임의 음성신호를 의미한다.

$$E_n = \sum_{k=0}^{N-1} S_n(k)^2 \qquad (1)$$

에너지를 이용하여 음성구간을 결정할 경우 잡음의 영향이 적은 입력음성에 대해서는 효과적으로 사용될 수 있으나, 그림(1)과 같이 잡음이 존재한다면 잡음신호와 음성의 시작부분을 주로 구성하는 무성음 구간과의 구별이 어렵게 된다. 또한, 무성음으로 끝나는 단어의 끝점 검출에서도 같은 어려움이 존재한다.

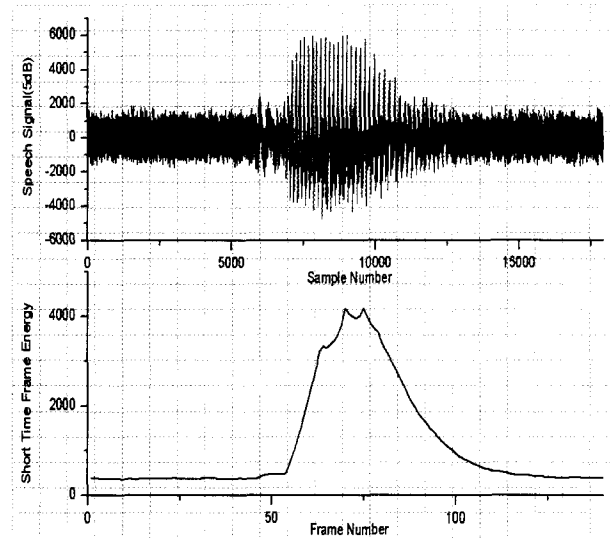


그림 1. 단구간 프레임 에너지

Fig 1. Short Time Frame Energy

2. 스펙트럼 밀도비교척도

단구간 프레임 에너지의 문제점을 고려하여 두 번째로 적용한 파라메터로는 식(2)와 같이 표현되는 스펙트럼 밀도비교 척도(spectral density comparison measure)를 사용하였다.

$$SD=\sum_{i=0}^{M-1}(S_i^2-N_i^2) \qquad (2)$$

여기서, S_i는 MFCC, N_i는 잡음의 추정치, M은 MFCC의 차수를 의미한다. 실제 환경에 존재하는 잡음의 영향이 음성에 적용되는 양을 파라메터에 적용하기 위하여 N_i 항을 정의하고 이런 잡음의 추정치는 미리 음성이 존재하지 않는 구간이라고 판단되어지는 구간 중에서 20 프레임을 선별하여 이의 평균치를 N_i로 추정한다. 이러한 파라메터의 적용은 입력 음성신호에 존재하는 잡음성분을 차감하는 효과를 얻을 수 있다. 그림(2)는 그림(1)에 비해 잡음성분이 차감된 효과를 볼 수 있다.

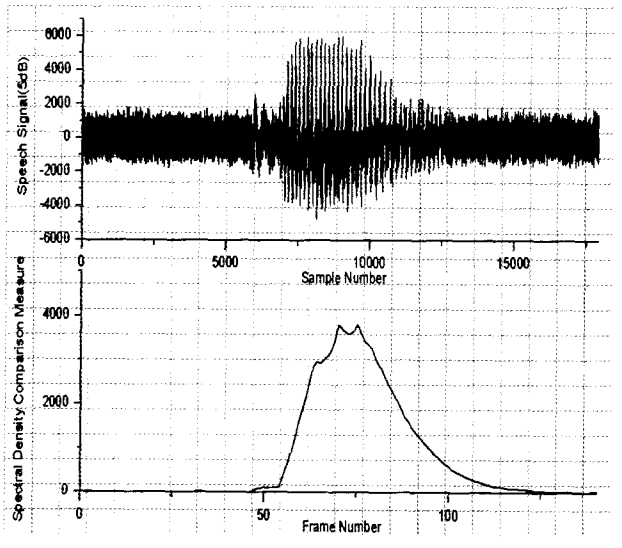


그림 2. 스펙트럼 밀도 비교 측정

Fig 2. Spectrum Density Comparison Measure

3. 선형결정함수

마지막으로 적용한 파라메터로는 식(3)로 표현되는 선형결정함수를 사용하였다.

$$y=\sum_{i=1}^{M-1}x_i w_i+w_0 \qquad (3)$$

x_i는 입력음성 각 프레임의 MFCC, M는 MFCC의 차수, w_i는 선형함수의 가중치, w_0는 성향(bias)을 의미한다. 여기서, 입력음성 x_i에 적절한 가중치 w_i를 적용하여 음성구간과 무음 또는 잡음구간으로 결정하기 위한 y를 출력한다. w_i를 학습하기 위하여 식(4)와 같이 표현되는 선형회귀(linear regression)를 사용하였다.

$$=(X^TX)^{-1}X^Ty \qquad (4)$$

여기서, X는 MFCC 벡터, y는 출력신호 벡터, 는 가중치 벡터를 의미한다. 식(4)를 사용하여 학습되어진 가중치 벡터는 선형결정함수 식(3)에 적용하여 얻어지는 출력값 y가 음성구간에서는 1을 출력하고 잡음이나 무음 구간에서는 0을 출력하도록 학습되어져야 한다. 가중치 벡터 를 학습하기 위하여 기존의 음성정보에서 음성구간과 잡음 또는 무음구간만을 상당량 수집하여 입력벡터 X가 음성구간이면 출력 y는 '1'이 되도록, 입력벡터 X가 잡음 또는 무음구간이면 출력 y는 '0'이 되도록 식(4)에 적용하여 가중치 벡터 를 결정하였다. 최적의 가중치 벡터를 구하기 위하여 음성구간으로 결정한 데이터에서 잡음 또는 무음구간과 단어의 도입부에 존재하는 무성음 구간이 유사한 점을 고려하여 무성음 구간의 도입부를 제외하였다. 그림(3)은 결정된 에 의해 잡음이 첨가된 음성이 입력으로 들어왔을 때 음성구간에서는 출력이 1에 가까워지고 잡음구간에서는 출력이 0에 가까워짐을 볼 수 있다. 이 파라메터는 잡음의 영향이 큰 음성신호에 대해서 효과적인 음성구간검출을 할 수 있으나 그 반대인 경우는 무성음 부분이 잡음으로 판단되는 단점이 있다.

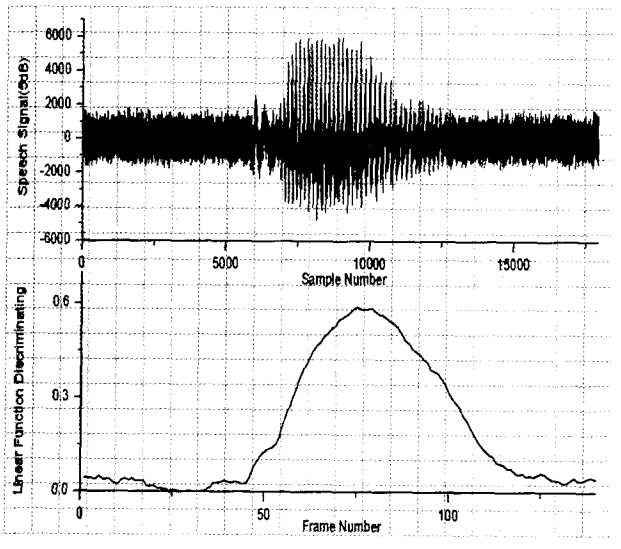


그림 3. 선형결정함수

Fig 3. Linear Discrimination Function

Ⅲ. 음성구간 결정함수

주변의 잡음이 존재하는 일상적인 환경에서 음성신호만을 추출하기 위해 주변잡음을 고려한 파라메터를 추가하여 각기 다른 특성을 지닌 파라메터를 복합적으로 적용한 다음 주변잡음에 강인한 특성을 보이는 음성구간 결정함수를 얻는다.

제안한 3가지의 파라메터를 구한 다음, 음성구간 검출에 필요한 정보만을 취하기 위하여 식(5)를 사용하여 이

동평균값을 구한다.

$$\overline{p_j}(n) = \alpha p_j(n) + (1-\alpha)\overline{p_j}(n-1),\ j=1,2,3 \quad (5)$$

여기서, $p_1(n)$ 은 n번째 프레임의 에너지, $p_2(n)$ 은 스펙트럼 밀도비교 척도, $p_3(n)$ 은 선형결정함수의 출력값을 의미하며 α값에 의하여 $\overline{p_j}(n)$ 의 변동량을 결정할 수 있다. 추출한 3가지의 특징 파라메터의 이동평균값을 사용하여 음성구간과 잡음 또는 무음구간을 결정할 수 있는 임계값을 결정하는 새로운 함수를 식(6)과 같이 정의한다.

$$T_j(n) = a_j\overline{p_j}(n) + b_j,\quad j=1,2,3 \quad (6)$$

연속음성신호에 다양한 잡음을 첨가한 혼합신호의 $\overline{p_j}(n)$ 로써 적절한 임계치 $T_j(n)$를 구하고 이 $\overline{p_j}(n)$ 와 $T_j(n)$로써 식(4)의 선형회귀를 사용하여 a_j와 b_j를 구한 다음, 음성구간 검출시 입력음성신호의 $\overline{p_j}(n)$ 에서 식(6)을 사용하여 음성정보의 유무를 결정하는 임계값 $T_j(n)$를 결정한다. 결정된 임계치는 주변잡음의 크기에 따라 각기 다른 가중치를 적용하여 합한 다음, 식(7)과 같이 표현되는 최종적인 음성구간 결정함수를 결정한다.

$$R(n) = \sum_{k=0}^{N_k-1}\sum_{j=1}^{3}\beta_j\ \delta_j(n+k) \quad (7)$$

식(7)에서 $\delta_j(n)$ 은 식(5)에서 구한 이동평균값과 식(6)에 의한 임계값 결정함수와의 비교에 의해 결정되는 이진값을 사용한다. 즉, $\overline{p_j}(n) > T_j(n)$ 인 경우에는 $\overline{p_j}(n)$ 의 값이 임계값 결정함수의 결과값 보다 큰 경우로 음성이 존재하는 구간임을 표현하는 것이기에 음성구간 검출함수의 결과 값에 해당 파라메터에 의해 음성구간으로 결정할 수 있도록 +1을 사용하고 반대의 경우는 -1를 사용한다. β_j는 음성구간 검출함수에 각기 다른 특성을 지닌 3종류의 파라메터가 적용되는 형태를 표현하는 것으로 식(8)과 같이 정의한다. 이것은 잡음이 음성신호에 섞인 정도에 따라 결정함수에서 각 파라메터의 가중 정도를 달리하는 방법으로 적용된다.

$$\begin{aligned}\beta_1 &= k_1(1/\ \overline{p_1}\ )\\ \beta_2 &= constant\\ \beta_3 &= 1-\overline{p_3}\end{aligned} \quad (8)$$

첫 번째 에너지의 경우 $\overline{p_1}$ 가 잡음이 클수록 증가하기 때문에 잡음이 적은 경우 에너지에 의한 음성구간 검출을 수행할 수 있도록 $\overline{p_1}$ 에 역수를 취하여 β_1을 얻는다. 여기서 입력음성은 파워 레벨을 정규화 한 음성을 사용한다. 두 번째 스펙트럼 밀도비교 척도의 경우는 주변잡음을 고려하여 미리 계산된 적절한 값을 실험적으로 결정하였고 세 번째 파라메터인 선형결정함수의 결과값은 0과 1사이의 값으로 잡음이 클수록 감소한다. 에너지의 경우와 반대로 잡음이 클 때 음성구간 검출에서 높은 성능을 가지도록 하기 위해 최대값인 1과의 차이값을 사용하였다. 비례상수 k_1 는 두 가중치 β_1, β_3의 척도를 일치시키기 위해 사용되었다. 식(8)의 조건으로 구한 $R(n)$이 '0' 이상의 값을 출력하면 음성구간이라 판단할 수 있다.

또한, 식(7)에서의 N_k 은 음성구간 결정함수를 얻기 위한 분석대상으로 취하는 연속적으로 인접한 프레임의 수를 의미한다. 음성으로써 의미 있는 구간은 최소 80ms는 되어야 하며, 단어 단위로 구간을 고려할 경우 단어와 단어 사이에 존재하는 묵음의 구간은 대략 40ms의 시간 동안 존재한다. 그러므로, 음성구간 검출을 위해서는 인접한 음성데이터에서 프레임의 간격을 8ms로 하고 80ms동안의 프레임을 계산하면 10개의 프레임이 된다. 10개의 연속적인 프레임 정보에서 음성구간 결정함수에서 음성구간이라 판단될 경우에 음성구간의 시작이라 결정하고 음절의 묵음구간을 고려하여 5개의 연속적인 프레임에서 잡음이나 무음이라는 정보가 존재하면 음성구간의 끝점이라고 결정하여 단어단위의 분할을 완료한다. 이러한 정보를 바탕으로 각 프레임에 음성구간 결정함수 $R(n)$를 적용하여 유효한 음성구간을 검출한다.

Ⅳ. 실험 및 결과

1. 분석조건과 분할

실험 데이터로는 실험실에서 PC를 사용하여 녹음한 연속 숫자음 데이터에 백색잡음을 각기 다른 SNR(5dB, 15dB, 25dB)로 첨가하여 사용하였다. 음성신호의 분석조건은 표1과 같다.

표 1. 분석 조건
Table 1. Analysis Conditions

A/D convert	16kHz, 16bit
window	hamming window
window length	24ms(384samples)
shifting period	8ms(128samples)
feature parameter	10th MFCC

자동분할실험과 DTW를 이용한 인식실험을 위하여 전처리 과정으로써 식(3)에서의 ω_i는 음성구간과 백색잡음구간을 각각 50개(화자5명의 숫자음 각10개)를 취하여 식(4)을 사용하여 결정하였으며, 식(6)에서의 a_j, b_j는 연속

음성신호에 다양한 종류와 크기를 가지는 잡음을 첨가하고 그때의 $\overline{p_j}$를 구한 다음 각 파라메터에 대한 최적의 임계치 T_j를 구한다. 여기서 구해진 $\overline{p_j}$와 T_j로써 식(4)의 선형회기를 적용하여 a_j와 b_j를 결정하였다.

이와 같은 전처리 과정에서 얻어진 w_i, a_j, b_j를 가지고 실시간 자동분할실험에서는 다음과 같은 과정에 의해 3종류의 SNR로 구성된 연속 숫자음 데이터에 대하여 음성구간 결정을 수행하였다. 먼저 입력음성에 대한 각 파라메터 p_j를 구하고 α를 0.1로 정하여 이동평균을 취한 다음 a_j, b_j에 의해 구해진 T_j와 $\overline{p_j}$로 식(7)의 δ_j와 식(8)의 β_j를 구하여 최종적으로 음성구간 결정함수 R을 얻게 된다.

그림(4)(5)(6)은 이상의 과정에 의해 계산된 $\overline{p_j}$와 결정함수 R을 도식화한 것이다. 그림(6)에서 보면 음성신호에 노이즈가 적을 때(25dB)는 프레임의 에너지인 p_1 값이 작아져서 가중치인 β_1의 값이 커짐으로 결정함수 R에 가장 큰 영향을 미치게 되어 음성구간 검출이 이루어진다. 반대 상황으로 음성신호에 노이즈가 증가할수록(5dB) p_3의 값이 작아져서 p_3의 가중치인 β_3의 값이 커지므로 결정함수 R에 미치는 영향이 커짐을 볼 수 있다.

각 파라메터의 음성구간 검출성능을 종합하여 보면 p_1은 잡음이 적을 때, p_2와 p_3는 주변잡음이 존재할 때에 음성구간 검출성능이 높아짐을 알 수 있다.

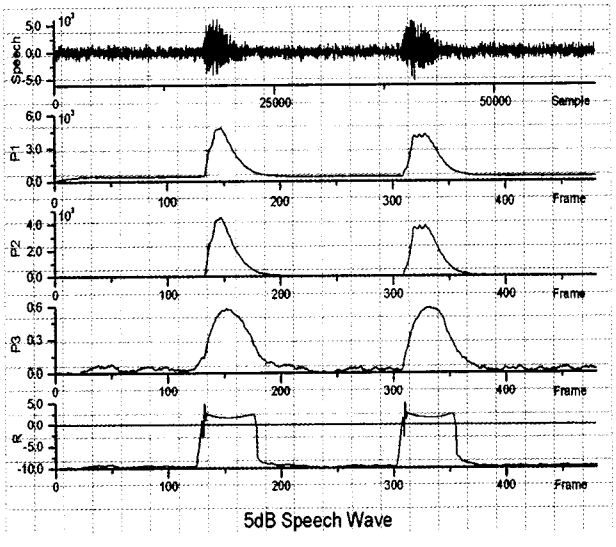


그림 4. 5dB SNR에서의 음성 구간 검출 결과

Fig 4. Segmentation result at 5dB SNR

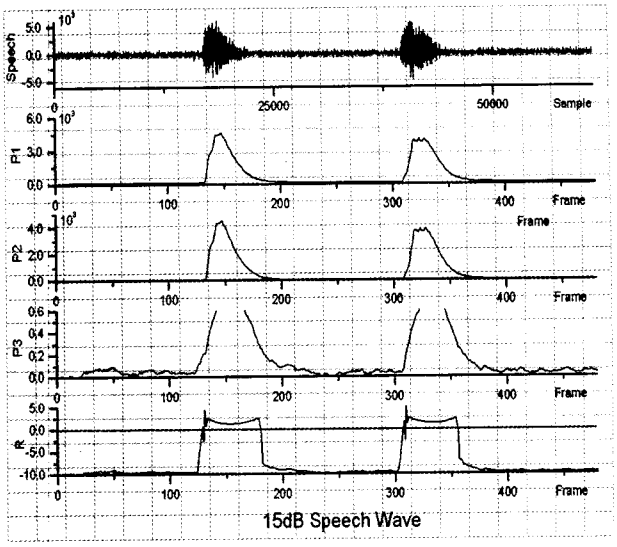


그림 5. 15dB SNR에서의 음성 구간 검출 결과

Fig 5. Segmentation result at 15dB SNR

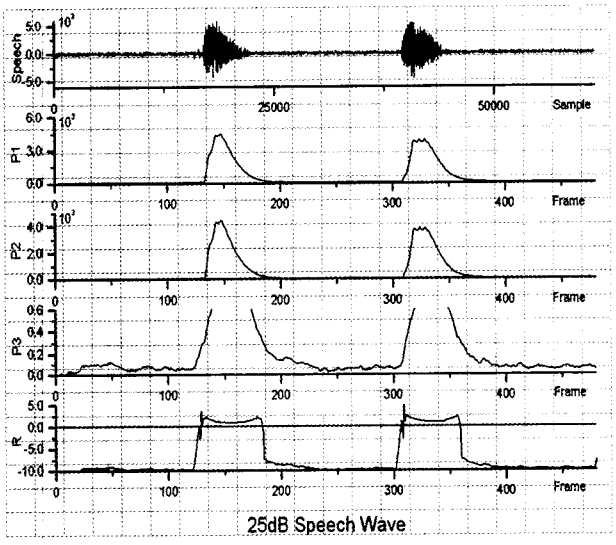


그림 6. 25dB SNR에서의 음성 구간 검출 결과

Fig 6. Segmentation result at 25dB SNR

2. 인식 실험

연속 숫자음을 자동 분할하여 DTW알고리즘에 의해 인식실험을 수행하였다. 인식실험은 화자 5명이 각각 2번 발성한 데이터를 사용하였으며 1회 발성분(각 숫자음 10단어)은 참조 데이터로 사용하고 나머지 1회 발성분(각 숫자음 10단어를 10회 발성)은 인식 데이터로 사용하였다.

표 2. 인식률

Table 2. Recognition rate

	참조데이터		인식데이터		인식률(%)					
					A	B	C	D	E	평균
실험 1	목측분할	5 dB	목측분할	5 dB	91	82	77	80	93	84.6
		15 dB		15 dB	92	89	76	88	98	88.6
		25 dB		25 dB	97	99	85	91	100	94.4
실험 2	목측분할	5 dB	자동분할	5 dB	73	57	36	69	70	61.0
		15 dB		15 dB	80	60	45	79	78	68.4
		25 dB		25 dB	71	71	50	79	71	68.4
실험 3	자동분할	5 dB	자동분할	5 dB						80.0
		15 dB		15 dB						85.4
		25 dB		25 dB						88.8

화자 5명(ABCDE)의 각 2회 발성분 전부에 대해 세 종류 SNR로 백색잡음을 첨가하여 각각 목측 분할과 자동분할을 수행하였으며 실험 결과는 표2와 같다.

실험1은 일반적으로 사용되어지는 방법으로써 참조데이터와 인식데이터에 대해 목측으로 분할 작업을 수행한 다음 인식실험을 수행하였으며, 실험2는 목측으로 미리 분할을 수행한 참조데이터에 대해 본 논문에서 제안한 방법으로 실시간 자동분할 하여 인식실험을 수행하였으며. 또한, 실시간 음성인식기에 적용하기 위한 모델로써 실험3은 참조데이터와 인식데이터에 대해 모두 자동분할을 수행하여 인식 실험을 수행하였다.

실험1의 경우 평균 89.2%의 인식율을 나타내었으며 실험2의 경우는 65.9%, 실험3의 경우 84.7%의 인식율을 나타내었다.

실험1의 결과는 실험2의 결과보다 모든 데이터 조합에서 인식 성능이 현저히 떨어짐을 알 수 있다. 이는 목측분할과 자동분할의 경우에서 분할한 음성데이터의 길이가 상이함으로 인해 발생하는 문제로 사료된다.

그러나, 실시간 음성인식시스템에 적용 가능한 모델인 실험3의 결과는 실험1의 결과와 비교하여 평균 4.5%의 인식율 저하를 나타내었다. 이는 목측으로 분할한 데이터 조합의 성능에는 미치지 못하였지만 잡음환경에서 동작하는 실시간 음성인식장치에 적용 가능할 것으로 사료된다.

Ⅳ. 결론

실시간으로 동작하는 음성인식기를 구현함에 있어서 가장 문제가 되는 음성구간의 분할과 주변잡음에 의한 영향을 최소화하는 방안으로서, 주변잡음이 존재하는 실제 환경에서 음성인식기를 적용한다는 가정을 두고, 각기 다른 분석기법을 조합하여 최종적으로 주변잡음에 강인한 특성을 나타낼 수 있도록 인식기를 구성하고, 제안한 단어단위의 음성분할 알고리즘을 적용하여 각기 다른 신호대 잡음비의 데이터에 대하여 두 경우로 나누어 실험하였다. 첫 번째로는 일반적인 방법으로 사용되는 목측분할에 의한 방법으로 단어단위의 분할을 수행하여 참조데이터로 사용하고 다시 목측분할법과 제안한 알고리즘을 사용하여 자동분할을 수행한 데이터를 사용하여 인식실험한 결과 목측분할과 자동분할의 경우의 결과가 평균 23.3%의 인식성능의 저하가 발생하였다. 두 번째로는 자동분할에 의한 방법으로 참조데이터와 인식데이터를 사용하여 인식 실험한 결과, 평균 4.5%의 성능 저하가 발생하였다.

전자의 경우는 인식알고리즘을 음성구간의 크기에 의해 인식성능이 좌우되는 DTW를 사용함으로 인해, 인식데이터를 결정할 때 제안한 알고리즘을 사용한 경우에서 음성데이터구간의 크기가 목측의 경우와 비교하여 일정하지 못한 이유로 상당한 인식 성능의 저하를 발생한 것으로 사료된다. 후자의 경우에서는 비록 평균 4.5%의 성능저하가 발생하였지만 충분히 적용 가능한 방법임을 확인할 수 있었다.

이러한 인식성능 저하의 해결책으로서는 인식방법으로 DTW가 아닌 HMM이나 ANN을 이용한다면 보다 나은 결과를 얻을 수는 있지만 제안한 알고리즘의 목적은 실시간으로 동작하기 위한 간단하고도 효과적인 방법을 사용함을 그 목표로 하기에 계산량과 성능이라는 두 가지의 조건을 만족시킬 수 있는 적절한 결정을 내려야 할 것이다. 또한, 실시간으로 동작하는 음성인식기에 제안한 알고리즘을 사용하여 비정상적인 주변잡음과 구간분할 시 음성의 도입부에 존재하는 무성음 성분의 일부 탈락에 의한 인식시스템의 성능하락을 보상할 수 있을 것으로 사료된다.

접수일자 : 2003. 1. 13 수정완료 : 2003. 3. 04

이 논문은 2001년도 정보통신(IT)사업 연구비에 의해 연구되었음

참고문헌

[1] M.Toma, A.Lodi, R,Guerrieri, "Word endpoints detection in the presence of non-stationary noise," ICSLP, vol.2, pp1053-1056, 2002

[2] R.Tucker, "Voice activity detection using a periodicity measure," IEEE Proceedings-I, vol. 139, no. 4, pp377-380, August 1992

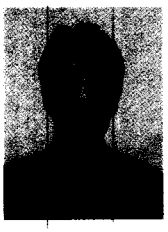
[3] N.B.Yoma, F.McInnes,M.Jack, "Robust speech pulse detecting using adaptive noise modelling," Electronics Letters, vol. 32, no. 15, pp.1350-1352, 1996

[4] J.A.Haigh, J.S.Mason, "Robust voice activity detection using cepstral features," IEEE TENCON,

pp. 321-324, China, 1993
[5] L.R.Rabiner, R.W.Schafer, "Digital processing of speech signals," Prentice Hall, 1978
[6] Richard O.Duda, Peter E.Hart, David G.Stork, "Pattern Classification(Second Edition)," A Wiley Interscience Publication, 2001
[7] A. Benyassine, E.Shlomot, H. Y. Su, D. Massaloux, C. Lamblin, J. P. Petit, "ITU-T Recommendation G.729 Annex B: a silence compression scheme for use with G.729 optimized for V.70 digital simultaneous voice and data applications," IEEE Communication Magazine, vol. 35, no. 9, pp. 64-73, Sept. 1997
[8] X.Huang, A.Acero, H.W.Hon, "Spoken Language Processing," Prentice Hall, 2001
[9] 표창수, 김창근, 허강인, "신경망을 이용한 HMM의 오인식보상에 관한 연구," 한국음향학회, 19권1(s)호, pp27-30, 2000
[10] 박정원, 김창근, 한학용, 허강인, "연속음성 인식장치를 위한 실시간 음성분할의 구현," 한국 신호처리 시스템학회 학술논문집, 3권1호, pp225-228, 2002

김 창 근(Kim Chang Keun)
準會員
1994년 2월 동아대학교 전자공학과 공학학사
1998년 8월 동아대학교 전자공학과 공학석사
2002년 2월 동아대학교 전자공학과 공학박사 수료
관심분야 : 음성신호처리, 음성인식

박 정 원(Park Jeong Won)
準會員
2002년 2월 동아대학교 전자공학과 공학학사
2002년 3월~현재 동아대학교 전자공학과 석사과정
관심분야: 음성신호처리, 음성인식



권 호 민(Kwon Ho Min)
準會員
2002년 2월 동아대학교 전자공학과 공학학사
2002년 3월~현재 동아대학교 전자공학과 석사과정
관심분야 : 음성신호처리, 음성인식

허 강 인(Hur Kang In)
正會員
1980년 2월 동아대학교 전자공학과 공학학사
1982년 2월 동아대학교 전자공학과 공학석사
1990년 2월 경희대학교 전자공학과 공학박사
1988년~1989년 일본 쓰꾸바대학 정보공학부 객원연구원
1992년~1993년 일본 도요하시대학 정보공학부 Post-Doc.
1984년~현재 동아대학교 공과대학 전기전자컴퓨터공학부 교수
관심분야 : 음성신호처리, 음성인식