논문 2006-43IE-2-5

은닉 마코프 모델과 켑스트럴 계수들에 따른 한국어 속삭임의 인식 비교

(Comparison of HMM models and various cepstral coefficients for Korean whispered speech recognition)

박 찬 응*

(Chan Eung Park)

요 약

본 논문에서는 모바일 환경에 따른 속삭임의 사용이 증가하는 데 따른 속삭임 인식을 위하여 음성인식에 많이 사용되고 있는 특징벡터들을 은닉 마코프 모델을 이용, 정상어 모델, 속삭임 모델, 정상어, 속삭임 통합 모델들에 인식 시험하고 결과를 분석하여 가장 적합한 인식 시스템을 찾으려고 하였다. 인식 시험을 통하여 속삭임의 인식은 정상어 모델로 인식하는 시스템은 낮은 인식률로 실용성이 없으며 속삭임 모델을 별도로 사용하는 것이 85% 이상의 가장 높은 인식률을 보였다. 또한 '정상어+속삭임' 모델도 인식률은 조금 떨어지나 가능성을 확인할 수 있었다. 특징벡터로는 속삭임 모델을 사용하는 경우 MFCC 혹은 PLCC를 사용하는 것이 거의 유사하게 높은 인식률을 얻을 수 있었으나 '정상어+속삭임' 모델을 사용하는 경우 PLCC를 특징벡터로 사용하는 것이 속삭임 인식에서 가장 좋은 결과를 보였다.

Abstract

Recently the use of whispered speech has increased due to mobile phone and the necessity of whispered speech recognition is increasing. So various feature vectors, which are mainly used for speech recognition, are applied to their HMMs, normal speech models, whispered speech models, and integrated models with normal speech and whispered speech so as to find out suitable recognition system for whispered speech. The experimental results of recognition test show that the recognition rate of whispered speech applied to normal speech models is too low to be used in practical applications, but separate whispered speech models recognize whispered speech with the highest rates at least 85%. And also integrated models with normal speech and whispered speech score acceptable recognition rate but more study is needed to increase recognition rate. MFCC and PLCC feature vectors score higher recognition rate when applied to separate whispered speech models, but PLCC is the best when applied to integrated models with normal speech and whispered speech.

Keywords : whispered, HMM, speech, recognition

Ⅰ. 서 론

속삭임은 성대의 떨림이 없거나 아주 적은 상태로 이야기 되는 비정상적인 형태의 음성으로 정의될 수 있다. 이런 속삭임에 대한 연구는 특히 1950년대부터 여러 분야에서 많은 이유로 계속되어 왔다. 음성과학에서는 음성처리를 위하여, 의학 분야에서는 의학적인 다양한 목적을 위하여, 법의학적인 측면에서는 녹음된 속삭임의 발성자를 정상 발음된 음성과 연관지음으로서 속삭임 발성자를 밝혀내려는 화자인식의 시도로부터 출발하여 연구되어 왔다[1]. 따라서 과거의 연구들은 정상어와 속삭임 사이의 연관 관계를 구명하는데 연구의 방향이 집중되었다. 그러나 근래에 들어와서 자동음답 서비스에서의 음성인식의 활용이 증가하고 또한 휴대전화가 거의 모든 개인에게 보급될 정도로 확산됨에 따라 공공

* 정회원, 인덕대학 정보통신전공
(Information and Communications Course, Induk Institute of Technology)
접수일자: 2006년3월27일, 수정완료일: 2006년6월10일

장소에서의 전화 통화의 빈도가 급속히 증대하였고 이에 따라 정숙성과 보안성이 요구되는 통화에서 속삭임의 사용이 증가하고 있으며, 따라서 속삭임의 인식에 대한 필요성 또한 증대되고 있다.

속삭임은 앞에서 정의된 바와 같이 성대의 떨림이 없을 뿐만 아니라 음압 레벨도 정상어에 비하여 상대적으로 아주 낮아 공공장소에서의 신호대잡음비도 더 낮다. 이런 이유로 속삭임에 대한 처리는 정상어와는 다른 접근 방법이 필요함에도 불구하고 이에 대한 연구는 상대적으로 잘 이루어지고 있지 않다. 외국에서는 속삭임에서의 운율정보 추출[2], 속삭임 모음에 대한 포만트 주파수의 추정, 정상어와 속삭임의 포만트 주파수 차이[3], 속삭임의 분석 및 인식[4]들에 대한 연구 결과들이 발표되었으나 국내에서는 우리말 속삭임에 대한 연구를 찾아보기가 어려운 실정이다.

본 연구에서는 한국어 고립단어 숫자음의 속삭임에 대하여 여러 가지 분석 방법을 통하여 얻어진 특징벡터들을 은닉 마코프 모델(Hidden Markov Model : HMM)을 이용한 음성인식에 적용하여 인식률을 비교하여 봄으로써 속삭임의 인식을 위한 효율적인 특징벡터와 인식방법을 찾아본다. 특징벡터는 LPCC, MFCC, FTCC, PLCC 들을 사용하고, 이들에 대하여 은닉 마코프 모델로 모델링한 정상어 모델, 속삭임 모델, 정상어+속삭임 모델들에 대하여 속삭임을 인식시켜 봄으로서 적합한 적용모델을 확인하고자 한다.

Ⅱ. 속삭임 발생 메커니즘

후두는 기관지 상단부에 위치하는 기관이며 이 기관의 목적은 폐로부터의 공기의 출입을 통제하는 역할을 한다. 그림 1은 후두의 앞부분을 보여주고 있는 데 성

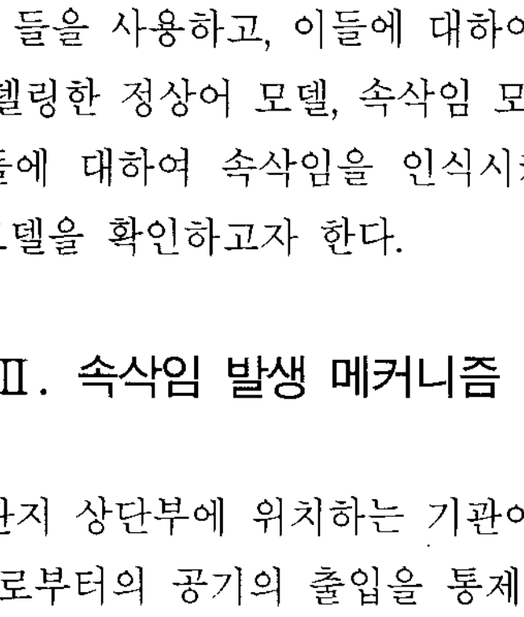


그림 1. 후두부[5]
Fig. 1. Diagram of larynx.[5]

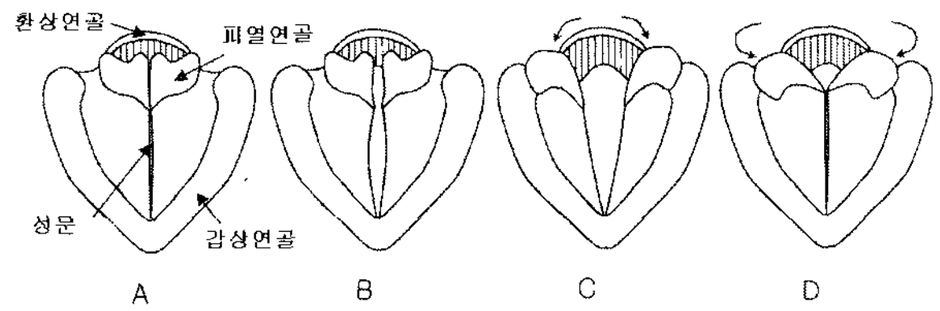


그림 2. 성문의 다른 형태들[6]
Fig. 2. Different configurations of the glottis.[6]

대라고 불리는 판이 후두 연골을 피열 연골2에 연결하여 주고 있다. 그리고 두개의 판 사이의 틈을 성문이라고 부른다.

성문의 모양과 크기를 조절하기 위하여 그림 2에서와 같이 피열연골이 이동, 회전할 수 있다. 유성음의 발성은 그림 2-A에서와 같이 끝부분의 윤상연골의 이완과 피질연골의 수축에 의하여 성문이 굳게 닫힌 상태에서 이루어진다. 그러나 속삭임의 발성은 성문의 좁은 틈새를 만들고 이를 통하여 공기를 통과시킴에 의하여 세찬 흐름을 만들어 생성된다.

이 틈새는 발음에 따라 좁게 찢어진 모양, V자 모양, Y자 모양 등을 갖게 되며 이 경우 성문의 면적은 0.1에서 0.4㎠의 크기를 갖는 것으로 보고 되었다[6].

Ⅲ. 음성분석

1. 선형예측 켑스트럴 계수(LPCC)[7]

가장 널리 알려진 선형예측분석은 여러 많은 실용적인 혹은 이론적인 연구의 기본이 되어 왔다. 선형예측분석을 통하여 선형예측계수(LPC)들을 구하고 이들 선형예측계수로부터 식(1)에 의하여 선형예측켑스트럴계수를 구한다.

$$c(n) = -a(n) - \frac{1}{n}\sum_{k=1}^{n-1} kc(k)a(n-k) \quad \text{for } n > 0 \tag{1}$$

2. 퓨리어 변환 켑스트럴 계수(Fourier Transform Cepstral Coefficients : FTCC)[8]

FTCC는 그림 3에서와 같은 과정에 의하여 구할 수 있으며 표현 식은 식(2)과 같다.

$$c(n) = \frac{1}{N}\sum_{k=0}^{N-1} log_{10}|S(k)|e^{(2\pi/N)kn} \quad \text{for } 0 < n \leqq p \tag{2}$$

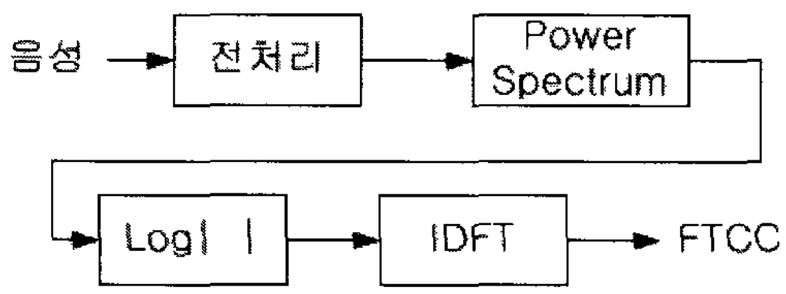


그림 3. FTC 분석 구성도
Fig. 3. Fourier Transform-derived Cepstral Analysis Block Diagram.

3. 멜 주파수 켑스트럴 계수(Mel Frequency Cepstral Coefficients:MFCC)[9]

인간의 음향 인지 특성중의 하나는 인지된 소리의 피치 혹은 주파수는 실제 물리적인 주파수와 선형적으로 연관되지는 않는다는 것이다. 인간에 의하여 인지된 소리의 피치 주파수를 mel이라는 단위로 나타낸다. 선형 주파수와 멜 주파수와의 관계를 그래프로 그린 것이 그림 4이다.

이러한 인간의 음향 인지 특성을 도입하여 음성을 분석하는 것을 mel주파수분석이라 하며 그림 5와 같은 과정에 의하여 MFCC를 구할 수 있다.

1) 전처리

프리엠파시스, 프레임화, 윈도우 씌우기 들의 전처리 과정을 입력 음성에 적용한다.

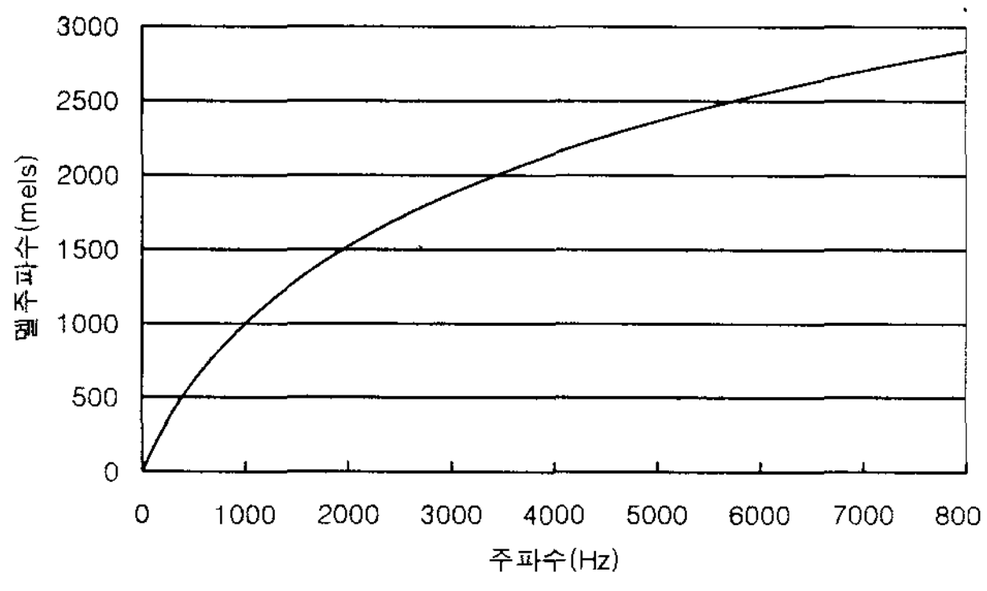


그림 4. 선형주파수 대 멜 주파수
Fig. 4. Linear Frequency vs. Mel Frequency.

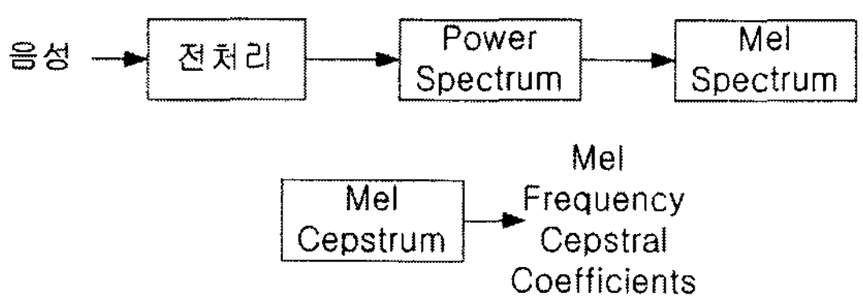


그림 5. MFC분석 구성도
Fig. 5. MFC Analysis Block Diagram.

2) 파워 스펙트럼

전처리 과정을 거친 음성을 프레임 단위로 DFT를 수행하여 파워 스펙트럼의 크기를 식(3)에 의하여 구한다.

$$S[k] = (real(X[k]))^2 + (imag(X[k]))^2 \qquad (3)$$

3) 멜 스펙트럼

파워 스펙트럼에 식(4)과 같이 해당 삼각 멜 가중치 필터를 각각 곱하여 멜 스펙트럼을 얻는다.

$$\tilde{S}[l] = \sum_{k=0}^{N/2} S[k] M_l[k] \quad l = 0, 1, ..., L-1 \qquad (4)$$

이 식에서 N은 DFT길이이고, L은 삼각 멜 가중치 필터의 총 개수를 말한다.

4) 멜 켑스트럼

멜 스펙트럼에 식(5)과 같이 자연대수를 취하여 IDCT를 적용하여 멜 주파수 켑스트럴계수를 얻는다.

$$c[n] = \sum_{i=0}^{L-1} ln(\tilde{S}[i] cos\left(\frac{\pi n}{2L}(2i+1)\right) \qquad (5)$$
$$c = 0, 1, ..., C-1$$

4. 퍼셉튜얼 선형예측 켑스트럴 계수(Perceptual Linear Predictive Cepstral Coefficients : PLCC)

선형예측분석은 파워 스펙트럼 $P(\omega)$의 평탄화된 스펙트럴 인벨로프를 구할 수 있다. 그러나 이러한 선형예측분석의 큰 단점 중에 하나는 선형예측 분석이 모든 주파수 대역에서 $P(\omega)$를 동일하게 근사화함으로서 인간의 청각 특성과 일치하지 않는다는 것이다. 즉 800Hz 이상에서는 주파수에 대한 분해 능력이 주파수가 증가함에 따라 감소하며, 대화시에는 음성의 가청 스펙트럼의 중간 대역에서 진폭 레벨에 대하여 더 민감하다는 특성을 갖는다는 것이다. 이러한 선형예측분석의 문제점을 보완하기 위하여 퍼셉튜얼 선형예측 분석이 연구되었다[10].

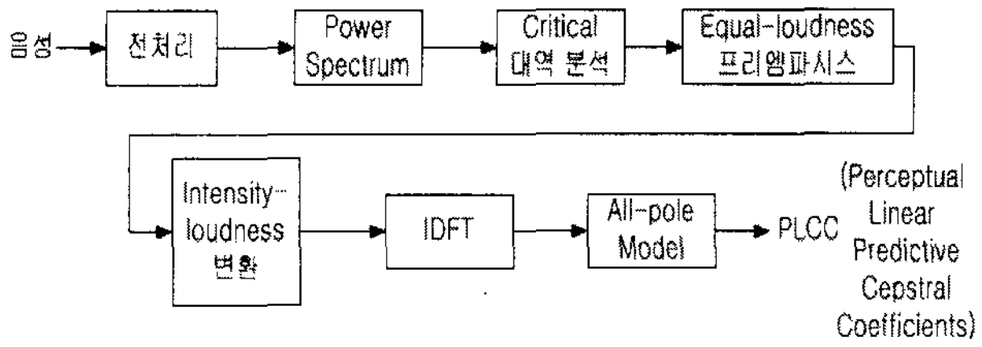


그림 6. 퍼셉튜얼 선형 예측 분석 구성도
Fig. 6. Perceptual Linear Predictive Analysis block diagram.

그림 6에 보이는 과정에 의하여 퍼셉튜얼 선형예측 분석이 이루어진다.

1) Critical 대역 분석

스펙트럼 $P(\omega)$의 주파수 ω를 식(6)을 적용하여 Bark 주파수 Ω로 변환하는 주파수 와핑(warping)을 한 후,

$$\Omega(\omega) = 6\ln\{\omega/1200\pi + [(\omega/1200\pi)^2 + 1]^{0.5}\} \quad (6)$$

와핑된 파워 스펙트럼을 식(7)으로 주어지는 critical 대역 커브 함수 $\Psi(\Omega)$와

$$\Psi(\Omega) = \begin{cases} 0 & \text{for } \Omega < -1.3 \\ 10^{2.5(\Omega+0.5)} & \text{for } -1.3 \leq \Omega \leq -0.5 \\ 1 & \text{for } -0.5 < \Omega < 0.5 \\ 10^{-1.0(\Omega-0.5)} & \text{for } 0.5 \leq \Omega \leq 2.5 \\ 0 & \text{for } \Omega > 2.5 \end{cases} \quad (7)$$

컨볼루션하여 critical 대역 파워 스펙트럼을 얻는다.

$$\Theta(\Omega_i) = \sum_{\Omega=-1/3}^{2.5} P(\Omega - \Omega_i)\Psi(\Omega) \quad (8)$$

2) Equal-loudness 프리엠파시스

인간의 다른 주파수들에 대한 민감도의 차이를 모의 실험 하여 얻은 함수를 $E(\omega)$라 하면 샘플링된 $\Theta[\Omega(\omega)]$는 식(9)에 의하여 프리엠파시스 된다.

$$\Xi[\Omega(\omega)] = E(\omega)\Theta[\Omega(\omega)] \quad (9)$$

3) Intensity-loudness 변환

청각에서의 파워 법칙에 따라 진폭에 대하여 3승근의 압축을 행하면 IDFT를 위한 과정이 완료된다.

$$\Phi(\Omega) = \Xi(\Omega)^{0.33} \quad (10)$$

Ⅳ. 인식 실험 및 결과

1. 인식모델

인식 대상은 고립 단어로서 '영', '일', '이', '삼', '사', '오', '육', '칠', '팔', '구', '공', '십', '백', '천', '만'의 15개의 숫자음들에 대한 정상어와 속삭임의 인식을 수행하였다. 인식 모델은 인식 성능과 유용성이 입증된 연속 은닉 마코프 모델(Hidden Markov Model with Continuous Mixture Density)[11][12]을 사용하였고, HMM 형태는 'left-to-right' 모델로 하였다. 연속 은닉 마코프 모델에서 각 단어의 모델은 식(11)으로 표현되며, 가우시안 혼합 확률 밀도는 식 (12)과 같다.

$$\hat{\lambda} = (A, \hat{B}, \hat{\mu}, \hat{U}) \quad (11)$$

여기서, A : 상태천이확률행렬, $\hat{B}$: 관찰확률함수, $\hat{\mu}$: 평균벡터, $\hat{U}$: 코배리언스 행렬을 나타낸다.

$$b_j(\mathbf{x}) = \sum_{k=1}^{M} c_{jk} N(\mathbf{x}, \mu_{jk}, U_{jk}) \quad (12)$$

여기서 $b_j(\,)$: 혼합(mixture)확률밀도, $\mathbf{x}$: 관찰특징벡터, c_{jk} : 혼합 밀도에 따른 가중치, $N()$: 정상분포확률밀도 M : 혼합수를 나타낸다.

모델링을 위한 음성 입력의 특징벡터 차수는 12차를 사용하였고, 상태 수는 s = 6, mixture 수는 M = 1, 코배리언스는 전코배리언스 행렬을 사용하였다.

2. 데이터베이스

실내 환경에서 성인 남자 20명, 성인 여자 20명이 '영', '일', '이', '삼', '사', '오', '육', '칠', '팔', '구', '공', '십', '백', '천', '만'의 15개 단어를 정상어 3회, 속삭임 3회씩 발음한 것을 16khz 샘플링, 16비트로 취득하여 각 단어당 정상어, 속삭임 각각 120개씩의 음성 데이터베이스를 구축하고 이들 음성데이터 각각에서 LPCC, FTCC, MFCC, PLCC의 특징파라메타 벡터들을 추출하였다. 각 스펙트럴계수들의 차수는 12차로 하였고, 프레임 길이는 25ms, 400 샘플, 프레임 이동은 10ms, 160 샘플로 하여 특징벡터들의 배열로 데이터베이스를 구축하였다.

따라서 15개의 각 단어 당 LPCC, FTCC, MFCC, PLCC 별로 120개의 특징벡터 배열들이 얻어졌고, 이들 중 각각의 90개는 모델링을 위한 훈련과정을 위하여 사용되었고 나머지 30개는 인식 시험을 위하여 사용되었다.

3. 실험결과

인식은 LPCC, FTCC, MFCC, PLCC의 각 특징벡터들에 대하여 정상어 모델, 속삭임 모델, 정상어 속삭임 통합 모델들로 모델링 하여 시험하였다. 실험 방법은 다음의 다섯 가지 경우를 대상으로 하여 인식 결과를 비교하였다.

각 특징벡터들 별로 표 1의 실험 방법을 통하여 얻어진 결과를 비교하여 보도록 한다.

표 1. 실험 방법
Table 1. Test Method.

경우	HMM 모델	인식 실험 대상
1	정상어	정상어
2	정상어	속삭임
3	속삭임	속삭임
4	정상어+속삭임	정상어
5	정상어+속삭임	속삭임

가. 특징벡터와 모델에 따른 대상 별 인식률

먼저 특징벡터 별 실험결과를 보면 LPCC를 특징벡터로 사용한 경우, 표 2에서의 실험 결과에서의 몇 가지 특징은, 첫째는 정상어 모델에 의한 속삭임의 인식은 42.3%의 아주 저조한 인식률을 보인다는 것이다. 이러한 인식률은 너무 저조하여 인식이라는 의미를 부여할 수조차 없을 것이다. 그러나 '영', '일', '이', '삼', '사', '오', '육', '칠', '팔', '구', '공', '십', '백', '천', '만'의 15개의 단어에 대한 인식률을 비교하여 보면 '사', '칠', '팔' 등과 같이 정상어에서 무성음으로 시작하는 어휘들의 인식률이 모음 혹은 유성자음으로 시작하는 다른 어휘들 보다 월등한 인식률을 보이고 있다. 이는 속삭임이 무성음과 유사한 특성을 갖는 것에 기인한 결과로 보인다.

둘째는 속삭임 모델을 사용하였을 경우 속삭임의 인식은 84.7%로 정상어 모델을 사용한 경우의 거의 2배에 이르렀으나 추가적인 인식률 상승 노력이 필요한 것으로 판단된다. 셋째, 정상어와 속삭임을 통합한 모델은 인식률에서 분리된 모델에 비하여 5-10%정도 저하되며 특히 속삭임을 인식 대상으로 하는 경우 인식률이 저조하였다.

FTCC를 특징벡터로 사용하는 경우는 LPCC의 경우와 유사하나, 모델별 인식률이 LPCC보다는 좀 더 평준화된 결과를 보여주었다. 그러나 속삭임을 인식 대상으로 하는 경우 정상어와 속삭임의 통합 모델보다 속삭임만의 모델의 경우가 8% 높은 인식률을 기록하였다.

MFCC의 경우 여러 음성인식 시스템에서 그 유용성이 입증되어 많이 사용되고 있으나 속삭임에 대해서는 LPCC나 FTCC와 유사한 결과를 보이고 있다. 인식 대상이 정상어인 경우 정상어 모델과 정상어, 속삭임 통합 모델에서 90% 이상의 인식률을 보이고 있으나 정상어, 속삭임 통합 모델의 경우가 정상어만의 모델보다 약 5% 낮은 인식률을 보였다. 속삭임을 인식 대상으로 하는 경우는 정상어, 속삭임 통합 모델의 경우가 속삭임만의 모델보다 8.4% 낮은 인식률을 보였다.

마지막으로 PLCC의 경우 인식 대상이 정상어인 경우 정상어 모델과 정상어, 속삭임 통합 모델에서 90% 이상의 인식률을 보이는 등 MFCC의 경우와 아주 유사한 결과를 보이고 있다.

모델별로 실험 결과를 분석하여 보면, 정상어를 HMM 모델링한 정상어모델에 정상어를 인식 실험한 경우, 그림 7에서 볼 수 있듯이 MFCC, PLCC를 특징벡

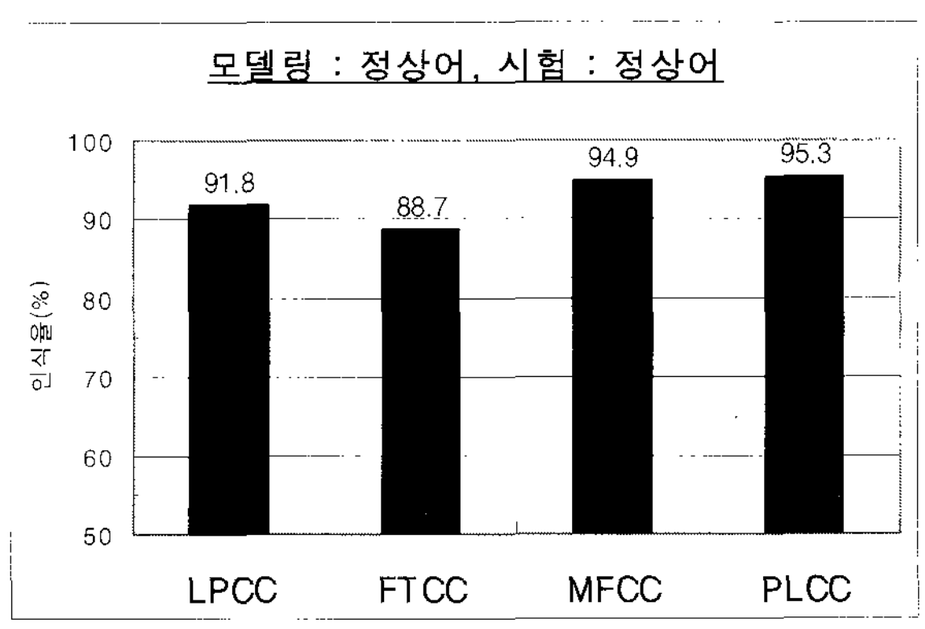


그림 7. 정상어 모델에 정상어를 대상으로 한 인식
Fig. 7. Recognition of normal speech applied to normal speech model.

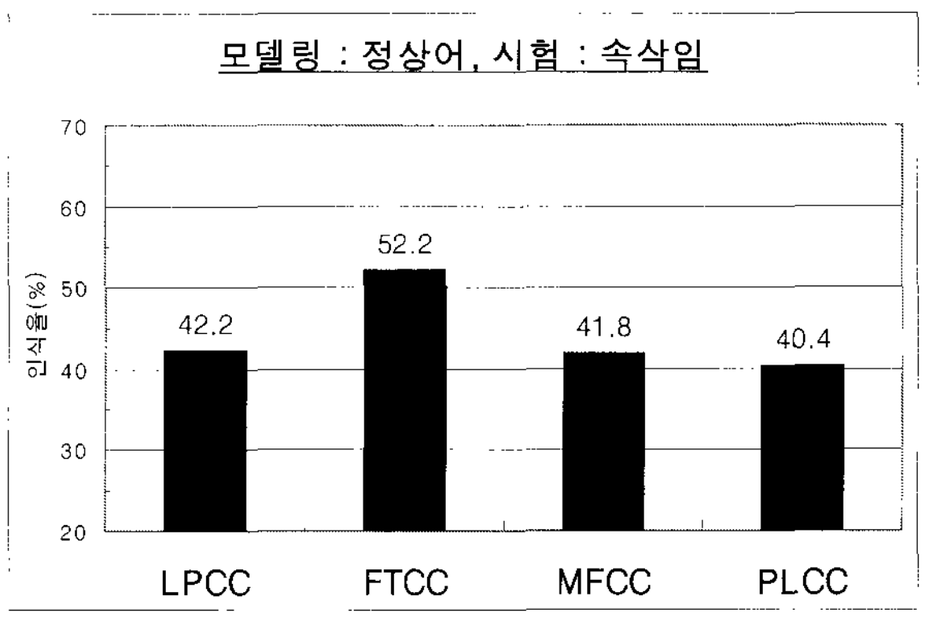


그림 8. 정상어 모델에서의 속삭임 인식
Fig. 8. Recognition of whispered speech applied to normal speech model.

표 2. 모델과 특징벡터에 따른 인식율
Table 2. Recognition rate of each model and feature vectors.

HMM 모델		정상어	정상어	속삭임	정상어 + 속삭임	정상어 + 속삭임
인식 대상		정상어	속삭임	속삭임	정상어	속삭임
인식률(%)	LPCC	91.8	42.3	84.7	87.3	75.0
	FTCC	88.7	52.2	86.0	86.4	78.0
	MFCC	94.9	41.8	86.4	90.0	78.0
	PLCC	95.3	40.4	86.0	90.9	81.3

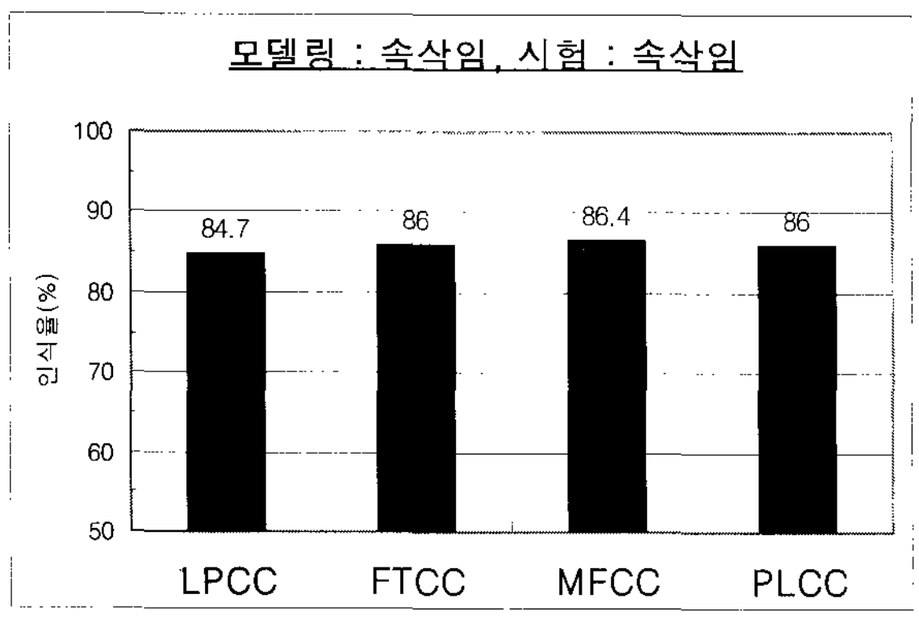


그림 9. 속삭임 모델에서의 속삭임 인식

Fig. 9. Recognition of whispered speech applied to whispered speech model.

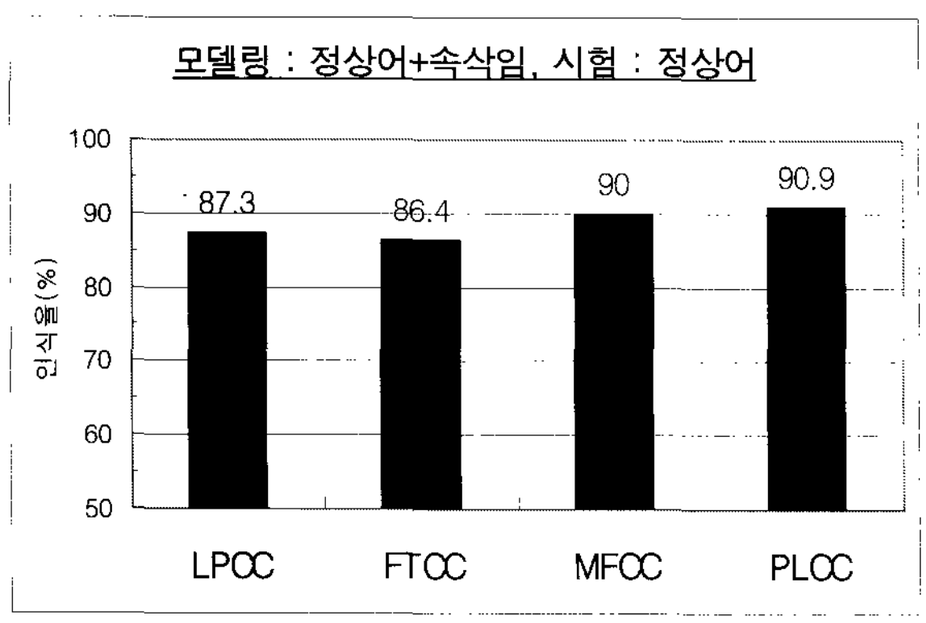


그림 10. 정상어와 속삭임 통합 모델에서의 정상어 인식

Fig. 10. Recognition of normal speeches applied to the model integrated normal speech with whispered.

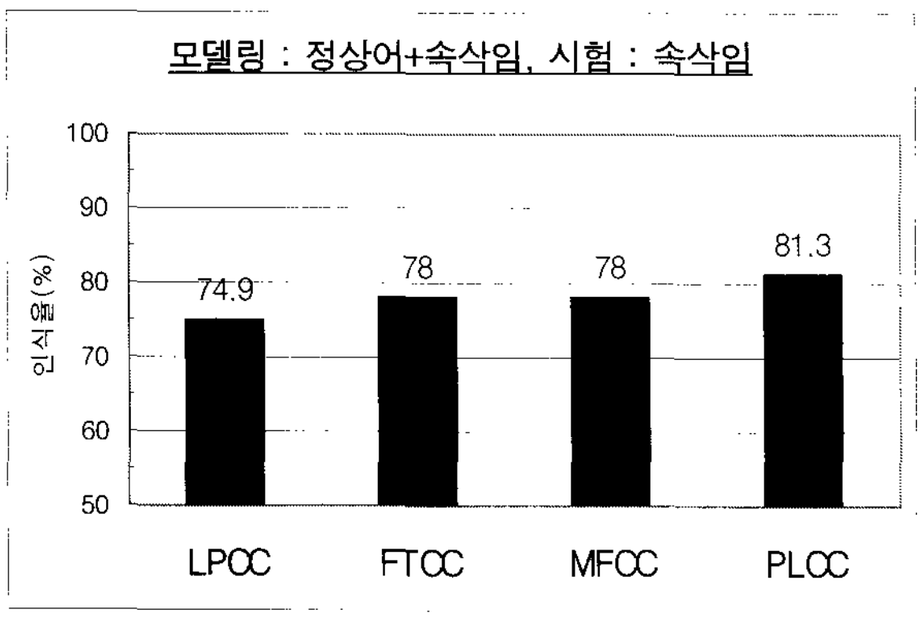


그림 11. 정상어와 속삭임 통합 모델에서의 속삭임 인식

Fig. 11. Recognition of whispered speeches applied to the model integrated normal speech with whispered speech.

터로 사용하였을 경우 인식률이 가장 뛰어난 결과를 얻을 수 있었다. 이 경우 FTCC를 특징벡터로 사용한 경우가 가장 인식률이 저조하였다.

정상어 모델에 속삭임을 인식 시험한 결과는 그림 8에서와 같이 인식률이 40~50% 정도로 매우 저조하였다.

정상어 인식에서 다른 특징벡터에 비하여 비교적 낮은 인식률을 보인 FTCC의 인식률이 속삭임을 인식 시험한 결과는 다른 특징벡터에 비하여 10% 이상 더 우수한 결과를 보였으나 이 경우에도 인식률이 50%대에 머물러 커다란 의미를 부여하기는 어렵다.

속삭임을 HMM 모델링한 속삭임모델에 속삭임을 인식 실험한 경우, 그림 9에서와 같이 4가지의 특징벡터에 대하여 인식률에서 큰 차이 없이 85~86% 정도의 인식률을 기록하였다. 단어별 인식률을 살펴보면 '이', '사', '오' 등에 오인식이 편중되어 있는 것을 볼 수 있다.

정상어와 속삭임을 함께 HMM 모델링한 '정상어+속삭임' 모델에 정상어를 인식 실험한 경우, 그림 10에서와 같이 MFCC와 PLCC의 특징벡터에 대하여 90% 이상의 인식률을 기록하였으나 LPCC와 FTCC에 대하여는 86~87%의 인식률을 보였다.

정상어와 속삭임을 함께 HMM 모델링한 '정상어+속삭임' 모델에 속삭임을 인식 실험한 경우, 그림 11에서와 같이 약 75%에서 81% 정도의 낮은 인식률을 기록하였으나 PLCC를 특징벡터로 사용한 경우가 다른 경우보다 3~6% 정도의 높은 인식률을 보였다.

나. 시험 결과 종합

시험 결과를 종합해 보면, 우선 속삭임에 대한 인식율의 대비를 위하여 시행한 정상어 모델에 대한 정상어 인식 시험은 특징벡터들 별로 90% 내외에서 95% 정도의 높은 인식률을 기록하였다. 특히 MFCC와 PLCC의 경우 인식률이 높았다. 그러나 시험에 사용된 모든 특징벡터에 대하여 정상어 모델에 의한 속삭임의 인식은 40~50%대의 아주 저조한 인식률을 보였으나 FTCC을 사용한 경우 다른 특징벡터를 사용한 경우보다 10%이상 우수한 결과를 보였다. 그러나 속삭임 모델을 사용하였을 경우 속삭임의 인식은 정상어 모델을 사용한 경우의 거의 2배에 이르렀으나 80% 중반대의 인식률을 보였다. 또한 정상어와 속삭임을 통합한 모델은 인식률에서 분리된 모델에 비하여 5-10%정도 저하되며 특히 속삭임을 인식 대상으로 하는 경우 인식률이 저조하였다.

정상어와 속삭임을 함께 HMM 모델링한 '정상어+속

삭임' 모델에 정상어를 인식 실험한 경우 80%대 후반에서 90%에 이르는 인식률을 보였으나 정상어 모델을 사용한 경우에 비하여 약 5% 정도 인식률이 저하되었다. 또한 이 모델에 속삭임을 인식 실험한 경우도 약 75%에서 81%의 인식률로 속삭임 모델의 경우보다 5~10% 정도 인식률이 저하되었다.

Ⅳ. 결 론

본 논문에서는 근래에 들어와서 공공장소에서의 전화 통화 빈도의 증대에 따라 통화에서의 정숙성과 보안성 요구가 증대되고 따라서 속삭임의 사용이 증가하는데 따른 속삭임 인식을 위하여 음성인식에 많이 사용되고 있는 LPCC, FTCC, MFCC, PLCC의 특징벡터에 따른 인식률을 은닉 마코프 모델을 이용, 시험하여 속삭임 인식에 보다 적합한 특징벡터를 찾고, 또한 정상어 모델, 속삭임 모델, 정상어, 속삭임 통합 모델들에 속삭임을 인식 시험하고 결과를 분석하여 가장 적합한 시스템을 찾으려고 하였다.

인식 시험을 통하여 속삭임의 정상어 모델에 대한 인식은, 정상어가 90% 이상의 높은 인식률을 기록한 것에 비하여, 40~50%대의 아주 저조한 인식률을 보였다. 이는 정상어 모델로 속삭임을 인식하는 것은 정상어와 속삭임 특성의 커다란 차이로 인하여 실용성이 없다고 할 수 있다. 또한 속삭임 모델에 대한 속삭임의 인식은 80% 중반대의 인식률을 보였고, 이는 인식률만 좀 더 향상시킨다면 실용적으로 사용할 수 있다는 것을 보여준다. 정상어와 속삭임을 함께 HMM 모델링한 '정상어+속삭임' 모델에 정상어와 속삭임의 인식 시험 결과는 정상어 모델과 속삭임 모델에 각각 정상어와 속삭임을 인식 시험한 결과보다 정상어의 경우는 약 5%, 속삭임의 경우는 5~10% 인식률이 떨어졌다. 따라서 동일 단어에 대하여 정상어와 속삭임을 각각 별도의 모델로 모델링 하는 것이 정상어, 속삭임 모두에서 높은 인식률을 얻을 수 있었으며 이 경우에 특징벡터로는 MFCC 혹은 PLCC를 사용하는 것이 거의 유사하게 높은 인식률을 얻을 수 있었다. 그러나 모델 수를 줄여 인식에서의 계산량의 감축이 필요한 경우에는 인식률이 다소 떨어지기는 하나 '정상어+속삭임' 모델을 사용할 수 있겠고 이 경우는 PLCC를 특징벡터로 사용하는 것이 더 높은 인식률을 갖는다는 것을 확인하였다.

향후 연구 과제는 본 연구에서 80% 중반의 인식률을 보이고 있는 속삭임 모델에 대한 속삭임의 인식률을 향상시킬 수 있는 알고리듬 혹은 특징벡터를 찾는 것으로써 속삭임 사용 빈도 증가에 따른 속삭임 인식 수요에 대처하는 것이 될 것이다.

참 고 문 헌

[1] S. T. Jovičić and M. M. dordević, "Acoustic features of whispered speech," *ACUSTICA-acta acustica,* vol. 82, pp. S228, 1996.

[2] Holmes J. N and A. P. Stephens, "Acoustic correlates of Intonation in whispered speech", *J. Acoust. Soc. Am.,* 73, S87, 1983.

[3] K. J. Kallail and F. W. Emanuel, "Formant Feature Differences Between Whispered and Voiced Sustained Vowels," *ACUSTICA-acta acustica,* vol. 84, pp. 739-743, 1998.

[4] Taisuke Itoh, Kazuya Takeda, and Fumitada Itakura, "Acoustic Analysis and Recognition of Whispered Speech," IEEE Int. Conf. on ASSP, vol. 1, pp. 389-392, 2002.

[5] J. L. Flanagan, *Speech Analysis Synthesis and Perception,* Springer-Verlag, New York, 2nd. edition, 1972.

[6] R-M. S. Heffner, *General Phonetics,* The University of Wisconsin Press, Madison, 1960.

[7] L. R. Rabiner and R. W. Shafer, *Digital Processing of Speech Signals,* Prentice-Hall, Englewood Cliffs, 1978.

[8] J. W. Picone, "Signal Modeling Techniques in Speech Recognition," in Proc. IEEE, vol. 81, no. 9, pp. 1215-1247, Sep. 1993.

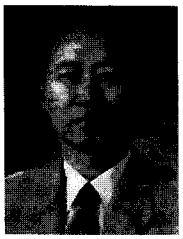
[9] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Trans. on Acoustic, Speech, and Signal Processing,* vol. 28, pp. 357-366, 1980

[10] H. Hermansky, "Perceptual linear predictive(PLP) analysis of speech," *J. Acoust. Soc. Am.*, pp. 1738-1752, 1990.

[11] L. R. Rabiner and B. H. Juang, "An Introduction to Hidden Markov Models," *IEEE ASSP MAGAZINE,* pp 4-16, Jan. 1986.

[12] L. R. Rabiner, B. H. Juang, S. E. Levinson, and M. M. Sondhi, "Recognition of Isolated Digits Using Hidden Markov Models with Continuous Mixture Densities," *AT&T Technical Journal,* Vol. 64, No. 6, pp. 1211-1234, July-August 1985.

저 자 소 개



박 찬 응(정회원)
1977년 서강대학교 전자공학과 학사
1989년 서강대학교 전자공학과 석사
1997년 서강대학교 전자공학과 박사
1984년~1992년 대우통신 종합연구소 수석연구원
1995년~현재 인덕대학 정보통신전공 부교수
<주관심분야 : 음성인식, 음성합성, 영상처리, 통신시스템>