

# Wavelet 특징 파라미터를 이용한 한국어 고립 단어 음성 검출 및 인식에 관한 연구

이 준 환<sup>†</sup> · 이 상 범<sup>††</sup>

## 요 약

본 논문에서는 한국어 고립 단어 음성을 대상으로 웨이브릿 변환을 이용한 특징 파라미터를 추출하여 음성 검출 및 인식 특징으로 사용하였다. 웨이브릿 변환을 이용한 음성 검출 결과, 기존의 영교차와 에너지를 이용한 음성 검출 방식에 비해 음성의 경계 부분에서 보다 정확한 음성 검출 결과를 나타내었다. 또한 웨이브릿 변환의 일반화된 형태인 웨이브릿 패킷 변환을 이용한 웨이브릿 패킷 계수에 MFCC를 적용한 특징 파라미터를 인식에 적용한 결과, 기존의 FFT를 이용한 MFCC 특징 파라미터와 대동한 인식 결과를 나타내어 음성 분석 및 인식을 위한 웨이브릿 변환을 이용한 특징 파라미터의 효용성을 입증하였다.

## A Study on Korean Isolated Word Speech Detection and Recognition using Wavelet Feature Parameter

Jun-Hwan Lee<sup>†</sup> · Sang-Burm Rhee<sup>††</sup>

## ABSTRACT

In this paper, feature parameters, extracted using Wavelet transform for Korean isolated word speech, are used for speech detection and recognition feature. As a result of the speech detection, it is shown that it produces more exact detection result than the method of using energy and zero-crossing rate on speech boundary. Also, as a result of the method with which the feature parameter of MFCC, which is applied to Wavelet packet coefficients using Wavelet packet transform of the generalized form of Wavelet transform, is applied to the recognition, it is shown that the result is equal to the result of the feature parameter of MFCC using FFT in speech recognition. So, it has been verified the usefulness of feature parameters using Wavelet transform for speech analysis and recognition.

### 1. 서 론

최근 정보의 범람과 각종 미디어의 발달, 많은 자동차 기계 시설에 대한 의존성의 폭발적인 증가로 인간과 기계 사이의 보다 자연스럽게 효율적인 인터페이스에 대한 많은 연구가 진행되고 있다. 특히 음성은 인

간만이 가지는 아주 간편하고도 빠른 의사 전달 수단으로서 음성을 이용한 인터페이스가 활용될 경우 이의 기술적인 과급 효과는 경제 및 사회적인 면에서 아주 획기적인 일이 될 것이며, 현재 컴퓨터 및 정보통신 업체들은 간단하게나마 편리한 인터페이스를 갖춘 음성 인식 기술을 컴퓨터, 통신기기, 가전제품, 게임기등에 탑재하기 위한 연구에 박차를 가하고 있다. 그러나 실생활에 적용키 위한 음성 인식 기술의 활용은 미미한 실정이며, 이것은 인간의 음성이 가지는 다양성을 정확히 분석, 표현할 수 있는 특징 파라미터 추출 및

\* 본 논문은 한국과학재단의 산학협력 연구지원(과제번호 : 97-2-11-0101-2)에 의한 연구결과임

† 정 회 원 : 극동대학교 전산정보학부 멀티미디어학과 교수

†† 정 회 원 : 단국대학교 공학부 컴퓨터공학전공 교수

논문접수 : 2000년 4월 26일, 심사완료 : 2000년 6월 7일

패턴 인식 기술에 대한 다양한 연구를 통해 개선되어  
져야 한다.

따라서 본 논문에서는 한국어 고립 단어 음성에 대하  
여 웨이브릿 변환(Wavelet transform)을 이용한 특징 파  
라미터를 추출 후, 음성 검출 및 인식을 수행하기 위한  
특징 파라미터로서의 가능성을 평가하였다[1, 4, 5]. 특히  
기존의 인식 특징 파라미터로 널리 사용되어지고 있는  
FFT(fast Fourier transform)를 이용한 MFCC(Mel-  
frequency cepstrum coefficient) 특징 파라미터와 인간  
의 청각과 유사한 주파수 분해 특성을 갖는 것으로 알  
려져 있는 웨이브릿 변환을 이용한 MFCC 특징 파라미  
터를 추출한 후, 연속 CHMM(CHMM: continuous hid-  
den Markov model) 인식 모델에 적용, 평가하였다[6,  
10-12].

실험 결과, 제안된 특징 파라미터가 화자 독립 학습  
데이터의 인식률의 경우, FFT를 이용한 MFCC 특징  
파라미터에 비해서 우위를 나타냈으며, 화자 독립 미  
지 데이터에 대해서는 거의 대등한 인식 결과를 나타  
내어, 향후 음성 분석 및 인식을 위한 특징 파라미터  
로서의 효용성을 입증하였다.

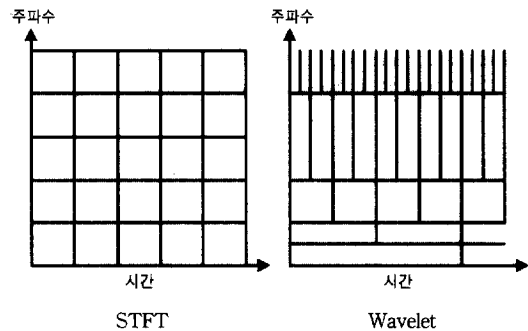
본 논문의 2장에서는 음성 검출 및 인식 특징 파라미  
터로 사용한 웨이브릿 변환에 대해 살펴보고, 3장에서  
는 웨이브릿 변환을 이용한 음성 검출 및 CHMM 인식  
기법 웨이브릿 패킷 필터 뱅크(Wavelet packet filter  
bank)를 이용한 MFCC 인식 특징 파라미터 추출 방법  
을 살펴본다. 4장에서는 제안된 알고리즘 및 특징 파라  
미터를 이용한 고립 단어 음성 검출 결과와 CHMM을  
이용한 인식 결과를 비교, 검토하였으며 5장에서는 결  
론을 나타내었다.

## 2. 웨이브릿 변환

일반적으로 음성을 분석하기 위해서 다양한 특징 파  
라미터들이 추출되어 사용되어지고 있다. 본 논문에서  
는 한국어 고립 단어 음성 신호 검출을 위해 다 해상  
도(multi-resolution) 특성을 갖는 웨이브릿 변환을 적  
용하였다. 또한 웨이브릿 변환의 일반화한 형태인 웨  
이브릿 패킷 변환을 이용한 인식 특징 파라미터를 추  
출하여 고립 단어 인식에 적용하였다.

실제 음성 신호를 이해하는 인간의 청각 체계는 저  
주파에 민감하고 고주파에는 덜 민감하게 작용한다고  
알려져 있으며, 음성 파형을 분석해보면 고주파 부분

보다는 상대적으로 저주파 부분에 정보가 집중되어 있  
다. 웨이브릿 변환(Wavelet transform)은 음성 신호를  
인간의 청각 특성과 유사한 높은 주파수 대역은 주파  
수 해상도가 낮고 시간 해상도는 높으며, 낮은 주파수  
대역은 주파수 해상도가 높고, 시간 해상도는 낮은 다  
해상도 특성을 갖고 있다(그림 1). 따라서 인간의 음성  
분석 특성과 유사한 음성 분석을 위해 저주파와 고주  
파의 시간 해상도가 다른 웨이브릿 변환을 이용하여  
효과적인 음성, 분석 및 인식을 수행하려는 시도가 많  
이 이루어지고 있다. 본 논문에서는 완전 재생 조건을  
갖는 Daubechies 20차 Orthogonal 기저(basis)를 사용  
한 이산 웨이브릿 변환을 이용하여 음성 신호를 분석  
하였다[1-3].



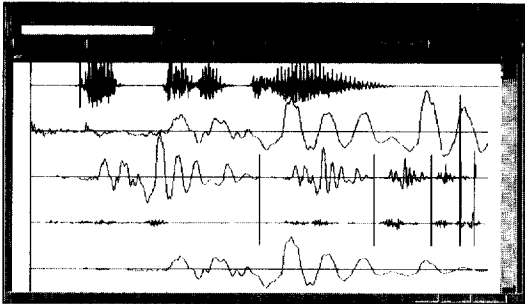
(그림 1) STFT(short time Fourier transform)와 Wavelet  
변환의 시간-주파수 해상도

웨이브릿 변환에 의해 음성 신호의 주파수 분해된  
필터 뱅크의 출력은 웨이브릿 변환의 특성인 Decima-  
tion 수행으로 인해 각 주파수 대역 출력 신호의 샘플  
수가 다르게 발생한다. 따라서 이산 웨이브릿 변환에  
의해 각각의 분해된 주파수 대역 신호들을 각 대역 신  
호 성분 외의 성분들을 제거한 후, 역 이산 웨이브릿  
변환을 수행하면 원래의 신호와 동일한 샘플 수를 갖  
는 필터 뱅크 수만큼의 다 해상도, 즉 주파수 분해된  
시변 신호들을 얻을 수 있다. 이것은 Octave 대역 필  
터의 이상적인 결과를 나타내며, 구현 가능한 필터의  
수는 웨이브릿 변환 트리(tree) 알고리즘의 Decima-  
tion 특성에 의해 식 (1)과 같이 제한된다. 웨이브릿 변환에  
의해 1 Octave 필터링 된 각각의 웨이브릿 주파수 대  
역들은 보다 세밀한 분석을 위해서 (그림 2), (그림 3)  
과 같이 저주파 영역에 대하여 1/2 Octave 웨이브릿  
필터링 된 후 사용될 수 있으며, 최대 발생 가능한 필

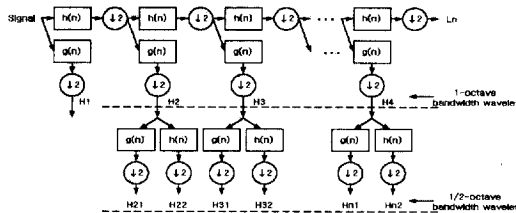
터 बैं크의 수는 식 (1), 식 (2)와 같이 계산되어질 수 있다[3-5].

$$1 \text{ Octave 필터뱅크 수} = \frac{\log_{10}(\text{분석 구간 샘플 수})}{\log_{10}(2) + 1} \quad (1)$$

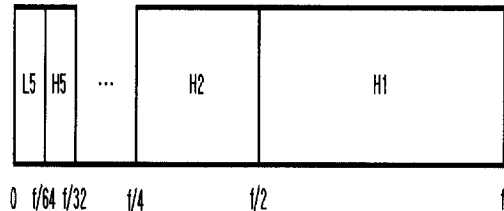
$$1/2 \text{ Octave 필터뱅크 수} = (1 \text{ Octave 필터뱅크 수} * 2) - 1 \quad (2)$$



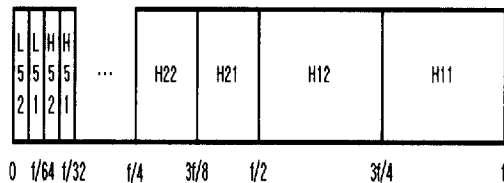
(그림 2) 1 Octave 웨이브릿 필터링 결과



(그림 3) 1, 1/2 Octave 웨이브릿 대역 분할



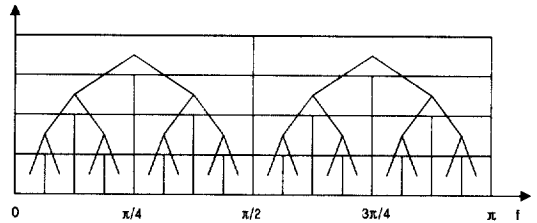
(그림 4) 1 Octave 웨이브릿 주파수 대역



(그림 5) 1/2 Octave 웨이브릿 주파수 대역

웨이브릿 패킷 변환은 (그림 3)과 같이 1 Octave, 1/2 Octave 웨이브릿 필터링을 통해 단순히 분해 대역의 저주파 대역 영역을 중심으로 계속하여 분할해 나가는 웨이브릿 변환과는 달리, 고주파 영역을 포함한 전 영역에 걸쳐 대역 분할이 수행되는 경우이다.

즉 웨이브릿 패킷 변환은 신호 해석에 있어서 더 넓은 범위의 요구를 만족시킬 수 있도록 웨이브릿 변환의 일반화된 형태이며 일반적인 형태의 웨이브릿 패킷 필터링된 트리 구조를 (그림 6)에 나타내었다[6].



(그림 6) 웨이브릿 패킷 트리

### 3. 음성 검출 및 인식 알고리즘

#### 3.1 웨이브릿 특징 파라미터를 이용한 음성 구간 검출 알고리즘

음성 인식을 수행하기에 앞서서 전처리 과정으로 잡음 및 묵음 영역을 제외한 음성 구간을 검출하는 과정이 수행되어진다. 특히 최종 음성 인식 시스템의 평가 시 음성 검출의 정확도는 인식 성능을 크게 좌우하므로 신뢰성 있는 음성 인식을 위해서는 정확한 음성 구간의 검출이 필히 선행되어야 한다[6, 7].

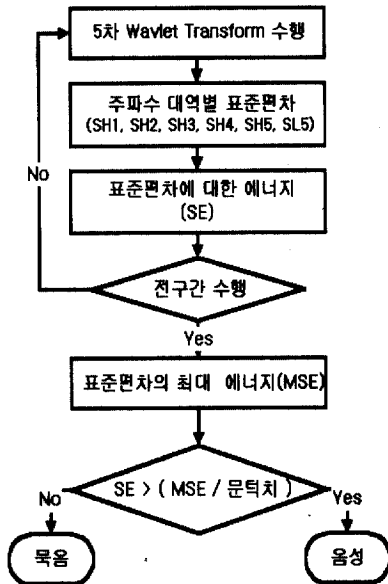
따라서 본 논문에서는 신뢰성 있는 음성 구간 검출을 위해 웨이브릿 변환을 이용한 음성 검출 알고리즘을 제안하여 적용하였다. 먼저 묵음 및 잡음을 제외한 음성 구간의 검출을 위해 전체 음성 데이터에 대하여 8ms를 하나의 분석 구간(frame)으로 하여 4ms씩 1/2 중첩시킨 형태로 프리엠퍼시스(pre-emphasis)후 식 (3), 창 함수(Hamming window)를 사용하여 분석하였다(식 (4)).

$$S(n) = S(n) - 0.95S(n-1) \quad (3)$$

$$W(n) = 0.54 - 0.46 \cos\left(2\pi \frac{n}{N-1}\right), \quad 0 \leq n \leq N-1 \quad (4)$$

각 프레임별 1 Octave 5차 웨이브릿 변환을 통해 필터링 된 각 분석 구간 단위의 주파수 분해된 6개의 대역(SH1, SH2, SH3, SH4, SH5, SL5) 각각에 대하여, 분석 구간내의 주파수 대역별 음성 신호의 변화 정도를 파악할 수 있도록 웨이브릿 계수의 표준 편차를 추출한 후, 각 표준 편차들에 대한 에너지를 구하였다. 구하여진 표준 편차의 에너지가 실험에 의해 측정된 임계값보다 크면 음성 구간, 작으면 묵음 구간이라 판별한다.

실험을 통하여 표준 편차에 대한 최대 에너지 값의 55%를 임계값으로 설정하였으며, 검출된 음성 구간에서 음성이 일정 시간을 지속하여야 한다는 특성을 이용하여 음성이라 판별된 구간이 8ms 이상 지속되지 못하는 경우에는 묵음 영역으로 분류함으로써 순간 발생하는 잡음 영역을 음성 구간으로 판별하는 오검출을 감소시켰다. (그림 7), (그림 8)은 음성 구간 검출 알고리즘의 순서도 및 이를 이용한 결과 화면을 나타내고 있다.

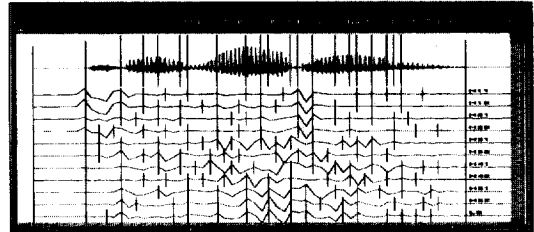


(그림 7) 음성 구간 검출 알고리즘

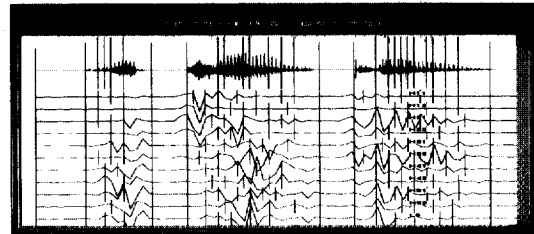
### 3.2 인식 알고리즘 및 웨이브릿 기반 인식 특징 파라미터

음성에서 추출한 특징 파라미터를 이용한 대표적인 인식 알고리즘으로는 신경망을 이용한 NN(neural net-

work), 시간적인 정합을 이용한 DTW(dynamic time warping), 확률적인 방법으로 잘 알려진 HMM(hidden Markov model)등이 있다. 그 중에서도 확률적인 방법으로 알려진 HMM은 1980년대 이후 활발한 연구가 진행되어져 오고 있으며, 현재 음성 인식의 대부분의 영역에 적용되어지고 있다[8, 9].



(a) '재활용' 음성의 음성 검출



(b) '못지않게' 음성의 음성 검출  
(그림 8) 음성 검출 결과 예

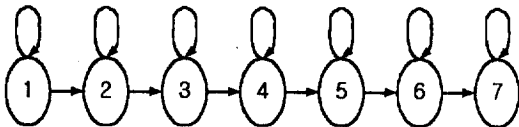
HMM은 음성 신호를 시변 랜덤 과정(time-varying random process)이라 가정하고, 비슷한 신호 특성들을 대표하는 상태(state)와 상태간의 천이, 즉 신호의 시변 특성을 나타내는 상태 천이 행렬(state transition matrix) A, 그리고 각 상태에서의 신호 특성을 나타내는 상태 관측 행렬(state observation matrix) B의 두 개로 구성된다. 여기에 초기에 발견될 상태의 확률을 표현하는 초기 상태(initial state)  $\pi$ 를 포함시켜 HMM을  $\lambda = (A, B, \pi)$ 로 표시한다. 이 때 각 상태에서 관찰되는 파형들의 유사한 안정적인(quasi-stationary) 특성을 나타낸 것이 상태 관측 행렬 B이다. 상태 관측 행렬 B를 얻기 위해서 관측 확률 밀도 함수가 사용되며 이것은 어떠한 상태 천이가 일어날 때 또는 어떠한 상태에 놓여 있을 때 관측 벡터가 나올 조건부 확률이다.

이러한 확률 밀도 함수를 코드북을 이용한 벡터 양자화를 통한 이산적인 값의 이산 HMM(DHMM : discrete

HMM)이나 연속적인 값의 연속 HMM(CHMM : continuous HMM)으로 표현할 수 있다. 또한 이산 HMM의 양자화 오차와 연속 HMM의 많은 계산량을 극복하기 위한 중간 모델로서 반연속 HMM(SCHMM : semi-continuous HMM)이 있다 특히 연속 HMM(CHMM)은 상태 관측 행렬 B의 관측 확률 분포를 이산 분포 HMM과 달리 연속적인 확률 밀도 함수로 표현하며, 이 때 음성 신호의 발생 확률 분포를 Gaussian, Poisson, Gamma, Cauchy 등 여러 가지 확률 밀도 함수를 사용하여 표현할 수 있다. 그러나 계산상의 편리함으로 인해 가우시안 분포 함수를 많이 사용하며, 여러 개의 정규 분포 함수에 가중치를 주고 합친 혼합 가우시안 확률 밀도 함수는 어떠한 확률 분포도 근사 시킬 수 있기 때문에 남녀의 음성 주파수 차이 등 다양한 변이성을 모델링하기에 적합한 장점을 지니고 있다.

연구 결과에 의하며, HMM 모델들 중에서, 연속 HMM 모델이 이산 HMM, 반연속 HMM보다 약 1.3 배, 1.15배 이상의 연산 시간을 더 소모하나, 인식률에 있어서는 두 가지에 비해 다소 뛰어난 성능을 나타내는 것으로 발표되고 있다[8, 10-12].

따라서 본 논문의 한국어 고립 단어 인식을 위한 인식 알고리즘으로 인식률을 고려한 CHMM 인식 모듈을 구성하여 적용하였으며, 이 때 상태 모델로 (그림 9)와 같이 7개의 좌우 구조의 상태 모델을 사용하고, 상태 당 16개의 혼합 가우시안 연속 확률 분포로 모델링 하였다.



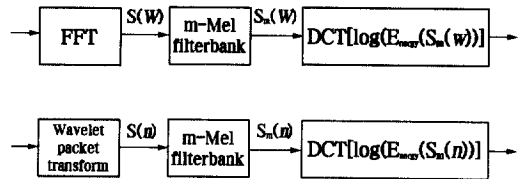
(그림 9) 7-state CHMM 초기 상태 모델

구성된 CHMM 인식 모듈에 적용키 위한 인식 특징 파라미터를 추출하기 위해 음성 분석 구간에 대하여 이산 최대 차수의 웨이브릿 패킷 변환을 적용하였다. 적용된 웨이브릿 패킷 변환은 고주파, 저주파 영역을 계속하여 주파수 대역 분할해 나가는 일종의 필터뱅크 방식으로서, 최대 차수의 대역 분할을 통해 최종적으로 발생하는 계수들은 일종의 주파수 영역 신호로 간주될 수 있다. 따라서 웨이브릿 필터뱅크의

출력들은 일반적으로 음성 인식시 많이 사용되어지고 있는 FFT를 기반으로 한 MFCC 특징 파라미터의 추출 과정과 비교해 볼 때 FFT 수행 과정을 대체 할 수 있다.

즉 웨이브릿 패킷 필터뱅크를 이용하여 최대 차수로 분해되어 추출된 웨이브릿 패킷 계수들은 멜 스케일 필터뱅크로 필터링 된 후, 멜 스케일링 필터뱅크에 의해 추출된 각 주파수 대역마다 에너지의 로그를 취한 뒤 DCT를 수행하면, FFT를 이용한 MFCC 특징 파라미터 추출 과정과 유사한 웨이브릿 패킷 MFCC 특징 파라미터를 추출할 수 있다.

본 논문의 실험에 사용된 특징 파라미터의 추출 과정을 (그림 10)에 나타내었다.

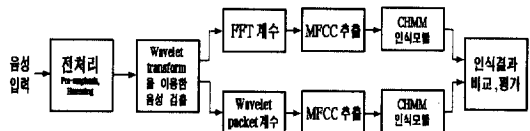


(그림 10) FFT 및 웨이브릿 패킷 변환을 이용한 MFCC 추출 순서도

#### 4. 결과 및 검토

본 논문에서는 웨이브릿 변환을 이용한 음성 검출을 수행한 후, FFT를 이용한 MFCC 특징 파라미터와 본 논문에서 제안, 적용한 웨이브릿 패킷 MFCC 특징 파라미터를 한국어 고립 단어 음성에 대한 인식 파라미터로 적용하여 인식률을 평가하였다.

실험에 사용된 한국어 고립 단어 음성은 데이터베이스화된 음성 시료 중 하나인 국어공학센터에서 구성된 남성화자 TIMIT 모델 고립 단어 PBW 음성에서 무작위 추출한 데이터이며, 전체적인 실험 과정을 살펴보면 (그림 11)과 같다.



(그림 11) 실험 시스템 구성도

실험은 Visual C++ 컴파일러를 이용하여 Pentium III 600MHz PC 환경하에서 수행되었으며, 실험에 사용된 한국어 고립 단어 음성 DB 및 녹음된 데이터의 사양, 전체 화자의 수 및 화자당 고립 단어 음성 데이터의 수를 <표 1>, <표 2>에 나타내었다. <표 3>에서는 인식하기 위해 사용된 한국어 고립 단어의 종류에 대하여 나타내었다.

<표 1> 음성 데이터의 조건

대상 어휘	데이터 형태	헤더 정보	샘플링 주파수	양자화 레벨
한국어 고립단어 TIMIT 모델 음성 DB	WAV	1024 byte	16KHz	16bit

<표 2> 음성 검출 및 인식을 위해 사용된 데이터의 형태 및 개수

구 분	음성 검출 데이터	학습 데이터	미지 데이터
화자의 수	20명	12명	8명
1 화자당 처리 단어의 개수 (화자당 2번 발생)	100개	100개	100개
전체 데이터의 갯수	4000개	2400개	1600개

<표 3> 고립 단어 음성 인식 실험에 사용된 어절 및 단어 음성 데이터의 종류

음성 단어의 종류	개수	내용
음성 단어의 종류	10개	계속하고, 과언이, 그에게, 낙화담, 당뇨병, 요컨대, 위대한, 자유와, 청와대, 컴퓨터
	30개	협동을, 특특히, 재할용, 요소와, 애초에, 쓸데없는, 세울을, 빼앗기고, 불안정한, 분석됐다, 부딪혀, 민방위, 무엇보다도, 못지않게, 멀티미디어, 남품대금, 꿩뚫어, 깨끗이, 교역을, 과일
	50개	헤메고, 헌법을, 해외증권, 이뤄져야, 웅크리고, 요즈음, 왕조의, 영등포, 연결되어, 국회의원, 업체의, 엄격한, 애기입니다, 수필집, 밤중에, 되겠지요, 동아일보, 내다봤다, 개연성을, 휩싸여
	70개	가요는, 개연성을, 거처하는, 겪어야, 고마워, 교육과, 교통사고, 국제청, 규모의, 그처럼, 기쁨을, 끌어안고, 끝끝내, 낮잠을, 내려온, 내뿜는, 내었다, 넘도록, 놓았을, 눈시울이
	100개	뉴욕에서, 대법원, 대통령, 더없이, 더이상, 도입된, 도처에, 동아일보, 동양의, 두번째, 뒤쪽에, 떠올랐다, 뚜렷이, 뛰었다, 뜻밖에, 마셨다, 마십시오, 마음과, 메우고, 명의로, 명확히, 못된다, 못살겠다, 바쁘게, 배우자, 뺏속에, 변화에, 병원에, 보컬선거, 부여와,

음성 검출의 경우, 4000개의 고립 단어 음성을 대상으로 실험 결과, 고립 단어 음성의 시작과 끝점뿐만 아니라 음성 발생 중간에 포함되는 묵음의 경우에도 정확한 형태의 끝점 경계를 찾아내었다. 이러한 결과는 먼저 동일 실험자에 의해 수행된 음성 신호 파형의 확대에 의한 수 작업 음성 구간 검출 결과와 비교해 볼 때, 전체 고립 단어 음성중 85%이상의 단어가 25ms 이내, 11%는 38ms이내, 4%는 60ms이내의 오차 범위내에서 검출되었으며, 일반적으로 많이 사용되는 에너지, 영 교차의 임계값 반복 적용에 의한 알고리즘을 구성하여 실험한 결과, 음성 검출 경계 영역 에러의 평균인 48ms보다 20ms이상 적게 나타나므로서 묵음과 음성의 경계 부분에서 보다 정확한 음성 검출 결과를 나타내었다.

또한 현재 많은 음성 인식 응용에서 사용되고 있는 FFT 주파수 영역을 이용한 16차 MFCC 특징 파라미터와 본 논문에서 적용한 웨이브릿 패킷을 이용한 16차 MFCC 특징 파라미터를 CHMM 인식 모듈에 입력하여 비교한 결과, 화자 독립 학습 데이터의 경우 웨이브릿 패킷 기반 MFCC 특징 파라미터가 FFT 기반 MFCC 특징 파라미터를 이용할 경우에 비해 단어 수의 확장에도 불구하고 평균적으로 약 0.10~0.25%의 인식을 향상을 나타내었다. 또한 최대 100단어 1600개의 화자 독립 미지 데이터에 대하여 FFT 기반 MFCC 특징 파라미터를 이용한 인식의 경우 97.94%, 웨이브릿 패킷 기반의 MFCC 특징 파라미터는 97.81%의 인식을 나타내어 거의 대등한 인식 결과를 나타내었다.

<표 4> FFT를 이용한 MFCC 특징 파라미터의 CHMM 인식 결과

	음성 데이터	전체데이터 갯수	인식데이터 갯수	오인식 갯수	인식률 (%)
화자 독립 학습 데이터 (단어당 24개)	10단어	240	239	1	99.58
	30단어	720	719	1	99.86
	50단어	1200	1197	3	99.75
	70단어	1680	1675	5	99.70
	100단어	2400	2392	8	99.67
화자 독립 미지 데이터 (단어당 16개)	10단어	160	159	1	99.34
	30단어	480	475	5	98.33
	50단어	800	785	15	98.13
	70단어	1120	1100	20	98.21
	100단어	1600	1567	33	97.94

〈표 5〉 웨이브릿 패킷을 이용한 MFCC 특징 파라미터의 CHMM 인식 결과

음성 데이터	전체데이터 개수	인식데이터 개수	오인식 개수	인식률 (%)	
화자 독립 학습 데이터 (단어당 24개)	10단어	240	239	1	99.58
	30단어	720	719	1	99.86
	50단어	1200	1198	2	99.83
	70단어	1680	1677	3	99.82
	100단어	2400	2394	6	99.75
화자 독립 미지 데이터 (단어당 16개)	10단어	160	159	1	99.34
	30단어	480	476	4	99.17
	50단어	800	788	12	98.50
	70단어	1120	1096	24	97.86
	100단어	1600	1565	35	97.81

5. 결 론

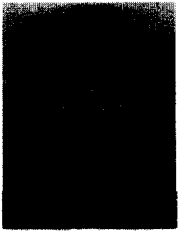
본 논문에서는 한국어 고립 단어 음성에 대하여 웨이브릿 변환에 의해 추출된 특징 파라미터를 음성 검출에 이용한 결과, 일반적인 영교차, 에너지에 의한 음성 검출 알고리즘에 비해 향상된 효율적인 음성 검출 결과를 나타내었다. 또한 최대 차수로 분해된 웨이브릿 패킷 개수 영역에 MFCC를 적용한 음성 특징 파라미터와 FFT 주파수 영역을 이용한 멜 스케일을 적용한 특징 파라미터를 각각 CHMM 인식 모듈에 적용한 결과, 화자 독립 학습 데이터의 경우 본 논문에서 사용한 웨이브릿 패킷 기반의 MFCC 특징 파라미터가 FFT 기반의 MFCC 특징 파라미터에 비해 평균적으로 약 0.2%의 인식률 향상을 나타내었다. 그러나 화자 독립 미지 데이터의 경우, 인식코자 하는 고립 단어 종류 및 개수의 변화와 확장에 따라 약간의 인식률 저하 및 상승이 발생하였지만, 특징 파라미터 상호간의 우위를 판별할 정도의 명확한 인식률 차를 발견할 수 없었다. 또한 특징 파라미터 추출을 위한 연산 시간에서도 대등한 결과를 나타내어 인식 특징 파라미터로서 웨이브릿 변환의 효용성을 검증하였다.

웨이브릿 변환을 이용하여 추출된 특징 파라미터는 기존의 FFT 기반의 특징 파라미터 추출 방법과 비교해 볼 때, 시간-주파수 영역의 다 해상도 특성으로 인해, 변환 단계에서 발생하는 분해된 주파수 대역 결과들을 음성 분석 및 인식 특징으로 동시에 사용할 수 있는 등 활용 방법의 개발에 따라 더욱 향상된 음성 분석 및 인식 결과를 나타내리라 사료된다. 향후 본

논문에서 적용한 웨이브릿 패킷 개수의 MFCC 특징 파라미터를 음성의 동적인 변화 특성을 잘 드러내는 차분 캡스트럼 형태의 특징 파라미터로 추출하여 다른 인식 특징 파라미터들과 독립적인 가중치를 적용한 혼합된 형태를 CHMM 인식 모듈에 적용할 경우, 인식률 향상이 기대된다.

참 고 문 헌

- [1] Olivier Rioul, Martin Vetterli, "Wavelets and Signal Processing," IEEE SP magazine, pp.14-38, Oct. 1991.
- [2] F. Hlawatsch, G. F. Boudareaux-Bartels, "Linear and Quadratic Time-Frequency Signal Representations," IEEE SP magazine, pp.21-33, April. 1992.
- [3] G. Strang and T. Nguyens, 'Wavelet and Filter Banks', Wellesley-Canbrige Press, 1996.
- [4] 김바울, 윤철현, 홍광석, 박병철, "청각 모델에 기초한 음성 특징 추출에 관한 연구", 대한전자공학회논문지, 제33권 제4호, 1996년 4월.
- [5] T. Tan, R. Lang, Heiko Schroder, Andrew Spray "Applying wavelet analysis to speech segmentation and classification," SPIE Wavelet Application, Vol. 2242, 1994.
- [6] 경북대학교, '웨이브릿 변환을 이용한 음성분석 및 응용', 한국과학재단 최종 연구 결과 보고서, 1999.
- [7] L. F. Lemel, "An improved endpoint detector for isolated word recognition," IEEE Trans. Acoust., Speech and Signal Processing, Vol.2, No.3, pp.406-412, 1994.
- [8] L. Rabiner and B. H. Juang, 'Fundamentals of Speech Recognition', Prentice Hall, 1993.
- [9] 연세대학교, '맨 머신 인터페이스를 위한 음성신호처리 응용연구', 한국과학재단 최종 연구 결과 보고서, 1998.
- [10] L. R. Rabiner and B. H. Juang, "An introduction to Hidden Markov Models," IEEE ASSP Magazine, pp.4-16, Jan. 1986.
- [11] L. R. Rabiner, "Tutorial on Hidden Markov Model and Selected Application in Speech Recognition," Proc. IEEE, Vol.77, No.2, pp.257-285, 1989.
- [12] X. D. Huang, Y. Ariki, M. A. Jack, "Hidden Markov Models for Speech Recognition," Edinburgh Univ., 1990.



## 이 준 환

e-mail : ljh@kdu.ac.kr

1994년 단국대학교 전자공학과  
(공학사)

1996년 단국대학교 대학원 전자  
공학과(공학석사)

1999년 단국대학교 대학원 전자공학과 박사과정 수료

2000년~현재 단국대학교 멀티미디어 산업기술연구소  
연구원

1999년~현재 극동대학교 전산정보학부 멀티미디어학  
과 겸임교수

관심분야 : 디지털 신호처리, 음성 및 영상처리



## 이 상 범

e-mail : sbrhee@dankook.ac.kr

1974년 연세대학교 전자공학과  
(공학사)

1978년 서울대학교 대학원 전자  
공학과(공학석사)

1986년 연세대학교 대학원 전자공학과(공학박사)

1984년 미국 IOWA대학교 컴퓨터공학과 객원교수

1979년~1999년 단국대학교 전자·컴퓨터공학과 교수

1997년~1999년 단국대학교 교무·연구처장

1997년~현재 단국대학교 멀티미디어산업기술연구소장

2000년~현재 단국대학교 공학부 컴퓨터공학전공 교수

관심분야 : 컴퓨터구조, 패턴인식, 디지털 신호처리