

한국어 고립 단어 음성의 자음/모음/유성자음 음가 분할 및 인식에 관한 연구

이 준 환[†] · 이 상 범^{††}

요 약

한국어의 경우 음소 형태가 아닌 한국어 음성 특유의 음운 변동 현상에 의해 음향학적으로 다른 형태의 음가를 생성한다. 그러므로 한국어를 이해하는 확장된 음성 인식 시스템의 구축은 한국어 특유의 음운 변동 특징에 대한 연구 및 규칙의 생성과 함께, 음소가 아닌 음가를 기반으로 음성을 인식 후, 후치리로 한국어 음운 변동 규칙을 적용하는 형태이어야 한다. 따라서 본 논문에서는 먼저 한국어 특유의 음운 변동 현상을 반영하는 텍스트 기반의 한국어 음운 변동 Rule-Based 시스템 구성 후, 시스템의 텍스트 음가 생성 결과에 기초하여 한국어 고립 단어 음성을 대상으로 한 음가 분할 예미 경계점의 검출과 검출된 경계점 사이의 비균일 블록 구간에 대한 병합 및 자음/모음/유성자음 음가 인식을 수행하여 한국어 음성에서의 음가 인식 가능성을 검토하였다.

A Study on Consonant/Vowel/Unvoiced Consonant Phonetic Value Segmentation and Recognition of Korean Isolated Word Speech

Jun-Hwan Lee[†] · Sang-Bum Lee^{††}

ABSTRACT

For the Korean language, on acoustics, it creates a different form of phonetic value not a phoneme by its own peculiar property. Therefore, the construction of extended recognition system for understanding Korean language should be created with a study of the Korean rule-based system, before it can be used as post-processing of the Korean recognition system.

In this paper, text-based Korean rule-based system featuring Korean peculiar vocal sound changing rule is constructed and based on the text-based phonetic value result of the system constructed, a preliminary phonetic value segmentation border points with non-uniform blocks are extracted in Korean isolated word speech. Through the way of merge and recognition of the non-uniform blocks between the extracted border points, recognition possibility of Korean voice as the form of the phonetic value has been investigated.

1. 서 론

현재 한국어 음성의 인식 단위로는 크게 음소(phoneme), 음절(syllable), 단어(word) 등으로 나눌 수 있다.

음절 형태를 인식의 단위로 사용할 경우, 음절의 위치를 비교적 쉽게 찾아낼 수 있는 장점을 가지고 있으나, 초/중/종성의 결합에 의한 음절의 발생 종류가 음소 형태에 비해 너무 많은 단점을 갖고 있다. 단어 혹은 그 이상의 형태를 인식의 단위로 사용할 경우 화자 발음 속도의 부분적인 변화와 잡음의 영향을 음절

※ 본 논문은 한국과학기술연구원 산학협력 연구지원(과제번호 · 97-2-11-0101-2)에 의한 연구결과임

† 경희대 · 서울대학교 공학부 컴퓨터공학과 교수

†† 서울대학교 공학부 컴퓨터공학과 교수
논문접수 · 2000년 4월 10일, 심사완료 · 2000년 5월 17일

이나 음소보다 덜 받고, 조음 결합 현상이나 일부 음소의 이음 현상을 무시할 수 있으므로 인식 알고리즘을 구현하기가 비교적 쉽고 인식률도 높은 장점을 갖는다. 그러나 인식 어휘의 수가 증가함에 따라 저장할 표준 패턴의 수와 기억 용량이 선형적으로 증가하는 단점으로 인해 대용량의 음성 인식에는 부적합하다. 음소를 인식의 단위로 사용할 경우, 음향학적으로 쉽게 분류되지 않으며 많은 음운 변동 규칙을 필요로 하는 단점이 있지만, 단어를 쉽게 음소로 표기할 수 있고 적은 기억 용량으로 대부분의 어휘에 적용할 수 있다는 장점으로 인해 활발한 연구가 진행되고 있다[1, 2].

그러나 한국어는 음성의 발생 시 음소 형태가 아닌 한국어 음성 특유의 음운 변동 현상에 의해 음향학적으로 다른 형태의 음가(phonetic value)를 생성하는 특성을 갖고 있다. 따라서 한국어 음성 언어를 이해하는 확장성을 갖는 음성 인식 시스템의 구축은 한국어 특유의 음운 변동 특징에 대한 연구 및 규칙의 생성과 함께, 음소가 아닌 음가(phonetic value)를 기반으로 음성을 인식한 후, 후처리로 한국어 음운 변동 규칙을 적용하는 형태이어야 한다.

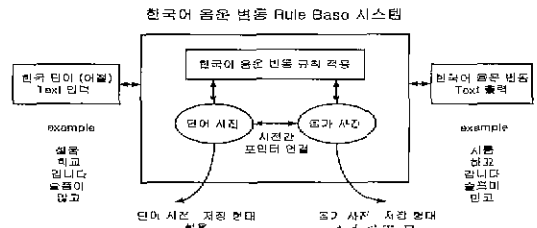
따라서 본 논문에서는 한국어 고립 단어 음성에서 자음/모음/유성자음 음가를 분할 및 인식하기 위한 근거로서 사용된 한국어 특유의 음운 변동 현상을 반영하여 텍스트 단어 및 음가 결과를 생성하는 Rule-Based 시스템에 대해 살펴본다[3]. 3장에서는 한국어 음운 변동 Rule-Based 시스템에 의해 추출된 발음 형태의 텍스트 한국어 음가를 기초로 하여 한국 국어공학센터에서 구성한 고본도 한국어 고립 단어 음성 DB를 대상으로 음성 집합과 자/모음/유성자음 음가 단위의 인식을 수행하기 위해 제안된 음가 경계점을 포함하는 비균일 블록 검출 및 인식 알고리즘에 대해 기술하였다[4-11]. 4장에서는 음가 분할 및 인식 결과에 의해 한국어 음성에서의 음가 인식 가능성을 검토하였으며 5장에서는 결론을 나타내었다

2. 한국어 음운 변동 Rule-Based 시스템

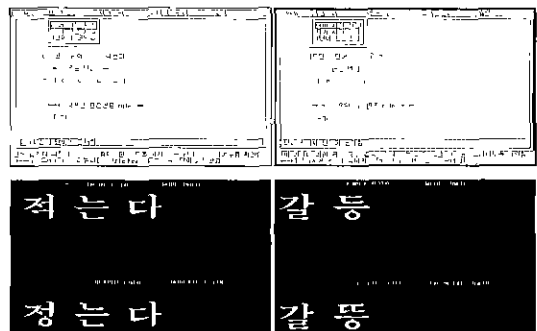
한국어는 음성의 발생시 표기 언어인 한글의 음소 단위 형태가 아닌 특유의 음운 변동 현상에 의해 음향학적으로 다른 형태의 음가(phonetic value)를 생성한다. 따라서 실질적인 한국어 음성 언어를 이해하는 확장성을 갖는 음성 인식 시스템의 구축은 디지털 신호 처리 측면에서 음소 형태의 구분이 아닌 발생 신호 형

태인 음가 신호를 기반으로 음성을 인식한 후, 후처리로 한국어 음운 변동 규칙이 적용된 형태의 인식 결과를 도출해야 한다.

따라서 본 논문에서는 한국어 고립 단어를 대상으로 음운 변동 규칙에 따라 텍스트 형태의 단어 및 음가를 생성할 수 있도록 한국어 표준어 규정의 표준 발음법을 이용한 한국어 음운 변동 Rule-Based 시스템을 구성하였다. 특히 한국어의 음운학적 특성상, 음운 변동의 대부분이 앞 음절의 받침과 뒤 음절의 첫소리가 합쳐서 일어나므로 받침별 27개 그룹의 규칙으로 설계된 한국어 음운 변동 Rule-Based 시스템을 구성한 후, 고립 단어 한국어 음절을 텍스트상의 자료 입력 과정을 통하여 검증하였다. (그림 1)은 구성된 한국어 음운 변동 Rule-Based 시스템의 전체적인 구성을 나타내며 (그림 2)는 한국어 음운 변동 규칙이 적용되어 생성된 텍스트 형태의 한국어 고립 단어 및 음가를 나타낸다[3].



(그림 1) 한국어 음운 변동 Rule Based 시스템의 구조



(그림 2) 한국어 음운 변동 Rule-Based 시스템이 적용된 텍스트 출력

본 논문에서는 (그림 2)와 같이 한국어 음운 변동 Rule-Based 시스템의 한글 단어 입력에 대한 텍스트 음가 표기 결과에 근거하여, 실제 한국어 음성에서 이에 대응되는 자음/모음/유성자음 음가로의 분할 및 인식 가능성을 평가하였다. 시스템의 음가 표기 결과에

기초한 분할, 인식된 자음/모음/유성자음 음가는 향후 자음/모음/유성자음 음가의 인식을 향상 및 개개의 음가를 인식하기 위한 특정 파라미터의 추출과 인식 알고리즘의 개발을 통하여 확장되어질 수 있으며, 구성된 한국어 음운 변동 Rule-Based 시스템은 개개의 음가 인식 예리 영역에 대한 후처리로 활용될 수 있다.

3. 한국어 자음/모음/유성자음 음가 분할 및 인식 알고리즘

31 음성 분석을 위한 특징 파라미터의 추출 및 기존의 인식 알고리즘

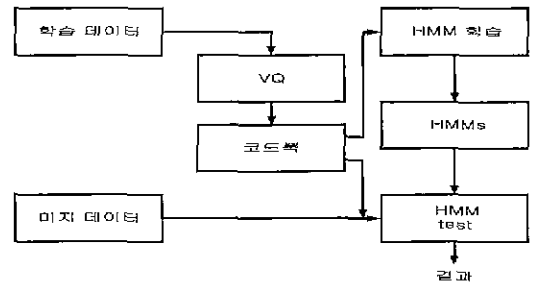
일반적으로 기존의 단어, 음절, 음소 형태의 음성 인식시, 음성을 분석하기 위한 다양한 특징 파라미터들이 추출되어 사용되어지고 있다. 본 논문에서는 한국어 음성 신호의 검출, 음가 분할, 인식을 위해 에너지, 영교차, MFCC(Mel Frequency Cepstral Coefficient) 등의 특징 파라미터외에 인간의 청각 특성과 유사한 주파수 분해 특성을 갖는 다 해상도 웨이브릿 계수의 표준 편차와 이를 이용한 MLR(Maximum Likelihood Ratio) 등 다양한 특징 파라미터를 추출하여 음성 분석 및 인식에 이용하였다[1, 2, 4-7].

음성에서 추출한 특징 파라미터를 이용한 기존의 대표적인 인식 알고리즘으로는 신경망을 이용한 NN(Neural Network), 시간적인 지합을 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, 사용하는 입력 관찰 심볼 및 확률 분포에 따라 이산적인 HMM(DHMM : discrete HMM), 연속적인 HMM(CHMM : continuous HMM) 그리고 반연속 HMM(SCHMM : semi-continuous HMM)으로 분류된다[8-11]

이산 HMM은 음성 신호에서 추출한 입력 관찰 벡터를 양자화하지 않고 그대로 표현하는 CHMM 및 CHMM과 DHMM의 중간 단계인 SCHMM보다 인식이 떨어지는 단점이 있으나, 이산적인 1차원 형태의 입력 벡터를 정보 손실이 가장 적은 코드워드(codeword)의 인덱스(index) 열로 양자화하므로써 학습과 인식 과정에서 학습 횟수 및 계산량이 적고, 양자화 적용 방식의 개선에 의해 정보 손실을 최소화 할 수 있다

따라서 본 논문에서는 자음/모음/유성자음의 음가 분

할 예비 경계점 검출 후, 추출된 비균인 구간의 자음/모음/유성자음의 음가 인식을 위해 기존의 단어 및 음소 인식에서 많이 활용되어지고 있는 LBG 벡터양자화를 통한 이산 HMM 방법을 기본 인식 구조로 채택하였으며, 인식기의 블록도는 (그림 3)과 같이 구성된다.



(그림 3) 벡터 양자화, HMM을 이용한 기본적인 인식 구조

32 음성 구간 검출 알고리즘

음소 혹은 음가 단위의 음성 인식을 수행하기에 앞서서 전처리 과정으로서 잡음 및 묵음 영역을 제외한 음성 구간을 검출하는 과정이 수행되어야 한다. 특히 최종 음성 인식 시스템의 평가시 음성 검출의 정확도는 인식 성능을 크게 좌우하므로 신뢰성 있는 음가 분할 및 인식을 위해서는 정확한 음성 구간의 검출이 선행되어야 한다[1, 2].

본 논문에서는 신뢰성 있는 음성 구간 검출을 위해 웨이브릿 변환을 이용한 음성 검출 알고리즘을 제안하여 적용하였다[5-7]. 먼저 묵음 및 잡음을 제외한 음성 구간의 검출을 위해 전체 음성 데이터에 대하여 16ms를 한 프레임(frame)으로 하여 8ms씩 1/2 중첩시킨 프레임 단위로 프리엠퍼시스(pre-emphasis)후(식 (1)), 창 함수(Hamming window)를 사용하여 분석하였다(식 (2))

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

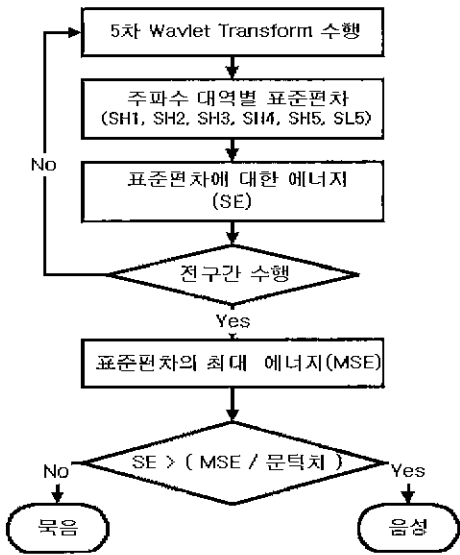
$$W(n) = 0.54 - 0.46 \cos\left(2\pi \frac{n}{N-1}\right), \quad 0 \leq n \leq N-1 \quad (2)$$

각 프레임별 1 Octave 5차 웨이브릿 변환 대역 필터링된 각 프레임 단위의 주파수 분해된 6개의 대역(H1, H2, H3, H4, H5, L5) 각각에 대하여, 프레임 단위 주파수 대역별 음성 신호의 변화 정도를 파악할 수 있도록 웨이브릿 계수의 표준 편차를 추출한 후, 각 표준 편차들에 대한 에너지를 구하였다. 구하여진 표

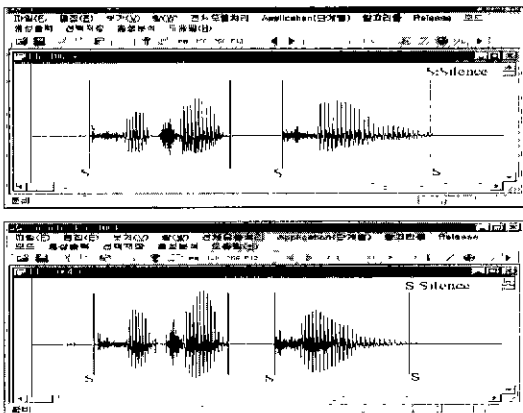
준 편차의 에너지가 실험에 의해 측정된 임계값보다 크면 음성 구간, 작으면 묵음 구간이라 판별한다.

본 논문에서는 실험을 통하여 표준 편차에 대한 최대 에너지 값의 55%를 임계값으로 설정하였다. 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 검출 에러를 제거하였다

(그림 4), (그림 5)는 음성 구간 검출 알고리즘의 순서도 및 이를 이용한 결과 화면을 나타내고 있다.



(그림 4) 음성 구간 검출 알고리즘



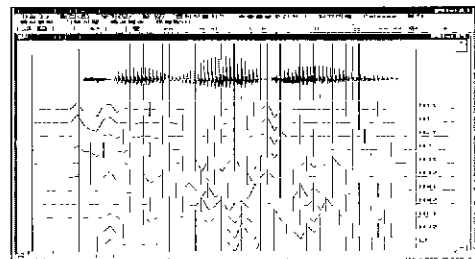
(그림 5) 음성 검출 결과 예

3.3 자음/모음/유성자음 음가 분할 및 인식을 위한 예비 경계점 검출 알고리즘

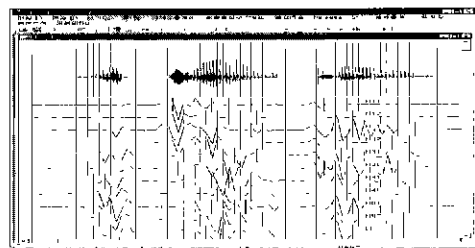
한국어 음성은 음성의 발생시 음소보다는 음가 형태의 자/모음 신호 발생 형태를 갖고 있으므로 음소보다는 음가 형태의 신호를 기준으로 한 인식이 필요하다. 따라서 한국어 음성의 음가 형태를 기반으로 한 인식을 수행하기 위해서는 전처리 과정으로서 인식 대상인 음가를 추출하는 음가 분할 과정이 요구된다.

이를 위해 본 논문에서는 한국어 음운 단동 Rule-Based 시스템을 구성한 후, 구성된 시스템의 결과에 근거하여 신호의 특징이 유사한 음가 경계점을 포함하고 있는 비균일 단위의 분할 예비 경계점을 선행 검출하고, 검출된 비균일 블록들에 대한 분석, 병합 및 인식 과정을 통해 향후 개개의 음가 단위 형태의 인식을 수행하기 위한 가능성을 모색하였다[3, 4].

먼저 자음/모음/유성자음 음가 형태의 신호 경계점에 근접한 음가 분할 예비 경계점을 검출하기 위하여, 검출된 음성 구간내에서 균일 프레임 단위로 1 Octave 5차 웨이브릿 분해된 각 주파수 대역별 신호에 대해 최하위의 저주파 영역을 제외한 상위 5개의 영역을 분해하는 1/2 Octave 웨이브릿 필터링을 수행하여 각 프레임을 전체 11개의 주파수 대역으로 세분하여 분석한다.



(a) 재촬영음성의 비균일 블록 검출



(b) '못지않게' 음성의 비균일 블록 검출

(그림 6) 음가 분할 예비 경계점을 갖는 비균일 블록의 검출

음성 검출 구간내의 각 균일 프레임 단위의 분석 구간에서 생성된 11개의 웨이브릿 분해 주파수 대역에 대하여 각 대역별 표준 편차를 추출하여 2차 미분을 수행한 후, 각 주파수 대역에서 표준 편차의 변화 정도가 급격한 피크(peak) 영역을 추출하면 (그림 6)과 같이 음가 분할을 위한 실질적인 경계점들을 포함한 분할 예비 경계점이 검출된다.

검출된 분할 예비 경계점은 실질적인 음기 경계점을 완전히 포함하고 있음을 실제 한국어 음성에 대한 헨드 세그멘테이션을 통해 검증하였다. 그러나 (그림 6)과 같이 1차 분석을 통해 검출된 음가 분할 예비 경계점들은 실질적인 음가 경계점들을 포함하고 있으나, 경계점들 사이에서 너무 많은 예비 경계점들이 검출되는 과검출 결과가 발생한다

따라서 너무 짧은 검출 구간을 갖는 분할 예비 경계점들을 인식을 위한 분석 대상에서 제외키 위해 분할 예비 경계점 사이의 비균일 단위 블록 구간들에 대하여 재 분석이 이루어진다. 먼저 신호의 관찰에 의해 한국어 음운 변동 Rule-Based 시스템의 텍스트 출력 음가 형태와 비교하여 실질적인 음가 분할 예비 경계점을 훼손하지 않는 형태의 구간별 병합을 수행한 후, 최저 개수의 자음/모음/유성자음 음가의 분할 예비 경계점을 다시 추출하는 2차 분석이 수행된다.

1차 검출된 분할 예비 경계점 사이의 비균일 구간에 대하여 다시 1 Octave 5차 웨이브릿 변환이 수행되고, 웨이브릿 변환 결과로 추출된 6개 웨이브릿 주파수 분해 대역의 계수들에 대하여 각 대역별 주파수의 변이 정도를 측정할 수 있도록 식 (3)과 같은 MLR(Maximum Likelihood Ratio)을 추출하였다[7]

또한 14차 MFCC 에너지, 14차 LPC 캡스트럼 에너지 및 영교차 등의 특징 파라미터를 비균일 구간 병합을 위한 분석 특징 파라미터로 추출하였다

$$MLR_k = \frac{N}{2} \left| \ln \left(\frac{\sigma^2}{\sigma_{noise}^2} \right) - \left(\frac{\sigma_k^2}{\sigma_{noise}^2} \right) \right| \quad (3)$$

MLR_k : k 주파수 대역에 대한 MLR 값

σ^2 : 해당 분석 구간에 대한 음성신호의 표준 편차

σ_k^2 : 웨이브릿 변환 후 k번째 주파수 대역 계수들의 표준 편차

σ_{noise}^2 : 묵음구간에 대한 웨이브릿 변환 후 k번째 주파수 대역 계수들의 표준 편차

N: 분할 주파수 대역의 개수

(그림 6)과 같이 과 검출된 1차 분할 예비 경계점 사이의 비균일 블록 구간은 MLR, MFCC 등의 특징 파라미터 추출과 분석을 통해 실질적인 음가 경계점을 훼손하지 않는 형태로 임계값 설정에 의해 블록 구간별 병합이 수행된다.

웨이브릿 변환을 이용하여 각 주파수 대역별로 추출된 MLR을 MH1, MH2, MH3, MH4, ML4라 정의하며, 식 (4)와 같이 산출된 DCV(dicesion consonant vowel), MFCC 에너지, LPC 캡스트럼 에너지, 영교차 특징 파라미터에 대하여 실험에 의한 임계값의 반복 적용을 통하여 비균일 구간별 병합을 수행하였으며 구간 병합 결과는 (그림 7)과 같다.

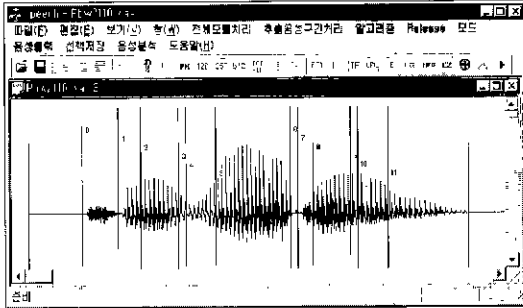
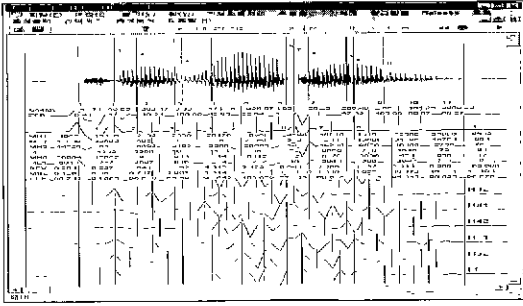
$$DCV = \frac{MH1 + MH2 + MF3 + MH4 + MF5}{MH4 + MF5 + ML5} \quad (4)$$

3.4 제안된 비균일 검출 구간의 자음/모음/유성자음 인식 알고리즘

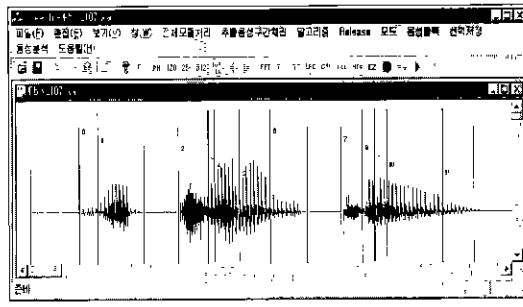
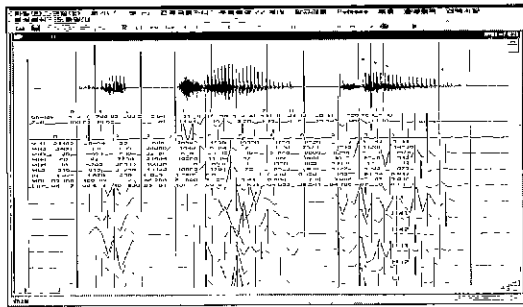
검출, 병합된 분할 예비 경계점 사이의 각 비균일 블록 구간은 VQ-HMM에 의한 최소 단위 음가인 실질적인 자음/모음/유성자음으로 인식하기 위해 비균일 구간별 1/2 중첩의 128 샘플 단위 분석 구간을 기준으로 하여 재 분석 후 자음/모음/유성자음 음가로 인식된다

검출된 비균일 블록의 자음/모음/유성자음 판별을 위한 인식 특징 파라미터로 분할 예비 경계점 사이의 분석 구간에 대해 16차 MFCC가 추출되었다. 코드북을 생성하기 위한 벡터 양자화 알고리즘으로 LBG 알고리즘이 사용되었으며, 구성된 LBG 코드북은 16 코드워드, 256 인덱스 레벨의 크기를 갖는다. 추출 특징 파라미터의 벡터 양자화된 관찰 벡터들은 HMM으로 입력되며 HMM의 초기 상태 모델로 (그림 8)과 같이 Left-to-Right 구조의 제한된 3-state HMM 모델을 사용하였다. Left-to-Right 모델은 상태의 천이가 제자리 아니면 오른쪽에 있는 상태로 이동하는 특징을 가지고 있으며, 신호의 특성이 시간에 따라 연속적으로 변화하는 음성을 모델링하기에 적합한 구조로서 일반적으로 음성 인식이 많이 사용되는 모델이다[8-11].

음가 분할 예비 경계점 사이의 비균일 검출 블록에 대한 자음/모음/유성자음 음가 인식시 이산 형태의 HMM 모델을 적용할 경우, 벡터양자화로 인한 특징 파라미터의 정보 손실에 의해 오인식의 확률이 커지는 단점이 발생한다. 또한 한국어 유성자음의 경우, 난류적인 자음의 특성보다는 모음의 주기적인 특성을 더 많이 내포하는

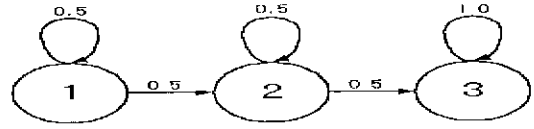


(a) '재활용' 음성의 '재활용' 음가 형태 최소단위 분할 예비 경계점 검출(화자 A)



(b) '못지않게' 음성의 '못지않게' 음가 형태 최소단위 분할 예비 경계점 검출(화자 B)

(그림 7) 웨이블릿 MLR을 이용한 DCV, MFCC 에너지, LPC 켄스트럼 에너지, 영교차를 이용한 비균일 구간의 병합



(그림 8) 3개의 상태를 갖는 left-to-right HMM 상태 모델

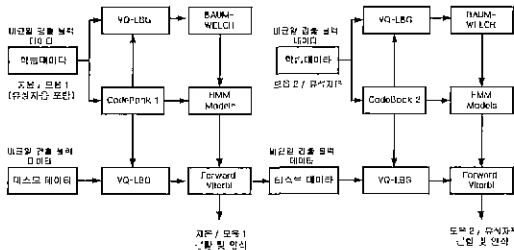
것으로 알려져 있으며, 이로 인해 자음/모음/유성자음 형태의 특정 파라미터의 추출 및 이를 이용한 단일 코드북 생성후 자음/모음/유성자음 형태의 VQ-HMM 모델을 적용하여 인식을 수행할 경우 각 형태에서 많은 인식 에러가 발생한다.

이러한 벡터 양자화 및 특정 파라미터의 유사성에 의한 비균일 블록의 오인식을 보완하고자 비균일 블록 형태의 학습 패턴에 대해 자음/모음1(유성자음 포함)의 형태로 구분된 데이터를 입력으로 하는 첫 번째 코드북 1을 생성후 자음/모음1(유성자음 포함)의 2가지 HMM 학습 모델을 먼저 생성한다. 이어서 학습을 위한 비균일 블록 데이터중 유성 자음을 포함하는 모음1 형태의 음기들에 대하여 유성자음을 포함하지 않는 모음2 음가와 유성자음 음가로 이루어지는 두 번째 코드북 2를 생성 후, 모음2/유성자음 형태의 HMM 모델을 다시 생성하여 전체적으로 2개의 코드북을 이용한 자음, 모음1, 모음2, 유성자음의 4개의 HMM 음가 모델(자음/모음/유성자음)을 생성하여 인식에 적용하는 일고리즘을 제안, 적용하였다.

추출된 비균일 블록 형태의 미지 입력 패턴은 1차적으로 먼저 코드북 1을 이용하여 양자화된 후, 생성된 자음/모음1(유성자음 포함) HMM 모델을 이용하여 자/모음1(유성자음 포함) 음가로 판별된다 모음1으로 판별된 미지 패턴은 다시 코드북 2를 이용하여 다시 양자화 된 후, 모음2/유성자음의 2가지 HMM 모델에 의해 해당되는 음가를 구분하는 다단 구조 형태의 학습 및 인식을 수행한다. 인식 결과 자음/모음/유성자음 형태를 인식하기 위한 단일 코드북을 구성한후 3개의 VQ-HMM 모델을 생성하여 인식에 적용한 경우에 비하여 본 논문에서 사용된 2개의 코드북, 4개의 HMM 모델을 생성하여 인식에 적용하였을 경우 각 단계에서의 누적된 인식 에러를 포함하여도 단일 형태의 코드북 이용에 비해 전체 인식률이 있어서 5~8% 이상의 향상된 결과를 나타내었다.

본 논문에서 자음/모음/유성자음 음가의 분할, 인식을 위해 제안하여 적용된 학습 패턴 및 미지 패턴에 대

한 다단 구조의 HMM 인식 구조를 (그림 9)에 나타내었다. 이때 미지 패턴에 대한 인식 결정 확률로서 Forward 및 Viterbi 알고리즘이 사용되었으며, Forward 결정 확률이 Viterbi 알고리즘에 비해 많은 계산 시간을 더 소모하였으니, 인식률에 있어서는 좀 더 나은 결과를 나타내었다.

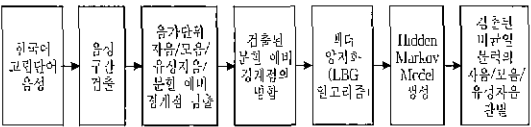


(그림 9) 제안된 자음/모음/유성자음 VQ-HMM 인식 알고리즘

4. 결과 및 고찰

본 논문에서는 한국어 특유의 음운 변동 현상에 의해 음가를 발생시키는 한국어 음운 변동 Rule-Based 시스템을 텍스트 입력력 형태로 구성한 후, 시스템의 텍스트 형태의 음가 결과에 기초하여 실제 한국어 음성에서 음가 신호를 인식하기 위한 자음/모음/유성자음 음가 분할 예비 경계점을 사전 검출한 다음, 검출된 경계점 사이의 비균일 블록을 자음/모음/유성자음으로 인식하므로써 한국어 음성 특유의 음가 인식 가능성을 평가하였다.

실험에 사용된 한국어 고립 단어 음성은 데이터베이스화된 음성 시료 중 하나인 국어공학센터에서 구성된 남성화자의 TIMIT 모델 PBW 음성으로서 음성 검출 및 자음/모음/유성자음 음가의 분할 및 인식에 대하여 실험하였다 전체적인 실험 과정을 살펴보면 (그림 10)과 같다.



(그림 10) 음가 분할 및 인식을 위한 시스템 흐름도

실험은 Visual C++ 컴파일러를 이용한 Pentium III 550 PC 환경하에서 수행되었으며, 음성 검출, 비균일 블록 검출 및 검출 블록에 대한 음가 분할 및 인식을

테스트하기 위해 사용된 한국어 고립 단어 음성 데이터의 사양, 전체 화자의 수 및 화자 당 고립 단어 음성 데이터의 수를 <표 1>, <표 2>에 나타내었다.

<표 1> 음성 데이터의 조건

대상 어휘	한국어 고립 단어 음성
헤더정보	1024 BYTE
샘플링 주파수	16KHz
양자화 레벨	16bit

<표 2> 음성 검출, 음가 분할 및 인식을 위해 사용된 데이터의 형태 및 개수

구분	음성 검출	자음/모음/유성자음 음가 분할 및 인식
처리 데이터	한국어 고립 단어	한국어 고립 단어 음성에서 추출한 음가 분할 예비 경계점 사이의 비균일 블록
화자의 수	10명	5명
1 화자당 처리 단어의 갯수	50개	1화자당 30개 단어 이상
전체 데이터의 갯수	500개	1040 개

음성 검출의 경우 화자 독립 500개의 고립 단어 음성을 대상으로 실험 결과, 99.8%이상의 검출 결과를 나타내었으며 일반적인 에너지 영교차에 의한 음성 검출시보다 묵음과 음성의 경계부분에서 보다 정확한 음성 검출 결과를 나타내었다. 음가 자동 분할을 위한 비균일 추출 블록의 경우 각기 다른 5명의 화자가 발생한 화자당 30개 이상의 고립 단어를 대상으로 비균일 블록을 추출한 후 실험되었다. <표 3>은 음가 분할 예비 경계점 검출 알고리즘을 통해 한국어 고립 단어 음성에서 추출된 비균일 블록들을 수작업에 의해 분류한 자음/모음/유성자음 형태의 비균일 블록의 종류 및 개수를 나타낸다.

<표 3> 음가 분할 및 인식을 위해 추출된 비균일 블록의 종류 및 개수

자음/모음/유성자음 인식을 위해 추출된 비균일 블록 데이터의 개수(전체 1040)	자음 (500)	ㄱ, ㅋ, ㆁ, ㄷ, ㅌ, ㄷ, ㅌ, ㅍ, ㅍ, ㅊ, ㅊ, ㅈ, ㅉ, ㅊ, ㅊ, ㅈ, ㅉ
	모음 (270)	음성자음 (270)
	모음 (270)	ㅣ, ㅏ, ㅑ, ㅓ, ㅕ, ㅗ, ㅛ, ㅜ, ㅠ, ㅡ, ㅝ, ㅟ, ㅛ, ㅜ, ㅠ, ㅡ, ㅝ, ㅟ, ㅛ, ㅜ, ㅠ, ㅡ, ㅝ, ㅟ, ㅛ, ㅜ, ㅠ, ㅡ, ㅝ, ㅟ

대 용량의 음성 인식을 위해서는 음성을 구성하는 최소 단위에 의해 적용되어지는 형태를 사용하여야 한다 따라서 본 연구에서는 한국어 고립 단어 음성에서

기존의 음소 형태를 기준으로 한 분류 및 인식이 아닌 음가를 기본 인식 단위로 사용하기 위하여 한국어 음운 변동 Rule-Based 시스템의 텍스트 음가 결과를 기초로 한 음가 분할 및 인식 가능성에 대하여 실험하였다. 음가 분할 및 인식 대상은 자음/모음/유성자음의 형태로 규정되었다. 그러나 기존에 보편적으로 사용되어지고 있는 DHMM 인식 기법인 단일 코드북에 의한 인텍스를 이용하여 자음/모음/유성자음 HMM 모델 생성에 의해 구분할 경우 미지 데이터에 대하여 81~85%의 낮은 인식률을 나타낸다. 또한 검출 구간이 너무나 짧은 형태이므로 중첩된 멀티섹션을 이용할 경우의 실험에서도 그다지 큰 인식률의 향상 결과를 가지오지 못하였다. 따라서 다단 구조의 2개의 코드북, 4개의 HMM 모델을 이용하는 (그림 9)와 같이 인식 구조를 제안하여 사용하였다. 먼저 자음의 난류적인 특성에 의한 자음 분류를 용이하게 하기 위하여 자음/모음1(유성자음 포함)의 1개의 코드북에 의해 2개의 HMM 모델을 생성하여 자음/모음1(유성자음)을 분리해낸 결과 95~97%대의 자음/모음1 인식률을 나타내었다. 다시 인식된 모음1 영역에서 유성자음을 제외한 모음과 유성자음을 인식하여 분리해내기 위해, 모음2/유성자음 형태의 1개 코드북을 이용한 HMM 모델을 생성하여 모음2/유성자음 영역을 분리해낸 인식 결과, 약 85~88%대의 모음1/유성자음 인식률을 나타내었다.

결과적으로 (그림 9)와 같이 제안된 2개의 코드북을 이용한 4가지 자음, 모음1, 모음2, 유성자음 HMM 모델을 생성하여 자음/모음/유성자음 인식에 사용할 경우, <표 4>와 같이 화자 독립 학습 데이터 비균일 블록 데이터에 대해서는 자음/모음/유성자음 HMM 인식시 전체 98.8%의 인식률을 나타내었으며, 또한 미지의 화자 독립 비균일 블록 데이터에 대해서는 전체 89.6%의 자음/모음/유성자음 인식률을 나타내어 기존의 단일 코드북을 이용한 미지 데이터의 인식 결과에 비해 단계별 누적 에러를 포함하여도 단일 코드북을 이용한 인식 결과에 비해 향상된 결과를 나타내었다.

전체적으로 비균일 블록 추출시, 추출된 비균일 블록의 길이가 너무 짧은 경우에 특정 파라미터가 해당 음가를 정확히 표현하지 못하게 되어 인식률의 저하를 가져오는 경향이 발생하였으며, 미지 패턴 특정 파라미터의 벡터양자화시 양자화 에러로 인해 인식률의 저하에 많은 영향을 미치는 것으로 판단된다.

특히 유성자음 및 모음 형태로 분류된 비균일 블록

음가들의 인식률 저하는 유성자음의 주기적인 특성으로 인해 모음과의 상호간의 구분 특징이 모호한 관계인 것으로 보이며, 보다 정확하게 모음 및 유성자음을 대표할 수 있는 특징 파라미터에 대한 연구 및 추출 후, 인식에 적용할 경우 인식률의 개선이 이루어질 수 있을 것이라 사료된다.

<표 4> 추출된 비균일 블록의 자음/모음/유성자음 음가 분할 및 인식 결과

학습/미지 데이터	비균일 블록 데이터	전체 데이터 갯수	인식 데이터 갯수	오인식	인식률(%)
화자 독립 학습 데이터	자음/모음(유성자음포함)	400/400	388/398	2/2	99.5/99.5
	모음/유성자음	198/200	196/197	2/3	98.9/98.5
각각의 유성자음	자음/모음/학습 데이터의 합계 및 인식률	400/200/200	388/196/197	3/4/3	97.0/98.0/98.5
전체 학습 데이터의 합계 및 인식률		800	791	9	98.8
화자 독립 미지 데이터	자음/모음(유성자음포함)	100/140	96/136	4/3	96.0/97.9
	모음/유성자음	69/67	60/65	10/8	85.5/88.1
각각의 유성자음	자음/모음/미지 데이터의 합계 및 인식률	100/70/70	96/57/62	4/13/3	96.0/81.4/88.6
전체 미지 데이터의 합계 및 인식률		240	215	25	89.6

따라서 추출된 비균일 블록이 보다 정확하게 해당 자음/모음/유성자음 음가를 나타낼 수 있도록 포만트를 이용한 다양한 특정 파라미터를 개발하여 추출하고, 이들을 혼합한 연속적인 관찰 벡터를 입력으로 하는 CHMM 인식 모듈을 이용하거나, 음성의 발생시 음절의 경계 위치를 확보하고 음절별 초성/중성/종성 구분 후에 자음/모음/유성자음 비균일 블록 형태를 학습 및 인식에 사용할 경우 더욱 높은 인식률을 나타내리라 사료된다.

5. 결 론

인간의 음성을 기계와의 정보 교환에 효율적으로 이용하기 위해 각 국에서 연구되거나 상용화되어지는 프로젝트 및 음성 관련 제품들은 자국의 언어 특성에 맞는 음성 특징들을 사용하여 수행되어지고 있으며, 각 나라마다 그 특질이 다르기 때문에 특정 시스템이 개발되었다고 해서 그대로 적용하여 사용할 수 없다.

한국어의 경우 한국어의 제 특징으로 인해 한국어 사용자 스스로 처리할 수 밖에 없으며, 이를 위해서는 음성 언어인 한국어와 표기 언어인 한글간에 발생하는 음운 변동 현상의 과학적 규명과 더불어 그 규칙을 합리적으로 설계한 후, 한국어 음성을 처리하는 형태의 인식 시스템이 구성되어야 한다. 본 논문에서는 한국어의 특유의 특성을 바탕으로 한 음운 변동 현상을 반영하여 텍스트 형태로서 단어 및 해당 음가를 출력하는 한국어 음운 변동 Rule-Based 시스템의 결과에 기초하여 실제 한국어 특유의 음운 변동 현상으로 인한 발생 형태의 음가 분할 및 인식의 가능성에 대하여 실험하였다.

자음/모음/유성자음 음가로의 분할 및 인식 결과, 인식률에 있어서 한국어 음가 형태를 기반으로 한 신호의 분할 및 인식 가능성을 입증하였으며 향후 개별 음가의 확장된 인식 및 인식 에러에 대한 후처리로서 한국어 음운 변동 Rule-Based 시스템을 적용할 경우 실질적인 한국어 음성 인식 시스템으로의 확장이 가능할 것으로 사료된다. 그러나, 음가 단위의 보다 객관적으로 정확한 분할 및 인식률을 산정하기 위해서는 대규모의 공통 음가 DB의 구성이 절실히 필요하며, 개개 음가의 인식률 확장을 위하여 보다 다양한 특징 파라미터와 인식 기법에 대한 개발 및 연구가 이루어져야 할 것이다.

참고 문헌

[1] L. Rabiner and B. H. Juang, "Fundamentals of Speech Recognition," Prentice Hall, 1993.
 [2] 오영환, "음성언어정보처리", 홍릉출판사, 1998
 [3] 아계영, "한국어 음운 변동 처리를 위한 효율적인 Rule Base System의 구성", 전자공학회 논문지, 제28권, 제12호, pp.9-18, 1991년 12월.
 [4] C. J. Weinstein, "A system for acoustic phonetic analysis of continuous speech," IEEE ASSP-23, pp. 54-67, Feb, 1975.
 [5] Olivier Rioul, Martin Vetterli, "Wavelets and Signal Processing," IEEE SP magazine, pp 14-38, Oct. 1991.
 [6] 김바울, 윤철현, 홍광석, 박병철, "칭각 모델에 기초한 음성 특징 추출에 관한 연구", 대한전자공학회논문지, 제33권, 제4호, 1996년 4월.
 [7] T. Tan, R. Lang, Heiko Schroder, Andrew Spray "Applying wavelet analysis to speech segmentation

and classification," SPIE Wavelet Application, Vol. 2242, 1994.

[8] R. M. Gray, "Vector Quantization," IEEE ASSP Magazine, April 1984.
 [9] L. R. Rabiner and B. H. Juang, "An introduction to Hidden Markov Models," IEEE ASSP Magazine, pp. 4-16, Jan 1986.
 [10] L. R. Rabiner, "Tutorial on Hidden Markov Model and Selected Application in Speech Recognition," Proc. IEEE, Vol 77, No.2, pp.257-285, 1989.
 [11] X. D. Huang, Y. Ariki, M. A. Jack, "Hidden Markov Models for Speech Recognition," Edinburgh Univ., 1990



이준환

e-mail : ljh@kdu.ac.kr

1994년 단국대학교 전자공학과 (공학사)

1996년 단국대학교 대학원 전자공학과(공학석사)

1999년 단국대학교 대학원 전자공학과 박사과정 수료

2000년~현재 단국대학교 멀티미디어 산업기술연구소 연구원

1999년~현재 극동대학교 전산정보학부 멀티미디어학과 겸임교수

관심분야 : 디지털 신호처리, 음성 및 영상처리



이상범

e-mail : sbhhee@dankook.ac.kr

1974년 연세대학교 전자공학과 (공학사)

1978년 서울대학교 대학원 전자공학과(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 멀티미디어산업기술 연구소장

2000년~현재 단국대학교 공학부 컴퓨터공학전공 교수

관심분야 : 컴퓨터구조, 패턴인식, 디지털 신호처리