

# 한국어 음성인식에서 음성의 특성을 고려한 음소 경계 검출

## Phoneme Segmentation in Consideration of Speech feature in Korean Speech Recognition

서영완\* 송점동\*\* 이정현\*\*\*  
Young-Wan Seo Jeom-Dong Song Jung-Hyun Lee

### 요 약

음소 단위로 구축된 음성 데이터는 음성인식과 음성합성 및 분석 등의 분야에서 매우 중요한 문제이다. 일반적으로 음소는 유성음과 무성음으로 구분된다. 이러한 유성음과 무성음은 많은 특징적 차이가 있지만, 기존의 음소 경계 검출 알고리즘은 이를 고려하지 않고 시간 축을 기준으로 이전 프레임과의 스펙트럼 비교만을 통하여 음소의 경계를 결정한다.

본 논문에서는 음소 경계 검출을 위하여 유성음과 무성음의 특징적 차이를 고려한 블록기반의 분류 알고리즘을 설계하였다. 분류 알고리즘을 사용하기 위한 스펙트럼 비교 방법은 MFCC(Mel-Frequency Cepstrum Coefficient)를 기반으로 한 거리 측정법을 사용하였고, 유성음과 무성음의 구분은 에너지, 영 교차율, 스펙트럼 비, 포먼트 주파수를 이용하였다. 본 논문의 실험결과 3~4음절 고립단어를 대상으로 약 79%의 정확도를 얻음으로써 기존의 음소 경계 검출 시스템보다 약 8%의 정확도 향상을 보였다.

### Abstract

Speech database built of phonemes is significant in the studies of speech recognition, speech synthesis and analysis. Phonemes consist of voiced sounds and unvoiced ones. Though there are many feature differences in voiced and unvoiced sounds, the traditional algorithms for detecting the boundary between phonemes do not reflect on them and determine the boundary between phonemes by comparing parameters of current frame with those of previous frame in time domain.

In this paper, we propose the assort algorithm, which is based on a block and reflecting upon the feature differences between voiced and unvoiced sounds for phoneme segmentation. The assort algorithm uses the distance measure based upon MFCC(Mel-Frequency Cepstrum Coefficient) as a comparing spectrum measure, and uses the energy, zero crossing rate, spectral energy ratio, the formant frequency to separate voiced sounds from unvoiced sounds. As the result of our experiment, the proposed system showed about 79 percents precision subject to the 3 or 4 syllables isolated words, and improved about 8 percents in the precision over the existing phonemes segmentation system.

## 1. 서 론

음성인식 및 음성분석 등의 음성신호처리 분야에서 정확한 음소분할은 중요한 문제 중의 하나

이다. 연속 음성 인식기에서 사용되는 분석의 기본 단위는 단어, 어절, 음소 등이 있으나, 음소 단위는 단어 및 음절 단위보다 그 종류가 작고 음향적인 특성을 인식기에 고르게 반영할 수 있기 때문에 많이 사용된다. 특히 40여 개의 음소를 갖는 한글에서는 이들 음소들을 인식단위로 사용함으로써 효율적인 음성 인식 시스템을 만들 수 있다. 그러나 음성을 음소 단위로 정확하게 분할한다는 것은 수 작업으로 하는 경우에도 쉽지 않은 작업으로 수 작업에 의한 음소 분할은 많은 시간

\* 정회원 : 인하대학교 전자계산공학과  
syw@nlsun.inha.ac.kr

\*\* 정회원 : 경문대학 컴퓨터정보과 조교수  
jdsong@kmc.ac.kr

\*\*\* 비회원 : 인하대학교 전자계산공학과  
jhlee@dragon.inha.ac.kr

이 소요되며, 일관성이 보장되지 않는 문제점을 안고 있다[1]. 따라서 음성을 정확히 음소 단위로 분할할 수 있는 자동 음소 분할기의 구현은 음성 인식 시스템의 인식률을 높일 수 있는 한 방안으로 연구되고 있다[2]. 이러한 음소는 일반적으로 유성음과 무성음으로 구분되어진다. 유성음과 무성음은 많은 특징적 차이가 있지만, 기존의 음소 경계 검출 알고리즘은 이를 고려하지 않고 매개변수 비교만을 통하여 음소의 경계를 검출한다.

본 논문에서는 한국어의 특성을 고려하여 유성음구간과 무성음구간에 차등 임계치를 적용한 음소 경계 검출기를 설계하였다. 음소의 특성상 유성음과 무성음, 무성음과 유성음구간은 많은 특징 차이가 보이므로 쉽게 구분을 할 수 있으나, 유성음과 유성음구간은 특징 차이가 적어 음소 구분을 하기 어렵기 때문이다. 본 논문에서 사용한 스펙트럼 비교 방법은 수정된 MFCC 거리 측정법을 사용하였고, 분석구간 결정을 위해 분류 알고리즘을 제안하였으며, 유성음과 무성음의 구분을 위해 영 교차율, 에너지, 스펙트럼 비, 포먼트 주파수를 사용하였다.

## 2. 기존의 음소 경계 검출

음소의 경계 검출 방법으로 LPC를 이용한 정규화된 오차법(Normalized Error)과 가장 보편적으로 알려진 Itakura-Saito 거리측정 방법 등이 있다. 그러나, 이러한 음소 경계 검출은 일부 음소들에 대해서는 음소의 경계를 잘 찾지 못한다는 단점과 무성음과 유성음의 음소 사이에서는 음소의 경계가 뚜렷하게 나타나지만, 유성음과 유성음 혹은 무성음과 무성음들 사이에서는 상호간의 조음 현상 때문에 정확한 경계를 찾기가 힘들다.

### 2.1 정규화된 오차법(Normalized Error)

입력된 음성신호를 선형 예측 계수를 통하여

구하게 된 오차는 식1과 같다.

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{i=1}^n a_i s(n-i) \quad \text{식(1)}$$

따라서, i번째 블록의 정규화된 오차  $V^i$ 는 식2와 같이 표현할 수 있다.

$$V^i = \frac{\sum_{n=0}^{N-1} e(n)^2}{\sum_{n=0}^{N-1} S(n)^2} \quad (\text{단, } n \text{은 블록의 크기}) \quad \text{식(2)}$$

정규화 된 오차  $V^i$ 가 크다는 것은 모형이 입력된 음성신호를 잘 표현하지 못한다는 의미이다. 따라서, 이전 블록의 매개변수로 이후 블록의 실측값을 예측하여  $V^i$ 가 일정 임계치(threshold)를 넘으면, 새로운 음소로 분리한다고 판단한다. 즉, 이전 블록의 매개변수가 이후 블록의 실측값을 예측하여  $V^i$ 가 크다는 것은 이전 블록과 이후 블록의 음성신호의 파형이 전혀 상이함을 의미한다는 것을 이용하여 음소가 분리된다고 판단한다[3].

### 2.2 Itakura-Saito 거리측정 방법

일반적으로 음성의 변화는 스펙트럼의 차이를 이용하여 측정할 수 있다. 그러나 스펙트럼만을 비교해서는 정확한 음성의 변화를 측정할 수 없다. 포먼트의 위치와 폭은 음성의 변화 요인 중 가장 큰 역할을 한다. Itakura-Saito 거리측정 방법이 이러한 요인을 고려한 측정법 중 가장 대표적인 방법이다. 수식은 식3과 같다[4,5].

$$D \left( \begin{matrix} - \\ i_1 \end{matrix}, \begin{matrix} - \\ j_1 \end{matrix} \right) = \frac{\sum_{k=0}^{N_{LP}} R_1^{(k)} R_2^{(k)}}{a_{LP2}^T R_2 a_{LP2}} \quad \text{식(3)}$$

$$R_a^{(k)} = \sum_{i=0}^{N_{LP}-k} a_{LP}^{(i)} a_{LP}^{(i+k)} \quad (0 \leq k \leq N_{LP})$$

( $R_a^{(k)}$ : LPC의 자기 상관 함수)

### 3. 제안한 분류 알고리즘

음성은 시간에 따라 변화하는 특성을 가지고 있으나, 20~30msec정도의 짧은 시간간격 동안은 변하지 않는다고 가정할 수 있다[6]. 따라서 이를 바탕으로 발음기관의 모양과 위치 정보를 포함하고 있는 성도 파라미터의 추출이 중요하다. 일반적으로 이런 시간적 특성을 고려하여 10msec로 중첩된 20msec프레임 간격으로 특징벡터를 추출한다. 본 논문에서는 입력된 음성신호에서 추출된 특징벡터를 분류 알고리즘에 스펙트럼 비교법을 적용하여 입력음성을 음소별로 분할하고 군집화하여 음소의 경계를 검출하였다. 또한 구간별 유성음과 무성음 판단을 위하여 에너지, 영 교차율, 스펙트럼 비, 포먼트 주파수를 이용하였다.

#### 3.1 수정된 거리측정 방법

FFT를 바탕으로 한 스펙트럼 비교법은 성분 서로 다른 파형의 스펙트럼 차이를 잘 나타낸다. 이러한 FFT를 이용한 거리측정 방법 중 본 논문에서는 인간의 청각 특성을 적용한 MFCC 거리를 적용하였다. 인간의 청각 특성은 1KHz정도까지의 저주파 영역의 신호에는 민감하여 선형 스케일을 보이지만, 고주파 영역의 신호에는 민감하지 못하여 로그 스케일의 특성을 보인다. 이런 특성을 고려한 것이 로그값을 취한 멜 스케일(mel scale)이다[7, 8]. 일반 선형 스케일을 멜 스케일로 변환은 식4와 같다.

$$mel\ frequency = 2595 \log_{10} \left( 1 + \frac{frequency}{700.0} \right) \quad \text{식(4)}$$

이러한 멜 스케일을 기초로 한 MFCC의 개념은 1980년도 Davis와 Mermelstein에 의해 도입되었고 현재 많은 인식 시스템이나 음향 코딩에 사용되고 있다. 본 논문에서 제시한 i번째 블록과 j번째 블록의 MFCC거리  $d_c^2(L)$ 는 식5와 같다.

$$d_c^2(L) = \sum_{n=1}^L \left( \sum_{k=i}^L \hat{c}_{kn} - \sum_{k=j}^L \hat{c}'_{kn} \right)^2 \quad \text{식(5)}$$

MFCC 거리측정방법은 FFT를 이용하여 측정된 방법이므로, 서로 다른 특성의 주파수를 동일하게 인식하는 오류율이 매우 적다. 그림 1은 멜 켈스트럼을 구하는 과정을 나타낸다[2, 11].



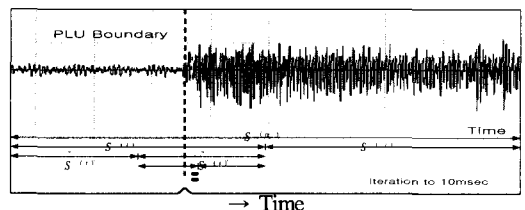
(그림 1) 음성 신호로부터 멜켈스트럼을 구하는 과정

#### 3.2 분류 알고리즘

일반적으로 스펙트럼의 차이를 이용하여 음성의 변화를 측정할 수 있다[11]. 그림 2는 본 논문에서 설계한 Top-Down 방식의 블록기반 분류 알고리즘에 대한 방법이다.

그림 2에서 보면, 우선 입력된 신호의 음성구간을 검출한 후 분석구간  $S(a)$ 을 결정한다. 결정된 구간을  $S(i)$ 와  $S(j)$ 로 분할한 후 매개변수를 비교하여 일정 임계치 이하이면 같은 음소가 연속적으로 발생되었다고 판단하고, 더 이상 분할을 하지 않고 다음 구간으로 이동을 한다. 이때, 만약 비교치가 임계치 이상이면 서로 다른 음소가 혼합되었다고 가정하고, 분할된 구간을 임계치 이하일 때까지 계속 분할을 수행한다. 음성의 정적 구간인 10msec까지 분할이 계속 되면 그 중간값을 취하고 분할을 멈춘다.

알고리즘 1은 본 논문에서 설계한 분류 알고리즘



(그림 2) 분류 알고리즘 방법

즘의 수행 단계를 나타낸다.

### 3.3 유성음과 무성음 분류 방법

음성은 크게 유성음과 무성음으로 분류할 수 있다.

#### (알고리즘 1) 분류 알고리즘

```

전체입력음성 : S[N],  FSIZE : 160
count = 0;
Range_Index[count]; //분할된Frame의 Index를 저장
Input();           //음성 입력
Assort_Split(0, N-1);
void Input()
{ ..
// Sampling : 16000 / 16 bit / mono
dat = _llseek(fin, 0L, FILE_END);
DataSize = dat; // 음성 전체 크기
int ndat = dat / 2; // 전체 음성데이터 개수
Fnum = ndat / 320; // 전체 프레임 개수
}
float *MFCC(int x, int y)
{ ..
Preemphasis(); //프리엠퍼시스 함수
Frm_Block(); //프레임단위로 나누기 위한 함수
Hamming(); //해밍 윈도우 함수
mfc = calc_mfcc(); //MFCC를 구하기 위한 함수
return *mfc; //구해진 MFCC의 포인터를 반환
}
void Assort_Split()
{ ..
fsize = Index2 - Index1;
if (fsize > FSIZE) {
Alpha = (Index2 - Index1)/2;
Beta = Index2;
mfc_1 = MFCC(Index1, Alpha);
mfc_2 = MFCC(Alpha+1, Beta);
mfcc_diff = mfcc_dist(mfc1, mfc2);
if(mfcc_diff > mfcc_dist_Threshold) {
Assort_Split(Index1, Alpha);
Assort_Split(Alpha+1, Index2);
}
else {
Range_Index[++count] = Index2;
}
}
else {
Range_Index[++count] = Index2
}
}
}

```

유성음의 경우는 일반적으로 주기 신호에 가깝고 에너지가 크며 저주파 영역에 에너지 성분이 집중된다. 무성음의 경우는 유성음의 경우와 반대로 잡음 신호에 가깝고 에너지가 상대적으로 작으며 고주파 영역에 에너지 성분이 집중된다[12]. 이러한 성질을 이용하여 유성음과 무성음을 분류하는 일반적인 방법으로 영 교차율과 에너지, 레벨 교차율, 정규화된 자기상관계수, 스펙트럼 에너지비 등이 있다. 본 연구에서는 영 교차율과 에너지 그리고 스펙트럼 에너지 비를 사용하여 유성음과 무성음 구간을 판단하고, 유성음에서 에너지가 집중적으로 나타나는 포먼트 주파수의 추출로 유성음과 무성음을 재차 판단한다.

#### ① 영 교차율(Zero Crossing Rate)

영 교차율이란 음성 신호 파형의 위상이 중심 축을 통과하는 횟수이므로 무성음의 경우 진폭이 크지 않고 불규칙적인 진동이 계속 있어 일정한 만큼만 중심 축을 통과하는 유성음보다 영 교차율이 크다.

#### ② 에너지(Energy)

영 교차율과는 반대로 유성음은 에너지가 무성음에 비하여 월등히 크기 때문에 임계치 이상 에너지가 크면 유성음이라고 간주할 수 있다.

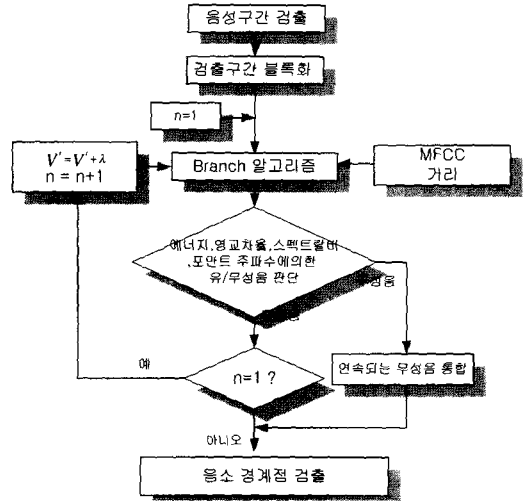
#### ③ 스펙트럼 에너지 비(Spectral energy ratio)

스펙트럼 에너지 비는 유성음과 무성음을 가장 확실하게 나타내는 특징을 가지고 있다. 확실한 유성음의 경우는 스펙트럼 에너지 비가 음수의 값을 나타내고, 확실한 무성음의 경우는 스펙트럼 에너지 비가 양수의 값을 나타낸다. 스펙트럼 에너지 비에는 저주파에서부터 고주파에 이르기까지 여러 밴드를 사용한다. 여러 가지의 밴드를 사용한 이유는 사람마다 같은 단어를 발성하여도 스펙트럼의 특성이 다르기 때문이다. 같은 사람이 같은 단어를 발성한다고 해도 상황에 따라 차이가 있을 수 있다는 것이다. 따라서 여러 밴드의 스펙트럼 에너지 비를 사용할 필요가 있다.

#### 4. 유성음과 무성음의 특성을 고려한 음소 경계 검출

음소 경계 검출은 입력음성을 음소 단위로 분류하기 위해 음소 경계 구간을 정확히 검출하기 위한 것이다. 기존의 음소 경계 검출기는 유/무성음을 구분하지 않고 스펙트럼 비교만을 통하여 음소의 경계를 찾으려고 한다. 그러나 이러한 음소 경계 검출은 그림 3에서의 예와 같이 무성음과 유성음의 음소 사이에서는 뚜렷하게 나타나는 경우가 있지만, 그림 4에서의 예와 같이 유성음과 유성음 혹은 무성음과 무성음들 사이에서 상호간의 조음 현상 때문에 정확한 음소 경계를 찾기가 어렵다.

본 논문에서는 앞 절에서 제시한 분류 알고리즘을 적용하여 1차 음소 경계 후보를 추출한 후, 포먼트 주파수를 이용하여 구간별로 유성음과 무성음을 판단한다. 이 때, 스펙트럼 비교 거리가 낮은 유성음 구간은 가변 임계치를 사용하여 2차 음소 경계 후보를 찾는다. 한국어의 특성상 연속되는 무성음은 없기 때문에 연속된 무성음은 하나의 음소로 통합한다. 그리고, 구간별 MFCC를 구하여 인접 구간의 거리를 측정하여, 측정된 결



(그림 5) 음소 경계 검출 흐름도

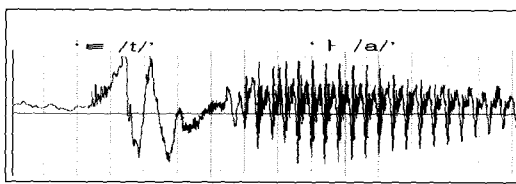
과가 임계치 이하이면 음소 경계 후보에서 제외시켜 최종 음소 경계를 검출한다. 그림 5는 전체적인 음소 경계 검출의 흐름도이다.

#### 5. 실험 및 평가

본 논문에서는 유성음과 무성음의 특성을 고려한 분류 알고리즘을 적용하여 음소의 경계를 검출하는 시스템을 구현하였다. 본 장에서는 유성음과 무성음을 고려하지 않고 음소의 경계를 검출하였을 때와, 본 논문에서 제시한 유성음과 무성음을 고려한 분류 알고리즘을 적용하였을 때를 비교한다. 그리고 실험 데이터를 입력하였을 때, 기존의 알고리즘과 본 논문에서 제시한 분류 알고리즘을 적용하였을 때의 결과치를 비교한다.

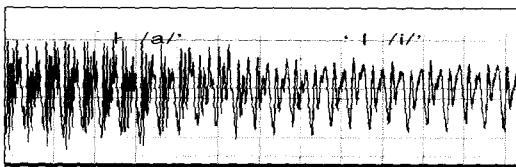
##### 5.1 실험환경

본 논문에서 제안한 음소 경계 검출의 성능을 평가하기 위해서 남녀 대학생 3명이 3개씩 발생한 고립단어를 기반으로 실험을 수행했다. 이때의 실험 환경으로 표본화 주파수는 16KHz, 그리고 A/D 양자화 해상도는 16bit이고 창 함수는 해밍



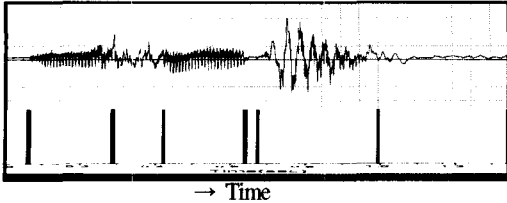
→ Time

(그림 3) 무성음과 유성음의 조음 (예: /타/)

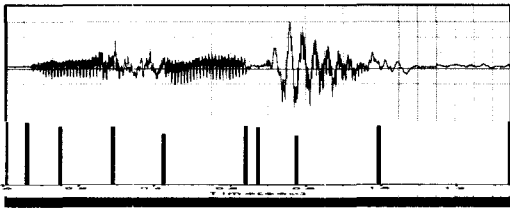


→ Time

(그림 4) 유성음과 유성음의 조음 (예: /아이/)



(그림 6) 유성음과 무성음의 특성을 고려하지 않은 음소 경계 검출(예: /인하대/)



(그림 7) 유성음과 무성음의 특성을 고려한 분류 알고리즘을 적용한 음소 경계 검출(예: /인하대/)

창을 사용하였다. 프레임 크기는 20ms, 창 이동 크기는 10ms간격으로 특징을 추출하였다.

그림 6에서 보면 유성음과 유성음 구간에서 음소의 경계가 검출되지 않았음을 볼 수 있다. 유성음이 연속되는 경우인 /i/와 /m/의 경우와 /d/와 /e/에서 음소의 구분이 되지 않았음을 알 수 있다.

음소의 특성상 연속되는 무성음은 존재하지 않고, 나머지 구간인 무성음과 유성음, 유성음과 무성음 구간을 살펴보면, 목음을 /sil/로 표현하였을 때, /sil/과 /i/, /m/와 /h/, /h/와 /a/, /a/와 /sil/, /sil/과 /d/, /e/와 /sil/부분은 음소의 경계가 잘 나타남을 볼 수 있다.

그림 7과 표 1은 본 논문에서 제안한 유성음과 무성음을 고려한 분류 알고리즘을 적용한 음소 경계 검출기의 예를 나타내었다. 입력음성은 “인하대”이고, 거리측정은 이전 구간과 이후 구간의

(표 1) 음소 경계 검출 결과 (예: /인하대/)

시간 순서	Time (hms)	distance (%)	시간 순서	Time (hms)	distance (%)
1	0.065	98.6	5	0.642	92.6
2	0.150	92.5	6	0.674	91.3
3	0.288	91.9	7	0.777	78.2
4	0.423	80.6	8	0.995	93.1

스펙트럼 차이를 나타낸 것이다.

## 5.2 실험결과

실험 결과 3~4음절 고립단어에 대한 음소 경계 검출의 결과는 표 1, 표 3과 같다. 제안한 음소 경계 검출 시스템은 기존의 시스템에 비해 약 8.3%의 정확도가 향상됨을 보였다. 음소 경계 검출 시 여성이 발성한 음성이, 남성이 발성한 음성보다 정확도가 높음을 보였는데, 그 이유는 남성이 발성한 음성보다 여성이 발성한 음성에 대하여 평균적으로 높은 주파수를 가지고 있기 때문이다. 이는 여성이 발성한 음성이 유성음과 무성음의 구분이 남성이 발성한 음성보다 잘 됨을 알 수 있다. 표 2와 표 3의 성능평가로는 전체 프레임 당 오류 프레임의 비율로써 정확도를 평가하였고, 그림 8은 음소 경계 검출의 성능 비교를 나타냈다.

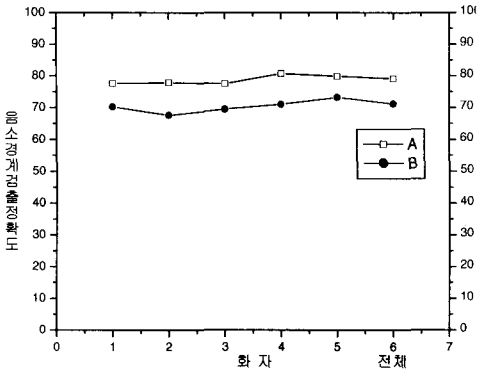
(표 2) 기존 음소 경계 검출 성능

화자	MPCC 거리측정 방법[11]											
	데이터1			데이터 2			데이터 3			평균		
	AF	EF	AC	AF	EF	AC	AF	EF	AC	AF	EF	AC
1	189	59	68.78	141	40	71.63	162	48	70.37	492	147	70.26
2	186	62	66.67	142	46	67.61	160	50	68.75	488	158	67.67
3	187	56	70.05	140	45	67.86	163	48	70.55	490	149	69.49
4	189	53	71.96	139	40	71.22	159	48	69.81	487	141	70.99
5	188	47	75.00	137	36	73.72	161	47	70.81	486	130	73.18
6	188	55	70.74	140	38	72.86	160	48	70.00	488	141	71.20
남자	562	177	68.50	423	131	69.03	485	146	69.89	1470	454	69.14
여자	565	155	72.57	416	114	72.60	480	143	70.21	1461	412	71.79

(표 3) 제안한 음소 경계 검출 성능

화자	분류 알고리즘에 의한 음소 경계 검출기											
	데이터 1			데이터 2			데이터 3			평균		
	AF	EF	AC	AF	EF	AC	AF	EF	AC	AF	EF	AC
1	189	49	74.07	141	28	80.14	162	33	79.63	492	110	77.64
2	186	47	74.73	142	29	79.58	160	32	80.00	488	108	77.87
3	187	43	77.00	140	32	77.14	163	35	78.53	490	110	77.55
4	189	37	80.42	139	27	80.57	159	30	81.13	487	94	80.70
5	188	40	78.72	137	26	81.02	161	32	80.12	486	98	79.83
6	188	42	77.66	140	27	80.71	160	33	79.37	488	102	79.10
남자	562	139	75.27	423	131	78.95	485	146	79.38	1470	328	77.69
여자	565	119	78.94	416	114	80.77	480	143	80.21	1461	294	79.88

(AF : 전체 프레임 EF : 에러 프레임 AC : 정확도)



(그림 8) 음소 경계 검출 성능 비교

## 6. 결 론

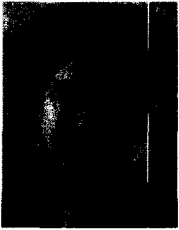
정확한 음소 경계 검출은 음성분석에 있어서 중요한 부분이다. 본 논문에서는 한국어 고립단어를 대상으로 음성신호를 음성의 최소단위인 음소 단위로 분할하는 알고리즘을 제안하였다. 음소 경계 검출 시스템을 스펙트럼 비교법만을 이용하여 구현한다면 적용범위가 넓기 때문에 에러가 발생할 확률이 높다. 따라서 본 논문에서는 자동 음소 경계 검출을 하는 알고리즘으로 유성음과 무성음의 특징 차이를 고려한 블록기반의 분류 알고리즘을 사용하였다. 스펙트럼 비교 방법은 MFCC를 기반으로 한 거리 측정법을 사용하였고, 유성음과 무성음의 구분은 에너지, 영 교차율, 스펙트럼 비, 포먼트 주파수를 이용하였다. 제안한 자동 음소 경계 검출의 최종 성능 평가 결과 3~4음절 고립단어를 대상으로 남자음성은 약 78%의 정확도를 보였고, 여자음성은 약 80%의 정확도를 보였다. 이는 고주파 성분이 강한 여성음성이 음소 경계검출이 잘 이루어짐을 보여준다. 이러한 결과는 기존의 음소 경계 검출 시스템에 비해 약 8.31%의 향상을 볼 수 있었다. 개선점으로, 본 논문에서는 고립단어를 대상으로 MFCC 거리측정을 통한 비교방법을 선택하였다. 따라서 비교방법 및 음성특성을 나타내는 다양한 방법을 복합적으로 적용하여 대 어휘 연속 음성에 대한 연구가 향후 필요하다.

## 참 고 문 헌

- [1] B. Eisen, H. G. Tillman, and C. Draxler, 'Consistency of judgments in manual labeling of phonetic segments: The distinction between clear and unclear cases', Proceeding of the ICSLP (Banff), pp. 871-874, 1992.
- [2] 김상훈, 이항섭, 김회린, '운율 분석용 DB 작성을 위한 자동 레이블러의 성능 평가 및 유용성', Proceeding of the SICOPS96, SESSION 3.6, 1996.
- [3] 김남수, '음성인식을 이용한 현장자료의 실시간 수집에 관한 연구', 석사학위 논문, 서울대학교 대학원 산업공학과, 1995.
- [4] R. A. Brieton, B. M. G Cheetham, M. C. Hall, 'A comparison of distance measures for speech segmentation in variable frame rate speech vocoding', Proceeding of the IEEE Colloquium, pp. 6/1-6/5, 1990.
- [5] F. Itakura and S. Saito, 'A statistical method for estimation of speech spectral density and formant frequencies', Electronics and Communications in Japan, 53A, pp. 36-43, 1970.
- [6] Hwang, Mei-Yuh, 'Subphonetic Acoustic Modeling for Speaker-Independent Continuous Speech Recognition', Tech Report No. CMS-CS-93-230, Computer Science Department, Carnegie Mellon University, Dec, 1993.
- [7] Hesham Tolba, Douglas O'Shaughnessy, 'Automatic Speech Recognition based on Cepstral Coefficients and A Mel-Based Discrete Energy Operator', Proceeding of the ICASSP, vol. 2, pp. 973-976, 1998.
- [8] Rivarol Vergin, Douglas O'Shaughnessy, Vishwa Gupta, 'Compensated mel frequency cepstrum coefficients', Proceeding of the ICASSP, vol. 1, pp. 323-326, 1996.

- [9] J. R. Deller, J. G. Proakis, and J. H. L. Hansen, 'Discrete-Time Processing of Speech Signals', Prentice Hall, 1987.
- [10] Lawrence Rabiner, Bing-Hwang Juang, 'Fundamentals of speech recognition', Prentice Hall.
- [11] L. R. Rabiner, R. W. Schafer, 'Digital processing of speech signals', Prentice Hall, 1978.
- [12] 이현복, '음성학과 언어학', 서울대학교 출판부, 1996.
- [13] J. D. Markel and A. H. Gray, Jr, 'Linear Prediction of Speech', Springer-Verlag, New York, 1976.

## ● 저자 소개 ●



### 서 영 완

2000년 인하대학교 대학원 전자계산공학과(공학석사)  
2000년~현재 삼성전자 재직중  
관심분야 : HCI, Network  
E-mail : syw@nlsun.inha.ac.kr



### 송 집 동

1989년 한국방송대학교 전자계산학과(이학사)  
1992년 인하대학교 산업대학원 정보공학과(공학석사)  
1999년 인하대학교 대학원 전자계산공학과(박사수료)  
1997년~현재 경문대학 컴퓨터정보과 조교수  
관심분야 : HCI, 인터넷 정보검색, 인공지능  
E-mail : jdsong@kmc.ac.kr



### 이 정 현

1977년 인하대학교 전자공학과(공학사)  
1980년 인하대학교 대학원 전자공학과(공학석사)  
1988년 인하대학교 대학원 전자공학과(공학박사)  
1979~1981 한국전자기술연구소 시스템 연구원  
1984~1989 경기대학교 교수  
1989~현재 인하대학교 전자계산공학과 교수  
관심분야 : HCI  
E-mail : jhlee@dragon.inha.ac.kr