

유성/무성/묵음 분류기와 주파수 스펙트럼을 이용한 음소 경계 검출 Phoneme Segmentation Using Voiced/Unvoiced/Silence Classifier and Spectral Information

이 상 래, 한 현 배, 한 민 수
(Sang-Rae Lee, Hyun-Bae Han, Minsoo Hahn)

한국정보통신대학원대학교 공학부

E-mail : leesl@icu.ac.kr

요 약

본 논문에서는 유성/무성/묵음 분류기와 주파수 스펙트럼 비교를 통하여 음소 경계 검출기를 구현하였다. 음소 경계 검출은 음성 인식, 합성 및 분석 등의 분야에서 매우 중요하다. 유성/무성/묵음 분류기를 이용하여 유성음으로 판별되는 구간은 스펙트럼 비교를 통하여 음소 단위로 세분화였고 무성음으로 판별되는 구간은 한국어의 음성 특성을 고려하여 하나의 음소 단위로 간주하였다. 유성음 구간에 대한 스펙트럼 비교는 수정된 Itakura-Saito distance measure 와 Euclidean MFCC(Mel Frequency Cepstrum Coefficients) distance measure 를 사용하였고 비교 프레임은 한 프레임을 건너 뛴 경우가 가장 결과가 좋았다. 최종적으로 평균 음소 길이 정보를 이용하여 음소의 경계로 검출된 구간을 더 세분하거나 통합하였다. 유성/무성/묵음 분류기의 경우는 사무실에서 녹음한 고립단어에 대하여 94.247%의 정확도를 보였고 음소 경계 검출의 경우는 72.8%의 정확도를 보였다.

I. 서 론

일반적으로 음소 분할은 음성인식, 음성 합성, 음성 분석 등의 음성신호 처리 분야에서 중요한 문제 중의 하나이다. 현재 다양한 분야에서 음성 응용 시스템들이 이용되고 있으며 이러한 음성 응용 시스템에서 가장 큰 문제 중의 하나는 입력 음성 신호를 정확하게 음소 단위로 분할하는 것이다. 즉, 음성 인식기의 인식물은 음소의 경계를 얼마나 정확하게 판별하느냐에 크게 의존하게 된다.

일반적으로 고립단어인식 시스템은 음성 검출기를 사용한다. 음성 검출기는 입력 신호에서 음성 신호를 분리해 내는 역할을 하며 표준 패턴 작성 시에는 전문가의 수작업에 의하여 행해지므로 소어휘 화자중속 인식 시스

템의 경우에는 손쉽게 실현 가능하나 대어휘 인식 시스템의 경우는 구현이 거의 불가능하다. 대어휘 인식 시스템의 경우는 표준 패턴의 양이 많아지므로 단순히 음성 검출기만을 이용하여 고립단어인식 시스템을 구현한다면 단어의 수에 따라 표준 패턴의 수는 비례해서 증가하므로 매우 비효율적인 것이다. 따라서 대어휘 인식 시스템의 경우는 음소단위 인식 시스템이 효율적이다.

음소단위 인식 시스템의 장점은 표준 패턴의 개수가 어휘 개수와 상관없이 일정하다는 것이다. 즉, 음소 또는 변이음에 대한 표준 패턴만 저장하고 있으면 어휘의 개수를 증가시켜도 새로운 어휘에 대한 표준 패턴을 추가할 필요가 없다.

그러나 음소 단위 인식시스템의 경우는 표준 패턴 작성이나 인식시 모두 입력 음성을 음소 단위로 분할하는 작업이 필요하다. 음성을 음소 단위로 정확하게 분할한다는 것은 수작업으로 하는 경우에도 쉽지 않은 작업이다. 따라서 음성을 정확히 음소 단위로 분할할 수 있는 자동 음소 분할기를 구현한다면 음성 인식기의 인식률을 높일 수 있을 뿐만 아니라 음성합성 기술의 구현시 기본 합성단위 DB도 손쉽게 구축할 수 있을 것이다.

본 논문에서는 이러한 자동 음소 분할기를 구현하고자 하였다. 이를 위하여 우선 입력 신호를 유성/무성/묵음 판별기를 통하여 유성음, 무성음, 묵음 구간으로 분류하였다. 연속된 유성음, 무성음으로 분류된 음성 신호는 유성음 구간만을 다시 스펙트럼 비교법을 이용하여 세분한 뒤 마지막으로 평균 음소 길이 정보를 이용하여 수정하므로써 최종적으로 음소의 경계를 검출하는 기법을 제안하였다.

II. 유성/무성/묵음 분류기

음성은 크게 유성음, 무성음의 두개의 카테고리로 분류할 수 있다. 유성음은 성대의 떨림을 수반하여 발생하는 신호로 비교적 주기적인 성질을 갖는다. 따라서

유성음의 경우는 에너지가 대부분 1 kHz 이하의 주파수 성분에 집중되고 주기 신호를 이용하여 여기된 상대적 으로 느리게 변하는 선형 시스템을 이용하여 모델링될 수 있다. 반면에 무성음의 경우는 성대의 떨림을 필요로 하지 않는다. 무성음의 경우는 성도에서의 제약에 의한 난류를 음원으로 발생한다. 따라서 유성음과 달리 고주파 성분이 많고 백색 잡음을 여기신호로 모델링될 수 있다. 이와 같이 유성음과 무성음은 발생 원인에 있어서 차이가 있으므로 이러한 차이를 이용하여 분류할 수 있다.

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. 본 논문에서는 다음과 같은 5 가지 특징을 이용하여 유성음과 무성음을 분류하였다.

1) 에너지(Log Energy)

$$E = 10 \times \log_{10} \left[1 + \frac{1}{N} \sum_{n=0}^{N-1} S^2(n) \right]$$

2) 영 교차율(Zero Crossing Rate)

3) 레벨 교차율(Level Crossing Rate)

: 문턱값으로 처음 5 프레임의 평균 에너지 이용

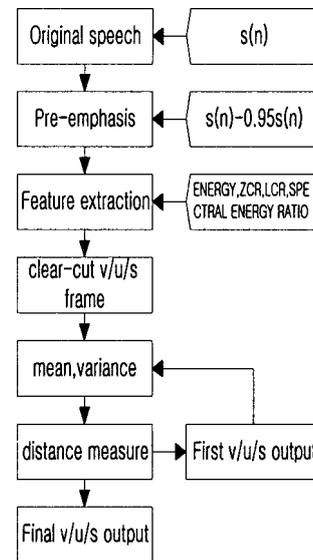
4) 정규화된 자가상관 계수(Normalized autocorrelation coefficient at unit sample delay)

$$\frac{\sum_{n=0}^{N-1} S(n)S(n-1)}{\sum_{n=0}^{N-1} S^2(n)}$$

5) 스펙트럴 에너지 비(Spectral energy ratio)

$$\log_{10} \left(\frac{\text{spectral band energy : 4000-4500Hz}}{\text{spectral band energy : 150-400Hz}} \right)$$

스펙트럴 에너지 비에는 150-400Hz, 800-1200Hz, 4000-4500Hz, 4500-5000Hz, 5000-5500Hz 의 다섯 밴드를 사용하였다. 다섯 가지의 밴드를 사용한 이유는 사람마다 같은 단어를 발성하여도 스펙트럼의 특성이 다르기 때문이다. 같은 사람이 같은 단어를 발성한다고 해도 상황에 따라 차이가 있을 수 있다. 따라서 여러 밴드의 스펙트럴 에너지 비율 사용할 필요가 있다.



[그림 2.1] 유성/무성/묵음 분류기의 흐름도

위의 5 가지 특징을 사용하여 유성음과 무성음을 분류한 알고리즘에 대한 흐름도는 그림 2.1 과 같다. 우선 유성음, 무성음, 묵음 분류에 사용될 특징 변수를 추출한다. 특징 변수의 추출은 pre-emphasis 된 신호 특성을 포함한다. 기본적으로 확실한 유성음과 무성음으로 판별되는 구간에서 각 특징 변수의 평균값과 표준편차를 구한 뒤 이를 바탕으로 나머지 구간의 유성음, 무성음 여부를 판단하였다. 확실한 유성음과 무성음의 구간을 설정하는 것은 스펙트럴 에너지 비율 이용하였다. 확실한 유성음의 경우는 스펙트럴 에너지 비가 음수의 값을 나타낼 것이고 확실한 무성음의 경우는 스펙트럴 에너지 비가 양수의 값을 나타낼 것이다. 그러나 이것은 모든 경우에 적용 가능한 법칙은 아니다. 따라서 위에서 제시한 5 가지 밴드를 모두 사용하여 확실한 유성음과 무성음 구간을 찾았다. 묵음은 처음 다섯 프레임을 묵음으로 간주하고 이 구간에서 평균값과 표준편차를 구하여 이용하였다. 이와 같이 1 차로 확실한 유성음, 무성음, 묵음 구간의 특징 변수에 대한 평균과 표준편차를 이용하여 다른 구간의 판별을 한 뒤 그 결과를 이용하여 유성음, 무성음, 묵음의 평균과 표준편차를 다시 구하였다. 2 차로 구한 평균과 표준 편차를 다시 같은 방식으로 사용하여 모든 구간에 대한 유성음, 무성음, 묵음 판별을 다시 수행하므로써 1 차만 수행한 결과보다 유성/무성/묵음 분류기의 성능을 높일 수 있었다.

III. 음소 경계 검출기

유성/무성/목음 판별기를 통해 유성음과 무성음으로 분류한 뒤 스펙트럼 비교법을 이용하여 음소 단위로 세분하게 된다. 그러나 일단 무성음으로 분류된 구간은 연속된 무성음이 발생하는 경우가 한국어에는 없으므로 하나의 음소로 간주할 수 있다. 그러나 유성음 구간에 대해서는 연속된 유성음에 대해 음소 단위로 더 세분하여야 한다. 연속된 유성음 구간을 세분하는 방법은 스펙트럼 비교를 이용하였다.

1. Itakura-Saito distance measure

일반적으로 스펙트럼의 차이를 이용하여 음성의 변화를 측정할 수 있다. 음성의 변화가 있다면 분명히 스펙트럼의 변화가 있을 것이다. 그러나 단순히 스펙트럼을 비교해서는 음성의 변화를 제대로 측정할 수 없다. 음성의 변화 요인 중 가장 큰 역할을 하는 것은 포먼트의 위치와 폭이다. 따라서 이러한 요인을 고려한 측정법이 필요한데 Itakura-Saito distance measure 가 대표적인 방법이다. 수식은 다음과 같이 표현된다.

$$D(\tilde{y}_1, \tilde{y}_2) = \frac{\tilde{a}_{L_1}^T R_1 \tilde{a}_{L_1}}{\tilde{a}_{L_2}^T R_2 \tilde{a}_{L_2}}$$

($R_1 : x_1$ 를 위한 LP 파라미터를 구하기 위해 사용된 자기상관 매트릭스)

분자항을 계산하기 위해 효율적인 형태로 바꾸면 다음과 같다.

$$D(\tilde{y}_1, \tilde{y}_2) = \frac{\sum_{k=0}^{N_{L_1}-1} R_1(k) R_2(k)}{\tilde{a}_{L_2}^T R_2 \tilde{a}_{L_2}}$$

$$R_1(k) = \sum_{i=0}^{N_{L_1}-k} a_{L_1}(i) a_{L_1}(i+k) \quad (0 \leq k \leq N_{L_1})$$

($R_1(k)$: LPC의 자기상관함수)

Itakura-Saito distance measure를 사용하면 유성음-무음, 유성음-무성음의 비교에서는 비교하고자 하는 프레임 간의 비교의 값이 크므로 쉽게 음소 변화의 위치를 얻을 수 있다. 그러나 유성음 구간에서의 비교는 상대적으로 값이 작고 로컬 피크(local peak)가 많기 때문에 이것에 대한 고려가 필요하다. 이러한 문제는 비교 프레임 간의 에너지 차이를 곱해줌으로써 많은 로컬 피크를 줄일 수 있었다. 그리고 에너지 차이가 큰 부분을 강조함으로써 음소의 경계를 강조할 수 있었다. 수정된

Itakura-Saito distance measure는 다음과 같다.

$$D(\tilde{y}_1, \tilde{y}_2) = |E_1 - E_2| \frac{\tilde{a}_{L_1}^T R_2 \tilde{a}_{L_1}}{\tilde{a}_{L_2}^T R_2 \tilde{a}_{L_2}}$$

본 논문에서는 수정된 Itakura-Saito distance measure를 사용함으로써 로컬 피크를 줄이고 음소의 경계가 되는 부분을 강조할 수 있었다.

2. Euclidean MFCC distance measure

Itakura-Saito distance measure는 기본적으로 LPC를 바탕으로 한 스펙트럼 비교법이기 때문에 로컬 피크가 적다는 장점이 있지만 상대적으로 스펙트럼의 변화가 작은 구간을 찾지 못하는 단점이 있다. 따라서 이러한 단점을 해결하기 위해 FFT를 바탕으로 한 스펙트럼 비교법을 도입하였다. 그러나 순수 FFT만을 이용하면 로컬 피크의 개수가 너무 많기 때문에 정확히 음소의 경계를 찾기 어렵다. 적당한 로컬 피크를 가지는 측정법을 도입하기 위해서는 인간의 청각 특성에 대하여 고찰할 필요가 있다. 인간의 청각 특성은 저주파 영역의 신호에는 민감하지만 고주파 영역의 신호에 대해서는 그렇지 못하다. 즉 1 kHz 정도까지는 선형 스케일(linear scale) 특성을 보이지만 1000Hz 이상에서는 로그 스케일(log scale)의 특성을 보인다. 이렇게 인간의 청각 특성을 고려한 것이 멜 스케일(mel scale)이다.

일반 선형 스케일을 멜 스케일로 변환하는 수식은 다음과 같다.

$$\text{mel frequency} = 2595 \log_{10}(1 + f/700.0)$$

이러한 멜 스케일을 기초로 한 MFCC(Mel Frequency Cepstrum Coefficient)의 개념은 1980년도에 Davis와 Mermelstein에 의해 도입되었고 현재 많은 인식 시스템이나 음향 코딩에 사용되고 있다. MFCC를 계산하는 과정은 다음과 같다.

1) 에너지 스펙트럼 계산

$$X(k) = \sum_{n=0}^{N-1} X(n)W(n) e^{-j2\pi \frac{kn}{N}}$$

$$W(n) = \text{Sc}(0.5 - 0.5 \cos(\frac{2\pi n}{N-1}))$$

$$P(k) = |X(k)|^2$$

($0 < k < N-1$, N : Hanning window의 크기)

2) 각 채널에 대한 에너지 계산

$$E_j = \sum_{k=0}^{K-1} \phi_j(k)P(k) \quad 1 \leq j \leq J$$

(채널의 수 J: 20 차)

3) MFCC 계산

$$C_j = A \sum_{k=0}^{K-1} \cos\left(j \frac{\pi}{J}(k+0.5)\right) \log_{10}(E_{k+1})$$

위의 과정을 통해 각 프레임에 대한 MFCC 를 계산한 뒤 Euclidean distance 를 계산하였다. i 번째 프레임과 j 번째 프레임에 대한 Euclidean MFCC distance 는 다음과 같다.

$$D_{MFCC}(i, j) = \sum_{k=1}^J (C_{ik} - C_{jk})^2$$

Euclidean MFCC distance measure 는 FFT 를 바탕으로 한 측정법이므로 Itakura-Saito distance measure 보다 음소 경계의 후보가 많았다. 따라서 두 가지 스펙트럼 비교법을 조합한다면 보다 정확한 음소의 경계를 찾을 수 있다.

3. 음소 경계 검출기

음소 경계 검출기는 음성을 발음의 최소 단위인 음소 단위로 자동으로 분류하는 것이다. 지금까지 대부분의 연구는 스펙트럼 비교법 만을 이용하여 음소의 경계를 찾으려고 하였다. 본 논문에서는 1 차적으로 유성/무성/목음 분류기를 이용하여 음성을 유성음, 무성음 구간으로 분류하였다. 무성음으로 분류된 구간은 한국어에는 연속된 무성음은 없기 때문에 하나의 음소로 분류할 수 있다. 따라서 무성음 구간에 대해서는 더 이상 스펙트럼 비교법을 이용하여 세분할 필요가 없다. 그러나 유성음 구간에 대해서는 연속된 유성음이 존재하므로 스펙트럼 비교법을 사용하여 세분하여야 한다. 스펙트럼 비교법으로는 앞에서 설명한 수정된 Itakura-Saito distance measure 와 Euclidean MFCC distance measure 를 이용하였다. 두 가지 스펙트럼 비교법을 조합하여 음소의 경계 후보를 찾은 뒤 마지막으로 평균 음소 길이를 이용하여 음소를 통합하거나 세분한 뒤 최종적으로 음소의 경계를 구하였다. 그림 3.1 은 음소 경계 검출기의 블록도이다. 음소 경계 검출기의 구현 과정을 정리하면 다음과 같다.

- 1) 유성/무성/목음 분류기를 이용하여 음성을 유성음과 무성음으로 분류
- 2) 무성음 구간은 하나의 음소 단위로 분류
- 3) 유성음 구간에 대해 스펙트럼 비교법을 이용하여 세분

4) 평균 음소 길이를 이용하여 음소를 통합하거나 세분

음소의 경계를 나누는 기준은 유성/무성/목음 분류기로 구분된 위치를 최우선으로 음소의 경계로 분류하였고, 두번째 우선 순위는 사용된 두 가지 스펙트럼 비교법에 의해 동일 위치에 음소의 경계를 찾은 경우로 하였다. 세번째 우선 순위는 두 가지 스펙트럼 비교법 중 어느 한 쪽에 의해 찾은 두 프레임 사이의 distance 값이 정해진 임계치 이상인 경우에 음소의 경계로 분류하였다. 이와 같이 음소 경계 후보에 우선 순위를 줌으로써 후처리 과정에서 음소를 통합하거나 세분할 때 참고 하였다. 즉, 음소의 경계 후보가 평균 음소 길이보다 적은 구간 안에 여러 개 있는 경우에는 음소 경계 후보를 나누는 우선 순위에 입각하여 우선 순위가 높은 위치를 음소의 경계로 분류하였다. 음소 경계 후보의 우선 순위 설정은 다음과 같다.

- 1) 유성/무성/목음 분류기를 이용하여 분류된 음소 경계 후보 위치
- 2) Itakura-Saito distance measure 와 Euclidean MFCC distance measure 에 의해 모두 음소 경계의 후보로 분류된 위치
- 3) Itakura-Saito distance measure 와 Euclidean MFCC distance measure 에 의해 어느 한쪽의 distance 값이 정해진 임계치 이상인 음소 경계 후보의 위치

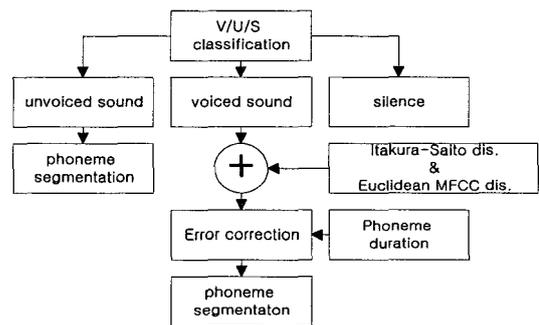


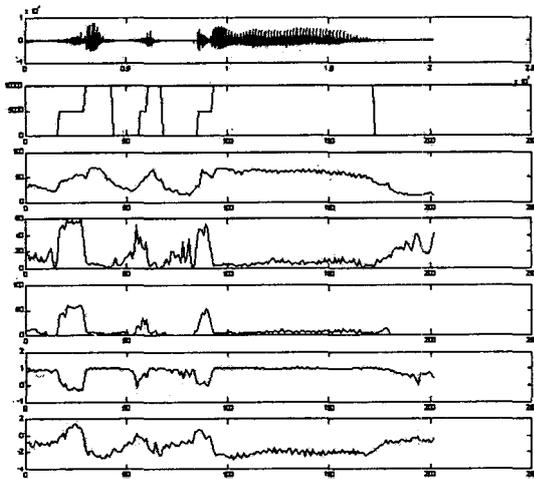
그림 3.1] 음소 경계 검출기의 블록도

IV. 실험 및 결과

유성/무성/목음 분류기와 음소 경계 검출기의 성능을 평가하기 위해 16 bit, 16 kHz 로 표본화된 각각 남녀 대학생 3 명이 20 개씩 발성한 고립단어를 음성 시료로

사용하였다. 알고리즘은 C 언어를 사용하여 구현하였다.

특징 파라미터 추출을 위한 분석 프레임의 길이는 에너지, 영 교차율, 레벨 교차율, 정규화된 자기상관 계수의 경우는 100 표본(10 msec)을 사용하였고 스펙트럼 에너지 비의 계산에는 256 표본(25.6 msec)을 사용하였다. 추출된 특징 변수의 예는 그림 4.1 과 같다. 발성 음성은 “소프트웨어”이고 위에서부터 차례대로 입력 음성 유성/무성/목음 분류 결과, 에너지, 영 교차율, 레벨 교차율, 정규화된 자기상관 계수, 스펙트럼 에너지 비율 나타낸다. 그림 4.1 을 통하여 앞에서 설명한 유성음과 성음의 특징을 확인할 수 있다.



[그림 4.1] 유성음/무성음/목음 판별기에 사용된 특징 변수 예 :/소프트웨어/

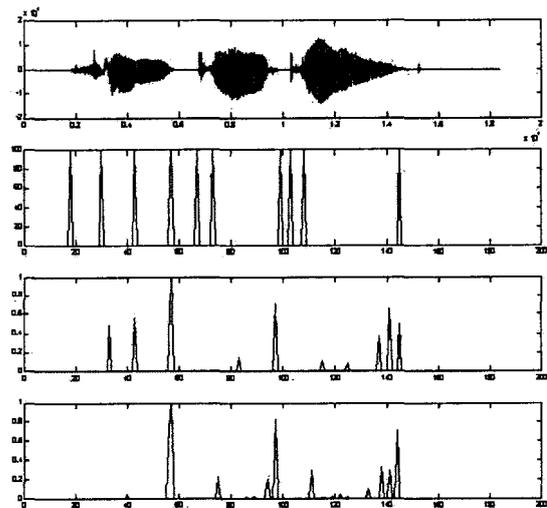
[표 4.1] 유성/무성/목음 분류기의 성능

Data Set	전체 프레임 수	에러 프레임 수	성능(%)
Mbtg	3546	113	96.813
Mcyj	3254	235	92.778
Mgsc	3246	200	93.839
Fcyj	3148	208	93.393
Fgbs	3242	174	94.633
Fhks	3416	212	93.794
남 자	10046	548	94.545
여 자	9806	594	93.942
전 체	19852	1142	94.247

유성/무성/목음 분류기는 앞에서 설명하였듯이 확실한

유성음, 무성음, 목음 구간에 대하여 각 특징 파라미터의 평균과 표준편차를 구한 뒤 이것을 표준 패턴으로 다른 구간에 대한 판별을 하였다. 유성/무성/목음 분류기의 성능을 향상시키기 위하여 1 차로 판별된 결과에 대해 동일한 과정을 적용하였다. 유성/무성/목음 분류기의 후처리 과정으로 무성음 구간에 일부 포함된 유성음, 혹은 유성음 구간에 일부 포함된 무성음을 수정함으로써 표 4.1 과 같은 결과를 얻었다.

그림 4.2 는 음소 경계 검출기의 예를 나타낸다. 발성 음성은 “생태계”이고, 위에서부터 차례 대로 입력 음성, 음소 경계 검출 결과, Euclidean MFCC distance 의 결과, Itakura-Saito distance 의 결과를 나타낸다. 앞에서 설명한 것처럼 Euclidean MFCC distance 의 결과가 음소 경계 후보로 분류한 개수가 Itakura-Saito distance 의 결과보다 많았다. 이 두가지 스펙트럼 비교법을 조합하여 음소 경계를 구하는 과정은 3.3 절에 설명한 내용과 같다. 스펙트럼 비교를 할 경우 비교 대상 프레임은 바로 이웃한 프레임과의 비교보다는 한 프레임 건너 뛴 프레임과 비교하였을 경우가 결과가 좋았다.



[그림 4.2] 음소 경계 검출 예 :/생태계/

표 4.2 는 본 논문에서 구현한 음성 경계 검출기의 결과를 나타낸다. 성능을 측정하기 위한 모든 조건은 유성/무성/목음 분류기에서 사용된 조건과 동일하다. 유성/무성/목음 판별기만을 사용하였을 경우 음소 경계 검출의 성능은 64.14%가 나왔고 음소 경계 검출기를 사용하였을 경우 음소 경계 검출의 성능은 72.80%가 나왔다. 결과적으로 본 논문에서 구현한 음소 경계 검출기는 유성/무성/목음 분류기만을 사용했을 경우보다 8.66 %의 성능 향상을 이룰 수 있었다.

[표 4.2] 음소 경계 검출기의 성능

Data Set	전체 프레임 수	에러 프레임 수	성능(%)	
			V/U/S 판별기	음소 경계 검출기
Mbtg	199	42	64.82	78.89
Mcyj	196	60	66.33	69.39
Mgsc	197	57	64.32	71.36
Fcyj	199	52	65.83	73.86
Fgbs	199	55	60.8	72.36
Fhks	199	58	62.81	70.85
남자	594	159	65.15	73.23
여자	597	165	63.15	72.36
전체	1191	324	64.14	72.80 (+8.66)

V. 결 론

본 논문에서는 한국어 고립단어 음성신호를 음소 단위로 분할하는 알고리즘을 제안하였다. 음소 경계 검출기를 스펙트럼 비교법만을 이용하여 구현한다면 적용 범위가 넓기 때문에 에러가 발생할 확률이 높다. 따라서 본 논문에서는 이러한 음소 경계 검출기를 유성/무성/목음 분류를 이용하여 1차로 분류한 뒤 스펙트럼 비교법을 이용하여 구현하였다. 즉, 유성/무성/목음 분류기로 분류된 결과 중 유성음 구간에 대하여서만 스펙트럼 비교법을 이용하여 세분화 하는 방식을 사용하였다. Itakura-Saito distance measure의 경우는 비교 프레임 간의 에너지 차를 곱해 줌으로써 많은 로컬 피크를 제거하였다.

최종 성능 평가 결과 유성/무성/목음 분류기의 성능은 94.25%였으며, 음소 경계 검출기의 성능은 72.8%였다. 이러한 음소 경계 검출 성능은 실제 사용하기에는 아직 많은 개선이 필요하며 이를 위해서는 유성/무성/목음 분류기의 성능을 향상 시킬 필요가 있다고 판단된다. 본 논문에서 구현한 유성/무성/목음 분류기의 경우 확실한 유성음, 무성음, 목음 구간을 찾는 과정에서 발생 고립단어에 유성음 구간이 약한 에너지를 갖는 경우 에러가 발생할 확률이 높았다. 따라서 이런 경우에 대한 개선책의 연구가 더 필요하다. 유성/무성/목음 분류기의 성능을 향상시킬 수 있다면 자연스럽게 음소 경계 검출기의 성능도 향상시킬 수 있을 것이다. 한편 본 실험을 위하여 사용된 어휘 수가 아직 적으므로 어휘수를 확장하여 성능 개선을 할 필요가 있으며 고립

단어 만이 아닌 연속음성에 대한 연구도 향후 이루어져야 할 것이다.

참고 문헌

- [1] L. R. Rabiner, R. W. Schafer, "Digital processing of speech signals," Prentice Hall
- [2] Lawrence Rabiner, Biing-Hwang, "Fundamentals of speech recognition," Prentice Hall
- [3] Hee-Il Hahn, Minsoo Hahn, "Voiced/unvoiced/silence classification of spoken Korean,"
- [4] Rivarol Vergin, Douglas O'Shaughnessy, Vishwa Gupta, "Compensated mel frequency cepstrum coefficients," Proc. ICASSP, vol.1, pp.323-326, 1996.
- [5] R. A. Brierton, B.M.G Cheetham, M.C.Hall, "A comparison of distance measures for speech segmentation in variable frame rate speech vocoding," Proc. IEE Colloquium, pp.6/1-6/5, 1990.
- [6] Joseph W. Picone, "Signal modeling techniques in speech recognition," Proc. IEEE, vol. 81, no. 9, pp.1215-1245, 1993.
- [7] Ludwig R. H. Everson, Walter T. Penzhorn, "Experimental comparison on several distance measures for speech processing applications," Proc. COMSIG, pp.12-17, 1998.
- [8] C. J. van der Merwe, J.A. du Preez, "Calculation of LPC-based cepstrum coefficients using mel-scale frequency warping," Proc. COMSIG, pp.17-21, 1991.