1999년도 한국음향학회 학술발표대회 논문집 제18권 제2(s)호

한국어 음성신호의 음절과 운율구 경계에 관한 연구

이기영, *송민석

관동대학교 정보통신공학과
*관동대학교 영어영문학과

A Study on Syllable's and Prosodic Phrase's Boundaries in Korean Speech Signal

Kiyoung Lee, *Minsuck Song

Dept. of Electronic Communication Engineering in Kwandong University
e-mail: kylee@mail.kwandong.ac.kr
* Dept. of English Language and Literature in Kwandong University
e-mail: mssong@mail.kwandong.ac.kr

※본 연구는 정보통신부에서 시행한 대학기초 연구지원 사업으로 수행된 것임.

Abstract

한국어의 연속 음성 인식을 위하여 운율구 단위로 경계를 검출하는 연구가 진행되고 있다. 그 과정의 일부로서 본 연구에서는 여러 음향 특징들을 조합하여 연속음성에서 음절 경계의 검출하는 방법을 제시하였으며, 연속 음성으로부터 한국어 운율구인 강세구의 경계를 운율 특징만을 이용한 패턴 비교 방법을 이용하여 검출한 것과 비교 검토하였다. 그 결과, 패턴 비교 방법으로 검출한 강세구의 경계를 음절의 경계와 일치되도록 정렬해줄 필요가 있음을 알 수 있었다.

1. 서 론

음성은 자연스럽고 편리하기 때문에 인간과 기계 사이의 통신수단으로 이용하기에 적절하지만, 음성언어는 그 자체가 불규칙하고 복잡하기 때문에 기계로 하여금 음성을 인식 또는 이해하게 하려는 연구가 매우 어렵게 진행되고 있다.

1980년대에 이후로는 고립단어를 인식하기 위하여 패턴매칭 방법인 DP 알고리즘[1,2]과 확률을 이용하는 HMM[3] 등이 연구되어 실용화되고 있다. 또한 음성인식의 범위를 넓혀 고립 단어 이상의 연속 음성으로 하려는 연구도 이미 이루어지고 있다.

연속 음성 인식 또는 이해 시스템의 접근 방법은 크게 bottom-up 접근과 top-down 접근이 있으며, 일반적으로는 이 두 가지 방법을 병행하여 사용하고 있다. 이를 해결하기 위한 방법으로 언어 모델을 적용하는 바이그램(bigram), 트라이그램(trigram), 유한상태 문법(finite state grammar), 무문맥 문법(context-free grammar) 및 LR 분석(LR-parsing) 등이 제안되어 연구되고 있으며, 이러한 방법을 이용하는 대어휘 연속음성 인식 시스템으로 Dragon 시스템[4], BBN Byblos 시스템[5] 등이 등장했다. 현재 국내에서도 음성인식 분야에서의 주요 연구 과제는 독립 단어 인식에서 연속 음성에 대한 인식으로 전환되고 있다[6,7]. 그러나 굴절 언어(inflected language)의 특성을 갖는 한국어에 대해 위와 같은 방법을 적용할 경우 음절 단위의 혼동이나 표준 언어 모델의 부족 등으로 인해 부적절하고 어려운 것으로 판단되고 있다[8]. 즉 다른 언어에 비교적 쉽게 적용할 수 있는 고립 단어 인식 기술과 달리 대어휘 연속 음성 인식 기술은 외국어 처리 기술을 그대로 한국어에 적용하기가 매우 어렵다.

본 연구에서는 한국어의 연속 음성 인식을 위한 bottom-up 접근 과정에 해당하는 음향 특징을 이용

한 음절단위의 경계검출 방법을 제시하고 본 연구자들이 제안한 바 있는 운율 정보만을 이용한 강세구의 경계검출[9,10]결과와 비교 검토하고자 한다.

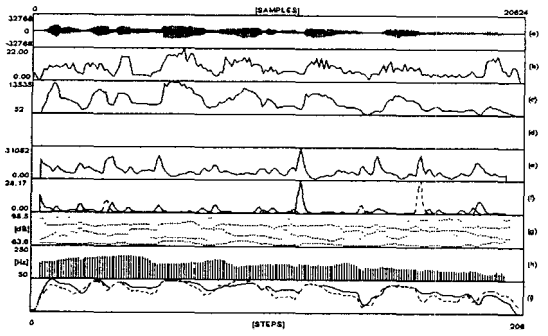

그림 1 연속 음성과 그의 음향 특징들

2. 음절 경계의 특징들

그림 1에는 연속 음성에서 음절의 경계 검출을 위한 특징들을 보이고 있다. (a)는 연속 음성의 파형이고 (b)는 영교차율, (c)는 0~2500 Hz까지 파형의 peak-to-peak(peak-to-peak), (d)는 1200~5000 Hz까지 파형의 peak-to-peak, (e)는 음성 파형의 프레임별 DFT의 유클리디안 거리, (f)는 전방향과 후방향에서 본 대수 우도비(log likelihood ratio), (g)는 포만트 궤적, (h)는 피치 궤적, (i)는 파형 에너지와 선형예측의 잔차 에너지이다.

(1) 영교차율

영교차율은 연속 음성의 마찰음을 검출하기 위한 것으로 에너지 특징만으로 무시하기 쉬운 음절의 경계를 파악하기 위하여 추출하였다.

(2) 대역 통과 파형의 peak-to-peak[11]

음성의 전 주파수를 포함하는 특징인 에너지는 모음과 자음의 특성을 모두 포함하고 있기 때문에 보다 정확한 음절의 경계를 추출하기에 부적당하여 모음 위주의 주파수 대역 0~2500Hz의 대역 통과 파형과 자음 위주의 1200~5000Hz 대역 파형의 프레임별 peak-to-peak를 각각 나타내었다. 여기서 대역 통과 필터는 FIR 필터이다.

(3) DFT의 유클리디안 거리

음성 파형의 DFT를 이용한 전력 스펙트럼은 해당 음성의 주파수와 피치 성분을 포함하고 있다. 따라서 이들의 연접한 프레임 사이의 유클리디안 거리는 음절들 사이에서뿐만 아니라 주파수 성분이 다른 음소들의 경계를 확인할 때 이용할 수 있다.

(4) 전방향과 후방향에서 본 대수 우도비[12]

Itakura 왜곡인 대수 우도비는 선형 예측 필터의 잔차 에너지의 비율로부터 얻을 수 있는 스펙트럼 왜곡의 측정 방법이다. 여기에서도 DFT의 유클리디안 거리와 마찬가지로 연접한 프레임 사이의 왜곡을 측정한다. 그러나 이 왜곡의 측정값은 예측된 필터에 입력되는 파형의 크기에 따라 달리 나타낼 수 있기 때문에 전방향과 역방향에서 본 대수 우도비로 나누어 스펙트럼 왜곡을 측정하였다. 전방향의 대수 우도비는 선형 예측된 필터에 다음 프레임의 파형을 입력시켜 잔차 에너지의 비율을 구하기 때문에 파형의 크기가 증가하는 부분의 스펙트럼 왜곡을 측정하기에 민감하며, 후방향의 대수 우도비는 이전 프레임의 파형을 입력시키기 때문에 파형의 크기가 감소하는 부분의 스펙트럼 왜곡을 측정할 때 민감한 측정값을 얻을 수 있다.

(5) 포만트 궤적

음성 파형이 포함하고 있는 포만트 주파수는 주로 서로 다른 모음 사이에 다른 모양의 궤적으로 나타난다.

(6) 에너지

음절의 경계를 검출할 때 에너지의 떨어지는 DIP를 구하는 방법으로 이용되고 있는 특징이다.

3. 음절과 강세구의 경계 검출

음절의 경계 검출이나 강세구의 경계 검출은 불특정 화자에게 적용할 수 있는 방법이며, 연속 음성을 인식하기 위해 사용되는 방법이다. 본 장에서는 음절 경계의 특징들을 이용하여 연속 음성에서 음절 경계를 검출하는 방법을 제시하고자 한다.

(1) 피크의 검출

한 음절은 반드시 모음을 포함하고 있기 때문에 해당 모음의 중심부에 음절의 핵이 있다. 이 부분을 검출하기 위해 모음 위주의 대역을 포함하는 0~2500 Hz 의 peak-to-peak에서 피크치를 검출한다. 또한 연속 음성은 여러 개의 음절로 구성되며 첫 음절부터 시작되므로 두 번째 음절 사이에 밸리(valley)가 존재한다. 따라서 피크를 검출하는 방법으로 한 피크를 구하면 그에 따른 밸리가 반드시 존재하며 마지막 음절에서는 피크와 그의 밸리가 음성의 끝점과 일치한다.

(2) 후보 DIP의 검출

이상의 과정에서 구한 밸리들은 꼭 DIP과 일치하지 않는다. sonorant가 연결되는 음절에서는 모음과 sonorant 사이에 밸리가 검출되기도 하기 때문이다. 따라서 검출된 피크 다음의 밸리를 후보 DIP라 명명한다. 0~2500 Hz 의 peak-to-peak에서 이 후보 DIP들을 검출하는 규칙은 다음과 같다.

① 피크의 70% 이하이다.
② 연속되는 값들은 이전 피크에서 밸리까지의 10%이상 상승한다.

(3) 음절의 경계 검출

음절의 경계는 이상에서 구한 후보 DIP들 중에서 결정한다. 음절의 경계를 검출하는 규칙은 다음과 같다.

① 연접한 피크사이에 후보 DIP의 크기가 양쪽의 어느 한 쪽의 피크보다 40%이하로 떨어지면 DIP로 한다.
② 한 피크 값이 전후의 각 피크 값보다 70% 이하로 떨어 지면 후보 sonorant로 한다.
③ sonorant 의 양쪽 후보 DIP 중에서 스펙트럼 왜곡이 작은 쪽을 DIP로 한다.
④ 모음이 연쇄적으로 발성되는 부분은 0~2500 Hz의 peak-to-peak에서 밸리가 없지만 포만트 궤적의 변화를 참조하여 음절의 경계로 한다.

(4) 강세구 경계 검출

문장 단위의 음성 데이터를 L (H L) H 패턴의 피치 궤적(pitch contour)에 의해 강세구의 경계를 검출한다. 여기서 강세구는 1993년 전선아가 제안한 한국어 운율구에서 적용하였으며 피치 궤적의 검출 방법으로 패턴 매칭 기법을 사용한다[9].

4. 실험 및 결과 고찰

(1) 실험 조건

본 실험에서는 표준어 발씨의 남자 5명과 여자 5명이 사전 지도 없이 서술형 문장 5개 씩 낭독하여 각자의 녹음기에 녹음한 것을 실험 음성 데이터로 하였다. 낭독할 서술형 문장 중에서 3문장은 15개의 음절, 다른 1문장은 14개의 음절, 나머지 1개 문장은 13개의 음절로 구성되어 5개의 문장에 포함된 음절의 수는 총 72개 이다. 컴퓨터 저장을 위해 KAY사의 Multi-speech를 이용하여 10kHz로 샘플링하였으며 분석 및 특징 추출을 위한 실험 조건은 다음 표와 같다.

표 1 분석 조건

프레임 간격	256msec
이동간격	100msec
선형예측차수	12

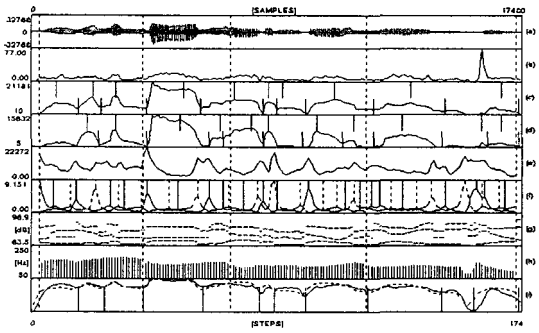

그림 2 음절과 강세구의 경계

(2) 음절 경계의 검출결과

본 연구에서 제시한 방법으로 음절 경계를 검출한 결과와 에너지에 의해 검출하는 기존의 방법으로 경계를 검출한 결과를 비교하였다.

표 2 음절경계 검출률 비교

	에너지	여러 특징
검출률	73%	78%

다음은 검출된 경계의 정확성을 비교 하였다. 그림 2는 제시한 방법에 의한 음절경계와 강세구의 경계를 보이고 있다. 그림 2의 (c)는 0~2500Hz의 peak-to-peak 이며 위쪽에서 내린 직선은 각 음절의 피크이고 아래쪽에서 올린 직선은 음절의 경계로 검출된 표시이다. 그림 2의 (i)는 에너지이며 아래쪽에서 올린 직선은 각 음절의 경계로 검출된 표시이다. 이들을 비교하기 위하여 경계의 차이 시간의 절대값들의 통계를 구한 결과 평균은 약 13msec, 표준편차는 약 6msec 였다.

본 실험의 프레임 간격을 10msec로 하였으므로 에너지에 의해 음절의 경계를 검출하는 기존의 방법과 큰 차이는 없었다. 그러나 음절의 경계가 음절사이에 휴지기로 나타나는 경우, 음절의 경계는 하나의 경계가 아니라 경계자체의 간격도 존재하며 두 번째 음절이 파열음이나 마찰음을 초성으로 할 경우에는 closure도 존재하므로 전 주파수 대역의

에너지만으로는 구분할 수 없음을 알 수 있다. 따라서 음절의 경계를 검출할 때 본 연구에서 제시하는 방법을 이용하여 음절 사이에 존재하는 휴지기나 closure 등을 검출해 준다면 보다 정확한 음절의 경계를 검출할 수 있을 것으로 사료된다.

(3) 음절과 강세구 경계의 비교

한국어의 강세구 경계는 주로 어절의 후치사의 경계와 일치해야한다. 그러나 강세구의 피치로 구성된 패턴으로 검출된 강세구의 경계는 후치사의 음절 경계와 거의 일치하지 않는다. 이들을 비교하기 위하여 강세구의 경계와 본 연구에서 제시한 음절의 경계와의 차이 시간의 절대값들의 통계를 구한 결과 평균은 약 35msec, 표준편차는 약 16msec 였다. 음절과 음절사이에 휴지기가 존재하는 경우에는 최대치가 80msec 인 차이 시간도 존재하였다. 따라서 강세구의 경계를 검출하기 위해서는 음절의 경계를 먼저 검출하여 피치 패턴으로 구한 강세구의 경계를 교정해줄 필요가 있음을 알 수 있었다.

4. 결 론

본 연구에서는 음향 특징들을 조합하여 연속음성에서 음절 경계의 검출하는 방법을 제시하였으며, 연속 음성으로부터 한국어 운율구인 강세구의 경계를 패턴 비교 방법을 이용하여 검출한 것과 비교 검토하였다. 그 결과, 에너지만을 이용한 음절의 경계 검출 방법보다 여러 특징을 이용하는 방법의 경계 검출율이 향상되었음을 알 수 있었으며, 운율 특징만을 이용한 패턴 비교 방법으로 검출한 강세구의 경계를 음절의 경계와 일치되도록 정렬해줄 필요가 있음을 알 수 있었다.

참고문헌

[1] Sakoe, H. and S. Chiba, 1978. "Dynamic Programming Optimization for Spoken Word Recognition." *IEEE Trans. Acoustics, Speech and Signal Processing*, Vol. ASSP-34, No.1, 52-59.

[2] Rabiner, L.R., A.E. Rosenberg and C. Myers, 1980. "Performence Tradeoffs in Dynamic Time Warping Algorithm for Isolated Word Recognition." *IEEE Trans. Acoustics, Speech and Signal Processing*, Vol. ASSP-28, No.6, 623-635.

[3] Levinson, S.E. L.R. Rabiner and etc., 1985. "Recognition of Isolated Digits using Hidden Markov Models with Continuous Mixture Densities." AT & T Tech. J., 64, No. 6, 1211-1234.

[4] Wegmann, Steven, Puming Zhan and Larry Gillick, 1999. "Progress in Broadcast News Transcription at Dragon Systems." *Proceedings of ICASSP '99*, Vol. 1, 33-36.

[5] Zavaliagkos, G. and etc., 1999. "Recent Experiments in Large Vocabulary Conversational Speech Recognition." *Proceedings of ICASSP '99*, Vol. 1, 41-44.

[6] Kim, Ho-Kyoung, Jae-In Kim and Myoung-Wan Koo, 1999. "A Study on Continuous Speech Recognizer based on Continuous Outout Probability Density." *Proceedings of ICSP '99*, Vol. 1, 329-332.

[7] Kwan, Oh-Wook, Kyuwoong Hwang and Jun Park, 1999. "Korean Large Vocabulary Continuous Speech Recognition of Newspaper Articles." *Proceedings of ICSP '99*, Vol. 1, 333-336.

[8] Waibel, Alex, Tanja Schultz, and Daniel Kiecza, 1999. " Data-Driven Determination of Appropriate Dictionary Units for Korean LVCSR." *Proceedings of ICSP '99*, Vol. 1, 323-327.

[9] Kiyoung Lee, Minsuck Song,"Automatic detection of Korean Accentual Phrase Boundaries," the Journal of the Acoustic Society of Korea,Vol.18,No.1E,pp.27-31,1999

[10] 송민석, 이가영, "악센트구를 이용한 한국어 연속음성에서의 조사 및 어미 인식," 한국음성과학회 제7회 학술발표회 논문집, pp.41-48, 1999

[11] R.A.Cole, Lily Hou,"Segmentation and Broad Classification of Continuous Speech," *Proceedings of ICASSP'88*, Vol. 1, 453-456

[12] 이기영, 배철수, 최갑석,"Likelihood ratio를 이용한 음소분류에 관한 연구," 한국음향학회지 7권 5호, pp.50-54, 1988