

낭독속도에 따른 강세구 경계 검출에 관한 연구

주장규, 이기영, *송민석

관동대학교 정보통신공학과
*관동대학교 영어영문학과

A Study on Detection of Accentual Phrase's Boundaries according to Reading Speeds

Jangkyu Ju, 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 지원하는 대학기초연구지원사업으로 수행된 것임.

Abstract

최근 운율 구조와 문장구조 및 음운규칙과 관련된 많은 언어학적 연구가 이루어져, 언어 이해 차원에서 의미 정보, 문장 구조 정보, discourse structure 등을 위한 운율 정보의 유용성이 입증되었으나, 이러한 결과가 최근의 음성인식 시스템에는 거의 적용되지 못하고 있다.

본 연구에서는 계층적인 방법을 기초로 하여 한국어의 연속음성으로부터 운율구를 검출하는 세그멘테이션법을 제안하였다. 우선, 입력된 음성으로부터 문장단위의 경계를 검출하기 위하여 휴지기를 이용하였으며 에너지, 휴지기의 지속시간 및 피치궤적을 참조하여 강세구의 경계를 검출하였다. 실험음성의 텍스트는 "만물상"이며, 남녀 각 2명의 표준어 화자가 빠른 속도와 보통 속도로 낭독한 음성데이터를 대상으로 비교하였다.

1. 서론

자연발화된 연속음성의 운율구 단위의 세그먼트와 레이블링은 연속음성의 인식에서 뿐만 아니라 음성합성 등의 음성신호처리 분야에서 필요한 처리 과정이다[1-3]. 왜냐하면, 인간은 음성언어의 경험

적 지식을 바탕으로 음성의 음향학적 정보인 단시간 스펙트럼의 정보를 청각을 통하여 인식할 뿐만 아니라 악센트, 억양, 에너지의 크기 및 pause 등의 운율정보를 인지하여 음성을 인식하고 이해하며 발성하기 때문이다.

언어학 분야에서는 운율 구조와 문장구조 및 음운규칙과의 관련성에 대한 많은 연구가 이루어져 [Selkirk, Nespor & Vogel, 전선아, Beckman and Pierrehumbert], 언어 이해 차원에서 의미 정보, 문장 구조 정보, discourse structure 등을 위한 운율 정보의 유용성이 입증되었으나, 이러한 결과가 최근의 음성인식 시스템에는 거의 적용되지 못하고 있다. 90년대 이후 미국의 Ostendorf[4,5], 독일의 프로젝트인 Verbmobil[6] 및 일본의 Shimodaira[7,8] 등의 연구 결과로 운율구 단위의 경계점 검출 및 레이블링 방법이 제시되면서 연속음성의 인식과 이해를 위한 작업으로 운율구 단위의 세그먼트방법의 많은 연구가 시도되고 있다. 한국의 경우 이기영[9-11]에 의해 운율 정보를 언어이해 시스템에 적용할 수 있는 가능성에 대한 연구가 계속적으로 이루어져, 한국어 중의성 문장의 음성을 이해하기 위하여 운율정보를 추정하거나[9,10] 운율패턴을 이용하여 낭독형 단일

문장의 음성에서 조사나 어미를 인식하였다[11]. 또 김진영[12]의 경우, 운율경계를 인식 단위로 하여 연속음성을 인식한 결과 계산량을 감소시키고 인식율을 증가시킬 수 있는 가능성을 확인하였다.

본 연구의 목적은 한국어의 연속음성인식 및 이해를 위한 실제적인 방법으로 피치 정규화 방법을 이용하여 한국어 음성의 강세구 단위의 경계를 자동 검출하는 방법을 제안 하고자 한다. 지난 논문 [10,11]에서는 평서문의 낭독형으로 발성한 단일 문장을 대상으로 피치패턴과 One-stage DP를 이용하여 강세구를 검출하였다. 그러나 이 방법만으로는 단일 문장이 아닌 연속 발성된 음성문장들을 운율 분석하기에는 불충분하였다. 따라서 본 연구에서는 우선 휴지기를 이용하여 연속음성을 억양구로 나누고 억양구 단위에서의 피치 궤적을 정규화하여 LH톤으로 나타나는 강세구의 경계를 검출하였으며 낭독속도에 따른 강세구의 검출율을 비교하였다.

2. 휴지기에 의한 억양구 검출

문장을 발성할 때 호흡의 길이나 의미의 전달을 목적으로 자연스런 휴지기를 가지며 발성을 진행한다. 이 휴지기는 억양구의 경계가 되며 그에 의해 문장의 의미도 달라질 뿐만 아니라 휴지기를 경계로 하는 억양구 단위를 연속음성의 인식에 이용하면 인식능력과 개선과 계산량의 감소에 효과적인 역할을 한다. 본 연구에서는 억양구를 자동 검출하기 위하여 에너지 특징으로 나타나는 휴지기를 검출하였다. 휴지기는 음절 경계의 검출과정에서 추출되는 DIP의 위치 중에서 검출된다. 그러나 음성의 휴지기는 문장의 경계나 억양구 단위의 경계에도 있지만 파열음의 앞/뒤에도 언제나 생성되는 요소이기도 하다. 따라서, 문장 단위나 억양구 단위 내에서 파열음의 존재로 인한 휴지기의 유무를 검토해 줄 필요가 있다. 예를 들면, 앞선 음절의 중성이 파열음이거나 현재 음절의 초성이 파열음인 경우 억양구 단위의 경계로서의 휴지기로 혼동하는 경우가 있다. 억양구 단위의 경계로 나타나는 휴지기는 파열음의 휴지기보다 훨씬 길게 나타나므로 이를 구별하기 위하여 파열음이 갖는 휴지기의 2 배길이를 억양구 단위 경계로서의 휴지기 최소 길이로 하여 문턱값으로 하면 올바른 억양구단위 경계의 휴지기를 검출할 수 있다.

3. 피치분석에 의한 강세구 검출

하나의 호흡단위 내에서 발성된 문장 음성의 피치궤적은 시간이 지날수록 점점 낮아지며 울곡의 폭도 차츰 좁아진다. 이는 생리적 현상 중에서 가장 뚜렷한 특징이다. 또한 한 문장의 연속음성은 여러 개의 강세구(accentual phrase)로 나뉘어 지기 때문에 피치궤적으로 나타나는 강세구의 특징을 보다 분명히 하기 위해서는 피치궤적의 정규화가 더욱 필요하다. 음성 파형에 대한 피치궤적, 시간에 따른 피치의 최소값으로 구성된 기저선(baseline), 또한, 기저선을 기준으로 한 정규화된 피치궤적, 정규화된 피치궤적의 미분궤적을 생각할 수 있다. 이 기저선은 거의 일직선 상에 존재하나 부방향의 기울기를 가진다. 이 기저선은 발성자나 음운학적 조건에 따라 다를 수 있으나 근사적으로 다음 식과 같이 나타낼 수 있다.

$$B(t) = \frac{p(T) - p(0)}{T} t + p(0) \quad (1)$$

여기서, $p(0)$ 은 음성의 시작점에서의 피치주파수이고 $p(T)$ 은 음성이 끝나는 시간에서의 피치주파수이며, T 는 음성의 시작점에서 끝점까지의 총 시간길이이다. 이상의 식(1)을 이용하여 재안한 정규화된 피치궤적을 식으로 나타내면 다음과 같다.

$$p(t) = p(t) - B(t) \quad (0 \leq t \leq T) \quad (2)$$

여기서, 식(2)를 최대치로 나누어 크기를 정규화한다. 또한 연속음성의 연속되는 LH톤의 경계를 검출하기 위하여 식(2)를 미분하면 다음과 같다.

$$\alpha(t) = \frac{dp(t)}{dt} \quad (3)$$

한국어 음성에서 강세구를 포함하는 억양구를 표시하는 주 요소는 경계 톤(tone)이다. 이것은 수반되는 접사나 의미에 따라 다양한 모양을 갖는다. 그 문장의 마지막 위치의 경계 톤은 정곡 동[12]이 지적인 것과 같이 휴지기(pause)은 억양구를 표시하는 중요한 요소이다. 강세구는 F0 음조곡선으로 표시된다. 하지만 F0 음조곡선은 focus, topic등과 같은 실

질적인 의미와 한국어 방언에 따라 다양한 패턴을 갖는다. 예를들면, 서울 방언에서는 강세구에 포함된 음절의 수에 따라 LH 또는 LHLH 톤의 패턴을 갖는다. 그러므로 강세구를 표시하기 위한 표준 패턴은 방언과 달라야 한다. 따라서 강세구를 자동 검출하기 위해서는 입력된 연속음성의 휴지기를 이용하여 억양구로 분류하고 각 억양구의 피치궤적을 이용하여 강세구를 추출한다.

문법적 어절의 경계와 운율적 어절의 경계는 매우 밀접한 관계가 있으므로 음절의 경계로 추출된 부분을 참조한다. 강세구 단위의 피치는 이상의 과정에 의해 기저선의 기울기를 평준화하고 크기를 정규화한 피치궤적과 그의 미분을 이용한다. 한국어 표준어에 있어서 강세구 단위의 경계에서는 그의 피치가 급격히 fall-rising하는 특성이 있으므로 정규화된 피치궤적의 미분은 강세구를 끝마치는 경계부위에서 최소치로 나타난다. 이를 기준으로 음절 분석 결과로 나타나는 음절 경계를 참조하여 가까운 위치로 강세구의 위치를 이동해 준다.

4. 실험 및 결과 고찰

4-1. 실험 음성 데이터

본 실험에서 사용한 텍스트는 21개 문장에 약 120개의 문법적 어절을 포함하는 조선일보의 “만물상”이며, 이들을 2명의 남성과 2명의 여성 표준어 화자가 낭독한 것을 실험의 음성 데이터로 사용하였다.

2명의 남성과 2명의 여성이 각각 보통 속도로 약 672.32초 동안, 또 빠른 속도로 약 571.48초 동안 낭독한 것을 테이프에 녹음하였다가 KAY사의 Multi-speech를 이용하여 컴퓨터에 저장하였으며 실험용 음성 데이터로 사용하였다. 또한 음성분석을 위한 프레임(frame)의 간격을 25.6 msec로, 이동간격을 10 msec로 하였다.

4-2. 실험 결과 및 고찰

한국어의 연속음성에서 운율구의 하나인 억양구를 자동 검출하기 위하여 휴지기를 검출할 수 있는 특징패러미터로 에너지를 이용하였다. 억양구 단위의 경계로 나타나는 휴지기를 파열음에서의 휴지기와 분리하기 위하여 파열음이 갖는 휴지기의 2 배

길이를 휴지기 길이의 문턱값으로 하였다. 또한 휴지기의 문턱 에너지는 최대 에너지와 배경잡음의 에너지의 중간 값으로 하였다.

그림 1은 연속음성에서 억양구를 검출한 예를 보이고 있다.

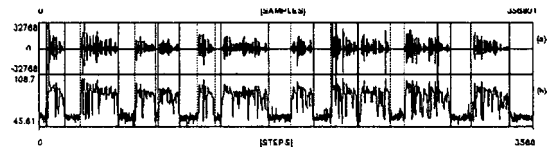


그림 1 휴지기를 이용한 억양구의 검출예

2명의 남성(M1, M2)과 2명의 여성(F1, F2)이 동일한 텍스트 “만물상”을 각각 보통 속도와 빠른 속도로 낭독한 음성 데이터를 대상으로 발성 속도에 따라 강세구의 검출률을 비교하였다. 여기서 보통 속도는 1초간 약 2개에서 3개까지의 음절을 낭독하는 속도로 텍스트 “만물상”의 낭독 시간은 남녀 평균 168.01초이며, 빠른 속도는 1초간 3개 이상의 음절을 낭독하는 속도로 텍스트 “만물상”의 낭독 시간은 남녀 평균 137.22초가 소요되었다.

그림 2는 보통 속도와 빠른 속도로 낭독한 연속음성으로부터 강세구를 검출한 결과를 비교하고 있다. 남녀 각 2명이 보통 속도로 낭독한 경우 강세구의 평균 검출률은 약 75%이며, 빠른 속도로 낭독한 경우 약 72%의 검출률을 얻을 수 있었다. 여기서, 빠른 속도로 낭독하는 경우 보다 보통 속도로 낭독한 음성의 경우 강세구를 검출률이 높은 것을 알 수 있었다. 또한 남성의 경우 빠른 속도에 비해 보통 속도에서 검출률이 높은 데 비해 여성의 경우는 오히려 빠른 속도에서 검출률이 높은 것으로 나타나고 있다.

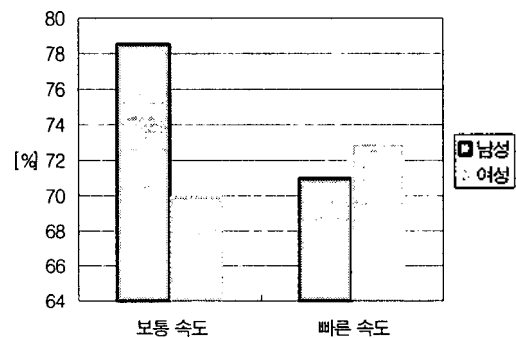


그림 2 낭독 속도에 따른 강세구의 검출률 비교

5. 결 론

이 연구에서는 한국어 연속 음성에서 운율단위인 억양구와 강세구를 자동 검출하는 방법을 제안하였다. 억양구를 검출하기 위해서는 입력된 문장 단위 이상의 음성 데이터로부터 파열음이 갖는 휴지기의 2 배의 길이를 갖는 휴지기를 추출하여 그 경계를 검출하며 음절 분석과 피치 분석을 이용하여 억양구에 포함되어 있는 강구를 검출한다. 또한 남녀 각 2명이 보통 속도로 낭독한 경우 강세구의 평균 검출률은 약 75%이며, 빠른 속도로 낭독한 경우 약 72%의 검출률을 얻을 수 있었다.

참고문헌

- [1] E. O. Selkirk, *Phonology and Syntax*, The MIT Press, Cambridge, Massachusetts, 1984.
- [2] M. Nespore and I. Vogel, *Prosodic Phonology*, Foris Publications, Dordrecht, 1986.
- [3] M. Beckman and J. Pierrehumbert, "Intonational structure in Japanese and English," *Phonology Yearbook 3*, ed. J. Ohala, pp. 255-309, 1986.
- [4] C. W. Wightman and M. Ostendorf, "Automatic labeling of prosodic patterns," *IEEE Trans. Speech, Audio Processing*, Vol. 2, No. 4, pp. 469-481, 1994.
- [5] C. W. Wightman and M. Ostendorf, "Automatic recognition of intonational features," in *Proc. of IEEE Int. Conf. ASSP*, pp. 1-221, 1992.
- [6] A. Kipp, M. Wessnick, and F. Schiel, "Automatic detection and segmentation of pronunciation variants in German speech corpora," in *Proc. of ICSLP96*, 1996.
- [7] H. Shimodaira and M. Kimura, "Accent phrase segmentation using pitch pattern clustering," in *Proc. of IEEE Int. Conf. ASSP*, pp. I-217, 1992.
- [8] H. Shimodaira and M. Nakai, "Prosodic phrase segmentation by pitch pattern clustering," in *Proc. of IEEE Int. Conf. ASSP*, pp. II-185, 1994.
- [9] Kiyong Lee and Minsuck Song, "Automatic recognition of sentence-final intonational patterns for Korean predicates," in *Proc. of the 12th Workshop for Speech Communication and Signal Processing*, Acoustic Society of Korea, pp.131-134, 1995.
- [10] Kiyong Lee and Minsuck Song, "Automatic recognition of intonational patterns for identifying Korean ambiguous sentences," in *Proc. of the 14th Conference of Acoustic Society of Korea*, Vol. 1, Acoustic Society of Korea, pp.72-75, 1995.
- [11] Kiyong Lee and Minsuck Song, "A study on recognition of Korean postpositions and suffixes in continuous speech," *Korean Journal of Speech Sciences*, Vol. 6, pp.181-195, 1999.
- [12] Sun-Ah Jun, *The Phonetics and Phonology of Korean Prosody*, Doctoral Dissertation, The Ohio State University, 1993.
- [13] Minsuck Song, *et al.*, "A theoretical model of spoken language processing," in *Proc. of the 1994 Conference of HCI(Human and Computer Interface)*, Korea Information Science Society, pp. 1-9, 1994.
- [14] Kiyong Lee, Minsuck Song, "Automatic detection of Korean Accentual Phrase Boundaries," *the Journal of the Acoustic Society of Korea*, pp.27-31, 1999