

한국어 연속음성에서의 조사 및 어미 인식에 관한 연구*

A Study on Recognition of Korean Postpositions and Suffixes in Continuous Speech

송민석** · 이기영***

(Minsuck Song · Kiyoung Lee)

ABSTRACT

This study proposes a method of recognizing postpositions and suffixes in Korean spoken language, using prosodic information. We detect grammatical boundaries automatically at first, by using prosodic information of the accentual phrase, and then we recognize grammatical function words by backward-tracking from the boundaries. The experiment employs 300 sentential speech data of 10 men's and 5 women's voice spoken in standard Korean, in which 1080 accentual phrases and 11 postpositions and suffixes are included. The result shows the recognition rate of postpositions in two cases. In one case in which only correctly detected boundaries are included, the recognition rate is 97.5 %, and in the other case in which all detected boundaries are included, the recognition rate is 74.8 %.

Keywords: accentual phrase, backward-tracking, continuous speech
recognition, grammatical function word, prosodic information

1. 머리말

음성은 자연스럽고 편리하기 때문에 인간과 기계사이의 통신수단으로 이용하기에 적절하지만, 음성 언어는 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(Sakoe & Chiba, 1978; Rabiner & Rosenberg, 1980)과 확률을 이용하는 HMM(Ferguson, 1980; Levinson & Rabiner, 1985) 등이 연구되어 실용화되고 있다. 또한 음성인식의 범위를 넓혀 고립 단어 이상의 연속 음성으로 하려는 연구도 이미 이루어지고 있으나, 연속

* 이 연구는 정보통신부에서 시행한 대학기초 연구지원 사업으로 수행된 것임.

** 관동대학교 문과대학 영어영문학과

*** 관동대학교 이공대학 정보통신공학과

음성에서 나타나는 단어 결합 법칙, 방언 또는 개인차에 따른 발음 변화 등 언어가 지니는 다양성으로 인하여 연속 음성 인식을 실현하기는 아직 어려운 실정이다.

이를 해결하기 위한 방법으로 언어 모델을 적용하는 바이그램(bigram), 트라이그램(trigram)(Shikano, 1987), 유한상태 문법(finite state grammar)(Reddy, 1980), 무문맥 문법(context-free grammar)(Ney, 1987) 및 LR 분석(LR-parsing)(Saito, 1989) 등이 제안되어 연구되고 있으며, 이러한 방법을 이용하는 대어휘 연속음성 인식 시스템으로 Dragon 시스템(Wegmann, 1999), BBN Byblos 시스템(Zavaliagos, 1999) 등이 등장했다. 현재 국내에서도 음성인식 분야에서의 주요 연구 과제는 독립단어 인식에서 연속 음성에 대한 인식으로 전환되고 있다(Kim & Koo, 1999; Kwan & Park, 1999). 그러나 위와 같은 방법을 적용할 경우 음절 단위의 혼동이나 표준 언어 모델의 부족 등으로 인해 부적절하고 어려운 것으로 판단되고 있다(Waibel & Schultz, 1999; 배재연, 신지영, 고도홍, 1999).

문장 단위 이상의 연속음성 인식에서는 독립단어 인식의 경우와는 다르게 문장 구조에 관한 통사 정보가 필요하다. 특히, 문장 성분의 경계 구분과 주요 성분인 주어, 목적어, 서술어 등의 문법 정보는 문장을 인식하고 이해하는 데 필수적인 정보이지만, 아직까지 연속음성에서 이러한 정보를 추출하는 기술이 개발되지 못하고 있다.

이 논문에서는 한국어 연속음성에서 이러한 통사 정보를 추출하기 위해 문법적 기능어인 조사 및 어미를 먼저 인식해 내는 방법을 제안하고자 한다. 이 방법에서는 문장 성분의 경계 구분 정보는 연속음성에서 운율 정보에 의해 나타나는 문장 성분의 경계를 알기 위해 Jun(1993)에서 제안된 운율구(prosodic phrase)인 강세구(accentual phrase)를 이용하며, 운율정보인 피치궤적(pitch contour)으로 나타나는 강세구를 자동검출하고(Lee & Song, 1997; Lee & Song, 1999; 한선희, 오미라, 1999), 그 경계점으로부터 각 강세구의 마지막 음절로 나타나는 문법적 기능어인 조사와 어미를 인식한다.

먼저 2절에서 통사 정보가 나타나는 언어적 특징을 영어와 한국어의 비교를 통해 알아보고 한국어에서 통사 정보를 알아내기 위해 필요한 요소와 이들의 특징을 검토한다. 3절에서는 2절에서 검토한 특징을 이용하여 문법적 기능어를 인식하는 알고리즘(algorithm)을 제안하고, 4절에서 실제 실험을 통해 이를 검증한다. 마지막으로 이러한 실험 결과와 앞으로의 연구 방향에 대해서 결론을 맺는다.

2. 한국어에 나타나는 문법적 기능어와 구 단위의 특성

한국어의 문법적 기능어는 명사, 동사, 형용사 등의 주요 어휘범주 단어들에 뒤 이어서 나타나는 후치사이다. 이들 기능어들은 자신이 붙어 있는 주요 어휘범주 단어들과 함께 문법적 구(grammatical phrase)를 형성한다. 문자 언어에서는 이 문법적 구 단위를 띄어쓰기로 표현한다. 다시 말해, 한국어의 후치사들은 자신이 붙어 있는 주요 어휘범주 단어들의 문법적 통사 정보와 함께 문장 구조 내에서의 문법단위 경계선 역할을 하고 있다. 따라서 현재까지 개발되어 사용하고 있는 음성인식 시스템과 같이 연속음성의 시작에서부터 끝 방향으로 단어를 인식하여 이러한 문법적 기능어를 먼저 찾아내기가 매우 어렵다(Waibel & Schultz, 1999).

텍스트 처리를 위주로 하는 자연어 처리 분야에서는 문장 구조를 분석(parsing)하기 위해 띄어 쓴 공백 문자를 찾아 거기서부터 역 방향으로 조사를 찾아내며, 서술어의 경우 형태소 분석을 수행하여 문장의 통사 구조를 파악하고 있다. 음성언어처리 분야에서도 이와 같은 과정을 이용한다면 문장 구조를 분석하기가 용이할 것이다. 즉, 문법적 구 단위의 경계점을 먼저 찾을 수 있다면, 거기서 역 방향으로 문법적 기능어를 찾아 문장 구조 분석을 수행할 수 있다(송민석 외, 1994; 이기영 외, 1995; 정국 외, 1996). 그러나, 문자 언어에서와는 다르게 음성 언어에서는 문법적 단위를 찾아내기가 쉽지 않다. 그것은, 경계점 역할의 요소가 문자언어의 공백 문자처럼 분명하게 나타나지 않고, 주로 피치 궤적의 변화, 길이 변화와 같은 운율 정보(prosodic information)에 의해 나타나며, 이러한 운율 정보는 여러 상황과 개인차에 따라 변화가 심하기 때문이다.

영어에서는 통사적 정보를 어휘 자체가 가지고 있으며 문장 내에서 어휘가 나타나는 위치에 의해 문장의 주요 성분을 파악할 수 있다. 즉, 동사 앞쪽 가까이 나타나는 명사가 주어가 되며 동사 뒤에 나타나는 명사가 목적어가 된다. 그러므로 독립단어 인식에서 사용한 기술을 확장하여 문장음성의 시작점으로부터 순차적으로 단어를 인식해 갈 경우 인식된 단어의 사전에 기록된 통사 정보와 문장 내에서의 그 단어의 위치에 의해 문장 구조를 파악할 수 있다.

그러나 한국어에서는 문장 주요 성분에 관한 통사적 정보가 조사와 어미 등의 문법적 기능어에 의해 나타나며 이로 인해 문장 내에서의 단어 위치는 문장 구조와는 전혀 상관이 없다. 다음 문장들은 영어와 한국어의 이런 차이점을 나타내는 예를 보이고 있다.

- (1) a. John loves Mary. (John이 주어, Mary가 목적어)
 b. Mary loves John. (Mary가 주어, John이 목적어)
 c. Mary John loves. (John이 주어, Mary가 목적어)
- (2) a. 철수가 영화를 사랑한다.
 b. 영화를 철수가 사랑한다.
 c. 철수가 사랑한다 영화를.

이런 이유로 영어를 위해 개발된 문장 단위 이상의 연속음성 인식 기술은 한국어에 적용하기 어려우며, 한국어 연속음성의 문장 구조 파악을 위해서는 우선 조사 또는 어미 등의 문법적 기능어를 찾아내는 것이 중요하다.

본 연구자들은 이미 한국어에 대해서, 운율 정보만을 이용해 자동으로 강세구(accentual phrase), 억양구(intonational phrase) 등과 같은 운율구(prosodic phrase)를 찾아내는 기술을 제안한 바 있으며(Lee & Song, 1997; Lee & Song, 1999), 실험결과 한국어의 운율구가 공백문자에 의해 분리되는 문법적 구 단위의 경계와 거의 일치함을 확인할 수 있었다. 따라서, 운율 정보를 이용하여 운율구 단위의 경계점을 찾아낸다면 그 경계점으로부터 역 방향으로 문법적 기능어를 찾아내는 것이 가능하다. 그림 1은 “은주는 마음이 아름다운 여인입니다”라는 문장의 음성 분석에서 나타나는 피치 궤적(pitch contour)이

L (H L) H 패턴을 보이는 운율구와 문법적 구 사이의 일치성을 보이는 예이다.

(3) a. 통사 구조:

[[은주]N는]TOPIC[[[[[마음]N이]SUBJ[아름다운]ADJ]R-CLAUSE [여인]N]NP입니다]PRED

b. 문법적 구의 구성:

[은주는]TOPIC/[마음이]SUBJECT/[아름다운]ADJECTIVE/[여인입니다]NOUNPREDICATE

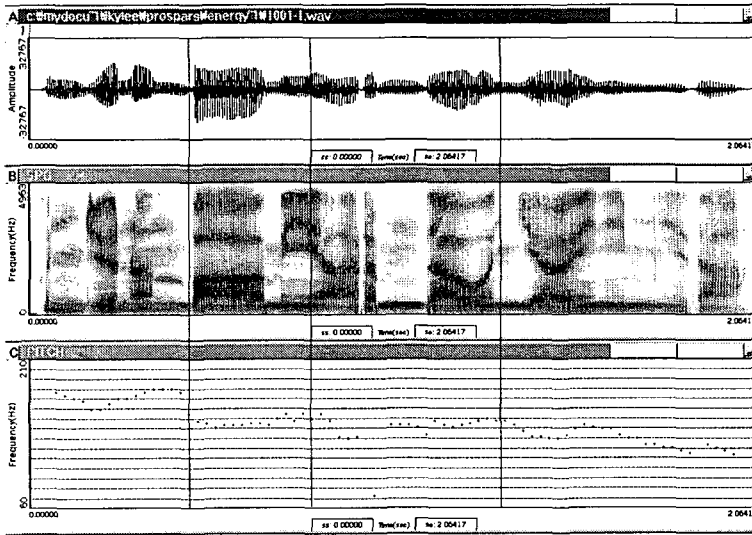


그림 1. “은주는 마음이 아름다운 여인입니다”의 피치궤적에 의한 강세구 구성

이러한 정보를 이용하면 다음과 같이 동일한 음소열로 구성되어 있으나 서로 다른 문법적 구를 형성하고 있는 두 문장 사이의 차이를 운율 특징에 의해 구분할 수 있다.

(4) a. “아버지가 방에 들어 가신다.”

b. “아버지 가방에 들어 가신다.”

그림 2에서 sentence 1과 sentence 2의 두 문장에 포함된 강세구의 피치궤적이 마지막 강세구를 제외하고는 L H 패턴으로 나타나며,¹⁾ 각 강세구의 마지막 음절이 중간에 나타날 때보다 지속시간이 긴 것을 알 수 있다.²⁾

1) 그림의 스펙트로그램 상에서 굵은 점선으로 나타나는 부분이 피치궤적임. 마지막 강세구의 경우는 억양구의 경계톤(boundary tone)이 나타나므로 LH 패턴이 나타나지 않는다.
 2) 각 문장의 세 번째 음절인 ‘지’를 비교해 보면 문장 1에서보다 문장 2에서가 더 길게 나타난다.

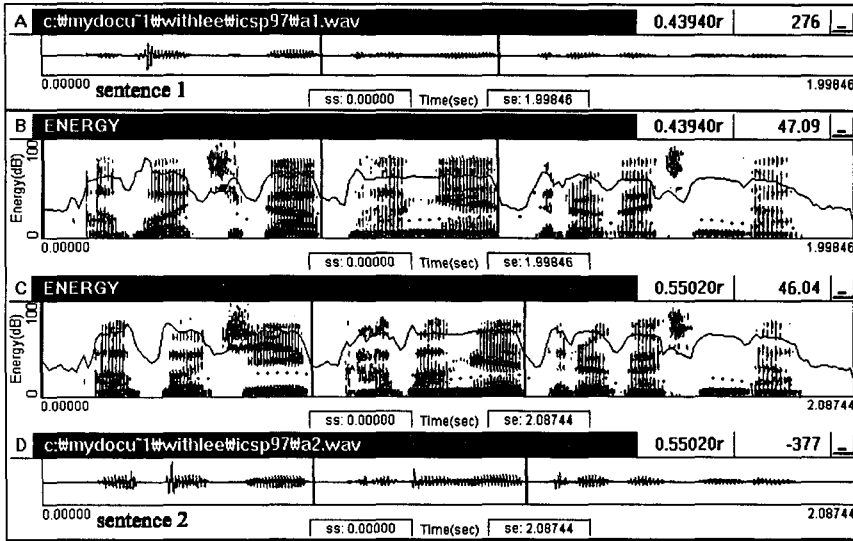


그림 2. “아버지가 방에 들어 가신다(sentence 1)”와 “아버지 가방에 들어 가신다(sentence 2)”의 운율 특징 차이

3. 한국어 문법적 기능어의 인식 알고리즘

한국어의 조사와 어미 등 문법적 기능어 인식을 위해서는 앞 절에서 설명한 것처럼 구 단위의 경계점을 먼저 찾아야 한다. 이를 위해, 운율구 단위로 Jun(1993)에서 제안한 강세구(accentual phrase)를 사용하여, 문장 단위의 음성 데이터를 L (H L) H 패턴의 피치 궤적(pitch contour)에 의해 강세구로 나누어 경계점을 검출한다. 이 때 피치 궤적의 패턴 매칭 기법으로는 DP 알고리즘(Ney, 1984)을 사용한다.³⁾

연속음성 데이터에서 강세구의 경계점을 검출한 후에는, 그 경계점에서부터 역 방향으로 추적하여 문법적 기능어를 인식한다. 그것은 앞 절에서 설명한 바와 같이 조사와 어미 같은 문법적 기능어들이 한국어에서는 구 단위의 마지막 부분에 나타나기 때문이다. 다음 그림 3은 한국어 문법적 기능어 인식을 위한 시스템 구조도이다. 여기서 AP는 강세구이다.

3) 패턴 매칭 기법으로 DP 알고리즘을 사용한 것은 이 연구가 가능성을 테스트해 보는 수준의 초기 연구 성격으로 한정된 데이터를 사용했고 이와 같은 소용량의 데이터를 위해서는 DP 알고리즘이 편리했기 때문이다. 앞으로의 연구에서 대용량의 음성 데이터를 사용하여 DP 알고리즘 이외에 퍼지, HMM 등 인식을 위해 제안된 다른 기법들을 이용하여 테스트해 보는 연구를 할 예정이다.

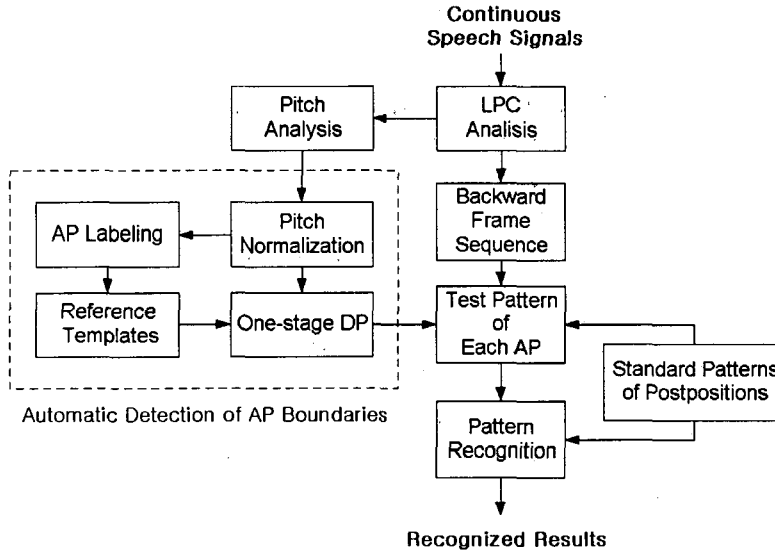


그림 3. 문법적 기능어 인식 시스템

3.1 피치궤적의 정규화

하나의 호흡단위 내에서 발생된 문장 음성의 피치궤적은 시간이 지날수록 점점 낮아지며 굴곡의 폭도 차츰 좁아진다. 이는 생리적 현상 중에서 가장 뚜렷한 특징이다. 또한 한 문장의 연속음성은 여러 개의 강세구(accentual phrase)로 나누어지기 때문에 피치 궤적으로 나타나는 강세구의 특징을 보다 분명히 하기 위해서는 피치궤적의 정규화가 필요하다.

그림 4는 한 서술형 문장에 대한 음성 파형의 피치궤적, 시간에 따른 피치의 최소값으로 구성된 기저선(baseline) 및 정규화된 피치궤적을 보이고 있다. 이 기저선은 거의 일직선 상에 존재하며 부 방향의 기울기를 가진다. 이 기저선은 발생자나 음운학적 조건에 따라 다를 수 있으나 근사적으로 다음 식과 같이 나타낼 수 있다.

$$B(t) = \frac{p_{end} - p_{begin}}{T} t + p_{begin} \quad (1)$$

여기서, p_{begin} 은 음성의 시작점에서의 피치주파수이고 p_{end} 는 음성이 끝나는 시간에서의 피치주파수이며, T 는 음성의 시작점에서 끝점까지의 총 시간길이이다.

이 연구에서는 식(1)을 이용하여 연속음성에 나타나는 기울어진 피치 궤적의 정규화 방법을 제안한다. 이 정규화 방법의 알고리즘은 다음과 같다. 먼저, 문장음성의 프레임 열의 수를 N 이라 하고 각 프레임별 피치의 값들을 $p(n)$ 이라 하자.

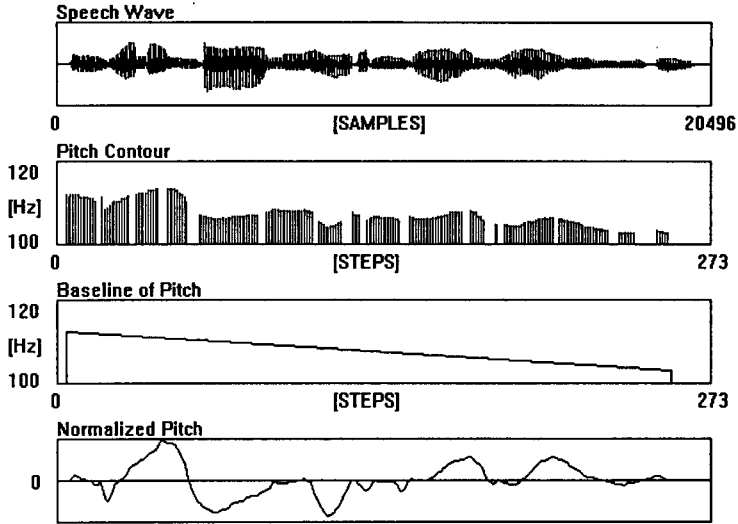


그림 4. 피치의 기저선과 정규화된 피치궤적

(a) 문장음성의 초반부 $N/4$ 에 해당하는 피치의 값들의 평균 P_{save} 를 구한다.

$$p_{save} = \frac{4}{N} \sum_{n=0}^{\frac{N}{4}-1} p(n) \quad (2)$$

(b) 문장음성의 후반부 $N/4$ 에 해당하는 피치의 값들의 평균 P_{favg} 를 구한다.

$$p_{favg} = \frac{4}{N} \sum_{n=0}^{\frac{N}{4}-1} p(N-n) \quad (3)$$

(c) 이상에서 구한 P_{save} 와 P_{favg} 사이를 선형내삽(linear interpolation)하여 부 방향으로 기울어진 근사 기저선의 식 $l(n)$ 을 구한다.

$$l(n) = \frac{n}{N} (P_{favg} - P_{save}) \quad (4)$$

(d) 근사 기저선 즉 식(4)의 $l(n)$ 을 이용하여 평균화된 피치궤적 $\bar{p}(n)$ 을 구한다.

$$\bar{p}(n) = p(n) - l(n) \quad (5)$$

(e) 식(5)에 의한 피치궤적 $\bar{p}(n)$ 의 각 값을 최대치로 나누어 크기를 정규화 한다.

강세구 단위의 피치패턴은 이상의 과정에 의해 기저선의 기울기를 평균화하고 크기를 정규화한 피치궤적을 이용한다.

3.2 한국어 강세구 경계점의 자동 검출법

한국어의 강세구 자동 검출 알고리즘(algorithm)으로는 Ney(1984)의 one-stage DP와 앞 절에서 제안한 정규화(normalization) 방법을 사용한다. 음성 데이터가 입력되면 먼저 자기상관함수(autocorrelation)에 의해 피치 궤적을 추출하고, 피치 궤적을 정규화 하여 이미 작성된 표준패턴과 one-stage DP를 이용하여 강세구의 경계점을 찾는다. 강세구의 표준 패턴 작성은 우선 하나의 문장을 다음과 같이 수 작업으로 강세구를 나누고 레이블링(labeling)한 후 나누어진 단위의 피치 궤적을 표준 패턴으로 삼는다.

(5) [은주는]TOPIC/[마음이]SUBJECT/[아름다운]ADJECTIVE/[여인입니다]NOUN PREDICATE

Jun(1993)에서 밝혔듯이, 한국어 강세구의 피치 궤적은 L (H L) H 패턴을 보인다. 음성 데이터의 피치 궤적을 앞 절에서 제안한 방법으로 정규화하면 단계적 하강(down-step) 모습을 갖는 평균화된 경사 각도가 추출되며, 이들은 더욱 분명한 LH 패턴을 보이게 된다. 또한 one-stage DP를 이용한 패턴 비교법을 사용하면, 연속한 패턴을 비교할 수 있게 되며, 표준 패턴 역시 패턴 비교를 위해 실제 음성 데이터에서 추출한 피치 궤적을 정규화하여 작성한다. one-stage DP에서 사용된 거리 측정식은 다음과 같다.

$$dAB(n, m) = (1 - \alpha) \cdot \{\bar{p}A(n) - \bar{p}B(m)\} + \alpha \cdot \{\Delta \bar{p}A(n) - \Delta \bar{p}B(m)\}, \quad 0 \leq \alpha \leq 1 \quad (6)$$

$$\Delta \bar{p}(n) = \partial \bar{p}(n) / \partial n \quad (7)$$

여기서 식(6)의 가중치 α 의 최적치는 Lee & Song(1999)의 논문 실험결과를 이용하여 0.2로 하였으며, 식(7)은 시간 변화에 따른 피치 변화 정도를 나타낸다.

3.3 조사 및 어미의 인식

조사 및 어미의 인식 과정에서는 그림 3과 같이 자동 검출된 강세구의 경계를 이용한다. 각 강세구 단위의 마지막 음절은 후치사로 조사 및 어미에 해당되므로 이를 인식하기 위해서는 검출된 강세구의 경계로부터 역 방향으로 추적할 필요가 있다. 즉, 각 강세구 단위로 작성되는 시험 패턴의 끝점은 각 강세구의 경계이며 마지막 음절의 끝점과 같다.

그러나 마지막 음절에 해당하는 조사나 어미의 시점은 알 수 없다. 따라서 각 시험 패턴의 끝점으로부터 역 방향으로 마지막 음절의 시점을 찾고 그 음절이 어떤 조사나 어미에 해당하는지 인식해 줄 필요가 있다.

이상의 역 방향 추적 방법으로 패턴 인식 방법인 DP 알고리즘(Sakoe & Chiba, 1978)을 확장하여 이용한다. 본래의 이 방법은 일반적으로 고티 단어 인식 방법으로 한 단어의 시험 패턴에 대한 시점과 끝점을 먼저 검출한 다음 시간축을 정규화하는 방법으로 이용되

고 있다. 이 연구에서는 강세구의 경계로 자동 검출되는 시험 패턴의 끝점은 알지만 끝점에서 역 방향으로 존재하는 마지막 음절의 시점 위치를 알 수 없으므로 시험 패턴의 끝점으로부터 역 방향으로 각 표준 패턴 길이의 1.5배 길이를 취하여 DP 알고리즘을 수행한다. 여기서 각 표준 패턴 길이의 0.5배에서 1.5배 사이에 해당하는 시험 패턴과의 정규화 거리 중에서 최소인 것을 각 표준 패턴과의 확장된 정규화 거리로 선택하며 그 경계를 인식 경계, RB(recognized boundary)로 저장한다. 이 과정을 그림으로 나타내면 다음과 같다.

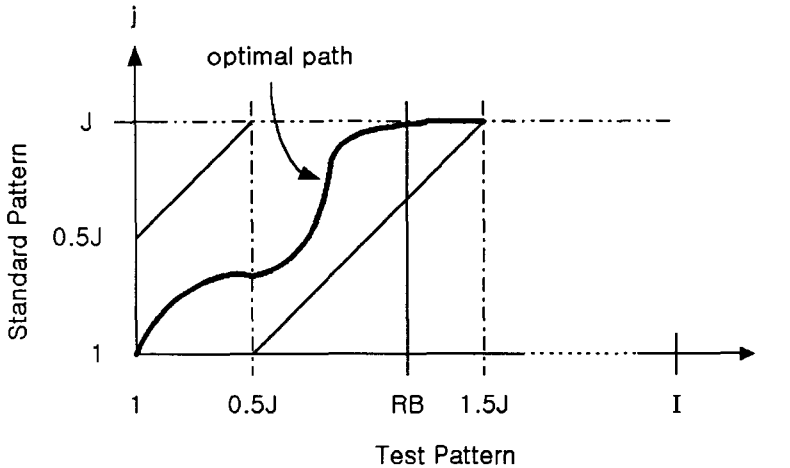


그림 5. 확장된 DP 알고리즘에 의한 인식 과정

그림 5에서 I는 시험 패턴의 총 길이이고 J는 표준 패턴의 총 길이이다. 또한 각 패턴은 미리 역 방향 시계열로 전개되어 있으므로 각 축의 1은 끝점이며 I와 J는 각 패턴의 시점을 나타낸다. k번째 표준 패턴의 확장된 정규화 거리 D_k 를 식으로 표현하면 다음과 같다.

$$D_k = \min d_k(i, 1) \quad (0.5J \leq i \leq 1.5J) \quad (8)$$

$$RB_k = \arg \min d_k(i, 1) \quad (0.5J \leq i \leq 1.5J) \quad (9)$$

식(8)과 같이 모든 표준 패턴의 확장된 정규화 거리를 구한 다음 전 표준 패턴의 확장된 정규화 거리 중에서 최소인 표준 패턴을 선택하여 인식하며 식(9)와 같이 그 때의 인식 경계, RB를 인식된 패턴의 경계로 한다. 이상의 과정에 의해 강세구 단위의 시험 패턴으로부터 음절 단위의 조사 및 어미를 인식한다.

4. 실험 및 결과

4.1 실험 데이터 베이스

데이터 수집에 사용된 자료는 20개의 평서문 문장으로, 네 개의 문장은 네 개의 강세구로, 또 다른 네 개의 문장은 다섯 개의 강세구로, 나머지 열 두 개의 문장은 세 개의 강세구로 구성되어 있다. 음성 데이터는 남성 10명, 여성 5명의 표준어 화자가 가능하면 자연스러운 발성을 하도록 아무런 사전 연습이나 읽는 방법에 대한 사전 안내 없이 읽은 낭독체 음성을 카세트 녹음기로 녹음하였으며 컴퓨터 저장을 위해 KAY사의 Multi-speech를 이용하여 10 kHz로 샘플링(sampling)하였다. 또한 음성분석을 위한 프레임(frame)의 간격을 25.6 msec로, 이동간격을 12.8 msec로 하였다. 이렇게 하여 실험에 사용된 자료는 남녀 총 15명의 표준어 화자가 발음한 1,080개의 강세구를 포함하고 있는 300개의 문장 음성 데이터이다.

인식에 사용한 조사 및 어미는 가능성을 확인하기 위한 연구이므로 형용사형 어미와 종결 어미를 포함하여 모두 11개의 단어가 사용하였다. 이들은 주격 조사 '-은, -는,' 주격 조사 '-이, -가,' 향진격 조사 '-로,' 처소격 조사 '-에,' 서술어의 형용사형 어미 '-ㄴ,' 종결형 어미 '-다, -요, -죠'이다.⁴⁾

4.2 실험 결과 및 고찰

연속음성의 문법적 기능어인 조사와 어미를 인식하기 위하여 그림 3의 문법적 기능어 인식 시스템에 의해 실험하였다. 실험 결과는 다음과 같이 연속음성의 강세구의 검출율과, 강세구의 자동 검출된 경계의 정확성에 대한 검토 및 강세구의 후치사인 조사 및 어미의 인식률로 나누어 기술하였다.

(1) 강세구의 검출율

연속음성이 입력되면 LPC 분석하는 과정에서 피치를 분석하고 정규화된 피치 궤적과 one-stage DP를 이용하여 강세구의 경계점을 자동 검출한다. 여기서 표준 패턴은 남자 2명과 여자 1명이 각각 발성한 1개 문장에서 수동으로 검출하여 레이블링한 강세구의 정규화된 피치 패턴으로 구성하였다.

다음 그림 5와 6은 one-stage DP와 식(6)을 사용하여 여자 5명과 남자 10명의 각 화자에 따른 강세구의 경계를 검출한 결과이다. 여성 화자 전체의 평균 검출율은 80.8 %이며 남성 화자 전체의 평균 검출율은 83.3 %이다.

4) 단, 사용된 원고에 '빨리'라는 부사가 포함되어 있는데, 이 경우 마지막 모음이 '이'로, 한국어의 주격조사 '이'와 구분되지 않는다. 따라서 이번 실험에서는 이것을 구분하지 않고 모두 '이'로 인식하였으면 일단 올바르게 인식한 것으로 처리하였다. 앞으로의 연구에서 자연어 처리 분야에서 수행하고 있는 어휘 분석이나, 조사를 인식하고 난 나머지 부분에 대한 단어 인식 부분을 포함한다면, 단어 전체가 부사인지 또는 명사에 조사가 붙어 있는 지를 판단하는 부분이 개발될 것으로 사료된다.

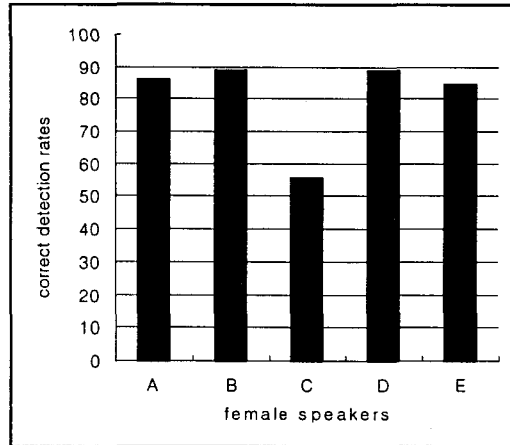


그림 6. 여성 화자별 검출율

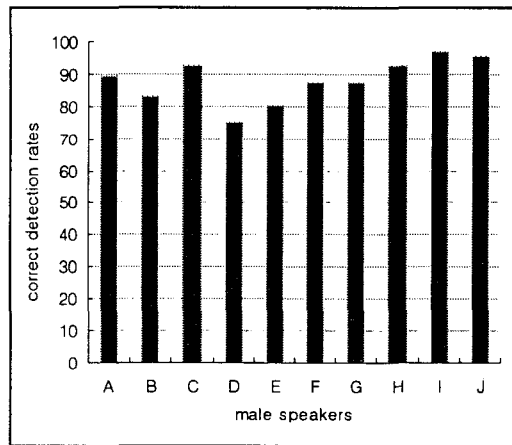


그림 7. 남성 화자별 검출율

여성화자 중에서는 C화자가, 또 남성화자 중에서는 D화자가 가장 낮은 강세구 검출율을 보이고 있다. 그 이유는 녹음 환경을 의식하여 낭독 음성이 매우 부자연스러웠으며 이에 따라 강세구의 피치 궤적이 LH 패턴으로 나타나지 않기 때문인 것으로 사료된다.

(2) 강세구 경계의 검토

한국어 강세구 경계의 자동 검출 방법에 의해 검출된 강세구의 경계는 각 강세구의 마지막에 위치한 후치사의 경계와 일치하여야 한다. 그러나 피치 패턴에 의해 검출한 강세구의 경계는 이 후치사의 경계와 대부분 일치하지 않는다(이기영, 송민석, 1999). 또한 본 연구에서 제시한 확장된 DP 알고리즘은 본래 시점과 끝점의 정확성에 의해 인식률이 크게 영향을 받는 방법이므로 제안된 본 연구의 강세구 경계의 자동 검출 결과를 실제 후치

사인 조사 및 어미의 경계와 비교하였다. 여기서 실제의 후치사의 경계는 수동으로 레이블링한 경계이다. 비교 방법은 자동 검출된 경계와 수동으로 레이블링한 경계와 오차 시간의 절대값들의 통계를 구하였다. 이들의 평균은 약 35.3 msec, 표준편차는 약 16.8 msec로 정확히 일치하지 않고 있음을 알 수 있었다.

이를 개선하기 위해서는 본 연구에서 제안한 강세구의 경계 자동 검출 방법과 음절경계의 추출 방법을 조합하여 강세구 경계의 정확성을 보완해 줄 필요가 있음을 알 수 있었다.

(3) 조사와 어미의 인식률

조사와 어미의 인식과정에 사용하는 패턴 비교 방법의 거리는 스펙트럼 왜곡 측정 방법인 대수우도비(log likelihood ratio; Itakura, 1975)를 이용하였으며 LPC 분석의 차수는 12차로 하였다. LPC 분석과정에서 추출된 프레임 단위의 특징 벡터인 LPC 계수는 후치사로 오는 문법적 기능어인 조사와 어미를 인식을 위해서 역 방향으로 정렬된다. 그 다음, 검출한 강세구의 경계점으로부터 역 방향으로 3.3절에서 제시한 확장된 DP 알고리즘을 이용한 패턴 매칭에 의해 조사와 어미를 인식한다. 여기서 표준패턴은 화자 종속 패턴으로 구성하였다.

표 1은 두 가지 경우의 조사 인식률을 나타내고 있다. 첫 번째 경우는 강세구 경계점을 올바르게 검출한 경우의 인식률(recognition rate of correctly detected boundaries: RRCDB)을 나타내며, 두 번째 경우는 강세구 경계점을 잘못 검출한 경우까지 포함한 전체 인식률(recognition rate of all detected boundaries: RRADB)을 나타낸다.

표 1. 조사와 어미의 인식률

	RRCDB	RRADB
인식률	97.5	74.8

이 연구의 조사와 어미의 인식인 우선 입력된 연속음성의 강세구 경계점을 검출한 후 이루어지기 때문에 피치 궤적에 의한 강세구 검출율에 매우 의존적이지만 한번 강세구의 경계가 정확히 파악되면 조사나 어미가 제한된 소규모 어휘이고 대부분 단음절로 이루어졌기 때문에 그 인식률이 상당히 높음을 알 수 있다.

5. 맺음말

이 연구에서는 한국어 연속 음성 언어에서 운을 정보를 이용하여 문법적 기능어인 조사와 어미를 먼저 인식해 내는 방법을 제안하였다. 입력된 문장 단위 이상의 음성 자료를 피치 궤적과 길이 등의 운을 정보를 이용해 강세구의 경계점을 찾은 후, 그 경계점으로부터 역 방향 추적으로 조사 및 어미를 인식하는 방법이다.

강세구의 후치사로 나타나는 조사와 어미의 끝점 경계를 검출하기 위하여 본 연구에서 제안한 자동 검출법으로 실험을 수행한 결과 평균 82 %의 검출율을 보였다. 또한 검출된

강세구의 경계를 이용하여 확장된 DP 알고리즘에 의해 조사와 어미의 인식을 수행한 결과 두 가지 경우의 조사 인식률을 얻을 수 있었다. 첫 번째 경우는 강세구 경계를 올바르게 인식한 경우의 조사 인식률로 97.5 %의 높은 인식률을 보였으며, 두 번째 경우는 잘못 인식한 강세구 경계점까지 모두 포함한 경우의 조사 인식률로 74.8 %의 인식률을 나타내었다. 따라서 제안된 방법은 기존의 연속음성 인식방법을 이용하지 않고 한국어 연속 음성 언어에서 운율 정보인 강세구를 이용하여 문법적 기능어인 조사와 어미를 인식할 수 있음을 확인할 수 있었다. 그러나 본 연구에서 제안한 강세구 경계의 자동검출방법은 피치 패턴을 이용하므로 실제로 후치사에 해당하는 조사나 어미의 음절 경계와 정확히 일치하지 않는다. 따라서 피치 이외의 음향 특징 파라미터를 이용한 음절 경계의 검출방법을 개발하여 강세구의 자동 검출된 경계와 일치 시켜줄 필요가 있음을 알 수 있었다.

이 연구에서 제안한 강세구 경계의 자동 검출 방법은 궁극적으로 인식된 문법적 기능어에 의해 문장의 문법적 구조를 분석할 수 있고, 문장 단위 이상의 연속 음성을 인식하는 것이 용이하다. 뿐만 아니라, 이 방법은 현재 개발되어 있는 독립 단어 인식 기술을 그대로 문장 단위 이상의 연속 음성 인식에 사용할 수 있도록 해 준다. 즉, 조사나 어미의 인식을 수행한 후 나머지 음성 자료를 그대로 독립 단어 인식 시스템의 입력(input)으로 사용한다면, 현재의 기술을 그다지 수정하지 않고도 한국어 연속음성 인식에 사용할 수 있을 것으로 사료된다.

마지막으로, 음성실험 연구에서 거의 항상 나타나는 문제점으로, 이러한 방법을 대용량의 음성 자료와 많은 화자에 의한 화자 독립적 인식 실험을 수행해 그 문제점을 파악하고 연속 음성을 인식할 수 있는 시스템을 개발하는 것이 앞으로의 연구 과제이다.

참 고 문 헌

- 배재연 · 신지영 · 고도홍. 1999. "음성 환경에 따른 한국어 폐쇄음의 음향적 특성." *음성과학* 5(2), pp. 139-159, 한국음성과학회.
- 석종원 · 손영호 · 배건성. 1999. "웨이브렛 변환을 이용한 피치검출." *음성과학* 5(1), pp. 23-34, 한국음성과학회.
- 송민석 외. 1994. "음성언어 인식의 이론적 모델." HCI '94 학술대회 발표논문집, *인간과 컴퓨터 상호작용 연구회 회보* 3(1), pp. 1-9, 한국정보과학회.
- 이기영 외. 1995. "한국어 중의성 문장을 구분하기 위한 억양패턴의 자동인식." *한국음향학회 학술발표대회 논문집* 14권 1(s)호, pp. 219-222.
- 이기영 · 송민석, 1999. "악센트구를 이용한 한국어 연속음성에서의 조사 및 어미 인식." *한국음성과학회 제7회 학술발표회 논문집*, pp. 41-48.
- 이기영 · 송민석. 1999. "한국어 음성신호의 음절과 운율구 경계에 관한 연구." *한국음향학회 학술발표대회 논문집*, 제18권 2(S)호, pp. 361-364.
- 정국 외. 1996. 음성 합성/인식 시스템 개발을 위한 한국어 운율 구조 및 담화 구조 연구. 한국통신 연구개발단 연구보고서.
- 한선희, 오미라, 1999. "한국어 억양구의 경계론." *음성과학* 5(2), pp.109-130, 한국음성과학회.
- Alex, W., Schultz, T. & Kieca, D. 1999. "Data-Driven Determination of Appropriate Dictionary Units for Korean LVCSR." *Proceedings of ICSP '99, Vol. 1*, pp. 323-327.

- Ferguson, J. D. 1980. "Hidden Markov Analysis: An Introduction." in Hidden Markov Models for Speech, Institute of Defense Analysis, Princeton, NJ.
- Jun, S.-A. 1993. *The Phonetics and Phonology of Korean Prosody*. Doctoral dissertation, The Ohio State University.
- Kim, H.-K., J.-I. Kim & M.-W. Koo, 1999. "A Study on Continuous Speech Recognizer based on Continuous Output Probability Density." *Proceedings of ICSP '99 Vol. 1*, pp. 329-332.
- Kiyohiro, S. 1987. "Improvement of Word Recognition Results by Trigram Model." *Proceedings of ICASSP '87, Vol. 3*, pp. 1261-1264.
- Kwan, O.-W., K. Hwang & J. Park, 1999. "Korean Large Vocabulary Continuous Speech Recognition of Newspaper Articles." *Proceedings of ICSP '99, Vol. 1*, pp. 333-336.
- Lee, K., & Song, M. 1997. "Automatic Segmentation of Korean Prosodic Phrases." *Proceedings of ICSP '97, Vol. 2*, pp. 747-751.
- Lee, K., & Song, M. 1999. "Automatic detection of Korean Accentual Phrase Boundaries." *The Journal of the Acoustic Society of Korea, Vol. 18, No. 1E*, pp. 27-31.
- Levinson, S. E., L. R. Rabiner, etc. 1985. "Recognition of Isolated Digits using Hidden Markov Models with Continuous Mixture Densities." *AT & T Tech. J.*, 64(6), pp. 1211-1234.
- Ney, H. 1987. "Dynamic Programming Speech Recognition using a Context-free Grammar." *Proceedings of ICASSP '87(1)*, pp. 69-73.
- Ney, H. 1984. "The use of a one-stage dynamic programming algorithm for connected word recognition." *IEEE Trans. Acoustics, Speech and Signal Processing, Vol. ASSP-32(2)*, pp. 263-271.
- Rabiner, L. R., A. E. Rosenberg. & C. Myers, 1980. "Performance Tradeoffs in Dynamic Time Warping Algorithm for Isolated Word Recognition." *IEEE Trans. Acoustics, Speech and Signal Processing, Vol. ASSP-28, No. 6*, pp. 623-635.
- Reddy, R. & Lowerre, B. 1980. "The Harpy Speech Understanding System." in Trends in Speech Recognition, Lea, Wanyne A.(ed.), Prentice-Hall.
- Saito, H., Kawabata, T. & Kita, K. 1989. "HMM Continuous Speech Recognition using Predictive LR Parsing." *Proceedings of ICASSP '89, Vol. 2*, pp. 703-706.
- Sakoe, H. & Chiba, S. 1978. "Dynamic Programming Optimization for Spoken Word Recognition." *IEEE Trans. Acoustics, Speech and Signal Processing, Vol. ASSP-34, No.1*, pp. 52-59.
- Steven, W., Zhan, P. & Gillick, L. 1999. "Progress in Broadcast News Transcription at Dragon Systems." *Proceedings of ICASSP '99, Vol. 1*, pp. 33-36.
- Zavaliagkos, G., etc., 1999. "Recent Experiments in Large Vocabulary Conversational Speech Recognition." *Proceedings of ICASSP '99, Vol. 1*, pp. 41-44.

접수일자: '99. 9. 19.

게재결정: '99. 10. 20.

▲ 송 민 석

강원도 강릉시 내곡동 522(우: 210-701)

관동대학교 문과대학 영어영문학과

Tel: (0391) 649-7218 (O)

Fax: (0391) 641-1010

e-mail: mssong@mail.kwandong.ac.kr

▲ 이 기 영

강원도 양양군 양양읍 임천리 산 7번지(우: 215-800)

관동대학교 이공대학 정보통신공학과

Tel: (0396) 670-3413 (O)

Fax: (0396) 672-9266

e-mail: kylee@mail.kwandong.ac.kr