

## 한국 표준어 연속음성에서의 억양구와 강세구 자동 검출\*

### Automatic Detection of Intonational and Accentual Phrases in Korean Standard Continuous Speech

이 기 영\*\* · 송 민 석\*\*\*  
Kiyoung Lee · Minsuck Song

#### ABSTRACT

This paper proposes an automatic detection method of intonational and accentual phrases in Korean standard continuous speech. We use the pause over 150 msec for detecting intonational phrases, and extract accentual phrases from the intonational phrases by analyzing syllables and pitch contours.

The speech data for the experiment are composed of seven male voices and two female voices which read the texts of the fable "the ant and the grasshopper" and a newspaper article "manmulsang" in normal speed and in Korean standard variation. The results of the experiment shows that the detection rate of intonational phrases is 95% on the average and that of accentual phrases is 73%.

This detection rate implies that we can segment the continuous speech into smaller units(i.e. prosodic phrases) by using the prosodic information and so the objects of speech recognition can narrow down to words or phrases in continuous speech.

**Keywords :** accentual phrase, continuous speech recognition, intonational phrase, segmentation of prosodic units

#### 1. 서 론

기계와 인간 사이의 가장 편리한 인터페이스로는 인간이 사용하는 자연어라는 점에는 모든 사람이 동의하고 있다. 그러나 인간의 언어는 워낙 다양하고 변이형이 많은 특징이 있어 아직 효과적으로 기계와 인간 사이의 인터페이스로 사용할 정도로 자연어를 처리하고 있는 시스템은 개발되지 못하고 있다. 그럼에도 아직 많은 연구들이 이러한 목적을 달

\* 본 연구는 한국과학재단 핵심연구(과제번호 981-0913-064-2)의 지원에 의하여 이루어 졌음.

\*\* 관동대학교 이공대학 정보통신공학과

\*\*\* 관동대학교 문과대학 영어영문학과

성하기 위하여 노력하고 있으며, 현재 개발된 인식시스템에 의하면 독립 단어를 이용한 인터페이스 개발은 상당한 진척을 이루고 있다.

이러한 음성인식 분야에서 현재 해결해야 하는 과제로 떠오르는 것은 문장 단위 이상의 연속음성 인식 분야이다. 문장단위 이상의 연속음성은 단독 발성된 고립단어 음성과 달리 지속시간, 억양, 크기 및 휴지기 등의 운율이라는 언어적 정보를 가지고 있다. 대화를 하거나 낭독을 하는 화자는 이 운율을 이용하여 의미, 감정, 의도 및 마음가짐 등을 전하며, 청취자는 화자가 발성한 음성으로부터 운율을 이용해 전체적인 의미를 파악한다[1, 2]. 또 하나 운율이 가지고 있는 중요한 정보는 발화된 문장의 문법적 경계에 관한 정보이다. 문자 언어에서는 이러한 정보가 띄어쓰기에 의한 공백에 의해 제공되나 소리가 연결되어 나타나는 문장 단위 이상의 발화에서는 운율에 의해 이 정보가 나타나게 된다.

본 연구의 목적은 발화된 문장 단위 이상의 연속음성에서 운율정보를 이용하여 문장보다 더 작은 단위를 자동으로 검출하는 방법을 제시하는 것이다. 문장단위 이하의 단위로 운율 특성에 근거한 강세구와 억양구를 사용하였으며[3], 가능성을 연구한 선행연구[4,5]를 바탕으로 필요한 음성 파라미터를 보충하고 실험 데이터를 신문 기사에서 발췌한 자료를 표준어 화자가 낭독한 음성으로 확장하여 본 연구의 결과를 테스트해 본다. 먼저 2절에서 한국어 강세구와 억양구의 운율적 특성에 대하여 논하며, 이를 바탕으로 3절에서 억양구와 강세구의 자동 검출 알고리즘을 제시하고, 4절에서 실험 및 결과에 대하여 논한다.

## 2. 한국어의 강세구와 억양구

사람이 사용하는 언어에서 문장보다 더 작은 단위로 생각할 수 있는 것은 우선 문법적 구 단위와 단어가 있다. 이것은 문장을 이루는 기본적인 단위로 이들에 의해 전체 문장의 구조가 형성된다. 문자 기반의 자연어처리 분야에서는 이러한 문법적 단위를 기준으로 문장을 분석하여 문장의 전체 구조를 파악한다. 그러나 이러한 문장들이 소리로 발화될 때는 사람의 호흡과 자연스럽게 조화되도록 문법적 단위와는 약간의 차이가 나는 새로운 단위로 재조정(readjustment)된다[1]. 따라서 음성 기반의 음성언어처리에서는 문법적 단위가 아닌 음향적 특성에 의해 나타나는 새로운 단위가 필요하다.

Nespor & Vogel(1986)은 운율의 특징과 음운규칙의 적용 범위 측면에서 전 세계 언어에 보편적으로 적용될 수 있는 운율 단위의 계층적 구조를 제안했다[6]. 이 구조는 음절(syllable), 음보(foot), 음운단어(phonological word), 접사 그룹(clitic group), 음운구(phonological phrase), 억양구(intonational phrase), 음운발화(phonological utterance) 등 일곱 개의 단위로 구성되어 있으며 이들을 나누는 기준은 문법적 기능이 아닌 순수한 음성, 음운적 측면을 고려한 것이다. 한편 Jun(1993)은 일곱 개 단위가 모든 언어에 공통적으로 나타나는 것이 아니라 각 언어마다 중요하게 작용하는 단위가 선택적으로 나타난다고 주장하고 한국어에는 강세구(accentual phrase)와 억양구(intonational phrase)가 중요하게 작용한다고 제시했으며[3], 정국 외(1996)에서는 한국어 낭독체 문장에서 그 타당성을 검증했다[7].

본 연구에서는 이러한 주장을 받아들여 한국어에 필요한 운율 단위로 강세구와 억양구를 설정하고, 이들의 음향, 운율적 특성을 분석하여 자연발화에서 자동으로 검출할 수 있는 시스템을 개발한다.

2.1 한국표준어의 억양구 특성

Jun(1993)에서 제시된 억양구(intonational phrase)는 Nespor & Vogel(1986)의 억양구에 해당한다. 이것은 더 하위 단위인 강세구(accentual phrase)의 연결로 이루어지며 마지막 부분에 의미 또는 문법적 기능에 따른 경계톤(boundary tone)이 나타난다.

하나의 억양구 내에서 나타나는 악센트구들은 전체적으로 낮아지는 피치 구조를 보이는 단계적 하향현상(down step)을 보임에 따라 억양구의 전체적인 피치 곡선이 점차 하향하는 모습으로 나타난다. 한 문장 내에 두 개 이상의 억양구가 나타날 경우 역시 두 번째 이후의 억양구가 갖는 피치 곡선이 첫 번째 억양구의 피치 곡선보다 낮은 특징을 보인다.

억양구 말의 경계톤은 의미에 좌우된다. 문장 말의 경우는 대략 평서문, 명령문, 의문문, 권유문의 유형에 따라 다른 경계톤이 나타나며 이러한 다른 경계톤을 이용하면 피치 곡선에 의해 문장의 종류를 운율 정보에 의해 구분할 수 있다[8,9].

2.2 한국 표준어의 강세구 특성

억양구를 형성하는 단위인 강세구는 주로 피치 궤적(f0 contour)에 의해 특징지어진다. 하나의 강세구가 세 음절 이하로 구성될 경우 주로 L H(low-high)의 피치 궤적으로 나타나며, 네 음절 이상일 경우 L H L H의 피치 궤적으로 나타난다[3,7].

하나의 억양구 내에서 여러 개의 강세구가 나타날 경우는 앞 소절에서 언급한 바와 같이 피치 궤적이 점차적으로 낮아지는 단계적 하향현상을 보인다. 또 억양구의 마지막 강세구에는 경계톤이 기본 강세구의 피치 궤적에 덮여짐으로써 기본 L (H L) H 피치 궤적이 아닌 경계톤의 피치 궤적으로 나타나게 된다.

이상을 종합하여 일반화된 억양구와 악센트구의 피치궤적을 나타내면 다음 그림과 같다.

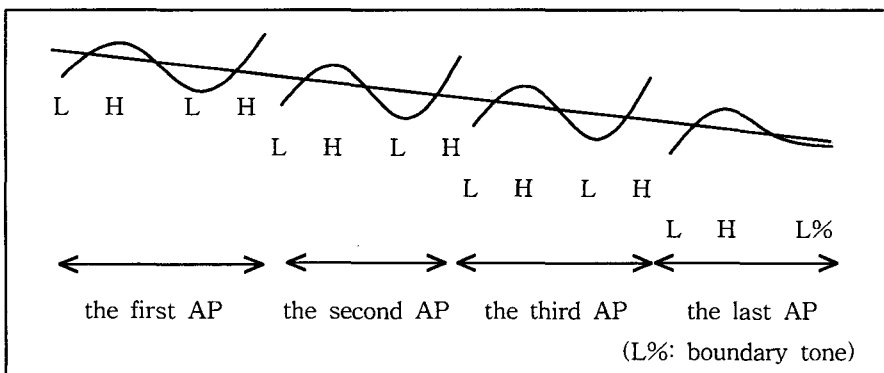


그림 1. 억양구와 악센트구의 기본적인 피치궤적

실제 발화에서는 이러한 일반화된 피치 궤적의 모습이 약간씩 변형된다. 변형을 가져오는 가장 주된 요인은 분절음이 갖는 고유의 피치 궤적에 대한 영향과 의미에 따른 변화를 들 수 있다. 먼저 한국어의 분절음 중에서 마찰음, 경음, 격음 등은 바로 뒤에 나타나는 모음의 피치를 변화시켜 다른 분절음들이 나타날 때보다 높게 변화시킨다. 따라서 표준어에서도 강세구의 기본 피치 궤적이 L (H L) H로 나타나는 것이 아니라 H (H) L로 나타나게 된다. 문장의 의미에 있어서는 특정한 강세구가 초점(focus)을 받을 경우 앞뒤의 피치 궤적보다 더 높은 피치 궤적을 갖게 된다.

이러한 특성을 고려하여 본 논문의 실제 음성 데이터 중에서 수작업으로 표시한 강세구의 예를 들어 보면 다음 그림과 같다.

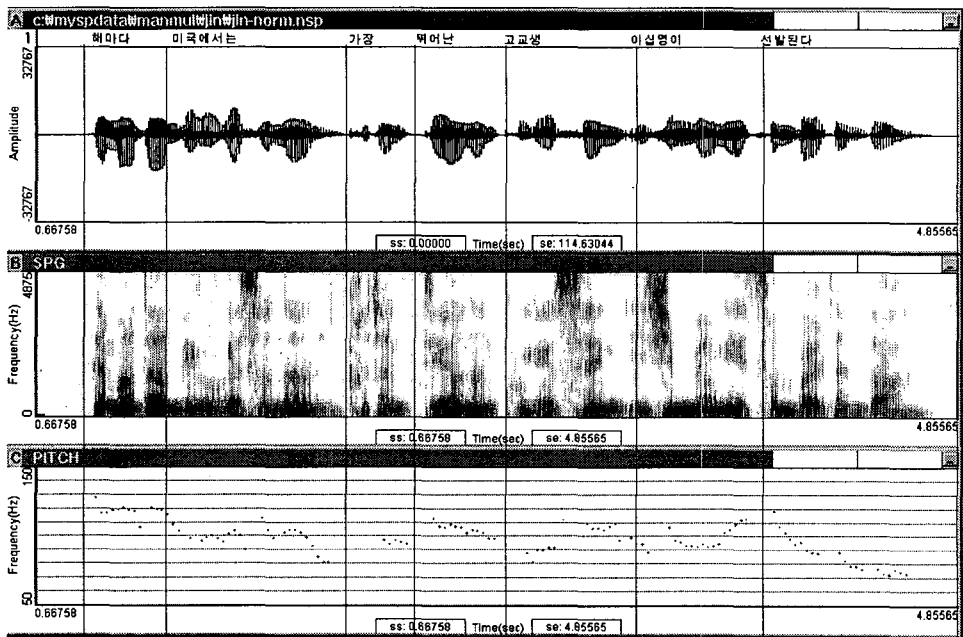


그림 2. “해마다 미국에서는 가장 뛰어난 고교생 이십 명이 선발된다”의 강세구 구성

### 3. 억양구와 강세구의 자동 검출

억양구 단위의 검출 알고리즘은 휴지기(pause)의 검출에 의해 이루어지며 대개의 경우 문장경계나 억양구 단위의 경계를 검출한다. 이전에 시행되었던 강세구 단위의 검출 알고리즘은 수작업에 의한 강세구 단위의 피치패턴과 DTW 알고리즘을 이용하여 미리 세그먼트된 억양구 안의 강세구 단위를 검출한 바 있다[4,5]. 그러나 운을 특징으로 표현되는 피치 궤적은 그 변화 범위가 광범위하기 때문에 패턴 매칭만으로는 경계를 추출하기에는 제한적이다. 따라서 본 연구에서는 우선 휴지기를 이용하여 연속 음성을 억양구로 나누고 억양구 단위에서 추출된 피치 궤적을 처리하여 L H 톤으로 나타나는 강세구의 경계를 검

출하며 문법적 어절과의 상관성을 높이기 위해 음절의 경계점을 참조하여 음절 경계와 강세구 경계점을 일치시켰다. 다음 그림 3은 무제한 연속 음성으로부터 억양구와 강세구를 검출하는 시스템 블록도이다.

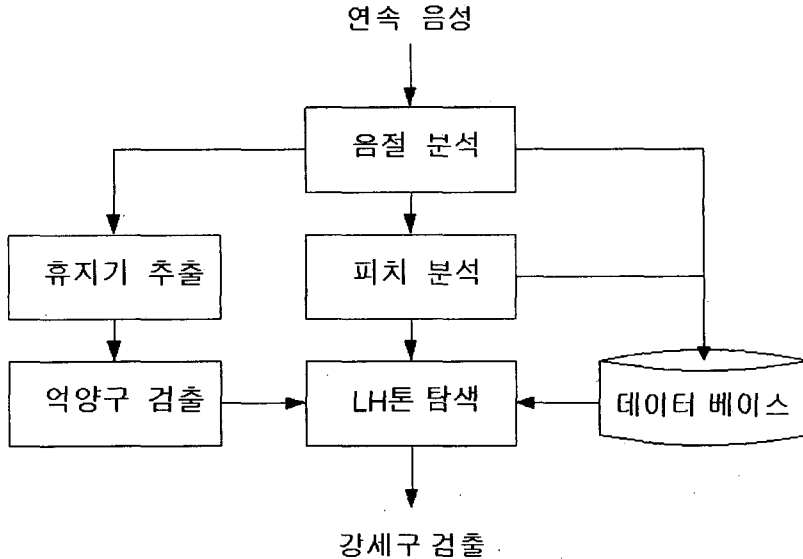


그림 3. 억양구와 강세구의 검출 시스템

### 3.1 음절 분석 특징 파라미터

그림 4에서는 연속 음성에서 음절의 경계점 검출을 위한 특징들과 피치 분석에 의한 강세구 검출의 예를 보이고 있다. 그림 4의 (a)는 연속 음성의 파형이고 (b)는 영교차율, (c)는 0-2500 Hz까지 파형의 peak-to-peak 궤적 및 핵과 DIP의 위치, (d)는 전방향과 후방향에서 본 프레임별 대수 우도비로, 전방향의 대수 우도비는 직선으로 또 후방향이 대수 우도비는 점선으로 구분하였다. (e)는 포만트(formant) 궤적, (f)는 피치 궤적, (g)는 피치 궤적의 기저선(base line), (h)는 기저선을 이용한 정규화 피치 궤적, (i)는 (h)의 미분 궤적이며 점선은 3.5 절에 기술할 미분 궤적을 이용한 강세구 경계의 위치, (j)는 음성 파형의 내용 “은주는 마음이 착한 여인이죠”이다.

#### (1) 영교차율

영교차율은 연속 음성의 마찰음을 검출하기 위한 것으로 에너지 특징만으로는 무시하기 쉬운 음절의 경계를 파악하기 위하여 추출한 것이다.

#### (2) 지역 통과 파형의 peak-to-peak 궤적[10]

음성의 전 주파수를 포함하는 특징인 에너지는 모음과 자음의 특성을 모두 포함하고 있기 때문에 보다 정확한 음절의 경계를 추출하기에 부적당하여 모음 위주의 주파수 대역 0-2500 Hz의 대역 통과 파형의 peak-to-peak를 각각 사용한다.

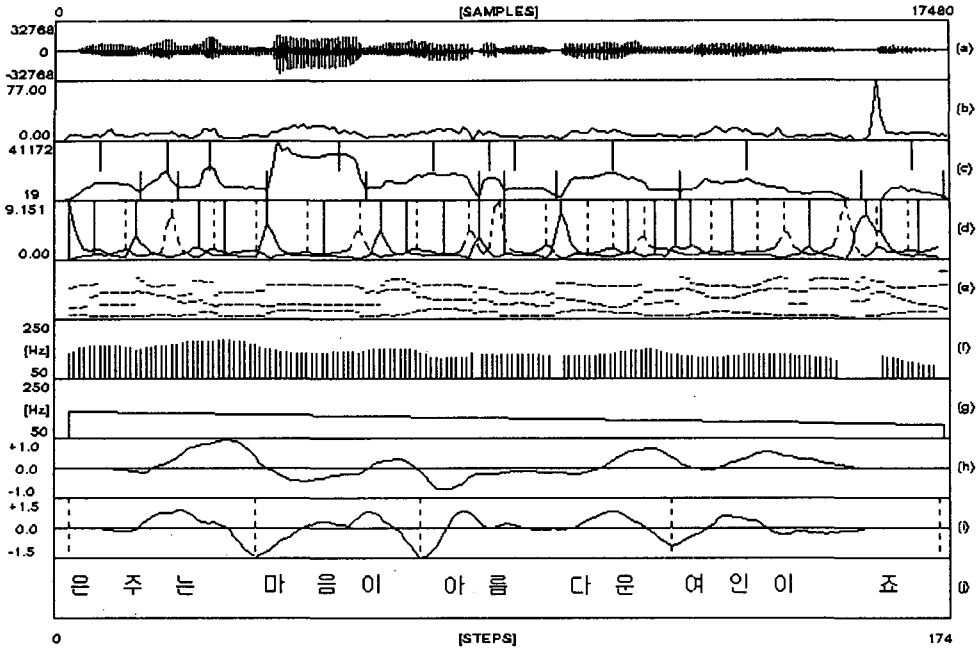


그림 4. 음절 분석과 피치 분석의 예

- (a) 음성 파형
- (b) 영교차율
- (c) 0-2500 Hz 대역 파형의 peak-to-peak 및 핵과 DIP의 위치
- (d) 전방향(직선)과 후방향(점선)에서 본 대수 우도비
- (e) 포먼트 궤적
- (f) 피치궤적
- (g) (f)의 기저선
- (h) 기저선을 이용한 정규화 피치 궤적
- (i) (h)의 미분 궤적
- (j) 음성 파형의 내용

(3) 전방향과 후방향에서 본 대수 우도비[11]

Itakura 왜곡인 대수 우도비[12]는 선형 예측 필터의 잔차 에너지의 비율로부터 얻을 수 있는 스펙트럼 왜곡의 측정 방법이다. 여기에서 대수 우도비는 연결한 프레임 사이에서 측정한다. 그러나 이 왜곡의 측정값은 예측된 필터에 입력되는 파형의 크기에 따라 달리 나타나기 때문에 전방향과 후방향에서 본 대수 우도비로 나누어 스펙트럼 왜곡을 측정하였다. 전방향과 후방향의 대수 우도비는 파열음의 파열 시작과 VOT 및 모음의 시작 부분을 분류하는데 사용한 바 있다[11]. 전방향의 대수 우도비는 선형 예측된 필터에 다음 프레임의 파형을 입력시켜 잔차 에너지의 비율을 구하므로 파형의 크기가 증가하는 부분에서의 스펙트럼 왜곡을 측정하기에 용이하며, 후방향의 대수 우도비는 이전 프레임의 파형을 입력시키기 때문에 파형의 크기가 감소하는 부분에서의 스펙트럼 왜곡을 측정하기에 용이하다.

(4) 포맷트 궤적

음성 파형이 포함하고 있는 포맷트 주파수는 주로 서로 다른 모음 사이에 다른 모양의 궤적으로 나타나며 포맷트 궤적이 불연속적으로 연결되는 부위가 음절의 경계인 DIP과 거의 일치하고 있다.

3.2 음절 경계점의 검출

음절 경계점의 검출은 불특정 화자에게 적용할 수 있는 방법이며, 특히 강세구의 경계점 문법적 어절과 정렬시키기 위하여 참조한다. 그림 5는 음절 경계점의 검출 과정을 보이고 있다.

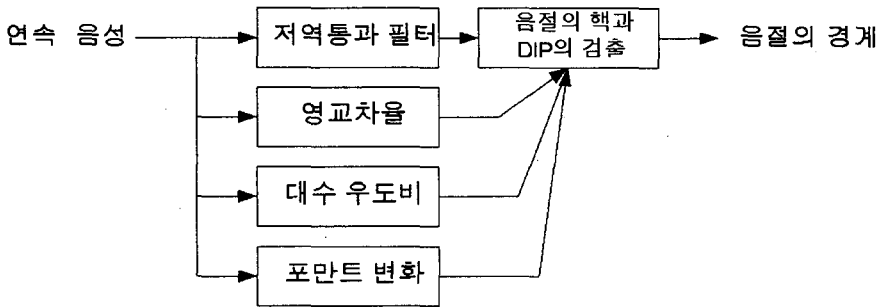


그림 5. 음절의 경계 검출 과정

(1) 음절 핵의 검출

한 음절은 반드시 모음을 포함하고 있기 때문에 해당 모음의 중심부에 음절의 핵이 있다. 이 부분을 검출하기 위해 그림 4의 (c)와 같이 모음 위주의 대역을 포함하는 0-2500 Hz의 peak-to-peak 궤적에서 피크를 이루는 곳을 음절 핵의 위치로 한다. 또한 연속 음성은 여러 개의 음절로 구성되며 음절과 음절 사이에 밸리(valley)가 존재한다. 따라서 피크를 검출하는 방법으로 한 피크를 구하면 그에 따라 다음 피크와의 사이에 밸리가 반드시 존재하며 마지막 음절에서는 피크와 그 음절의 밸리가 음성의 끝점과 일치한다. 단 두 음절이 연결되어 있는 부분의 핵의 위치는 후반부 음절의 피크를 이루는 곳을 핵의 위치로 하며 영교차율이 큰 부분은 한국어의 초성 마찰음으로 간주하여 무조건 음절이 시작되는 DIP으로 한다.

(2) 후보 DIP의 검출

이상의 음절 핵의 검출 과정에서 구한 밸리들은 꼭 DIP과 일치하는 것은 아니다. 공명음(sonorant)이 연결되는 음절에서는 모음과 공명음 사이에 밸리가 검출되기도 하기 때문이다. 따라서 검출된 피크 다음의 밸리를 후보 DIP이라 명명한다. 이 후보 DIP들을 검출하는 규칙은 다음과 같다.

- ① 음절 핵에서 피크의 70% 이하이며 동시에 이전 피크에서 밸리까지의 10% 이상 상승한다.

- ② 전방향이나 후방향 대수 우도비가 큰 위치이다.
- ③ 포만트 궤적의 불연속 위치이다.

### (3) 음절 경계점의 검출

음절 경계점은 이상에서 구한 후보 DIP들 중에서 결정한다. 음절 경계점을 검출하는 규칙은 다음과 같다.

- ① 연결한 피크사이에 후보 DIP의 크기가 양쪽의 어느 한 쪽의 피크보다 40% 이하로 떨어지면 DIP으로 한다.
- ② 한 피크 값이 전후의 각 피크 값보다 70% 이하로 떨어지면 후보 공명음으로 한다.
- ③ 공명음의 양쪽 후보 DIP 중에서 스펙트럼 왜곡이 작은 쪽을 DIP으로 한다.
- ④ 모음이 연쇄적으로 발생되는 부분은 0-2500 Hz의 peak-to-peak에서 밸리가 없지만 전방향과 후방향의 대수 우도비의 변화나 포만트 궤적의 불연속으로 나타나는 후보 DIP을 참조하여 음절의 경계점으로 한다.

### 3.3 휴지기에 의한 억양구의 검출

자연 발화에서는 호흡의 길이나 문법적, 의미적 단위의 경계점을 전달할 목적으로 자연스런 휴지기가 나타난다. 이 휴지기는 억양구의 경계가 되며, 이를 연속 음성의 인식에 이용하면 인식 성능의 개선과 계산량의 감소에 효과적이다.

언어학적 관점의 억양구 단위 모두에 휴지기가 뒤따르는 것은 아니다. 휴지기가 뒤따르지 않는 억양구의 경우는 실제 발화에서 대개 피치 패턴의 변화와 마지막 음절의 길이 조절에 의해 억양구 단위의 경계를 표시하게 된다.

이러한 휴지기를 동반하지 않는 억양구의 경계는 본 연구에서 대부분 3.2절과 3.4절에서 서술한 알고리즘에 의해 강세구로 검출된다. 그러나 본 연구의 주된 목적은 실제 발화에서 문장 단위 이하의 더 작은 단위를 검출하는 것을 목적으로 하며, 언어학적 강세구와 억양구가 본 연구의 강세구와 억양구에 정확히 일치되지 않는다 하더라도 일단 단위의 경계점으로 인식되지만 한다면 추후 자연어 처리의 문장 구조 분석 과정에서 문법 단위의 종류를 구분해낼 수 있으므로 문장의 의미를 파악하는 데는 지장이 없을 것이다. 따라서 본 연구에서는 억양구를 자동 검출하기 위하여 에너지 특징으로 나타나는 휴지기만을 이용하며 나머지는 모두 강세구 검출 알고리즘으로 처리한다.

휴지기는 음절 경계의 검출 과정에서 추출되는 DIP의 위치 중에서 검출된다. 그러나 음성의 묵음 구간은 문장의 경계나 억양구 단위의 경계에도 있지만 초성 파열음의 시작 부분과 종성 파열음의 끝 부분에 나타나는 요소이기도 하다. 따라서, 문장 단위나 억양구 단위 내에서 파열음의 존재로 인한 묵음구간과 휴지기를 구분해줄 필요가 있다. 예를 들면, 앞선 음절의 종성이 파열음이거나 현재 음절의 초성이 파열음인 경우 억양구 단위의 경계로서의 휴지기로 혼동되는 경우가 있다. 억양구 단위의 경계로 나타나는 휴지기는 파열음의 묵음 구간보다 대부분 훨씬 길게 나타나므로 이를 구별하기 위하여 파열음이 갖는 묵음 구간의 두 배 길이를 억양구 단위 경계로서의 휴지기 최소 길이로 하여 문턱값을 정하면 억양구 단위 경계의 휴지기를 대부분 검출할 수 있다.



### 3.4 피치 분석

하나의 억양구 단위 내에서의 피치 궤적은 시간이 지날수록 점점 낮아지며 굴곡의 폭도 차츰 좁아진다. 또한 하나의 억양구는 여러 개의 강세구로 나뉘어지기 때문에 피치 궤적으로 나타나는 강세구의 특징을 보다 분명히 하기 위해서는 피치 궤적의 정규화가 필요하다. 그림 4의 (f)는 음성 파형에 대한 피치 궤적, (h)는 시간에 따른 피치의 최소값으로 구성된 기저선(baseline), (i)는 기저선을 기준으로 한 정규화된 피치 궤적, (e)는 정규화된 피치 궤적의 미분 궤적을 보이고 있다. 이 기저선은 거의 일직선 상에 존재하나 부방향의 기울기를 가진다. 이 기저선은 발생자나 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  $p(T)$ 는 음성이 끝나는 시간에서의 피치 주파수이며,  $T$ 는 음성의 시작점에서 끝점까지의 총시간 길이이다. 식(1)을 이용하여 제안한 정규화된 피치 궤적을 식으로 나타내면 다음과 같다.

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

여기서, 식(2)을 최대치로 나누어 크기를 정규화 한다. 또한 음성의 연속되는 L H 톤의 경계점을 검출하기 위하여 식(2)을 미분하면 다음과 같다.

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

### 3.5 강세구의 검출

강세구는 2.2 절에서 설명한 바와 같이 L H 또는 L H L H 톤의 피치 패턴을 갖는다. 그러나 이러한 피치 패턴은 방언마다 다르며, 따라서 강세구를 표시하기 위한 표준 패턴도 방언마다 달라야 한다. 본 연구에서는 실험 대상을 표준어로 한정한다.

강세구를 자동 검출하기 위해서는 먼저 3.3 절에서 설명한 바와 같이 입력된 연속 음성의 휴지기를 이용하여 억양구로 분류한다. 또한 각 억양구의 피치 궤적과 식(1)을 이용한 피치 궤적의 기저선에 의해 식(2)와 같이 정규화하고 정규화한 피치 궤적의 미분값을 이용하여 강세구의 후보 경계를 추출한다. 한국어 표준어에 있어서 강세구 단위의 마지막 부분에 나타나는 H 톤과 다음 강세구의 L 톤 사이의 시간적 거리가 매우 가깝고 급격한 단절이 생기는 특성이 있으므로 정규화한 피치 궤적의 미분값은 그림 4 (i)에서 점선으로 표시한 부분과 같이 강세구의 단위마다 급격한 하향곡선과 상향곡선으로 이루어지는 계곡점이 나타난다. 따라서 이 부분에서 최소점의 위치를 강세구 단위의 후보 경계로 하고 문법적 어절 단위와 운율적 강세구 단위는 서로 매우 밀접한 관계가 있으므로 그림 5의

과정으로 추출된 음절 경계점에서 가장 가까운 강세구의 후보 경계를 강세구의 경계로 한다.

#### 4. 실험 및 결과

##### 4.1 음성 데이터

본 실험에서 사용한 텍스트는 33개의 어절을 포함하는 일곱 개 문장으로 구성된 이숙우화 “개미와 베짖이,” 그리고 21개 문장에 약 120개의 문법적 어절을 포함하는 조선일보의 “만물상” 자료이다. “개미와 베짖이”의 경우 5명의 표준어 사용 남성 화자가 보통 속도로 낭독한 것과, “만물상”의 경우 2명의 남성과 2명의 여성이 각각 보통 속도와 빠른 속도로 낭독한 음성을 카세트 테이프를 녹음한 후, KAY사의 Multi-speech를 이용하여 16 KHz의 샘플링레이트(sampling rate)로 AD하여 실험용 데이터로 사용하였다. 음성분석을 위한 프레임(frame)의 간격은 25.6 msec로, 이동간격은 10 msec로 하였다.

이렇게 하여 실험에 사용된 데이터는 남녀 총 9명의 표준어 화자가 낭독한 1,330개의 문법적 어절을 포함하고 있는 234개의 문장 음성 데이터이다. 실험 결과는 다음과 같이 연속 음성의 억양구의 검출률과, 강세구의 검출률을 비교, 검토하였다.

##### 4.2 실험 결과 및 고찰

###### (1) 억양구의 검출 결과

3.3 절에서 기술한 바와 같이 억양구 단위의 경계로 나타나는 휴지기를 파열음에서의 묵음구간과 분리하기 위하여 약 150 msec를 억양구 단위 경계로서의 최소 휴지기 길이의 문턱값으로 하였다. 또한 휴지기의 문턱 에너지는 최대 에너지와 배경잡음의 에너지의 중간 값으로 하였다.

그림 6은 연속 음성에서 억양구를 검출한 예를 보이고 있으며, 그림 7은 본 실험의 9명 화자가 낭독한 연속 음성 데이터에서 억양구를 검출한 결과이다. 결과는 평균 95.36%의 검출률을 보이고 있다. 여기서 잡음의 에너지가 문턱 에너지를 초과하거나 낭독 중에 화자가 호흡이나 무의식적으로 끊어 읽음으로 해서 나타나는 휴지기 때문에 오류로 검출된 억양구는 오검출된 것으로 하였다.

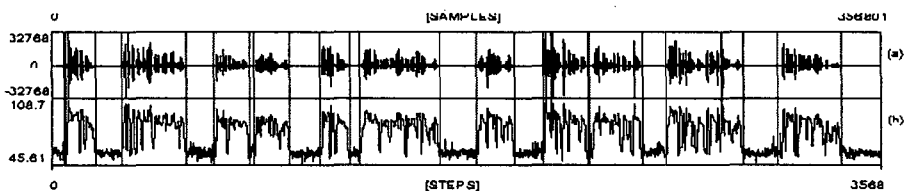


그림 6. 휴지기를 이용한 억양구의 검출 예

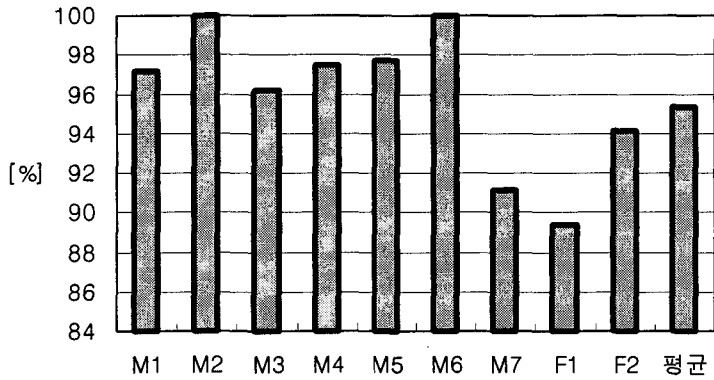


그림 7. 억양구의 검출률

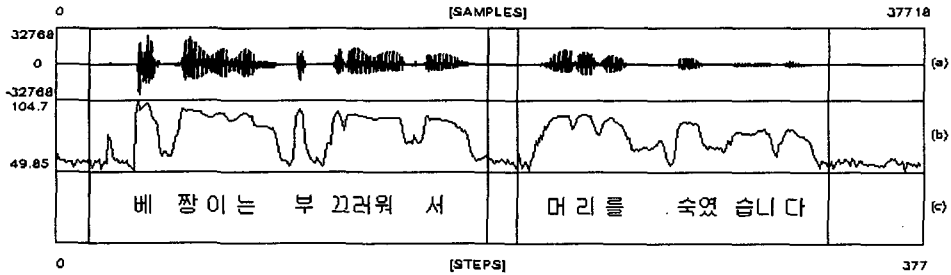


그림 8. 한 문장 내의 두 억양구

그림 8에서는 에너지 특징을 이용하여 한 문장의 음성을 두 개의 억양구로 분리한 예를 보이고 있다. (a)는 “베짜이는 부끄러워서 머리를 숙였습니다.”라는 한 문장의 음성 파형이고, (b)는 이 음성 파형의 에너지, (c)는 음성 파형에 대한 내용을 보이고 있다. 여기서 “베짜이는 부끄러워서”와 “머리를 숙였습니다” 사이에 150 msec 이상의 휴지기가 나타나므로 억양구로 나눌 수 있다. 따라서 하나의 문장이라도 두 개 이상의 억양구로 나뉘어 질 수 있음을 알 수 있다.

(2) 강세구 자동 검출 결과

다음 그림 9는 일곱 명의 남성(M1-M7)과 두 명의 여성(F1, F2)이 낭독한 음성 데이터로부터 강세구를 검출한 결과를 검출률과 삽입률로 나누어 보이고 있다. 검출률은 낭독 텍스트의 문법적 어절을 기준으로 강세구의 자동 검출을 수행하였을 때 정검출된 결과와의 비율이며, 삽입률은 하나의 강세구를 두 개 이상의 강세구로 오류 검출된 강세구의 비율이다. 그 결과 평균 73.4%의 검출률과 평균 32.4%의 삽입률을 얻을 수 있었다.

강세구의 검출률이 억양구에 비해 비교적 낮은 것은 2.2 절에서 언급한 바와 같이 강세구 톤이 모두 L H 톤의 형태가 되는 것이 아니라 앞 뒤 자음의 영향과 의미적인 초점에 따라 다른 형태로 변형되어 낭독하였기 때문이다. 또한 L H L H 톤을 갖는 네 음절 이상으로 구성된 강세구에서 앞쪽의 L H 톤을 하나의 강세구로 잘못 추출된 것으로 인한

삽입률도 적지 않은 것을 알 수 있었다.

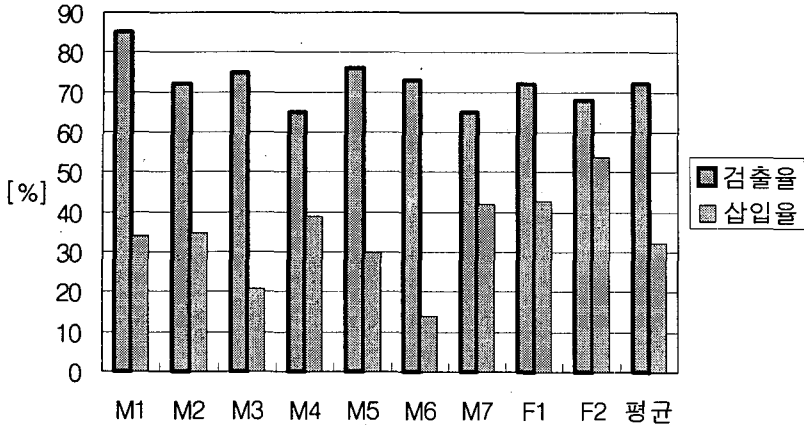


그림 9. 강세구의 검출률과 삽입률의 비교

(3) 낭독 속도와 남·여 간의 비교

두 명의 남성(M1, M2)과 두 명의 여성(F1, F2)이 동일한 텍스트 “만물상”을 각각 보통 속도와 빠른 속도로 낭독한 음성 데이터를 대상으로 발생 속도에 따라 강세구의 검출률을 비교하였다. 보통 속도로는 “만물상”의 전체의 낭독시간이 남녀 평균 168.01초이며, 빠른 속도로는 남녀 평균 137.22초가 소요되었다.

그림 10은 보통 속도와 빠른 속도로 낭독한 연속 음성으로부터 강세구를 검출한 결과를 비교하고 있다. 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 있다. 또한 남성의 경우는 빠른 속도에 비해 보통 속도에서 검출률이 높은 데 비해 여성의 경우는 오히려 빠른 속도에서 검출률이 높은 것으로 나타나고 있다.

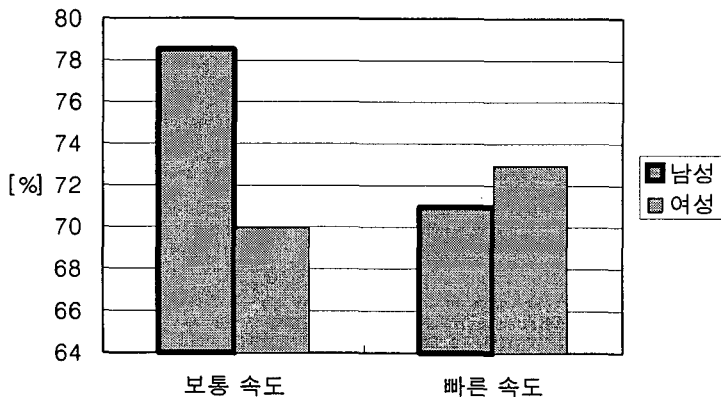


그림 10. 낭독 속도에 따른 강세구의 검출률 비교

#### (4) 음절 경계와의 비교

L H 톤에 의해 검출된 강세구의 경계는 각 강세구의 마지막에 위치한 음절의 경계와 일치하지는 않았다. 더구나 피치 궤적에 의해 검출한 강세구의 경계는 이 마지막 음절의 경계와 대부분 일치하지 않았으므로 음절 분석 결과로 나타나는 음절 경계를 참조하여 가까운 위치로 강세구의 위치를 정렬시켜 주었다. 또한 피치 분석에 의해 자동 검출된 강세구의 경계와 음절 분석 과정에서 얻은 음절 경계와의 오차 시간들을 구하여 그 절대값들의 통계를 구하였다. 이 오차 시간들의 평균은 약 25.3 msec, 표준편차는 약 10.8 msec로 정확히 일치하지 않고 있음을 알 수 있다.

이를 개선하기 위해서는 본 연구에서 제안한 강세구의 자동 검출 방법과 보다 향상된 음절 경계의 추출 방법을 조합하여 강세구 경계의 정확성을 보완해 줄 필요가 있는 것으로 사료된다.<sup>1)</sup>

### 5. 결 론

이 연구에서는 한국어 연속 음성에서 운율 단위인 억양구와 강세구를 자동 검출하는 방법을 제안하였다. 억양구를 검출하기 위해서는 입력된 문장 단위 이상의 음성 데이터로부터 150 msec 이상의 길이를 갖는 휴지기를 추출하여 그 경계를 검출하며, 음절 분석과 피치 분석을 이용하여 억양구에 포함되어 있는 강세구를 검출한다.

실험에 사용한 자료는 이숙우화 “개미와 베짚이”와 조선일보의 “만물상”으로 하였으며 표준어를 사용하는 남성 7명과 여성 2명이 보통 속도로 낭독하도록 하여 억양구와 강세구를 검출하였다. 그 결과 억양구는 평균 95%의 검출률을 보였으며, 한 문장 내에 두 개 이상의 억양구가 포함된 경우도 정확히 검출되었다.

강세구는 평균 73%의 검출률을 보였으며 문법적 어절에 해당하지 않는 오류 강세구의 검출률인 삽입률은 평균 32%였다. 이와 같이 강세구의 검출률이 억양구의 검출률에 비해 비교적 낮은 것은 강세구의 톤 형태가 모두 L H 톤의 형태가 아니라 앞 뒤 자음의 영향과 의미적으로 초점인 경우 등과 같이 기본 형태에서 변형된 형태로 나타나기 때문이며, 강세구의 삽입률은 어절의 음절 수에 따라 네음절 이상으로 구성된 L H L H 톤의 강세구에서 앞의 L H 톤을 하나의 강세구로 검출하였기 때문이었다. 그러나 전체적으로 70% 이상의 강세구 검출률을 보이는 것은 강세구 인식에서 상당히 의미있는 인식률로 보이며, 운율 정보를 이용하여 강세구를 검출할 수 있는 가능성이 충분히 있는 것으로 사료된다.

1) 이상의 실험 결과들이 실제 인간의 강세구와 억양구의 인식 능력과 비교되기 위해서는 위에서 사용한 음성 데이터에 대한 청취 실험 결과와의 비교 검토가 필요하다. 그러나 본 연구는 자연 발화된 문장의 기계에 의한 자동적 이해 시스템 개발의 일부분에 해당하는 연구로서 강세구와 억양구의 타당성에 대한 언어학적 측면의 연구가 아닌 언어학적 연구 결과가 공학적으로 적용 가능한가에 대한 연구이므로 문법적 구조에 의한 단위를 근거로 강세구 경계를 설정했으며 청취 실험을 생략하였다. “만물상” 자료에 대한 청취 실험은 정국 외(1996)에서 같은 자료로 이미 수행한 바가 있으므로[7], 이 문헌을 참고할 것.

특히 앞으로의 연구에서 자음이 앞 뒤 모음의 피치에 미치는 영향과 의미적인 초점이 갖는 피치 패턴의 영향을 고려할 경우 상당히 높은 강세구 검출률을 얻을 수 있을 것으로 기대된다.

이 연구에서 제안한 억양구 및 강세구의 자동 검출 방법을 문장 단위 이상의 연속 음성의 인식 과정에서 전처리 작업으로 사용한다면 억양구의 경계 검출에 의해 인식량의 범위를 감소시키고 특히 문법적 어절과 거의 일치하는 강세구의 검출에 의해 문장 음성을 문법적 구조로 분석하는데 용이할 것으로 사료된다.

### 참 고 문 헌

- [1] E. O. Selkirk. 1984. *Phonology and Syntax*. Massachusetts: The MIT Press.
- [2] M. Beckman and J. Pierrehumbert. 1986. "Intonational structure in Japanese and English," *Phonology Yearbook 3*, pp. 255-309.
- [3] Jun, Sun-Ah. 1993. *The Phonetics and Phonology of Korean Prosody*. Ph.D. Dissertation. The Ohio State University.
- [4] Lee, Kiyong and Minsuck Song. 1997. "Automatic Segmentation of Korean Prosodic Phrases." *Proceedings of ICSP '97, Vol. 2*, pp. 747-751.
- [5] Lee, Kiyong and Minsuck Song. 1999. "Automatic detection of Korean Accentual Phrase Boundaries." *The Journal of the Acoustic Society of Korea, Vol. 18, No. 1E*, pp. 27-31.
- [6] M. Nespor and I. Vogel. 1986. *Prosodic Phonology*. Dordrecht: Foris Publications.
- [7] 정국 외. 1996. 음성 합성/인식 시스템 개발을 위한 한국어 운율 구조 및 담화 구조 연구. 한국통신 연구개발단 연구보고서.
- [8] 이기영 외. 1995. "한국어 중의성 문장을 구분하기 위한 억양패턴의 자동인식." *한국음향학회 학술발표대회 논문집, 14권 1(s)호*, pp. 219-222.
- [9] Tae-Yeoub Jang, Minsuck Song, Kiyeong Lee. 1998. "Disambiguation of Korean Utterances Using Automatic Intonation Recognition." *Proceedings of ICSLP '98*.
- [10] Cole, R.A. and Lily Hou. 1988. "Segmentation and Broad Classification of Continuous Speech." *Proceedings of ICASSP'88, Vol. 1*, pp. 453-456.
- [11] 이기영, 배철수, 최갑석. 1988. "Likelihood ratio를 이용한 음소분류에 관한 연구." *한국음향학회지, 7권 5호*, pp. 50-54.
- [12] Gray, A. H. and J. D. Markel. 1976. "Distance Measures for Speech Processing." *IEEE Trans. Acoustics, Speech, Signal Proc., ASSP-24(5)*.

접수일자: 2000. 5. 21.  
게재결정: 2000. 6. 03.

▲ 이 기 영

강원도 양양군 양양읍 임천리 산 7번지  
관동대학교 이공대학 정보통신공학과(우: 215-800)  
Tel: (0396) 670-3413 (O), Fax: (0396) 672-9266  
e-mail: kylee@mail.kwandong.ac.kr

▲ 송 민 석

강원도 강릉시 내곡동 522  
관동대학교 문과대학 영어영문학과(우: 210-701)  
Tel: (0391)649-7218 (O), Fax: (0391)641-1010  
e-mail: mssong@mail.kwandong.ac.kr

## 부 록 : 실험 자료

## 1. 개미와 베짚이

햇볕이 쨍쨍 내리쬐는 여름날입니다. 개미들이 땀을 흘리며 부지런히 일하고 있습니다. 그러나 베짚이는 시원한 나무그늘에서 노래를 부르며 놀고 있습니다. 추운 겨울이 되었습니다. 개미들이 따뜻한 방 안에서 맛있는 음식을 먹고 있을 때 추위와 배고픔에 지친 베짚이가 찾아 왔습니다. 베짚이는 부끄러워서 머리를 숙였습니다.

## 2. 조선일보 만물상(1993년 5월 27일자)

해마다 미국에서는 가장 뛰어난 고교생 이십 명이 선발된다. 상금은 약 이백만 원밖에 안된다. 그러나 전국의 고등학교를 대표한 구백 명 가까운 후보 중에서 선발된 이들은 미국의 내일을 이끌어 나갈 자랑스런 청소년들이다. 이들은 물론 전 과목에서 A 학점을 받을 만큼 학교 공부도 잘 한다. 그러나 그들은 과학이든 예술이든 어느 한 분야에서 뛰어난 재능이 따로 있어야 한다. 그 뿐 아니라 남을 이끌어 갈 수 있는 리더쉽도 있어야 한다. 사회 봉사도 정신도 철저해야 한다. “우리는 집단과 팀워크의 시대를 살고 있다.” 이렇게 한 심사위원은 말했다. 한 여학생은 매일 다섯 시간씩 피겨 스케이트를 연습하는 운동 선수이기도 하다. 그러면서도 동네 철물점에서 산 도구를 가지고 인조 다이아몬드를 만들어 내는 데 성공했다. 동시에 이웃 어린이들의 과학을 개인지도하고 있다. 또 다른 학생은 간장병의 새 치료법을 발명했다. 에이즈에 걸린 어린이들을 위로하기 위해 동물원에서 큰 잔치를 주최한 학생도 있다. 한 학생은 또 십이 세기 음악 양식을 써서 본격적인 뮤지컬을 작곡했다. 이년 전 어선을 타고 월남에서 망명해 온 여학생은 미술상을 받은 바도 있다. 그러면서 마을 사회단체에서 자원봉사하고 있다. 지난 몇 해 동안 한국계 학생들도 자랑스럽게 두어 명씩 으레 끼이기도 했다. 그 만큼 미국은 다양하다. 그게 미국의 힘이 되는 것이기도 하다. 엇그제 대한민국 학생 발명전에서 한 초등학교 육학년생이 잔디깎는 기계로 대통령상을 받았다. 대견스러운 일이다. 그러나 것처럼 뛰어난 능력을 우리네 학교 교육이 앞으로 과연 얼마나 더 길러줄 수 있는 지 딱한 생각이 든다.