

운율구 단위의 연속음 인식

The Continuous Speech Recognition with Prosodic Phrase Unit

강 지 영*, 엄 기 완*, 김 진 영*, 최 승 호**,
(Ji Young Kang*, Ki Wan Eom*, Jin young Kim*, Seong ho Choi**)

* 한국과학재단 지정 전남대학교 고품질전기전자부품 및 시스템 연구센터의 연구비 지원에 의해 연구되었음.

요 약

일반적으로 사람은 말을 할 때 어절들은 몇몇의 구로 그룹핑하여 발음함으로써 발화한다. 이것은 듣는 사람으로 하여금 발화의 의미와 의도를 잘 파악하도록 도와준다. 특히, 이러한 목적으로 발화자는 무의식적으로 운율정보(억양, 장단, 리듬 등)를 적절히 사용하게 된다.

본 논문에서는 발화된 문장에서 운율경계를 인식의 단위로 하는 음성인식방법에 대하여 제안한다. 즉, 발화된 문장을 운율구단위로 나누는 방법을 제안하고 나누어진 단위에 따라 연속음 인식실험을 수행하였다. 인식실험결과 연속음인식 시간의 감소를 관찰할 수 있었으며, 물론 음성인식률도 20-10%정도 증가하였다.

ABSTRACT

Generally, a speaker structures utterances very clearly by grouping words into phrases. This facilitates the listener's recovery of the meaning of the utterance and the speaker's intention.

To this purpose, a speaker uses, among other things, prosodic information such as intonation, pause, duration, intensity, etc. The research described here is concerned with the relationship between the strength of prosodic boundaries in spoken utterances as perceived by untrained listeners(Perceptual boundary strength, PBS)-In this paper, the preceptual boundary strength is used as the same meaning of the prosodic boundary strength-and prosodic information. We made a rule determining the prosodic boundaries and verified the usefulness of the prosodic phrase as a recognition unit.

Experiments results showed that the performance of speech recognition(SR) is improved in aspect of recognition rate and time compared with that using sentences as recognition unit.

In the future we will suggest the methods that estimate more appropriate boundaries and study more various methods of prosody assisted SR.

I. 서 론

음성인식기술은 1970년 초부터 연구되어 왔지만 현재 까지 사람의 음성을 정확히 인식할 수 있는 시스템은 개발되어 있지 않다. 지금까지 개발된 시스템은 다음과 같은 네가지의 제약성을 가지고 부분적으로 음성을 어느정도 인식하고 있는 실정이다. 첫 번째로 화자 독립성, 두 번째로 연속음성 인식, 세 번째로 대용량 단어 인식시스템의 구성, 마지막으로 무제한 문법(Unconstrained Grammar)이 바로 그것인데, 본 논문에서는 음성인식의 제약성중의 하나인 무제한 문법을 세우기에 앞서 인식시 계산량을 줄

일 수 있고, 단위 문법의 생성 단위가 될 수가 있는 운율구 단위의 인식을 제안하고, 인식실험을 통해서 효율성을 입증하였다.

그러면 운율구 단위 인식의 유용성을 인식적인 측면과 화자적인 측면에서 살펴보겠다. 먼저 인식적인 측면에서 보면, 음성인식 시스템은 또한 언어 종속적이기 때문에 한국어의 특수성이 고려되어야 하는 것은 당연할 것이다. 연속음성 인식시 굴절어인 영어의 경우 단어를 인식대상으로 삼지만, 교착어인 우리말의 경우 구를 기본 인식대상으로 삼아야 한다. 그러므로 어휘의 규모가 엄청나게 증가한다. 따라서 언어처리에서 문장 내에서의 단어(혹은 구)의 위치에 따라 인식대상의 단어(또는 구)를 한정함으로써 인식시 계산량을 줄이고 인식률을 높일 수 있을 것이다. 따라서, 본 논문에서는 운율구(prosodic phrase)단위

* 전남대학교 전자공학과
** 동신대학교 정보통신공학과
접수일자: 1999년 1월 18일

의 음성인식의 기초연구로서 운율구 경계를 찾는 방법을 제안했는데, 운율구 단위는 주로 운율경계강도가 커서 운율정보에 의하여 결정하는 것이 가능하기 때문이다.

그리고 화자적인 측면을 살펴보면, 일반적으로 대화 속에서 화자는 의미를 표현하기 위해서 단어들 이 묶일 수 있다는 정도를 나타내거나 다른 차별된 견해를 가져올 수 있는 초 분절적인 수단(suprasegmental mean)을 이용한다. 즉, 화자는 휴지기(pause)와 speech melody, 그리고 임시적인 단어들의 조직화에 의해서 단어들 또는 단어의 그룹들 사이의 경계를 강조한다. 그리고 청원자는 구어적인 메시지의 구조를 이해하는데 도움을 주는 단어들 사이의 경계들을 차례대로 인식하게 된다. 일반적으로 초 분절적이거나 또는 운율적인 특징의 적절한 사용은 발화의 의미와 화자의 의도를 이해하는데 도움을 주고 의사소통의 과정을 용이하게 한다. 본 논문에서는 이러한 운율의 의사 소통적인 기능측면에서의 발화문의 초 분절적인 구조화에 주된 관심을 가지고, 인식의 기본 단위로써 운율경계를 도입하였다.

본 논문에서는 운율구 단위의 음성인식을 위한 방안으로서 다음과 같은 방안을 제시한다.

- 첫째, 피치정보와 휴지기 정보를 사용하여 운율구의 개수와 경계를 대략 정한다.
- 둘째, 시간에 따른 에너지 정보와 스펙트럼 변화를 계산하여 이 정보를 바탕으로 운율구 경계를 교정한다.

위의 방안에서 피치정보와 휴지기 정보를 사용하는 이유는 이들이 운율의 불연속 점들을 잘 반영하기 때문이다. 운율구 경계를 찾기 위한 기본 아이디어는 구문론적, 음운론적인 휴지기의 실험적인 분석과 피치정보의 분석을 통한 임계값을 정하는 것이다.

우리가 살펴본 운율구 단위의 음성인식의 전처리구조는 다음 [그림 1.1]과 같이 표현할 수 있다. 그림에서 휴지기 정보와 피치정보에 의한 운율구 경계 예측시 발생한 경계지점의 오차는 에너지 윤곽선과 스펙트럼 변화도를 추출하여 교정되어지는 것을 보여주고 있다.

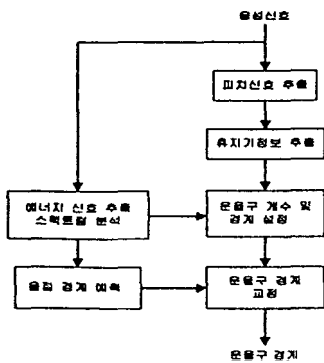


그림 1.1. 운율구 예측 모형
Fig. 1.1. Prosodic phrase prediction model.

그리고 본론에서는 운율구 단위 음성인식이 가져오는 효과를 두가지 측면에서 검증하였다. 인식 시간과 인식률이 바로 그것이다. 본 실험에서는 HMM(Hidden Markov Model)을 통한 학습을 통해서 다중화자의 발음을 모델링 한 후에 임의의 테스트문장을 선정해서 인식률을 체크하였다.

II. 운율정보 추출방법

2.1 피치(pitch)신호 추출방법

본 연구에서 사용한 Pitch추출 방법은 LPC inverse filtering와 AMDF(Average Magnitude Difference Function)에 기초하는 알고리즘이다[9].

2.2 음절 경계정보 추출방법

운율구 경계 바로 앞음절 내에서 피치의 변화율이 급격히 변화하는 모습을 관찰 할 수 있다.

음절의 경계정보를 위해서는 음소간의 경계를 알아내야 하는데, 음소단위의 경계를 정하는 것은 쉽지 않은 작업이다. 음절의 경계를 정하는 방법으로 에너지 윤곽선 정보(fixed frame방법, Pitch간격에서 최대값을 이용하는 방법)와 스펙트럼 변화율에 따른 정보를 이용하였다.

▶ 에너지 윤곽선 정보

에너지 윤곽선 정보를 추출하기 위한 첫 번째 방법으로 fixed frame방법을 보면, 음성신호를 일정한 frame으로 자르고 그 프레임 내에서 에너지 분포 곡선을 그리고 그 분포 곡선으로부터 음절의 경계를 예측하는 방법인데, 이 방법은 경계 예측시 smoothing이 필요하다.

또다른 에너지 윤곽선 추출 방법은 Pitch간격에서 최대값을 이용하는 방법이 있는데, 음성의 피치간격마다 발생하는 음성의 최대값을 연결한 윤곽선을 이용하는 방법으로, 에너지 윤곽선의 deep을 음절의 경계로 예측한다.

▶ 시간에 따른 스펙트럼 변화율

시간에 따른 스펙트럼의 변화를 이용한 음절길이 정보 추출법은 LPC residual 신호로부터 프레임당 스펙트럼 변화율의 정도를 잇는 곡선을 그리고, 피크값들을 체크해서 음소의 경계로 정한다. 음절의 길이정보는 음소의 길이정보를 더하면 된다.

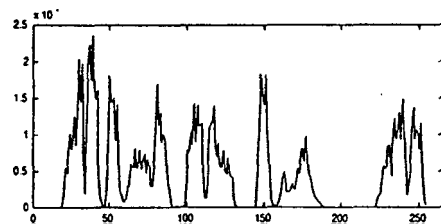


그림 2.1. fixed frame방법을 이용한 에너지 윤곽선 추출
Fig. 2.1. Energy contour extraction using the fixed frame method.

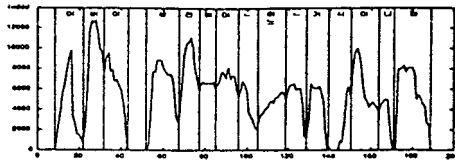


그림 2.2. Pitch간격에서 최대값을 이용하는 방법으로 음절 길이정보 추출

Fig. 2.2. Syllable duration extraction using the maximum value in the pitch period.

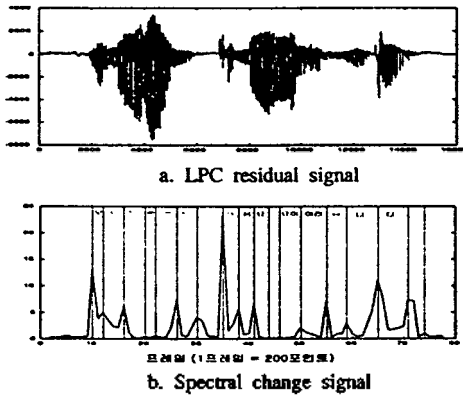


그림 2.3. 스펙트럼 변화율을 이용한 음절 길이정보 추출
Fig. 2.3. Syllable duration extraction using the spectrum variation rate.

2.3 음절경계 추출 알고리즘

음절 경계정보는 음절 내에서의 피치 변화율에 따른 운율경계 예측 시 이용되어진다. 또한 음소단위의 음성인식을 위한 기본 아이디어로서 향후 더 많은 연구가 필요할 것으로 보인다. 본 논문에서 제시한 음절경계 추출 알고리즘은 2.2절에서 살펴봤던 피치 간격에서의 최대값을 이용한 에너지 윤곽선 검출과 시간에 따른 스펙트럼 변화 정보를 이용하였는데, 아래의 [그림 2.4]에서 그 흐름을 간략히 보여주고 있다.

알고리즘의 기본 아이디어는 구해진 에너지 윤곽선의 Null점을 찍어서 [그림 2.2] 음절의 경계로 삼고, 시간에 따른 스펙트럼 변화율을 구해서 Peak점을 검출해서[그림 2.3-b] 일치하는 경계만을 음절의 경계로 예측하는 것이다.

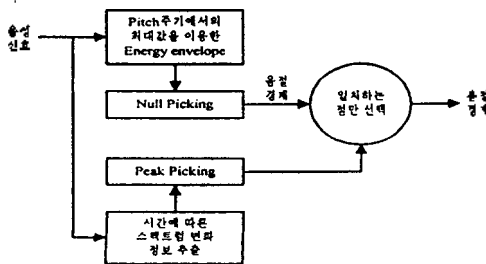


그림 2.4. 음절길이 추출 알고리즘
Fig. 2.4. Syllable duration extraction algorithm.

2.4 휴지기 정보 추출 방법

음성신호로부터 휴지기 정보를 추출하기 위해서는 에너지나 ZCR이나 에너지 정보를 이용해서 음성구간을 추출함으로써 가능하다는 것은 널리 알려진 사실이다. 본 논문에서는 휴지기 정보를 추출하기 위해서 2.2절에서 구해왔던 Pitch간격에서 최대값을 이용한 에너지 분포 곡선을 이용하였다. 추가적으로 각 음성의 태깅파일을 수정해서 휴지기를 두 가지의 휴지기로 구분했다. 구문론적 휴지기와 음운론적 휴지기가 그것인데, 우리가 일반적으로 끊어서 끊어 구문론적 휴지기가, 한 어절 내에서 묵음이 발생하는 경우에 묵음 부분을 하나의 작은 휴지기로 간주해서 표시한 것이 음운론적 휴지기이다. 구해진 구문론적 휴지기와 음운론적 휴지기를 이용해서 전체 휴지기 임계값을 결정했다.

실험을 위해서 표준말을 쓰는 남자화자가 녹음한 100개의 문장을 사용했는데, 그 발화문장들에 대해서 음운론적 휴지기와 구문론적 휴지기를 구한 히스토그램을 [그림 2.5]에 나타내었다. 이 그림에서 얻어진 두 곡선이 만나는 점 350포인트를 휴지기 임계값으로 결정했다.

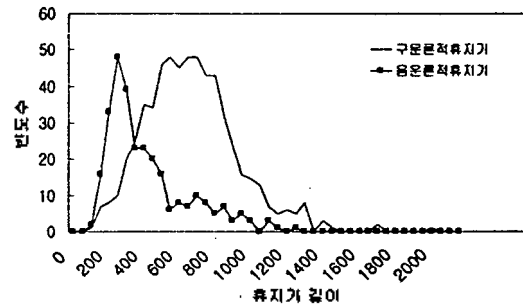


그림 2.5. 휴지기 히스토그램
Fig. 2.5. pause histogram.

III. 운율구 단위 추정 방법

화자가 발화문의 의미를 명확하게 전달하기 위해서 길이와 피치, 그리고 휴지기를 조절하게 되는데, 앞장에서는 이러한 정보들을 추출하는 방법에 대해서 알아보았고, 이번 장에서는 추출된 운율정보들을 바탕으로 운율구 경계를 결정하는 이론을 제안하고 실험적으로 검증하였다.

강한 운율적 불연속성은 청취테스트 결과 마치 긴 휴지기가 어절사이에 존재하는 것처럼 청취자는 판단하는데, 실제 음성 데이터를 살펴보면, 긴 휴지기가 존재하는 경우가 많지만 그렇지 않은 경우도 종종 발생한다. 그러한 경우에는 피치가 하나의 음소에서 HL의 형태를 보이면서, 음소의 길이가 길어지는 경향을 보인다. 따라서 본 연구에서는 이러한 사실을 바탕으로 하여 운율구 경계를 찾는 방법을 제안하였다. 그 방법은 다음과 같다.

첫째, 주어진 음성 데이터를 분석하여 긴 휴지기가 존재하는 구간을 찾아 이를 일차적으로 운율구의 경계로 정한다.

둘째, 피치 차이값을 조사하여 그 값이 급격히 변화했다면(실험에서 결정된 임계값 이상) 그곳을 운율구 경계로 정한다.

셋째, 결정된 운율구 경계 내에서 HL의 톤이 존재하는 지를 체크하여, 만약 이러한 톤이 발생하면, 이를 운율구 경계라 생각하여 자른다.

3.1 휴지기 길이정보를 이용한 운율경계 예측

2.3절에서 구했던 휴지기 정보를 이용하여 긴 휴지기가 나타나는 경계를 운율구의 경계로 정한다. 긴 휴지기가 함은 이전 실험에서 얻은 구문론적 휴지기와 음운론적 휴지기의 일치정보다 같거나 큰 휴지기를 의미한다.

긴 휴지기가 나타나는 샘플문장에 대해 분석한 그림의 예는 [그림 3.1]와 같다. 이미 언급한 것과 같이 휴지기를 결정할 때 본 논문에서는 Pitch주기에서 최대값 정보를 이용해서 얻은 에너지 윤곽선을 이용했는데, 이 방법을 이용했을 때 에너지가 0이 나와야 하는 지점에서 정확히 0이 나오지 않는 것을 관찰 할 수 있었다. 그래서 에너지의 크기 임계값을 정해서 그 값보다 작으면 0으로 간주하도록 하였다. [그림 3.1]에서 가로 점선은 그 임계값을 나타내고 있고, 세로 점선은 긴 휴지기에 의한 운율경계를 나타내고 있다.

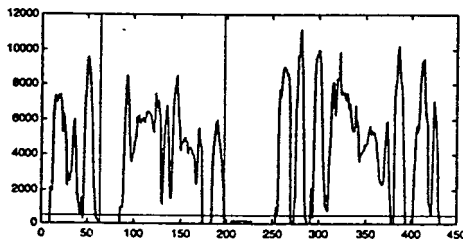


그림 3.1. 휴지기에 의한 운율경계 예측
Fig. 3.1. Prosodic boundary prediction using the pause information.

표 3.1. 휴지기 정보를 이용한 운율구 예측 결과(10문장)
Table 3.1. The prediction results of prosodic phrase using the pause information.(100 sentence)

문장 번호	운율구 개수	삽입	삭제
1	2	0	0
2	3	0	1
3	3	0	2
4	2	0	0
5	2	0	1
6	6	0	2
7	3	0	0
8	5	0	4
9	4	0	3
10	3	0	0
계	33	0	3
비율	%	0/33	16/33
		0	48.5

[표 3.1]은 남자화자가 발음한 10개의 문장에 대해서

휴지기 정보만을 이용해서 운율구 경계를 추정했을 때의 삽입, 삭제율을 보여주고 있다.

위의 표를 보면 휴지기를 이용해서 운율경계를 예측하면 삽입에러는 나타나지 않고 삭제 에러만 나타나는 것을 확인할 수 있다. 휴지기는 대체적으로 강한 경계를 잘 반영하기 때문이다.

3.2 피치신호의 급격한 변화를 이용한 운율경계 예측

두 번째 제안한 방법은 운율경계에서의 피치 차이 값들이 갖는 통계적인 값을 정함으로써 운율경계를 정하는 또 하나의 방법이다. 이를 위해서 먼저 100개의 문장에 대해서 음소 태깅을 한 후에 그 태깅 파일 속에 운율경계를 마킹 하였다. 이때 마킹 하는 기호는 나름대로 정해서 이용했는데, 휴지기에 의한 운율경계일 경우에는 PA 기호를 썼고, 피치 차에 의한 운율경계는 PD를 사용하였다. 피치 임계값을 결정하기 위한 시뮬레이션과정을 간단히 설명하면 다음과 같다.

- ▶ 태깅파일에서 운율경계에 해당하는 곳의 마킹 기호를 검사하고 그중에서 PD마킹인 것들의 위치를 포인트 단위로 저장해놓는다.
(bound.mbf: Master Boundary File)
- ▶ 저장된 PD마킹의 위치 정보를 얻어서 그 지점에서의 좌우의 피치 차이값을 구해서 파일로 저장해놓는다.
(sorted.pdf: Pitch Difference File)
- ▶ 저장된 값을 토대로 각각의 값들을 피치 임계값으로 주고 100개의 문장에 대해서 운율경계를 구한다. 단, 휴지기에 의한 운율경계는 이미 구해졌다고 가정한다.

[그림 3.2]는 경계에서의 피치의 차이값에 따른 삽입

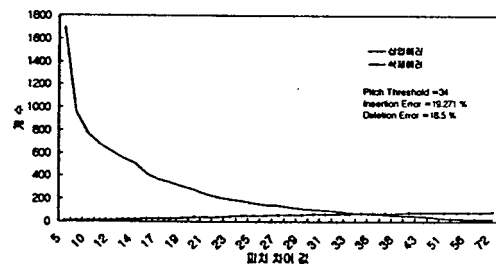


그림 3.2. 피치임계값 결정을 위한 삽입삭제율 비교곡선
Fig. 3.2. The comparing curve of insertion and deletion rates for the threshold of pitch.

삭제율 보인 그림이고 [표 3.2]는 최적의 임계값에 따른 삽입 및 삭제율을 보인 것이다.

표 3.2. 피치 차이값에 의한 임계값 결정
Table 3.2. The threshold determination using the difference pitch.

운율구 총개수	409
삽입 에러	19.27 %
삭제 에러	18.5 %
피치 임계값	34

구해진 운율경계들 중에 보면 피치 차이가 크지만 실제로 경계가 되지 못하는 부분들이 존재하게 된다. 바로 무성음구간인데 이런 부분은 아직까지는 오류로서 남게 된다. 그래서 무성음구간에서 크게 나타나는 ZCR(Zero Crossing Rate)을 이용해서 이 부분을 운율경계에서 제외시키도록 한 것이 [그림 3.3]이다. 그림에서 보면 운율경계로 결정되어진 지점에서 ZCR이 높다면 그 곳을 운율경계에서 제외시키고 있다. 이것을 앞의 실험에 적용해서 새롭게 시뮬레이션을 한 결과가 [그림 3.4]이다.

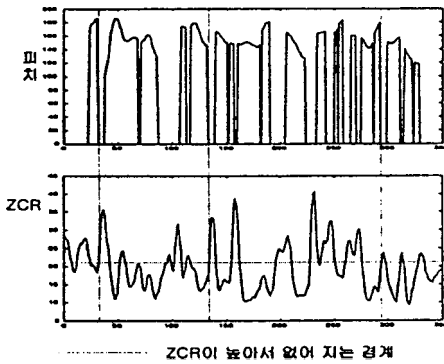


그림 3.3. ZCR보상을 하고 있는 그림
Fig. 3.3. The ZCR compensation.

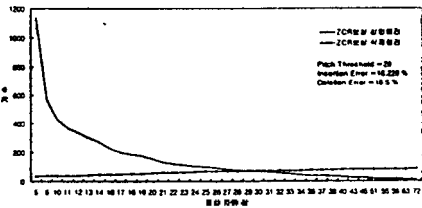


그림 3.4. ZCR보상후에 삽입율과 삭제율
Fig. 3.4. Insertion and deletion rates after ZCR compensation.

[표 3.3]을 보면 우 하향 곡선을 그리는 삽입에러 곡선이 더욱 급격하게 줄어들어서 [표 3.2]보다 피치 차이 임계값이 더 작아졌음을 알 수 있다.

표 3.3. 피치 차이값에 의한 임계값 결정(ZCR 적용)
Table 3.3. The pitch difference threshold determination using the ZCR.

운율구 총개수	409
삽입 에러	18.2 %
삭제 에러	18.5 %
피치 임계값	29

또 ZCR로 경계 보상한 후에도 남아 있는 삽입오류를 유발하는 경계들 중에는 어절의 경계가 되는 부분들도 많이 남아 있다. 운율구의 경계도 또한 특정 어절의 경계라고 말할 수 있으므로 어절의 경계는 삽입오류에서 제외시키는 것이 바람직할 것이다. 그래서 어절의 경계를 삽입오류에서 제외시킨 삽입, 삭제율 비교 그래프는 [그림 3.5]와 같다.

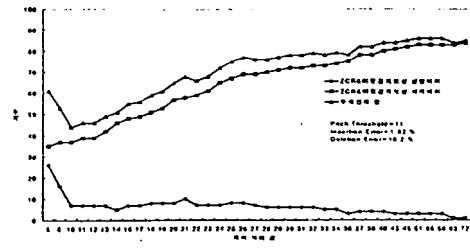


그림 3.5. ZCR과 어절경계 보상후에 삽입율과 삭제율
Fig. 3.5. Insertion and deletion rates after ZCR and ujeol boundary compensation.

삽입에러가 전체적으로 급격히 감소했음을 볼 수 있다. 비록 두 그래프가 어떤 점에서 만나고 있지는 않지만 삽입에서 그래프가 어느 순간부터 점점 일정해지는 것을 볼 수 있는데, 이 지점을 구하기 위해서 두 그래프의 합 의 그래프의 최소값을 피치차이 임계값(Pitch difference threshold)으로 정했다. 결과적으로 11이라는 값을 임계값으로 정할 수 있었다. 위의 그래프를 보면 10이 가장 적합한 값인데, 실험을 해본 결과 11이 더 좋은 성능을 보였다.

위의 그래프는 휴지기를 이용한 운율경계는 모두 찾았다고 가정하고, ZCR과 어절 경계 정보를 추가적으로 이용해서 임계값을 결정한 것인데, 100문장에 대해서 운율경계 일치율은 삽입에러가 약1.82%, 삭제 에러가 약 10.2%로 상당히 개선되고 있는 것을 볼 수 있다.

표 3.4. 피치 차이값에 의한 임계값 결정(ZCR과 어절경계 정보 적용)
Table 3.4. The pitch difference threshold determination using the ZCR and ujeol boundary information.

운율구 총개수	409
삽입 에러	1.82 %
삭제 에러	10.2 %
피치 임계값	11

3.3 음절내의 피치변화율을 이용한 운율경계 예측

마지막으로 제안한 방법은 운율구 경계의 한음절 내에서 HL의 톤이 존재하는 지를 체크하여, 만약 이러한 톤이 발생하면, 이를 운율구 경계라 생각하여 자른다.

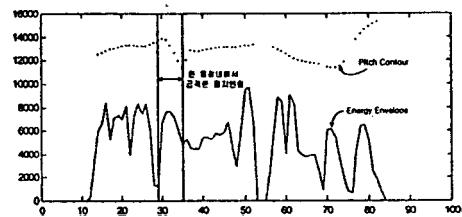


그림 3.6. 한 음절내의 피치의 변화가 큰 곳에서 나타나는 강한 운율경계
Fig. 3.6. A strong prosodic boundary at pitch differential peak in a syllable.

즉, 2.2절에서 했던 에너지 정보와 시간에 따른 스펙트럼 변화를 계산하여 음절의 경계를 예측하고, 한 음절 내에서 피치의 변화 폭이 심한 부분을 찾아서 운율구 경계로 정하는 것이다.[그림 3.6]

위에서 제안한 휴지기 정보 및 피치악센트 정보를 이용한 운율구 개수 예측모델의 신뢰도 검증을 위해서 남자화자가 읽은 100개의 문장에 대해서 그 결과를 [표 3.5]에 표기했다.

표 3.5. 휴지기 정보, 피치 정보를 이용한 운율구개수 추정(100문장)
Table 3.5 The estimation of prosodic phrase using pause and pitch informations.(100 sentence)

삽입경계 개수	삭제경계 개수	운율구 총개수	삽입 에러율	삭제 에러율
8	42	409	1.9	10.2

한 음절 속에서 피치변화폭이 큰 형태는 100개의 문장 중에서 5개를 관찰할 수 있었는데, 정한 규칙에 의해서 4개까지 찾아 낼 수 있었다. 그래서 운율구 예측의 삭제 에러율을 10.2%에서 9.2%로 줄일 수 있었다. 본 실험에서 사용한 남자화자의 발음은 한 음절 속에서 피치의 변화가 큰 형태가 많이 나타나지 않아서 에러율의 감소가 적었지만, 샘플로 몇 개의 여자화자의 목소리를 관찰해본 결과, 급격한 피치 변화의 패턴이 많이 발견되었다. 그래서 여자 화자의 경우에는 삭제 에러율의 감소에 더 좋은 효과를 보일 수 있을 것으로 기대 된다.

3.4 실험 요약

실험 처음에 나타났던 삽입에러는 어절경계 정보와 ZCR 정보를 포함시킴으로서 상당량을 줄일 수 있었다. 하지만, 삽입에러가 많다는 것은, 주어진 문장을 강한 운율구 이상으로 세분하여 자를 수 있는 단서를 보여주는 것으로서, 이를 잘 이용하면 오히려 음소단위 또는 음절단위의 음성 인식에 도움이 될 수 있을 것이다. [그림 3.7]에서는 이 장에서 했던 실험들을 도식적으로 나타내었다.

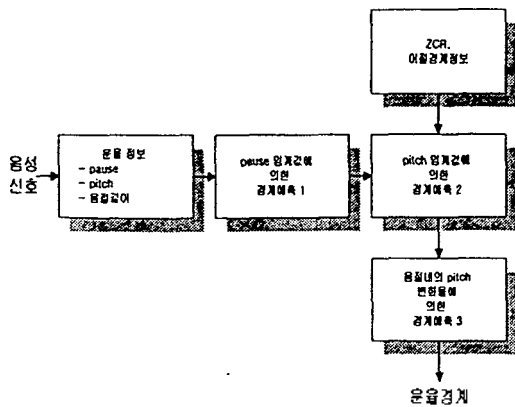


그림 3.7. 운율경계 추출 과정
Fig. 3.7. Prosodic boundary extraction block diagram.

다음 장에서는 지금까지의 실험을 통해서 예측한 운율구 경계를 바탕으로 음성인식 실험을 통해서 운율구단위의 음성인식과 문장단위의 음성인식이 보여주는 인식률과 복잡성에 따른 인식시간의 차이에 대해서 살펴보겠다.

IV. 운율구 단위의 음성인식 실험

화자 독립 음성인식을 위해서 많이 쓰이고 있는 학습 방법의 하나인 HMM(Hidden Markov Model)을 이용한 인식을 구현하고 지금까지 구했던 운율구 단위의 인식이 실제로 어느 정도의 효과를 가져오는지 실험을 하였다.

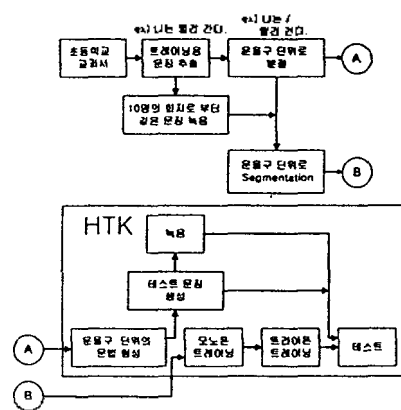


그림 4.1. HTK(Hidden Markov Model Tool Kit)를 이용한 운율구 단위 인식 테스트 모형
Fig. 4.1. Prosodic phrase unit speech recognition model using the HTK.

인식 실험에 사용된 텍스트는 초등학교 5학년 국어 교과서를 이용했는데, 총 문장은 32문장이고, 이 문장들을 운율구로 나누면 총 119개의 구를 얻을 수 있었다. 이 텍스트는 앞서 운율구 추출 시 사용했던 텍스트와는 다른 텍스트이다. 각 인식 단위의 학습을 위해서 남자화자 5명의 발성 데이터를 이용하였다. 즉, 문장 단위 인식을 위해서 32문장 각각을 5명의 화자에게 읽혀서 문장을 학습시켰고, 운율구 단위 인식을 위해서는 119개의 운율구를 5명의 화자에게 읽힌 후, 각 운율구를 학습시켰다. 학습은 HMM을 이용한 Monophone 학습, Triphone 학습, Tied-state Triphone 과정을 거쳤다.

실험을 하면서 word network의 사이즈에 따라서 인식 시간의 차이 심한 것을 알 수 있었다. word network에서는 단어들간에 다음 단어가 나올 확률을 모두 계산하게 되는데, 단어가 많으면 많을수록 기하 급수적으로 그 확률의 경우의 수가 많아지게 되고, 복잡도가 증가하게 되어 word network파일의 크기가 커져서 인식하는데 시간이 많이 걸리는 것이었다. 그래서 해결책으로 word network에 bigram을 적용하기로 했다. bigram을 적용하게 되면 word network의 사이즈를 효과적으로 줄일 수 있어서 인식시간을 4초 정도로 약1/45정도의 시간 단축을 가

저울 수 있었다. 즉, bigram을 이용하면 문법구조상에서 다음에 올 단어와의 확률만 구함으로서 필요 없는 확률 계산을 하지 않게 되어 계산시간이 단축되고 word network의 사이즈도 훨씬 감소되는 것을 알 수 있었다. 하지만, bigram으로 인식을 하게 되면 word network에 확률이 없는 단어 배열에 대해서는 전혀 인식을 할 수 없는 단점이 있다. 그러나 본 실험에서는 운율구 내에서 단어의 배열이 바뀌는 일은 없기 때문에 bigram을 이용해서 word network를 구성하는 것도 무리가 없었다.

아래의 [표 4.1]에는 두가지 단위에 대해서 인식실험 결과를 나타내고 있다. 이 표에서는 운율구단위로 인식 테스트를 했을 때와 긴 문장을 그대로 인식 테스트했을 때 걸리는 시간과 인식률을 보이고 있는데, 여기서 인식률은 단어의 정확도가 아니고 인식단위 각각에 대해서 정확도를 가르킨다.

첫 번째 실험에서는 문장단위로 읽어서 5명까지 학습시켜 가면서 학습시킨 문장을 문장단위로 읽어서 테스트를 했고, 두 번째 실험에서는 운율구단위로 읽어서 5명까지 학습시켜 가면서 학습시킨 문장을 운율구단위로 읽어서 테스트했다.

[표 4.1]을 보면 인식률이 운율구 단위의 인식이 훨씬 우수하다는 것을 알 수 있다. 그리고 인식 시간은 주로 모델의 크기에 의해서 영향을 받는데, 더 많은 사람들로 부터 학습을 하게 되면 모델이 더욱 커져서 시간이 많이 걸리는 것은 당연하겠지만, 인식하려는 단위의 문장이나 구가 길수록 찾아야 하는 모델이 많아서 시간이 더 많이 걸린다는 것도 있겠다. 아래의 표를 보면 문장단위 인식을 위해서 소요되는 시간이 같은 학습화수에서 더 크고, 학습량이 많아질 수록 더욱더 커진다는 사실을 확인할 수 있다.

표 4.1. HTK를 이용한 운율구 단위 인식기의 성능 비교
Table 4.1. The performance of prosodic phrase unit speech recognition model using the HTK.

트레이닝 화자수	운율구 단위 인식 (단위: 초)	인식률 (%)	긴 문장을 그대로 인식 (단위: 초)	인식률 (%)
1	3.05	75	4.1	55
2	3.05	80	4.1	66.6
3	3.25	86.5	4.3	76.4
4	3.25	88	4.3	80
5	3.30	93.4	4.3	81

결론적으로 운율구 단위의 음성인식을 하게 되면 인식해야할 문장의 길이가 짧아지므로 문법의 복잡성 감소에 따른 계산량의 감소에 의해서

- 인식 속도가 빨라지고,
- 인식률이 높아진다.

는 것을 위의 실험을 통해서 알 수 있었다. 음성 인식기의 제약성중의 하나인 무제한 문법의 구현은 사실상 더 작은 단위로의 구분이 없다면 문장의 길이에 따라서 문법의 종류도 무한적으로 나오게 되므로 더욱 어렵게 될 것이다. 이러한 점으로 미루어 볼 때 본 논문에서 제안한 운율구단위의 인식은 비록 세부 문법구조는 HTK에서 제공하는 task grammar를 사용했지만, 인식 단위의 문법으로서 적합한 구조로 생각되어 진다.

V. 결 론

본 논문에서는 끊어 읽는 단위를 강한 운율경계로서 결정짓고, 휴지기 정보 및 피치 정보를 사용한 추정방법을 제안하고, 100개의 발화문장을 대상으로 그 성능을 검토하였다. 제안한 방법이 모든 발화문들에 대해서 운율구의 경계를 완벽하게 찾는 알고리즘으로서 아직 부족한 점이 많지만, 실험에 쓰인 문장들에 대해서는 어절경계 보상 없이 삽입에러 18.2%, 삭제에러 18.5%이고, 어절경계를 보상하면 삽입에러 1.82%, 삭제에러 10.2%로서 좋은 성능을 보여주었다. 다만 실험에서 정한 통계값인 휴지기 임계값과 피치 임계값은 실험에 사용한 100문장에 대해서는 적합할 수 있겠지만, 발화문의 속도나 크기가 다른 발화문 들에 대해서는 경계추정에 실패할 것으로 본다. 그래서 정규화된 휴지기정보와 피치정보를 바탕으로 임계값을 결정하는 것이 요구된다. 그리고 운율경계 예측시 경계의 보정에 사용되었던 시간에 따른 에너지 외곽선 정보와 스펙트럼 변화정보를 이용한 음절의 경계 추정법은 음소단위 인식이나 자동 Labeling, 음성인식 타자기 등 다양한 분야에 쓰일 수 있으므로 좀더 깊이 있는 연구가 요구된다.

무제한적인 음성인식의 제약성으로 작용하고 있는 무제한 문법구조를 제안하고자 한다면 단위 문법구조를 조금이라도 단순화시킬 수 있는 아이디어가 필요하게 되는데, 이러한 필요성을 만족시키면서 인식시 복잡성의 감소와 계산량의 감소로 인한 인식 시간 단축과 인식률의 증가를 가져오는 운율구 단위의 인식은 연구의 가치가 있다고 하겠다.

참 고 문 헌

1. ETRI, "다중 매체 환경에서의 대화체 음성 번역 통신 기술 개발," 정보통신부, 1996.
2. 한국통신 연구개발본부, "음성 언어 시스템 개발을 위한 한국어의 운율구조 및 담화구조 연구," 1997.
3. 김선미, "한국어의 리듬단위와 문법구조," 서울대학교 박사 학위 논문, 1997.
4. 전남대학교 공업기술 연구소, "한국어 운율구조 연구," 1996.
5. KAIST AI, "음성언어 시스템 개발을 위한 한국어의 운율구조 및 담화구조 연구," 1996.
6. 전남대학교, "Korean Prosody Information Processing for Speech Recognition and Synthesis," 1998. 6. 31.
7. A.J.Hunt, "Models of Prosody and Syntax and their

