운율구 경계현상 분석 및 텍스트에서의 운율구 추출

Analysis and Prediction of Prosodic Phrage Boundary

김 상 훈*, 성 철 재*, 이 정 철*
(Sang Hun Kim*, Cheol Jae Seong*, Jung-Chul Lee*)

요 약

이 논문에서는 첫번째로 운율구 경계를 자동으로 추출하고자 할때 인간의 지각과 관련하여 어떠한 자질을 이용하는 것이 좋은가 하는 점을 밝혀 보았다. 운율구 경계의 유형은 크게 경계가 없는 강도(zero), 약한 경계 강도(minor break strength), 강한 경계 강도(major break strength) 3 단계로 정하는 것이 통계적으로 의의가 있으며 지속시간의 정보가 강한 경계 유형의 결정에 중요하게 작용하는 자질이었음을 알 수 있었다. 두번째로는 운율구 경계현상 분석결과를 바탕으로 운율구 경계의 경계 강도를 약한 경계 강도(zero를 포함)와 강한 경계 강도로 나누고, 2단계의 경계 강도를 텍스트상에서 문장성분의 bigram, trigram을 정보를 이용하여 자동으로 할당하였다. Bigram의 경우 Test-I, Test-II 텍스트 DB에 대해 각각 46.0%, 38.2%의 강한 경계 강도 예측정확률과 22.8%, 8.4%의 삽입오류율의 성능을 내었으며, Trigram인 경우 Test-I, Test-II 텍스트 DB 각각에 대해 58.3%, 42.8%의 강한 경계 강도 예측정확률과 30.0%, 11.8%의 삽입오류율를 나타냈다.

ABSTRACT

This study aims to describe, at one aspect, the relativity between syntactic structure and prosodic phrasing, and at the other, to establish a suitable phrasing pattern to produce more natural synthetic speech. To get meaningful results, all the word boundaries in the prosodic database were statistically analyzed, and assigned by the proper boundary type. The resulting 10 types of prosodic boundaries were classified into 3 types according to the strength of the breaks, which are zero, minor, and major break respectively. We have found out that the durational information was a main cue to determine the major prosodic boundary. Using the bigram and trigram of syntactic information, we predicted major and minor classification of boundary types. With bigram model, we obtained the correct major break prediction rates of 46.0%, 38.2%, the insertion error rates of 22.8%, 8.4% on each Test-I and Test-II text database respectively. With trigram model, we also obtained the correct major break prediction rates of 58.3%, 42.8%, the insertion error rates of 30.8%, 11.8% on Test-I and Test-II text database respectively.

I. 서 론

운율구는 화자의 자연스런 발성에 따라 형성되는 단위로서 지속시간, 억양, 휴지 등의 음향적 변화로 경계지어지며, 적절한 운율구 할당(prosodic phrasing)은 음성합성의 자연성을 크게 향상시킬 수 있다[1]. 특히 운율구경계를 추출함으로써 복합적으로 일어나는 운율현상을 운율구 경계에서의 운율 현상, 운율구 내(단어간)에서의 운율현상, 운율구간의 영향 등으로 단계화해 순차적으로 모델링할 수 있게 된다. 또한 운율구는 음성인식 결과의 정확한 의미를 출력할 수 있는 단위로 이용될 수 있다[2]. 운율구를 운율제어모델의 기본 단위로 할 경우 다음과 같은 연구가 필요하다.

- 운율구를 기술하는 태그 세트(tag set) 정의
- 운율구를 결정하는 자질 추출
- 문장구조와 운율구간 비교 방법
- 세밀한 경계 강도 할당
- 형태소 해석을 이용한 문장성분 추정
- 문장상에서 운율구 추출
- 운율구내 운율 분석
- 운율구 단위의 지속시간, 억양, 휴지 모델링

이에따라 본 논문은 운율구를 기술하는 태그 세트 정의, 운율구를 결정하는 자질 추출 및 문장상에서 운율구 추출 연구결과를 기술하고자 하며 이를위해 첫번째로 운

*한국전자통신연구소 음성언어연구실
접수일자: 1996년 8월 14일

율구 경계에서의 운율 현상을 분석하고 운율구 경계 유형을 설정하는 자질(feature)을 소개하고자 한다. 다음으로는 운율경계에서의 끊김의 정도(Break strength : Tone and Break Indices 시스템에서 Break Indices를 말함)를 기계적 레이블링과 인간의 지각에 의한 레이블링으로 할당하고 양자를 비교함으로써 향후 음성신호상에서 운율구 경계강도의 자동할당에 대한 가능성을 살펴볼 것이다[3]. 마지막으로 실제 TTS에 적용하기 위한 방법으로 품사정보와 bigram, trigram을 이용한 텍스트에서의 운율구 경계 강도를 할당하는 방법을 소개한다. 이 결과는 문장상에서 운율구로 phrasing 하는데 사용될 것이며, 지속시간 모델링의 유용한 자질로 이용될 것이다.

Ⅱ. 운율구 경계 현상 분석

2.1 운율구 경계의 정의 및 음향요소

음성데이타의 효율적 분석을 위한 첫단계는 각 문장의 내부를 음성적으로 의미있는 단위로 끊어주는 것이다. 이를 영어로는 "prosodic phrasing", 우리말로는 "운율구 만들기"라는 용어로 표현하고자 한다. 문장의 문법적인 분석(syntactic analysis)과정에서 각 문장의 성분들에 의미있는 역할을 부여하여 이들을 또 다른 대단위로 무리지어가는 과정을 "phrasing"이라고 하듯이 운율의 흐름을 단위별로 구분하여 무리짓는 과정도 이에 맞추어 운율구라 이름짓는 것이다[4]. 경계는 이러한 운율구의 끝부분을 말한다. 음성자료에서 경계의 모습은 다양하게 나타나며 통사적, 의미적인 단위와 밀접한 관계를 가지면서도 어느 정도는 일탈된 모습을 보여준다. 경계를 어떻게 구분하며 분석하느냐가 운율구 분석의 첫단계이며 이 장에서는 이러한 경계를 설정하는데 있어 주도적인 역할을 할 수 있는 자질집합을 기술하고자 한다. 일반적으로 경계현상은, 통사단위 혹은 의미단위의 경계가 주어진 환경에 따라 운율적인 경계로 실현될 때, 그리고 생리학적인 숨쉬기의 한계 및 호흡과 관련지어 나타나는 지각적인 끊김의 현상을 말한다. 경계의 운율적인 부분은 쉼(pause)과 관련하여 다양한 모습을 보여준다. 묵음구간 뿐 아니라 묵음구간이 없이 마지막 음절의 장음화만으로 이루어지기도 하고, 억양의 변화로 나타나기도 하며 음질(voice quality)의 변화로 나타나기도 한다.

Lehiste[5]는 문장 경계와 단락 경계의 인지에 관한 연구에서 휴지의 길이, 후두음화(laryngealization)의 여부, 경계앞 음절의 장음화 등을 경계의 운율요소로 파악하였고 Strangert & Zhicite[6]는 스웨덴어의 경계유형을 다루면서 들숨(inhalation)의 존재, 묵음구간, 경계 앞 음절의 장음화, 경계전후의 기본주파수 변화, 음질변화 등을 상관된 음향 요소로 들고 있는데 경계앞 음절의 장음화를 제외하면 경계의 통사적 위계가 높을수록 상관된 음향요소들이 강하게 나타난다고 하였다.

여러개의 낱말이 모여 구를 이루고 이들이 다시 절로 그리고 문장으로 묶여져 나가는 것과 같이, 부려쓰인 음성언어에서도 끊어읽기 및 숨쉬기의 과정을 통하여 나름대로의 단위를 만들어 나가는 단위구성의 방법이 존재한다. 운율구 단위는 화자의 입장에서의 끊어읽기, 그리고 청자의 입장에서의 끊김의 청각적 지각을 이용하여 만들어지며 호름이 끊어지는 바로 그 부분이 운율구의 경계가 된다. 종래, 억양과 관련된 많은 논의에서 억양구(intonational phrase) 단위 분석이 한 나라 언어의 운율구조 분석에서 가장 중요한 역할을 담당해왔다. 현재까지도 음성, 음운론적인 분석의 출발점은 억양곡선이라는 것이 주지의 사실이다. 그러나 실제 부려쓰인 음성언어의 역동적인 모습을 관찰하기 위해서는 억양곡선의 분석만으로는 부족함이 많은 것이 사실이다. 근래의 많은 논의에서 지속시간이 경계인지의 중요한 자질로 자리잡아 가고 있는 것이 그 반영이라고 할 수 있다. 적어도 운율구 경계가 강할때의 주요자질은 우선적으로 경계앞 음절의 장음화와 관련되어 있다는 것이 많은 저작들에서 밝혀지고 있다[7].

2.2 운율구 경계의 레이블링

2.2.1 음성 데이타베이스

운율구 경계현상 분석에 사용된 운율 DB는 1명의 여성화자(아나운서)가 1회 발성한 156 문장으로 구성되어 있다. 156 문장은 다양한 장르의 문장으로부터 추출하였으며 녹음시 문장간 문맥을 제거하기 위해 무작위로 문장을 섞어 발성하였다. 이 문장들은 방음실에서 녹음되고 음성학자에 의해 분절과 레이블링이 이루어졌으며,유성음부에는 피치동기(pitch-synchronous)로 피치값이 표시되어 있다. 피치값은 먼저 자동으로 찾고, 오류에 대해 수작업으로 수정하였다. 이 운율 DB의 통계적 분석을 위해 음절단위로 각 음소의 지속시간, 어절내 음절의 위치, breath group내 어절의 위치, 음절의 음운환경, 모음의 평균 피치, 모음부의 피치 개수를 계산한 데이타 구조를 만들었다.

2.2.2 기계적 경계

전체 156 문장의 운율 DB에서 중성과 종성에 대한 평균지속시간 통계를 냈으며, 어절내 위치에 따른 영향을 고려하여 어말 음절(final)과 이외의 음절(initial, medial)로 구분하여 평균치를 추출하였다. 억양에 대한 변화를 분석하기 위해 어말 음절의 모음에 대한 평균 $F\phi$와 이 음절을 포함한 문장의 평균 $F\phi$를 추출하였다. 위와 같은 두가지 값들을 구한 다음 아래의 10가지 기준안을 근거로 하여 경계의 유형을 분류하였다. 예를 들어 Mlabel-1(이후 Mlabel-? : machine label?로 표기함)의 경우는 중성(모음)과 종성 음소(유성자음)의 지속시간이 음성 데이타베이스에서 추출한 평균 지속시간보다 10% 이상 길며,

모음의 평균 $F\phi$ 값이 그 모음이 속한 문장의 평균 $F\phi$ 값보다 10% 미만의 경우를 나타낸다. 특히 유성 종성자음은 모음과 같이 지속시간 변화에 민감하게 영향을 받는 부분이기 때문에 지속시간 비율 계산에 포함하였다. 또한 음절위치에 따라 지속시간 분포가 다르므로 이 연구에서는 어말음절(모음 + 종성유성자음)의 지속시간 비율을 어말음절 평균 지속시간에 대한 비율로 구하였다. 그 외 위치에서의 음절은 그외 위치에서 구한 지속시간 평균치로 나누어 신장율을 구한다. 예를들면 어절의 첫음절은 기타음절에서 구한 평균치를 적용한다. 또한 음절의 평균 지속시간은 음절위치로 구분되어 구한 평균 중성 지속시간과 평균 종성 지속시간을 적용한다. 즉 음절구조가 V, CV 형태인 경우 평균 중성 지속시간을 적용하고, 음절구조가 VC, CVC 형태인 경우 평균 중성 지속시간과 평균 종성 지속시간을 더해서 구한다.

그리고 $F\phi$의 경우 모음의 안정구간의 음의 높이가 억양의 지각에 가장 중요한 요소이기 때문에 유성 종성자음은 제외하였다. 이 논문에서 사용한 10%의 $F\phi$의 변화량은 인간이 지각할 수 있는 변화량이라고 할 수 있다. 즉 피치의 경우 10% 변화가 음악에 있어 거의 한 음계의 차이를 나타내어 지각적으로 확연히 구별이 되는 정도라고 말할 수 있다. 참고로 음악에서의 한 음계의 차이는 12.5%의 높이의 차이에 해당한다. 지속시간의 경우 평균치에 비해 10% 증감에 따라 나눈 이유는 주로 지속시간의 장음화(lengthening), 단음화(shortening)을 충실히 반영하고자 함이다. 즉 10% 기준으로 10% 이상 증가이면 장음화의 확률이 높고 10% 이하 감소면 단음화의 확률이 높다. 또한 -10%〈변화율〈10% 구간을 normal 구간으로 둠으로서 장음화가 단음화로 단음화가 장음화로 반영되는 것을 줄이고자 하였다.

2.2.3 지각적 경계

운율 데이타베이스를 대상으로 음성학 전문가가 지각적 경계매김 작업을 수행하였다. 이는 실제 끊어짐이 지각되는 부분의 운율적인 현상은 어떠하며 기계적 매김과 어느 정도 일치되는지 그리고 어떠한 자질을 설정해야 효과적으로 경계단위를 매겨나갈 수 있는지 등을 알아보기 위해서이다. 이 논문에서는 지각에 의한 경계의 정도를 아래와 같이 4단계로 구분하였다.(이후 Plable-?:perception lable1?로 표기함)

- 경계가 지각되지 않는 부분(Plable-1)
- 경계가 아주 조금 지각되는 부분(Plable-1:weak minor)
- 경계가 어느 정도 지각되는 부분(Plable-2:strong minor)
- 경계가 확실히 지각되는 부분(Plable-3:major)

표 1. 경계유형 설정의 기준(X:don't care)

Label	Pause	Duration	Intonation
Mlabel-0	50msec 이상	X	X
Mlabel-1	50msec 미만	lengthening	falling
Mlabel-2	"	lengthening	level
Mlabel-3	"	lengthening	rising
Mlabel-4	"	normal	falling
Mlabel-5	"	normal	level
Mlabel-6	"	normal	rising
Mlabel-7	"	shortening	falling
Mlabel-8	"	shortening	level
Mlabel-9	"	shortening	rising

2.2.4 경계 유형의 분석

지각적 경계와 기계적 경계의 대응관계는 표 2와 같이 구하였다.

표 2. 지각적 경계와 기계적 경계와의 관계(%)

기계/지각	Plabel-0	Plabel-1	Plabel-2	Plabel-3
Mlabel-0	0(0)	0(0)	2(0.5)	367(99.5)
Mlabel-1	1(3.4)	0(0)	3(10.3)	25(86.2)
Mlabel-2	40(13.6)	8(2.7)	20(6.8)	226(76.9)
Mlabel-3	12(13.8)	9(10.3)	13(14.9)	53(60.9)
Mlabel-4	8(61.5)	0(0)	3(23.1)	2(15.4)
Mlabel-5	81(56.3)	19(13.2)	24(16.7)	20(13.9)
Mlabel-6	57(51.8)	17(15.5)	33(30)	3(2.7)
Mlabel-7	82(89.1)	1(1.1)	8(8.7)	1(1.1)
Mlabel-8	405(78.8)	43(8.4)	39(7.6)	27(5.3)
Mlabel-9	252(64.3)	67(17.1)	62(15.8)	11(2.8)
개수	938	164	207	735

지속시간이 그 평균값보다 10% 늘어났고 $F\phi$가 하강, 평탄, 상승의 패턴을 나타내는 기계적 레이블 Mlabel-1, Mlabel-2, Mlabel-3의 경우, 경계가 확실히 있다고 지각된 Plabel-3에 대해 86.2%, 76.9%, 60.9%의 대응관계를 보여주고 있다. 이로부터 강한 경계의 지각에서는 지속시간의 변화가 $F\phi$ 정보보다 더욱 중요하게 작용함을 알 수 있다. 기계적 레이블과 지각적 레이블을 상호 비교하여 그 사이에 어떤 공통점이 도출된다면 기계적인 레이블링의 효율성이 보장될 수 있을 것이다. 표 1로부터 Mlabel-0으로 표시된 50msec 이상의 묵음구간이 온 경우 높은 비율로 경계를 인식하고 있다. Mlabel-1, Mlabel-2, Mlabel-3의 경우에도 가장 백분율이 높은 열은 모두 강한 경계 부분으로 나타나고 있다.

기계적 기준이 인간의 지각적 판단과 반드시 동일한 것은 아니다. 사람의 지각적 판단도 상호 다를 수 있다. 요는 기계적인 레이블링이 이루어지기 위해서 어느 정도의 지각적 한계점을 수용하느냐는 판단이며 또한 기계적인 작업을 용이하게 하기 위한 자질집합을 어떻게 구성하느냐는 것이 관건이다. 이후의 논의에서는 경계지각에 참여하리라고 짐작되는 짝을 이룬 상대적인 음향요소들을 자질요소로 기술하려고 한다. 경계 좌우 음절들의 음향변수들의 차이값이 출발점이다. 이러한 논의를 통해서

지금까지의 결과들이 통계적으로 어떠한 의의를 갖는지 파악될 수 있을 것이다.

2.3 경계지각에 참여하는 음향자질의 효용성

경계를 어부는 남말 마지막 음절을 기준으로 하여 그 앞,뒤 음절들에서의 지속시간과 $F\phi$ 고저값의 차이가 경계인지에 참여할 것이라는 가설을 세웠다. 위에서 살펴본 두가지 지각 자료를 분석하기 위하여 경계 앞, 뒤 음절과 경계음절 자체와 음향변수의 비율을 다음과 같은 4가지 비율로 구분하여 파악하였다.

$$\bullet\ Dur_{\frac{PNT}{BND}} \equiv \frac{PENULTIMATE\ SYLLABLE\ DURATION}{BOUNDARY\ SYLLABLE\ DURATION}\Big|_{in\ WORD}$$

$$\bullet\ F\phi_{\frac{PNT}{BND}} \equiv \frac{PENULTIMATE\ SYLLABLE\ F\phi}{BOUNDARY\ SYLLABLE\ F\phi}\Big|_{in\ WORD}$$

$$\bullet\ Dur_{\frac{PST}{BND}} \equiv \frac{POST\ BOUNDARY\ SYLLABLE\ DURATION}{BOUNDARY\ SYLLABLE\ DURATION}\Big|_{between\ WORD}$$

$$\bullet\ F\phi_{\frac{PST}{BND}} \equiv \frac{POST\ BOUNDARY\ SYLLABLE\ F\phi}{BOUNDARY\ SYLLABLE\ F\phi}\Big|_{between\ WORD}$$

위 4가지의 비율 각각에 대한 값을 지각 경계 관찰 결과에 대해서 모두 조사해 보았으며 위 자질들의 분포 및 평균값에서 차이가 나는지를 통계를 이용하여 밝혀보았다. 즉 4가지의 자질들이 의미를 가지려면 레이블 상호간 서로 다른 데이타 분포 및 평균값을 보여줘야 한다는 것이다. 이 4가지 각각에 대한 값을 지각의 경계 관찰 결과에 대해서 동일한 자질의 평균값이 서로 다른 경계 유형간에 차이점이 있는지를 알아보기 위해 t-TEST를 실시하였으며 분포에서의 차이를 알아보기 위해 등분산의 가설검정에 이용되는 F-test도 실시하였다($p<0.05$). 각 자질들이 레이블 상호간을 구분시켜줄 수 있는지를 파악하기 위한 통계 결과는 다음과 같다. t-TEST는 unpaired(두 집단이 상호 분산이 다름을 가정), 2-tail로 시행하였다.

표 3. 4가지 자질값에 대한 평균, 표준편차, 분산(m : mean, σ : standard deviation, σ^2 : variance)

label/feature	item	$Dur_{\frac{PNT}{BND}}$	$F\phi_{\frac{PNT}{BND}}$	$Dur_{\frac{PST}{BND}}$	$F\phi_{\frac{PST}{BND}}$
Plabel-0 (938개)	m	1.157	0.996	1.277	0.959
	σ	0.681	0.114	0.619	0.121
	σ^2	0.463	0.013	0.383	0.015
Plabel-1 (164개)	m	1.031	0.547	1.250	0.989
	σ	0.507	0.085	0.584	0.119
	σ^2	0.257	0.007	0.341	0.014
Plabel-2 (207개)	m	0.983	0.959	1.145	0.999
	σ	0.492	0.088	0.458	0.198
	σ^2	0.242	0.008	0.210	0.039
Plabel-3 (735개)	m	0.661	0.993	0.757	1.110
	σ	0.298	0.120	0.312	0.166
	σ^2	0.089	0.014	0.097	0.028

모든 음향자질이 경계음절의 지속시간과 $F\phi$값을 분모로 하여 산출된 비율이므로 1.0보다 작은 평균값은 경계음절 부분의 음향변수값이 더 큼을 의미한다. 평균값만으로 비교하면 Plabel-0와 Plable-3는 $F\phi_{\frac{PNT}{BND}}$ 자질값을 제외한 모든 자질들에서 많은 차이를 보인다. Plabel-1과 Plabel-2 사이에서는 $Dur_{\frac{PNT}{BND}}$ 의 차이가 관찰되는데 이는 경계 유형의 구별요소로 지속시간이 중요하게 작용했음을 말해준다. Plabel 상호간의 통계적 차이는 다음과 같이 나타난다.

특이한 것은 $F\phi_{\frac{PNT}{BND}}$ 자질이 Plabel-0와 Plabel-3를 구분하는데 별 통계적 의의를 가지지 못한다는 사실이다. 또한 약한 경계 강도를 Plabel-1과 Plabel-2로 나누는 것이 의미가 별로 없다는 것을 알 수 있다.

2.4 경계현상 분석 및 결과

지금까지 살펴본 통계적 절차에서 평균값의 차이를 검정하는 t-TEST보다 분산에서의 차이를 검정하는 F-test가 우선적으로 작용함을 알아야 한다. 분포가 다르더라도 평균값에서는 비슷한 결과가 나올 수 있기 때문이다. 더 중요한 것은 자질값 1.0을 기준으로 하여 1.0보다 작은 쪽과 1.0보다 큰 쪽 중 어느 쪽으로 분포가 더 치중되어 있는가 하는 것이다. 위 4가지 자질의 특성을 파악하는데 기본적으로 작용하는 것은 경계음절을 기준으로 그 앞, 뒤 유절간 음향변수들에서의 '차이'이기 때문이다. 1.0을 기준으로 한 데이타 분포를 살펴보면 다음과 같다.

표 4. 각 레이블간의 4가지 자질값에 대한 t-TEST & F-test($p<0.05$, $p^*>0.05$)

label/feature	test	$Dur_{\frac{PNT}{BND}}$	$F\phi_{\frac{PNT}{BND}}$	$Dur_{\frac{PST}{BND}}$	$F\phi_{\frac{PST}{BND}}$
Plabel-0와 Plabel-1	t	0.006	0	0.587*	0.003
	F	0	0	0.360*	0.848*
Plabel-0와 Plabel-2	t	0	0	0.001	0.006
	F	0	0	0	0
Plabel-0와 Plabel-3	t	0	0.652*	0	0
	F	0	0.129*	0	0
Plabel-1와 Plabel-2	t	0.359*	0.201*	0.062*	0.569*
	F	0.697*	0.707*	0.001	0
Plabel-1와 Plabel-3	t	0	0	0	0
	F	0	0	0	0
Plabel-2와 Plabel-3	t	0	0	0	0
	F	0	0	0	0.001

표 5. 4가지 자질의 1.0을 기준으로 한 분포(개수)

label/feature	item	$Dur_{\frac{PNT}{BND}}$	$F\phi_{\frac{PNT}{BND}}$	$Dur_{\frac{PST}{BND}}$	$F\phi_{\frac{PST}{BND}}$
Plabel-0	1.0 <	465	428	580	276
	=	8	49	5	34
	1.0 >	465	461	353	628
Plabel-1	1.0 <	63	36	102	74
	=	4	5	0	3
	1.0 >	97	123	62	87
Plabel-2	1.0 <	82	48	122	94
	=	1	10	3	6
	1.0 >	124	149	82	107
Plabel-3	1.0 <	83	296	142	526
	=	1	31	9	16
	1.0 >	651	408	584	193

표 4에서 Plabel-0와 Plabel-3 사이의 $F\phi_{\frac{PNT}{BND}}$ 자질이, 두 집단 사이의 분산과 평균값에서 그 차이점이 통계적으로 검정되지 못했으나, 표 5의 $F\phi_{\frac{PNT}{BND}}$ 행을 참고하면 Plabel-0와 Plabel-3 사이에 많은 차이점이 있다는 것을 알 수 있다. Plabel-3의 $F\phi_{\frac{PNT}{BND}}$ 경우, 1.0미만이 전체의 약 56%, 1.0을 초과한 경우가 40.3% 정도 분포하므로 경계 음절이 경계 앞 음절보다 $F\phi$ 값이 어느 정도 높다는 것을 알 수 있다.

위의 표 5로부터 분석결과를 정리하면 다음과 같다.

1. Plabel-0에서는 경계뒤 음절의 $F\phi$ 값이 경계음절보다 더 낮은 경우가 많다.
2. 약한 운율구 경계 설정에 도움되는 자질은 $Dur_{\frac{PNT}{BND}}$, $F\phi_{\frac{PNT}{BND}}$, 그리고 $Dur_{\frac{PNT}{BND}}$ 라고 할 수 있다.
3. 강한 운율구 경계에서는 거의 모든 자질이 의미를 가진다.
   - $Dur_{\frac{PNT}{BND}}$ 자질은 거의 90%에 가까운 확률로 경계음절이 그 앞 음절보다 길어지는 경우를 보여준다.
   - $F\phi_{\frac{PNT}{BND}}$ 는 55.5% 정도로 경계음절이 그 앞음절보다 $F\phi$ 값이 높은 경우가 많음을 보여준다.
   - $Dur_{\frac{PST}{BND}}$ 의 경우 79.5%로 경계음절이 경계뒤 음절보다 길어진다.
   - $F\phi_{\frac{PST}{BND}}$ 의 경우는 72% 정도로 경계음절이 경계뒤 음절보다 낮은 경우를 보였다.

지금까지 살펴보았듯이 경계의 유형은 경계가 없는 강도(zero), 약한 경계 강도(minor break strength), 강한 경계 강도(major break strength)의 3단계로 정하는 것이 통계적으로 의의가 있다. 그리고 지속시간의 정보가 경계 유형의 결정에 중요하게 작용하는 자질이었음을 알 수 있었는데 특히 강한 경계 강도와 경계가 없는 강도를 레이블링하는데 중요한 작용을 하였다. 강도가 약한 경계는 경계앞 음절 장음화가 별로 일어나지 않으면서 $F\phi$의 주요한 변화가 나타나는 부분에서 확인될 가능성이 높다고 판단된다. 4가지 자질에 대한 통계(t-TEST, F-test)결과는 4가지 자질 모두 의미를 갖고 있으며 단지 경계의 유형을 3단계로 구분하는 것이 더 바람직하다는 결과를 제시한다. 결국 기계적 레이블링의 Mlabel-4, Mlabel-5, Mlabel-6과 Mlabel-7, Mlabel-8, Mlabel-9 부분의 재분류가 필요하다는 것인데 Mlabel-6과 Mlabel-9를 하나로 묶어서 약한 경계 강도 유형으로 자리잡게 하는 것이 대안이 될 수는 있겠다. 경계현상에 영향을 미치는 4가지 자질은 향후 ToBI 시스템에 의한 운율 DB 확보에 있어 자동으로 ToBI(Tone and Break Indices)를 레이블링할 수 있는데 기여하리라 생각된다[3].

Ⅲ. 문장내에서의 운율구 경계 추출

3.1 개 요

이 장의 목적은 TTS(Text-to-Speech)[10][11]에서의 운율처리를 위해서 음성신호에서 나타나는 운율구 단위와 텍스트상의 품사정보 사이의 상관관계를 설정하고자 하는 것이다[9]. 일반적으로 이러한 상관관계는 문장구조(통사론적 구조)와 밀접한 관계를 가지고 있으나, 통사적 구조와 운율 구조가 반드시 일치하는 것이 아니다. 따라서 최근에는 음향적으로 실현되는 운율구 경계와 문장구조와의 관계를 통계적으로 찾아내는 연구가 수행되고 있다. 이에 따라 본 연구에서는 156개의 낭독체 문장으로 구성된 운율 DB와 태깅된 텍스트 DB를 바탕으로 bigram, trigram을 이용한 운율구 경계(prosodic phrase)를 추출하고자 한다. 일반적으로 문장상에서 구단위 파싱을 위해 사용하는 파서는 무제한 합성기와 비교해 볼 때 훨씬 복잡한 시스템이며 이 파서로 한 문장을 처리하는데 시간이 많이 걸릴 뿐더러 출력도 복수개의 문장분석 결과가 생성될 수 있다. 이에 제안하는 운율구 경계 추출 방법은 구현이 용이하고 실시간에 만족할 만한 운율구를 파싱할 수 있다는 장점이 있다고 하겠다[1]. 우선 본 연구에서는 13개의 태그 세트를 이용하여 태깅된 텍스트 DB로부터 어절간 bigram 확률을 알아보고, 이를 확장하여 trigram 확률을 구한다. 사용된 bigram, trigram은 가장 기본적인 운율구 경계 강도(break strength)를 예측하는 방법이며, 향후 이의 성능 개선 방법 및 실제 TTS의 적용 여부에 대해 설명하겠다.

3.2 운율구 경계 강도 할당

운율구 경계에서의 운율(지속시간, 억양) 모델링을 위해 우선 운율 DB의 각 어절 경계에서 일어나는 운율 현상을 분석하여 운율구 형성의 자질을 찾고, 이 자질의 통계 결과로부터 자동으로 운율구 경계 강도를 할당하는 기계적 매김(Mlabel)과 음성 전문가의 지각(perception)에 의한 지각적 매김(Plabel)으로 운율구 경계 강도를 정의하였다. 기계적 매김에 의해서는 9개의 운율구 경계 강도로, 음성학 전문가에 의해서는 4개의 운율구 경계 강도로 나누었다. 기계적 매김과 음성학 전문가에 의한 매김 중 약하게 또는 강하게 운율구가 형성됨에 따라 약한 경계 강도와 강한 경계 강도로 정의하고, 텍스트상에서 이 2단계의 경계 강도를 자동으로 할당하였다. 특히 주요하게 나눌 수 있는 약한 경계 강도와 강한 경계 강도는 기계적 매김과 음성학 전문가에 의한 매김이 거의 일치한다. 따라서 2 단계로 단순화해 레이블링한 데이타베이스는 일관성이 있으며 훈련시 데이타의 특성을 잘 반영하리라 생각된다.

3.3 실험에 사용된 텍스트 DB 및 태그 세트

대량의 데이타로부터 통계적인 방법으로 운율구 경계

강도를 추출하고자 이에 필요한 텍스트 DB를 구축하였다. 구축된 텍스트 DB는 운율구 경계현상 분석에 사용된 운율 DB 156 문장과 KBS 뉴스문장 등으로 구성되어 있다. 운율 DB의 텍스트는 약 2,200개의 어절로 구성되어 있으며 정의된 태그 세트로 태깅(tagging)되어 있다. 그리고 KBS 뉴스문장은 훈련용이 약 5,300여개, 평가용이 약 1,200여개로 되어 있으며, 끊어읽기 부분이 표시되어 있다. 이 끊어읽기는 발화시 강한 경계 강도와 유사하다고 가정한 상태에서 trigram을 적용하기 위해 구축되었다. 또한 구축된 텍스트 DB는 어절간 긴밀도, 문장성분 및 형태소 해석이 되어 있다.

텍스트에서 추출한 정보는 품사정보 혹은 품사정보에 의한 문장성분이 될 수 있다. 이 연구에서는 2가지 태그 세트를 정의하여 사용하고 있는데 Group-I은 형태에 의해 주로 결정되는 태그 세트이며 실사와 허사를 합쳐 52개로 이루어져 있다. 하나의 어절이 실사+허사로 태깅된다면 최소한 100개 이상이 되는 태그 정보가 발생하고 이 정보를 이용한 trigram을 적용한다면 $100^3 = 1{,}000{,}000$개의 단일한(unique) trigram이 발생한다. 따라서 52개의 태그 세트를 유사한 유형으로 그룹화하여 태그의 수를 줄였다. 다시 정의한 태그 세트는 표 6과 같다. 그룹화된 태그 세트는 모두 13개로 이로부터 발생할 수 있는 trigram은 $13^3 = 2197$개가 된다.

표 6. 형태에 의한 태그 세트 Group-I

심볼	의미
SU	주격
OB	목적격
EU	관형격
LO	위치격,방편격
ET	비교, 보조, 특수, 접속
TV	호격
PE	종결의미
PC PI	접속의미 내포의미
A1	접속, 시공, 의도 부사
A2	방식, 정도 부사
A3	부정, 의성의태, 연결 부사
KW	관형사, 명사(홀로 쓰일때)

표 7. 문장성분에 의한 태그 세트 Group-II

심볼	의미
SU	주어
PR	서술어
OB	목적어
AN	체언+관형어
AE	용언+관형형 어미
LO	위치어
CO	대비어
QU	인용어
AV	부사어
AB	독립어
IN	방편어

태그 세트 Group-II는 문장의 통사적 분석에 의한 문장성분이 된다. 이 11개로 이루어진 태그 세트는 운율구의 형태를 분석하는데 사용하였다.

3.4 강한 운율구 경계에 의한 운율구 형태

우선 156 문장에 대해 지각적 경계와 거의 일치하는 기계적 경계매김(Mlabel-0, Mlabel-1, Mlabel-2, Mlabel-3)으로 레이블링된 강한 경계 강도를 추출하며, 총 818개의 운율구를 구하였다. 다음의 예는 2장에서 설명한 기준에 의해 한 문장이 2개의 운율구로 나누어진 경우를 보이고 있다.

성교육의 교육 목표를 구체적으로 요약해 보자.

운율구 1

1]성-교육-의	AN(N-N-JPT)
2]교육	AN(N)
3]목표-를	OB(N-OPT)

운율구 2

1]구체-적-으로	IN(N-SF-IPT)
2]요약-해-	PR(N-SF-CT)
3]보-자	PR(AU-PE)

표 8. 운율구당 음절수와 그 빈도

음절수	1	2	3	4	5	6	7	8	9	10	11	12
빈도	19	66	71	45	50	62	82	59	70	63	52	42
음절수	13	14	15	16	17	18	19	20	21	22	23	24
빈도	23	37	13	15	10	10	8	7	4	4	2	4

운율구를 문장성분으로 나타내면 이로부터 문장구조와의 관계를 추측할 수 있다. 발생한 총 818개의 운율구 중 325개의 서로 다른 문장구조 형태가 발생했으며, 이중 3개 이상 같은 문장성분의 연속으로 발생한 운율구는 약 53%에 해당하며 각 운율구당 음절수와 그 빈도는 표 8에 나타내었고, 운율구의 형태는 표 9에 나타내었다. 표 9는 325개중 76가지만 보여주고 있다.

3.5 Bigram을 이용한 운율 경계 강도 예측

Bigram은 어절간 긴밀도를 알아보는 1 차적인 척도이다. 13개의 태그 세트로 발생할 수 있는 가능한 bigram 개수는 169개이며, 실제 훈련용 텍스트 DB에서는 138개의 bigram이 발생하였다. 5,300여개로 구성된 훈련용 텍스트 DB로부터 약한 경계 강도 및 강한 경계 강도에 대한 bigram 확률을 구하고, 이로부터 평가용 데이타 및 운율 DB에 경계 강도 예측을 한 결과는 표 10과 같다. 여기서 Test-I은 평가용 1,241개의 어절에 구성된 태깅된 텍스트 DB이며, Test-II는 운율 DB에 태깅된 텍스트 DB를

표 9. 운율구의 형태

순서	형태	빈도	%	순서	형태	빈도	%
1	AN	61	9.20	2	AV	34	5.13
3	AN+SU	33	4.98	4	SU	27	4.07
5	PR	27	4.07	6	AB	23	3.47
7	AN+AN	23	3.47	8	AN+LO	21	3.17
9	OB+PR	20	3.02	10	AN+OB+PR	20	3.02
11	SU+PR	17	2.56	12	LO	16	2.41
13	CO	13	1.96	14	AN+SU+PR	11	1.66
15	PR+PR	11	1.66	16	OB	10	1.51
17	AN+OB	10	1.51	18	AN+PR	9	1.36
19	AN+AN+SU	9	1.36	20	AN+AN+LO	9	1.36
21	OB+PR+PR	8	1.21	22	AN+OB+PR+PR	6	0.90
23	LO+PR	6	0.90	24	SU+AN	5	0.75
25	AN+AN+AN+LO	5	0.75	26	AN+AN+AN	5	0.75
27	AN+LO+PR	5	0.75	28	AN+IN	5	0.75
29	AN+AN+AN+SU	4	0.60	30	AN+AN+SU+PR	4	0.60
31	IN+PR	4	0.60	32	AN+SU+AV	4	0.60
33	IN	4	0.60	34	AV+PR+PR	3	0.45
35	SU+AV	3	0.45	36	AN+AN+LO+PR	3	0.45
37	AN+AV	3	0.45	38	AN+CO	3	0.45
39	IN+PR+PR	3	0.45	40	AN+OB+AN	3	0.45
41	AN+AN+OB	3	0.45	42	AN+OB+AN+PR	3	0.45
43	AN+LO+AE	3	0.45	44	QU	3	0.45
45	AN+AN+AN+AN	3	0.45	46	AV+PR	3	0.45
47	OB+AE+LO	2	0.30	48	LO+AN	2	0.30
49	OB+IN+AE	2	0.30	50	AN+AN+AN+PR	2	0.30
51	AN+SU+AN	2	0.30	52	LO+AN+LO	2	0.30
53	SU+AN+PR	2	0.30	54	SU+AN+OB+AN+LO	2	0.30
55	AN+LO+IN	2	0.30	56	AE	2	0.30
57	SU+SU	2	0.30	58	AB+AN+AN+SU	2	0.30
59	OB+AN+LO	2	0.30	60	AN+AN+IN	2	0.30
61	OB+AV+PR	2	0.30	62	AN+AN+OB+PR	2	0.30
63	AV+LO+PR	2	0.30	64	AN+LO+OB+PR+PR	2	0.30
65	AN+LO+LO	2	0.30	66	AN+SU+AN+AN	2	0.30
67	OB+AE	2	0.30	68	IN+AE	2	0.30
69	PR+AN	2	0.30	70	AN+AN+PR	2	0.30
71	LO+OB+PR	2	0.30	72	LO+AV	2	0.30
73	AV+AV+PR+PR	2	0.30	74	AN+SU+PR+PR	2	0.30
75	PR+AN+SU+PR	2	0.30	76	SU+AV+PR	2	0.30

나타낸다. 표 10의 correct major break score는 실제 DB에서 강한 경계 강도가 발생했을때 예측도 강한 경계 강도를 나타낼 때를 말하며, 삽입오류(insertion error)는 실제 약한 경계 강도가 강한 경계 강도로 예측되었을 때, 그 오류 개수와 실제 DB에서의 강한 경계 강도의 총수의 비를 나타낸다. Modified correct score는 태깅된 텍스트 DB에서 강한 경계 강도 개수와 약한 경계 강도 개수의 비율이 다를 때를 고려한 값이다[1]. 즉 전체 어절쌍(word pair)의 개수와 약한 경계 강도 개수의 비율을 MINOR BREAK RATIO, 강한 경계 강도 개수의 비율을 MAJOR BREAK RATIO라고 하면,

$$\bullet\ PERCENT\ CORRECT = \frac{\#\ of\ correct\ break\ prediction}{\#\ of\ total\ breaks}$$

$$\bullet\ correct\ major\ break\ score = \frac{\#\ of\ correct\ major\ break\ prediction}{\#\ of\ total\ major\ breaks}$$

$$\bullet\ MAJOR\ BREAK\ RATIO = \frac{\#\ of\ major\ break}{\#of\ total\ breaks}$$

$$\bullet\ MINOR\ BREAK\ RATIO = \frac{\#\ of\ minor\ break}{\#of\ total\ breaks}$$

$$\bullet\ Modified\ Correct\ Score = \frac{PERCENT\ CORRECT\text{-}MINOR\ BREAK\ RATIO}{MAJOR\ BREAK\ RATIO}$$

가 된다.

강한 경계 강도와 약한 경계 강도를 나누기 위해 적절한 임계치(또는 bigram 확률값)의 설정이 필요하다. 이 값에 따라 correct major break 및 삽입오류율이 변화되므로 이 연구에서는 bigram과 trigram의 성능을 서로 비교하기 위해 임계치를 0.5로 일정하게 하였다. 특히 bigram을 사용한 운율구 경계 추출 결과는 속성이 비슷한 태그로 그룹화하는데 이용될 수 있다.

표 10. Bigram의 성능(단위=%, 임계치:0.5)

	Correct major break score	Insertion error	Percent correct	Modified Correct Score
Train	47.9	20.5	82.9	27
Test-I	46.0	22.8	82.9	23
Test-II	38.2	8.4	74.9	30

표 11. Trigram의 성능(단위=%, 임계치=0.5)

	Correct major break score	Insertion error	Percent correct	Modified Correct Score
Train	63.0	23.1	85.9	40
Test-I	58.3	30.0	84.0	28
Test-II	42.8	11.8	75.6	38
*Taylor	70.0	22.0	89.0	50

3.6 Trigram을 이용한 운율 경계 강도 예측

Trigram은 연속하는 3개의 태그 정보를 이용하여 경계 강도를 예측한다. 일반적으로 어절간 경계 강도는 두개의 어절관계에 의해서 결정될 수도 있지만 선행하는 어절에 의한 영향에 따라 더 정확하게 경계가 결정된다. 다시말해서 두개의 어절관계만 볼때 약한 경계 강도가 올 수 있으나 선행하는 단어가 더 밀접하다면 약한 경계 강도가 올 수 있는 두 어절 사이에 강한 경계 강도가 올 수 있다. 따라서 두 어절 및 선행하는 어절간의 관계를 고려한 trigram을 사용하는 것이 더 정확한 경계 강도를 예측하는 방법이 될 수 있다.

실제 trigram을 이용한 경계 강도 예측의 결과는 표 11에 나타나 있다. 이 논문에서는 훈련용 텍스트 DB에서 발생하지 않은 trigram이 평가용 DB에서 발생했을때(약 40개)는 제외하였다.

3.7 실험결과 분석

Bigram과 trigram을 사용한 운율구 경계 강도 예측은 어절의 문장성분 관계를 이용한 방법으로 어절간 긴밀도에 있어 문장성분간의 관계가 중요하게 작용함을 가정으로 하고 있으며, bigram, trigram의 강한 경계 강도 정확률에서도 알 수 있듯이 어절간 긴밀도에 있어 문장성분간의 관계가 어느 정도 일관된 규칙을 가지고 있음을 보여준다. 그러나 삽입오류도 상당히 발생하는 것으로 보아 오류가 빈번히 발생하는 trigram에 대해 태그 세트를 더 세밀하게 분류, 확장할 필요가 있다.

Trigram 확률치를 이용한 운율구 경계 강도 예측의 성능은 영국의 에딘버러 대학의 Paul Taylor가 발표한 연구결과와 간접적으로 비교될 수 있다[1]. Taylor의 경우 18,131개의 훈련용 데이타와 6,420개의 평가용 데이타를 이용하였으며, trigram 외 phrase distance probability을 이용하여 운율구 경계 예측 성능을 향상시켰다.

Ⅳ. 결 론

본 연구에서는 첫번째로 운율구 경계를 자동으로 추출하고자 할 때 인간의 지각과 관련하여 어떠한 자질을 이용하는 것이 좋은가 하는 점을 밝혀 보았다. 운율구 경계의 유형은 크게 경계가 없음, 강/약한 경계 강도의 3단계로 정하는 것이 통계적으로 의의가 있으며 지속시간의 정보가 경계유형의 결정에 중요하게 작용하는 자질이었음을 알 수 있었다. 운율구로 경계를 나누기 위해 지속시간, 억양의 변화량 및 휴지를 고려한 기준은, 이 기준에 의해 경계지어진 단위가 의미적 경계로 나누어지는 바 상당히 타당한 기준이라고 생각되나 일부 단위는 오류를 보이고 있어 좀 더 세밀한 기준이 마련되어야 할 것이다. 즉 운율구 경계의 자동 레이블링을 위해서는 자질집합을 보다 확실하게 정의하는 것이 앞으로의 과제이며 가능하다면 보다 많은 자질들을 고안하는 작업도 필요하다.

두번째, 운율구 경계현상 분석결과를 바탕으로 경계 강도를 크게 약한 경계 강도와 강한 경계 강도로 나누고, 텍스트상에서 경계 강도를 각 어절 경계에 자동 할당하는 방법을 소개하였다. Bigram, trigram에 의한 운율구 경계에서의 경계 강도 할당은 성능에 있어 아직은 높은 오류를 보이고 있지만 기본적으로 운율 DB의 확보와 태그 세트의 보완 및 운율구내의 음절개수 정보를 이용한다면 좀 더 정확률을 높일 수 있다.

참 고 문 헌

1. Eric Sanders and Paul Taylor, "Using Statistical Models to Predict Phrase Boundaries for Speech Synthesis," in *EUROSPEECH'95* Spain, pp. 1811-1814, 1995.
2. Ostendorf, "Parse scoring with prosodic information: an analysis and synthesis approach," in *Computer Speech and Language*, pp. 193-210, JULY 1993.
3. Wightman, Ostendorf, "Automatic Labeling of Prosodic Patterns," in *IEEE Trans. on Speech and Audio Processing* VOL. 2, NO. 4, pp. 469-481, OCTOBER 1994.
4. 성철재, "한국어 리듬의 실험음성학적 연구", 서울대학교 박사논문, pp. 57-118, 1995.
5. *Lehiste, I., "The Perception of Duration within Sequences of Fϕ* Intervals," Journal of Phonetics 7, pp. 313-316,1979.

6. Strangert, E. and Zhi, M.J., "Pause Patterns in Swedish : A Project Presentation and Some Data," STL-QPSR 1/1989, pp. 27-31, 1989.
7. 성철재, 김상훈, "경계(Boundary) 신호의 지각적/음성학적 분석-운율구 단위설정과 관련하여," 한글학회, 한글 232, 1996.
8. F. Emerard, L. Mortamet, and A. Cozannet, "Prosodic processing in a text-to-speech synthesis system using a database and learning procedures," in *Talking Machines : Theories, Models, and Designs*, North-Holland, pp. 225-254, 1992.
9. Andrew J. Hunt, "Syntactic Influence on Prosodic Phrasing in the Framework of the Link Grammar," in *EUROSPEECH'95* Spain, pp. 997-1000, 1995.
10. J.C.Lee, and S.H.Kim, and Minsoo Hahn "Intonation Processing for Korean TTS Conversion Using Stylization Method," in *Proc. ICSPAT95*, pp. 1943-1946, 1995.
11. S.H.Kim and J.C.Lee "Korean Text-to-Speech System Using TD-PSOLA," in *Proc. SST94*, pp. 587-592, 1994.

▲김 상 훈(Sang Hun Kim) 1967년 10월 1일생

1990년 2월 : 연세대학교 전기공학과 학사
1992년 2월 : KAIST 전기 및 전자공학과 석사
1992년 3월 : 한국전자통신연구소 음성언어연구실 연구원

▲성 철 재(Cheol Jae Sung) 1965년 7월 19일생

1988년 2월 : 서울대학교 언어학과 학사
1991년 2월 : 서울대학교 언어학과 석사
1995년 2월 : 서울대학교 언어학과 박사
현재 : 충남대학교 언어학과 교수 및 한국전자통신연구소 음성언어연구실 위촉연구원

▲이 정 철(Chung Cheol Lee)

현재 : 한국전자통신연구소 음성언어연구실 선임연구원