1998년도 한국음향학회 학술발표대회 논문집 제 17 권 제 1(s) 호

연속음성 인식 및 합성을 위한 운율 경계강도 예측 모델

강 평 수*, 김 진 영*, 홍 재 희**
전남대학교 전자공학과*
송원대학교 전자공학과**

Prosody Boundary Index Prediction Model for Continuous Speech Recognition and Speech Synthesis

Pyung-Su Kang*, Jin-Young Kim*, Jae-Hee Hong**
Dept. of Electronics Engineering, Chonnam Univ., Kwangju, Korea, 500-757*
kpyung@dsp.chonnam.ac.kr kimjin@dsp.chonnam.ac.kr
Dept. of Electronics Engineering, SongWon Univ., Kwangju**

요약

본 연구에서는 연속음 인식과 합성을 위한 운율 경계 강도 예측 모델을 제안한다. 운율 경계 강도는 음성 합성에서는 운율구 사이의 휴지기의 길이 조절로 합성음의 자연도에 기여를 하고 연속음 인식에서는 인식과정에서 나타나는 후보문장의 선별 과정에 특징변수가 되어 인식률 향상에 큰 역할을 한다. 음성학적으로 발화된 문장은 큰 경계 단위로 볼 때 운율구 형태로 이루어졌다고 볼 수 있으며 구의 경계는 문장의 문법적인 특징과 관련을 지을 수 있게 된다. 본 논문에서는 운율 경계 강도 수준을 4로 하고 문법적인 특징으로는 트리구조 방법으로 결정된 오른쪽 가지의 수식의 깊이(rd)와 link grammar방법으로 결정된 음절수(syl), 연결거리(torig)를 bigram 모형과 결합하여 운율적 경계 강도를 예측한다. 예측 모형으로는 다중 회귀 모형과 Marcov 모형을 제안한다. 이들 모형으로 낭독체 200 문장에 대해 실험한 결과 76%로 경계 강도를 예측할 수 있었다.

Ⅰ. 서론

합성음의 자연도를 향상시키기 위해서 운율 정보는 길이, 세기, 피치에 주로 관심이 집중되어 비교적 많은 성과가 얻어졌으나 구문론적인 면과 결합하여 운율구 단위 정보를 이용한 연구는 비교적 적다[1]. 또한 연속음 인식에 운율 정보를 적용한 예는 거의 없는 상태이다. 국어와 같이 자유로운 문장 순서를 가진 언어의 경우는 일반적인 구문 구조 문법보다 종속 문법을 통하여 분석할 때 더 효과적인 것으로 알려져 있다[2]. 이러한 종속 문법적인 해석 방법을 통하여 얻어지는 문장 성분간 지배와 종속 관계를 청각적으로 측정된 운율 경계 강도와의 관계를 살핌으로써 음성 공학에 이용하고자 한다. 종속 문법적인 해석의 한 방법으로 Hunt는 구문론적인 표현이 간단한 link grammar를 이용하였다[3]. 이는 문법적으로 관련있는 문장 성분을 연결하고 그 관계를 설명하는 레이블을 부여하였다. 이 경우 link에 의해 표현된 표면-문법 관계는 문장 내 운율 경계 강도를 가지고 운율구를 이룬다고 보고하고 있다. link grammar 방법을 이용하여 link가 이루어지는 문장 성분들 간의 거리와 POS를 대상으로 하여 운율 경계 강도를 예측한 연구도 있다[2]. 이는 구문 구조 정보, POS 레이블링, 성분 길이를 Marcov 모형화하여 Viterbi algorithm으로 문장내 운율구의 경계강도를 결정하였다. 구문론적인 정보를 보다 정확히 관찰하기 위하여 순수한 문장 성분 레이블링 방법도 시도되었다[4]. 즉 다양한 구문론적인 현상을 밝히기 위해 레이블의 수를 늘려서 연구를 하여 비교적 정확한 구문론적인 관계와 운율구와의 관계를 밝혔으나 레이블의 수가 늘어남으로 인하여 대량의 corpus가 필요하고 문법적인 요소와 운율구 형성에 대한 전체적인 시각을 방해하는 한계를 가지게 되었다. 다른 연구로는 bigram과 trigram만으로 경계 강도를 예측하기도 했으나 문장 내 문법적인 상관 관계를 고려

하지 않아서 높은 에러율을 보이고 있다[5]. 본 연구에서는 종속 문법 해석 방법인 link grammar 방법, 트리 구조 해석 방법, 그리고 bigram 방법을 다중 회귀 분석 방법과 Viterbi 과정을 이용하여 운율 경계 강도를 예측한다. 이를 위해 운율구를 결정하는 구문론적인 변수를 찾아 내며 이들 변수를 이용하여 운율구를 예측하는 모형을 제안하고 실험을 한다. 본 연구를 통하여 구문론적인 정보와 운율 정보를 연결시킬 수 있게 하여 합성음의 휴지기의 위치와 그 길이를 결정할 수 있게 한다. 또한 문장을 구문 단위로 처리함으로써 비교적 긴 문장이나 복잡한 문법 구조를 간단하게 처리하여 동일한 음소열을 가진 다른 어의의 문장들에 대한 정확한 연속음 인식을 가능하게 해 준다.

Ⅱ. 운율구를 형성하는 특징 변수와 휴지기 강도

일반적으로 발화자는 일상의 대화에서 문법적인 틀을 생각하며 문장을 발음하기 보다는 의미 전달을 의식하여 단어 열을 결정하게 된다. 그러한 화자의 의사 전달 과정에서 강조, 삽입, 간접 전달을 실으면서 특정 위치에서 일정한 길이의 휴지를 가져 오게 된다. 그러한 휴지는 운율구를 형성하게 되며 운율구 자체가 독립적으로 특정 피치 패턴의 형성, 운율구 내 끝 모음의 삼음화와 같은 음향학적인 특징을 가지게 된다. 일반적으로 모든 화자에 공통적인 운율구 형성 규칙을 찾는다는 것은 어려운 일이다. 하지만 의미 전달과정에서 나타나는 휴지의 규칙을 문법적으로 관련있는 문장 성분을 연결함으로써 얻어지는 변수들과 트리 구조로부터 얻어지는 변수와의 관련성을 관찰함으로써 예측 모형을 만들고자 한다.

① Link grammar와 트리 구조로부터 얻어지는 특징 변수

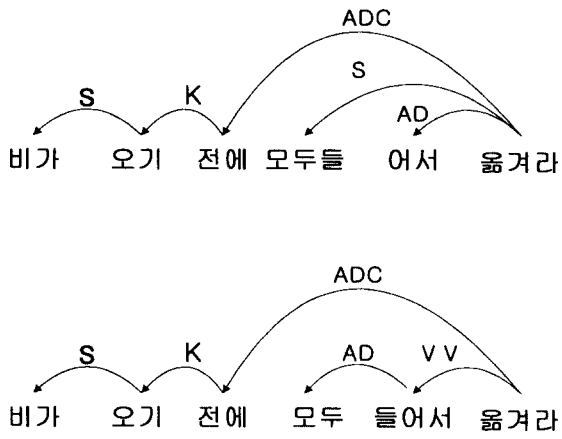

그림 1 . link grammar 방법으로 분석한 두 문장

표 1.문장 1에 대한 구문 분석

문장1	Word	dep	rd	link	par	torig	syl	pau
0	비가	3	0	1	S	0	2	0
1	오기	2	0	2	K	0	2	0
2	전에	1	1	5	ADC	2	7	2
3	모두들	1	1	5	S	1	5	1
4	어서	1	0	5	AD	0	2	0
5	옮겨라	0						

표 2 . 문장 2에 대한 구문 분석

문장2	Word	dep	rd	link	par	torig	syl	pau
0	비가	3	0	1	S	0	2	0
1	오기	2	0	2	K	0	2	0
2	전에	1	2	5	ADC	2	7	2
3	모두	2	0	4	AD	0	2	0
4	들어서	1	0	5	VV	0	3	0
5	옮겨라	0						

그림 1은 link grammar 방법으로 분석한 문장이고 표 1과 2는 link grammar 방법으로 구문 분석한 결과이다. Hunt의 경우 link grammar를 형성하는 구문론적 변수로써 주로 연결관계가 있는 문장 성분사이의 거리를 세는 방식을 이용하였으나 국어의 경우는 영어의 경우와 달리 후위 수식하는 경우가 없기 때문에 많은 변수들이 적용되지 않고 국어에 맞는 변수들만 채택하고 트리구조 해석 방법으로 결정된 오른쪽 가지의 수식의 깊이(rd)를 변수로 채택하였다. 이와 같은 방법으로 분석된 문장은 한 눈에 문장의 구조를 알아볼 수 있으며 또한 휴지기 강도와 나타난 구문 변수들의 크기가 매우 비슷함을 관찰할 수 있다.

각 구문 변수들에 대한 구체적 설명은 다음과 같다.

rd : 오른쪽 수식 가지의 깊이
parse : 단어의 연결 관계를 나타내는 문장 성분
pause : 휴지기 강도
to.right : 연결 고리 안에 나타나는 단어의 수
syl : 연결 고리 안에 나타나는 음절의 총 수

그림 1의 두 문장은 동일한 음소열을 가진 다른 어의의 분장이다. 그림 1와 위 분장의 경우 '모두들'에서 끊어 읽게 되는 반면 아래 문장의 경우 '모두'와 '들어서'를 이어서 읽는 경향이 있다. 연속음 인식에 이들 문장을 적용하면 두 문장은 대체적으로 오인식을 하게 된다. 하지만 위에서와 같이 운율구 경계 위치를 결정하게 되면 오인식의 가능성은 훨씬 줄어들 것을 예상할

수 있다. 또한 강한 운율구를 느끼게 하는 '전에' 부분을 운율구의 경계로 삼아서 하나의 독립된 인식 단위로 만들 수 있게 된다.

② 구문론적 특징 변수들과 휴지기 강도

위의 방법에 의해 각 구문론적인 특징 변수들의 크기와 휴지기 강도가 어떤 관계가 있는 지를 다음 그림 2가 보여준다. 즉, rd와 to.right 변수들은 그 크기에 따라 휴지기 강도가 Square Root모형으로 증가하며 syl의 경우 선형적으로 증가하다가 일정한 값에 이르면 포화 상태에 이른다는 것을 관측할 수 있다.

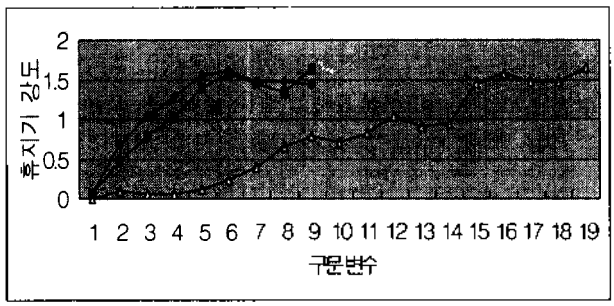

그림 2. rd, to.right, syl 과 휴지기 강도

Ⅲ. 운율구를 예측하기 위한 통계적인 모델

구문론적인 특징들인 수식의 깊이(rd), 음절수(syl), 연결거리(to.right)들이 단조증가 함수 형태를 가짐을 관측하고 이들을 다중 회귀 모형을 통하여 적합시킨다. 회귀 모형으로는 선형 모형, 로그 모형, Square Root 모형을 적합시킨 결과 Square Root 모형이 가장 높은 결정 계수를 가지고 경계 강도 구분현상을 가장 잘 설명하고 있음을 보여 이를 예측 모형으로 채택하였다 (식 1).

$$B.I = a*\sqrt{rd} + b*\sqrt{tor} + c*syl \quad (식\ 1)$$

또한 구문론적인 특징 변수들이 관련되어 있다는 가정으로 그 패턴 별 확률을 구하고 Bigram 확률과 결합하여 Viterbi Algorithm을 수행한다. 패턴은 수식의 깊이정도, 음절수 , 연결 거리 정도를 조합하여 만들어 낸다. 즉,

$$\Phi(w_{1\ldots n-1}) = \arg\max_{b_{1\ldots n-1}} [P(b_{1\ldots n-1}|w_{1\ldots n-1})*P(bigram_i)]$$

$$= \arg\max_{b_{1\ldots n-1}} [P(w_{1\ldots n-1}|b_{1\ldots n-1})*\frac{P(bigram_i)}{P(w_{1\ldots n-1})}]$$

$$\cong \arg\max_{b_{1\ldots n-1}} [P(w_{1\ldots n-1}|b_{1\ldots n-1})*P(bigram_i)]$$

(식 2)

여기에서 b_i는 i번째 단어에서 휴지기 경계강도이고 w_i는 i번째 단어 경계에서 결합 구문 변수들의 패턴이며, $bigram_i$ 는 i번째 단어 경계에서의 문장 성분에 대한 bigram이다. 특징 변수들의 값은 패턴의 개수가 지나치게 커지는 것을 막기 위해서 변수들을 비규칙적으로 양자화한 후 모델 적용을 했다. 문장 성분의 종류는 16가지로 하였으며 문장 성분에 대한 레이블링은 자동화 방법을 택할 수도 있지만 정확한 구문 분석을 위하여 손으로 직접 구문 분석을 하였다. 녹음 문장은 전문 아나운서의 음성으로 이뤄진 낭독체 음성 데이터 200문장을 이용하였으며 휴지기 강도를 매기기 위한 청취 테스트는 연구자 2명이 실시하였다.

Ⅳ. 실험 결과 및 평가

경계 강도의 수준도 예측 모델의 성능에 큰 영향을 미친다. 영어의 경우 7 단계까지 시도하고 있으나 국어에서 일반적인 수준으로 사용되는 4 단계를 채택했다. 운율 경계 강도를 연속음 인식에 적용할 경우 높은 경계 수준을 요구하지 않으므로 에러율을 줄이기 위해 3 수준도 사용하였다. 표 3은 다중 회귀 모형으로 운율 경계 강도를 예측한 결과이고 표 4는 구문 변수들의 결합 확률 분포와 bigram을 이용하여 예측한 결과이다. 다중 회귀 모형으로 분류한 에러율은 0.28이며 각 휴지기 강도 별 분류율은 [0.901 , 0.403 , 0.428 , 0.459]와 같은 결과를 얻었으며, Viterbi algorithm 의해 얻은 에러율은 0.24 이며 각 휴지기 강도별 분류율은 { 0.911 , 0.320 , 0.564 , 0.712}이다.

표 3 . 다중 회귀 모형에 의한 휴지기 강도 분류

관측 \ 예측	0	1	2	3
0	1677	156	20	6
1	181	176	62	18
2	47	142	198	76
3	7	29	104	118

표 4 .Viterbi algorithm에 의한 휴지기 강도 분류

관측 \ 예측	0	1	2	3
0	1694	105	59	2
1	185	140	102	10
2	51	73	261	78
3	6	10	58	183

운율 경계 강도 1의 경우는 비교적 에러율이 높은데 이는 청취 테스트의 과정에서 생긴 오류라고 생각된다.

운율적인 경계 강도를 예측하기 위해서는 다양한 구문론적인 현상을 반영하면서도 문장 성분들간의 연결 관계를 반영하여야 한다. 또한 구문론적인 특징들이 서로 관련되어 있다는 가정도 가지고 있어야 한다. 이러한 가정을 가지고 회귀모형을 적용하여 경계 강도를 예측할 경우 72%의 예측률을 Viterbi 수행한 결과는 76%의 예측률을 보였다. Bigram을 추가할 경우 큰 개선이 이뤄지지 않은 것은 모든 Bigram이 운율 경계 강도 예측에 도움이 되는 것은 아니기 때문인 것으로 보인다. 실제로 보다 정밀한 구문 분석을 위해 레이블링의 개수를 늘려 휴지기 강도를 예측한 연구에 의하면 특정한 구문론적인 연결 관계만이 강한 운율 경계를 가진다고 보고하고 있다[4]. 즉, bigram이나 trigram 부분만을 따로 독립시켜 운율 경계 강도와의 관련성을 찾은 후 다른 구문론적인 특징들과 결합하는 방법을 모색해 볼 필요가 있다.

V. 결론

위의 실험 결과로부터 구문론적인 특징 변수들을 이용하여 휴지기 강도를 예측할 수 있음을 알 수 있다. 그림 1은 위의 결과를 연속음 인식에 적용할 수 있는 한 예이다. 통계적인 방법으로 연속음 인식을 할 경우 인식될 후보 문장을 여러 개 선출한 후 구문 정보로부터 휴지기 강도를 예측하여 가장 높은 가능성을 가진 문장을 선택하게 되는 것이다. 이처럼 입력된 음성 신호로부터 자동으로 운율 경계 강도가 결정되어 질 수 있다면 구문론적 관계와 음향학적인 관계를 설명할 수 있는 모형을 통하여 음성합성과 연속음 인식의 효율을 높일 수 있으리라 예측할 수 있다. 좀 더 많은 음성 데이터와 다양한 화자를 대상으로 위의 모형이 적용되는지는 앞으로의 과제이다. 이러한 구문론적인 정보를 이용하여 음성 합성에 이용한 연구는 많이 있으나 음성 인식에 적용한 연구는 드물다. 위의 결과는 비교적 긴 문장이나 자유로운 어순으로 발화된 문장을 운율 경계 강도에 상응하여 쪼갠 후 각 쪼개진 운율구 단위로 구문론적인 정보와 관계를 짓는다거나 그 단위로 인식하는 방법으로 연속음 인식 모형을 만드는 것을 가능하게 한다.

※ 본 논문은 정보통신부 97년도 대학 기초 지원 사업비에 의해 이루어졌습니다.

7. 참고 문헌

[1] A.J.Hunt, A generalized Model for Utilising Prosodic Information in Continuous Speech Recognition, Speech Technology Research Group, University of Sydney, 1995.

[2] Y.J.Kim, S.H.Lee, Y.H.Oh, Relationship Between Prosodic features and Defendency Relation. ICSP'97, 1997.

[3] A.J.Hunt, Models of Prosody and Syntax and their Application to Automatic Speech Recognition, Ph.D thesis. University of Sydney, 1995.

[4] 김선미, "한국어의 리듬 단위와 문법 구조-음성 합성에서 리듬 구현의 자연성 향상을 위한 음성·언어학적 연구", 문학 박사 학위 청구 논문, 서울 대학교, 1997.

[5] ETRI, "다중 매체 환경에서의 대화체 음성 번역 통신 기술 개발", 정보 통신부,1996.