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

한국어 문음성 변환기의 음운지속시간 제어에 관한 연구

A Study on Segmental Duration Control for the Korean TTS

김인영 , 이양희

동덕여자대학교 전자계산학과

E-mail : inyoung@dongduk.ac.kr, yhlee@www.dongduk.ac.kr

요 약

자연스러운 한국어의 음성합성을 위해서는 음운의 지속시간의 제어가 매우 중요하다. 본 연구에서는 POW3848 어절에 대한 음성 데이터에 대해 음운 세그먼트, 음운 라벨링, 품사 태깅을 행한 음성 데이터베이스를 구축하여 한국어 음운의 지속시간을 변화 시키는 시간 특징을 통계적으로 분석하였다. 이 시간 특징들 중 변화 폭이 큰 요인들을 제어요소로 각 음운의 고유길이를 최대한 배제하고 단지 음운 발성 환경의 영향에 의한 지속시간 변화만을 고려하는 정규화 지속시간에 대한 회귀트리로 한국어 음운 지속시간을 모델화 하였다. 제안된 음운 지속시간 모델을 실시간 제어 알고리즘으로 구현하여 평가한 결과, 음운 지속시간 예측오차의 88% 정도가 25ms이내 이었고 예측치와 관측치 간의 다중 상관계수는 0.92정도로 평가되어, 제안된 모델의 타당성이 입증되었다.

1. 서 론

본 논문에서는 일반적인 음운지속시간의 변화 요인을 분석하고 정도 높은 지속시간 제어 규칙을 생성하기에 충분한 POW(Phonetically Optimized Word) 3848 어절 음성데이터에 대해 음운 세그먼트, 음운 라벨링, 품사 태깅을 행한 음성데이터 베이스를 구축하고, 이 데이터베이스를 사용하여 일반화된 음운 지속 시간을 통계적 방법에 의해 자동적으로 모델화한다. 여기에서 사용되는 데이터베이스에는 음운이 편중되어 있지 않고, 또한 음운 환경 및 문법정보를 포함하고 있어 일반화된 음운 지속시간 모델화에 충분한 데이터라고 생각된다. 이러한 음성 데이터를 통해서 한국어 음운의 지속시간을 변화 시키는 시간 특징을 통계적으로 살펴보고, 분석된 시간 특징들 중 변화 요인이 큰 특징들을 제어요소로 사용하여 한국어 음운 지속 시간을 모델화한다. 이 때 규칙의 최적화 및 규칙의 자동생성을 위해 회귀 트리 모델을 사용하였다. 이 방법에서 통계적 방법으로는 제어요소 간의 의존 관계를 표현할 수 있다. 2절에서는 음운지속시간 제어규칙을 생성하기 위한 음성 데이터베이스에 대해 기술하고, 음운지속시간의 변화요인에 대해 분석결과를 기술한다. 또한 각 음운의 고유 길이를 최대한 배제하여 음운 환경에 의해서만 영향을 받는 지속시간 변화요인을 적용하기 위해 각 음운의 Zscore에 대한 회귀 트리 모델화에 대해 설명한다. 3절에서는 세그먼트지속시간 예측 알고리즘을 구현하며, 그 알고리즘 타당성 평가에 대해 기술한다.

2. 음운 지속시간 제어규칙 생성

2.1 음성 DB 구축

통계적인 방법으로 일반화된 규칙을 생성하기 위해서는 다양한 경우를 포함하는 많은 양의 데이터가 요구된다. 보다 일반적이며 정교한 음운지속시간 제어 모델을 생성하기 위하여, 다양한 음운 환경을 고려하는 충분히 많은 자연음성을 분석하여 음운지속시간 변화에 영향을 미치는 요인을 추출하여야 한다. 음운 지속시간을 변화 시키는 요인을 크게 음운 환경과 문법적인 요인으로 나누어 생각할 수 있다. 음성DB는 음운환경에 의한 변화를 분석하기

위해 음운 단위로 나누어 세그먼트 되어야 하고 문법적인 요인에 의한 변화를 분석하기 위해 품사 태깅이 필요하다. 본 연구에서는POW(Phonetically Optimized Word) 3848 어절을 남성, 여성화자 각각 1명이 발성한 음성 데이터를 음운별로 세그먼트, 음운 라벨링 및 음운별 품사 태깅하였다. 음성 데이터 베이스는 다음과 같은 정보를 포함한다.

- 음운레벨 세그멘트(시간정보)
- 음운 라벨링(음운 기호)
- 어절에 대한 음운레벨 품사 태깅(품사)

2.2 음운지속시간 변화에 대한 통계적 분석

구축된 음성 데이터베이스를 이용하여 세그먼트의 지속시간 변화에 크게 영향을 미치는 요인을 분석하기 위하여 통계적인 방법을 사용한다. 각 음운의 고유지속 시간의 영향을 배제시킨 순수한 음운환경에 의한 세그먼트 지속시간 변화 요인을 분석하기 위하여 본 연구에서는 식(1)과 같은 Zscore를 사용하여 음성DB내 세그먼트들의 지속시간을 정규화 하였다.

```
Zip  =  (Xip - Mp) / SDp          ------ 식 (1)
 Zip  : 음운 p의 지속시간에 대한 i번째
        세그먼트의 관측치
 Mp   : 음운 p의 지속시간에 대한 평균치
 SDp  : 음운 p의 지속시간에 대한 표준편차
```

1) 조음 양식에 의한 영향

조음양식이나 조음 위치가 유사한 음운들의 정규화 분포를 고려하여 정규화 음운 지속시간이 유사한 분포를 갖는 음운들을 22종류로 분류하였다. 이렇게 분류된 조음 양식이나 조음 위치에 따른 음운들의 정규화 지속시간이 유사하기 때문에 음운의 지속시간 변화에 유사하게 영향을 미친다.

2) 음절 유형에 의한 영향

우리말의 음절유형을 다음의 8가지로 분류할 수 있다. (CV, CVC, V, VC, SV, SVC, CSV, CSVC) . 음절유형에 따라 분석한 결과 모음의 지속시간이 변화한다. 종성이 있는 CVC, VC, SVC, CSVC의 음절 유형에서 모음의 음운 지속시간이 짧아지는 경향이 있고 초성 없는 음절유형에서는 모음지속시간이 평균보다 길어지는 경향이 있다. 또한 초성을 갖으며 종성이 없는 음절 유형에서는 평균 모음 지속시간을 유지하는 경향이 있다.

3) 어절 내 음절수에 의한 영향

음운수에 의한 모음의 지속시간 변화는 음운의 수가 1-5개 까지는 음운수에 따라 음운지속시간 변화가 크게 감소하며, 6-21개까지는 음운수에 따른 지속시간이 작게 변화하는 경향이 있다. 이러한 현상은 어절 내 음운 수가 음운 지속시간의 변화에 큰 요인이 되고 있음을 나타낸다.

4) 어절 내 음절 위치에 의한 영향

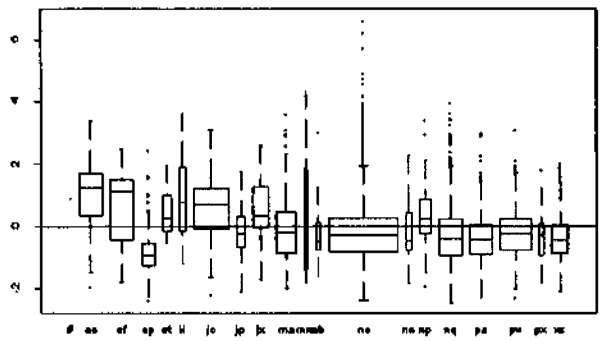
7개 음운을 갖는 어절에 있어서 모음의 위치에 따라 모음의 지속시간이 규칙적으로 변화하지 않았다. 따라서 이러한 요인은 지속시간 제어 요소로는 부적합하다.

5) 앞,뒤 인접음운에 의한 영향

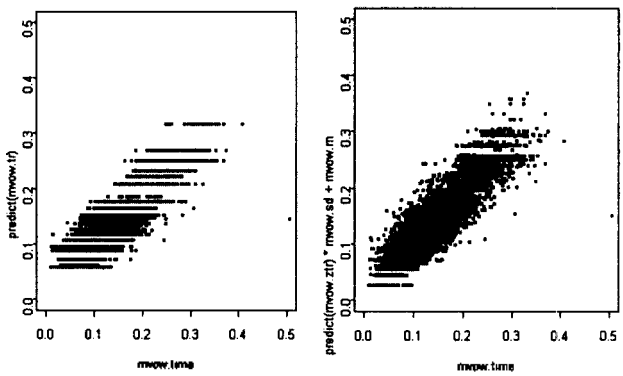
음운 발성 환경에 따른 지속시간 변화를 분석하기 위하여 고려하는 앞 음운과 뒤 음운의 영향을 분석한다. 이때 전후 음운은 조음 양식과 위치로 분류하여 분석한다. 앞 음운에 의한 모음의 정규화 지속시간의 분포와 뒤 음운에 의한 모음의 정규화 지속시간의 분포를 비교해보면 뒤 음의 영향이 앞 음의 영향보다 크게 나타남을 알 수 있었다. 따라서 뒤 음운의 영향은 음운지속시간 제어 규칙을 생성하는데 가장 중요한 특징 요소가 된다.
특히 V_C, U_C, BDG, ktpc들의 앞에 오는 음운(모음)의 지속시간이 짧아지고, 또한 PAUSE(#)의 앞에 오는 음운(모음) 즉 마지막음운지속시간은 길어짐을 알 수 있다.

6) 문법적 요소에 의한 영향

문법적인 요인에 대한 음운지속시간을 분석하기 위하여 한국어의 품사를 다음 [그림 1]과 같이 19품사로 분류하였다. [그림 1] 에 나타난 것과 같이 품사별 모음의 지속시간 변화를 분석한 결과는 다음과 같다. 품사가 ec (연결어미), ef (종결어미), jc (격조사), np (지시대명사)의 경우 음운지속 시간이 길어지는 반면, ep(선어말어미), 고유명사를 제외한 내용어인 경우에는 음운 지속시간이 짧아지는 경향이 있다.



[그림 1] 품사별 정규화 음운지속시간



[그림 2] 예측치와 관측치간의 분포

2.3 정규화 회귀트리 구현

음운의 고유지속 시간의 영향을 배제시키고 순수한 음운환경에 의한 세그먼트의 지속시간을 예측하기 위하여 각 세그먼트의 지속시간을 Zscore로 정규화 하였다. 각 세그먼트의 정규화 지속시간은 음운의 고유지속시간을 제외한 음운환경에만 의존하여 변화하게 된다. 따라서 지속시간 변화 요인에 의해서만 분류되기 때문에 보다 정교하게 예측이 가능하도록 정규화 지속시간에 대해 회귀트리로 모델화 하였다.

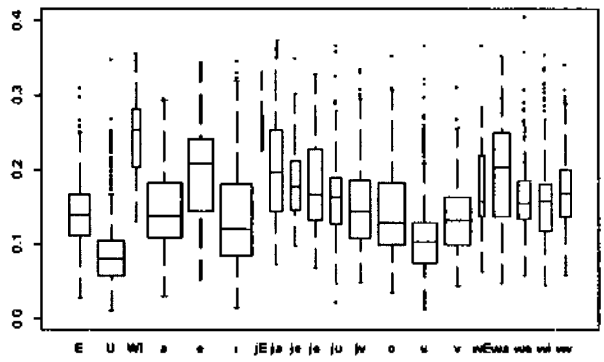
2.4 생성된 규칙의 평가

생성된 규칙의 타당성을 확인하기 위하여 관측치와 예측지간의 오류정도를 평가하고 오류 분석을 행한다. 이 때 관측치와 예측치간의 오류정도를 다중상관 계수로 평가한 결과는 [표 1]과 같다.

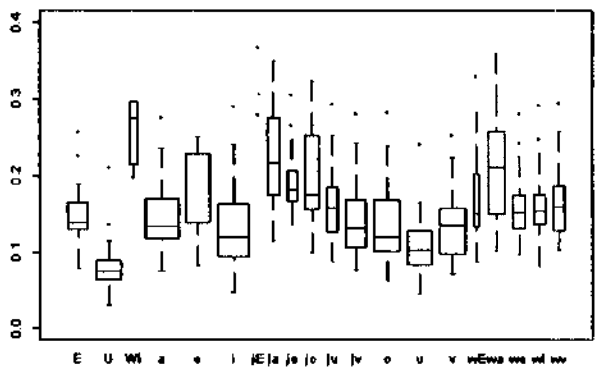
[표 1]

	지속시간 회귀트리	정규화된 회귀트리
다중상관 계수	0.921(남성) 0.904(여성)	0.929(남성) 0.906(여성)
예측오차 25ms 이내	87.0%(남성) 82.8%(여성)	88.2%(남성) 84.6%(여성)

정규화된 지속시간에 의한 예측과 지속시간에 의한 예측의 오류 정도를 평가하기 위하여 다음[그림 2]를 살펴보면 정규화된 지속시간의 예측(오른쪽)이 일반 지속시간에 의한 예측(왼쪽)보다 정교하고 오류 정도의 차가 더 작게 나타난다.

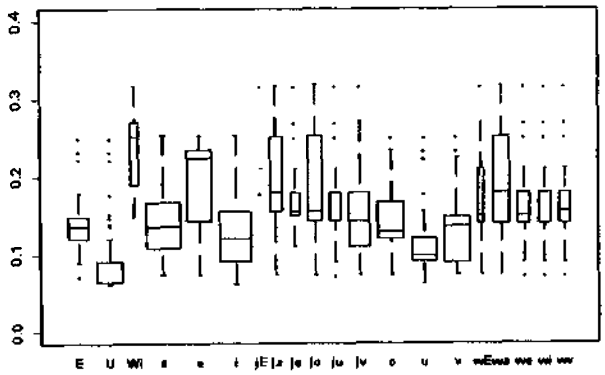


[그림 3] 모음의 음운지속시간 분포



[그림 4] 정규화 지속시간 회귀트리에 의한 예측 모음지속시간 분포

모음의 세그먼트의 관측치의 분포와 일반 회귀트리에 의해 예측치의 분포, 정규화된 세그먼트 지속시간에 의한 지속시간 예측치의 분포는 이들 [그림 3,4,5]에서 보듯이 정규화 트리에 의한 예측이 관측치 분포와 유사하게 나타남을 알 수 있다.



[그림 5] 비 정규화 음운지속시간 트리에 의한 예측 모음 지속시간 분포

3. 세그먼트 지속시간 예측 알고리즘

세그먼트 지속시간 예측 시스템은 다음과 같이 크게 3 부분으로 구성된다.

1) 특징 파라미터 변환 시스템 : 입력으로 부터 음소열과 각 음소의 문법적 정보 즉 품사정보을 받아서 각 음소의 조음 양식 및 위치, 그 음소가 포함된 어절내 음운의 수, 어절내 위치, 음절 유형 ,인접 음운 정보, 품사 정보등을 출력한다. 2) 음운 지속시간 제어규칙 : 1)로부터 출력된 정보들을 회귀트리의 특징 제어 요소로 사용하여 음운 지속 시간 제어 규칙을 생성한다. 즉 음소의 발성 환경에 대한특징 열을 입력으로 하여 회귀트리로부터 정규화 지속시간을 예측한다. 3) 세그먼트 지속 시간 예측 : 예측된정규화지속시간으로부터 세그먼트 지속시간으로 변환에서는 정규화 지속시간을 다음 식 (2)와 같은 방법에 의하여 세그먼트 지속시간을 구한다.

$$DURip = Mp + (Zip \times SDp) \quad \text{----- 식 (2)}$$

여기에서,

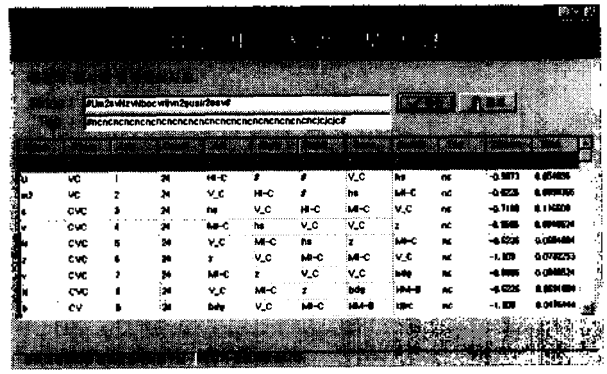
DURip : p음운의 i번째 세그멘트의 예측 지속시간

Mp : 음운p의 평균지속시간

Zip : p음운의 i번째세그멘트의 예측 정규화 지속시간

SDp : 음운p의 지속시간의 표준편차

세그먼트 지속 시간 예측 알고리즘 구현의 프로그램 실행 예는 다음 [그림 6]과 같고 처리 시간도 실시간 처리에 충분한 정도로 판정 되었다. 청취 실험 결과, 어절 단위에서는 자연스러운 음성이 합성 되었으나 문장 단위 음성 합성에서는 다소 부자연스러운 부분이 나타났다. 이는 어절 단위 음성 DB로부터 구한 제어 규칙이기 때문이라 생각된다.



[그림 6] 예측 알고리즘 구현의 프로그램 실행 예

4. 결 론

본 논문에서는 음운발성 환경 및 문법적 요인을 고려한 음운지속시간 변화을 분석하여 다음과 같은 경향을 발견하였다. 조음양식 및 조음위치, 음절의 유형에 따른 고유 지속시간 분포를 갖는다. 어절내 음운수가 증가함에 따라 음운지속시간이 감소하나, 1-6개에서는 다소 크게 감소하고, 그 이상에서는 완만하게 감소한다. 어절내 음운의 위치에 의해서는 지속시간의 변화가 불규칙적이며 앞,뒤 인접음운의 영향에 있어서 뒤음운의 영향이 더 크다. 한편 본 연구에서는 한국어 음운지속시간을 정규화 음운지속시간에 대한 회귀트리로 모델화하였다. 또한 실시간 제어 알고리즘을 구현하여 평가한 결과 음운 지속시간 예측 오차의 88%정도가 25ms이내 이었고, 예측치와 관측치간의 다중상관계수는 0.92정도로 제안된 음운 지속시간 모델의 타당성이 입증되었다. 금후의 연구과제로는 문장 레벨에 있어서 구문정보를 포함하는 음성 데이터베이스구축과 제안된 모델을 문장레벨의 지속시간 예측을 위한 모델로의 확장이다.

[참고 문헌]

[1] Leo Breiman, Jerome H, Friedman, Richard A, Olshen,"CLASSIFICATION AND REGRESSION TREE". Wadsworth, Inc. 1984.

[2] N. Kaiki, K. Takeda and Y. Sagisaka, " Linguistic properties in the control of segmental duration for synthesis" ,Talking machines : Thoeries, Models, Designs, pp 255-263, 1992.

[3] M.D.Riley, "Tree-based modelling of segmental duration", Talking machines : Theories, Models, Designs, pp 265-273, 1992.

[4] W. N. Campbell, "Syllable based segmental duration", Talking machines : Theories, Models, Designs, pp 211-223, 1992.

[5] Y.N.Sung ,B.I. Kim, Y.H. Lee ,"Tree-based Modeling on Korean segmesntal duration," Proceedings of ICSP'97, Vol.1 of 2 , pp 223-228, 1997.

본 연구는 한국전자통신연구소의 1997년도 수탁과제 연구지원비에 의해 연구 되었습니다.