제15회 음성통신 및 신호처리 워크샵(KSCSP '98 15권1호)

음운지속시간의 정규화와 모델링

김인영 , 정지혜 , 이양희

동덕여자대학교 전자계산학과

A Normalization and Modeling of Segmental Duration

Inyoung Kim , Jihye Jung , Yanghee Lee

Dept.of ComputerScience DongdukWomen's Univ.

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

요 약

본 논문에서는 한국어의 자연스러운 음성합성을 위해 280문장에 대하여 남성화자 1명이 발성한 문음성 데이터를 음운 세그먼트, 음운 라벨링, 음운별 품사 태깅하여 음성 코퍼스를 구축하였다. 이 문 음성 코퍼스를 사용하여 음운환경, 품사 뿐만 아니라 구문구조에 의하여 음운의 지속시간이 어떻게 변화하는가에 대하여 통계적으로 분석하였다. 음운지속시간을 보다 정교하게 예측하기 위하여, 각 음운의 고유 지속시간의 영향이 배제된 정규화 음운지속시간을 회귀트리를 이용하여 모델화하였다. 평가결과, 기존의 회귀트리를 이용한 음운지속시간 모델에 의한 예측오차는 87%정도가 20ms이내 이었지만, 정규화 음운 지속시간 모델에 의한 예측오차는 89%정도가 20ms이내로 더욱 정교하게 예측되었다.

1. 서 론

기존의 음운지속시간을 모델화 하는 방법들 중에는 통계적인 방법으로 선형회귀모델(Linear Regresst -ion), 적의 합 모델(Sum of product), CART (Classifica -tion And Regression Tree)를 사용한 모델 등이 제안되었다[1],[2],[3]. 선형회귀를 이용한 모델[1]에서는 제어요소간의 의존 관계를 고려할 수 없다는 단점이 있고, Sum of product 모델[2]은 음운지속시간에 영향을 미치는 요소간의 상호작용을 분석하기 위해 많은 노력을 요한다. 회귀트리를 사용한 모델은 변수들이 갖는 영역을 제어요소의 의존관계에 의해 분할하는 것으로 선형회귀모델과 다르게 비선형성을 표현할 수 있기 때문에, 영어, 한국어등 다양한 언어에 대하여 사용되고 있다[3],[4],[5]. 회귀트리를 사용한 한국어의 음운지속시간 모델에 있어서, 문헌[4]경우 어절 데이터를 사용한 모델로 문 음성에 대한 모델로는 불충분하고, 문헌[5]의 경우 지속시간 변화 요인 만을 고려하지 않고 음운의 고유지속시간까지도 포함하여 회귀트리로 지속시간을 모델화하므로 정교한 지속시간의 예측이 불충분하다. 따라서 본 논문에서는 통계적으로 처리하기에 충분한 문음성 코퍼스를 구축하고, 이 음성 코퍼스를 사용하여 음운의 지속시간을 변화시키는 시간특징을 통계적으로 분석한다. 또한 이 시간 특징들 중 변화폭이 큰 요인들을 제어요소로 각 음운의 고유길이를 최대한 배제하고 단지 음운 발성 환경의 영향에 의한 지속시간 변화만을 고려하는 정규화 지속시간을 회귀트리로 모델화 한다.

2. 음운 지속시간 정규화와 통계적 분석

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.

[표 1] 음성 데이터

데이터	280문장
화자	남성 단일화자
어절수	4,856
음운수	33,723
	자음 : 18,359, 모음 : 15,364

본 연구에서는 [표 1]과 같은 문음성 데이터를 음운세그먼트, 음운라벨링, 음운별 품사 및 문법정보를 태깅한 음성 데이터베이스를 구축한다. 이 문음성 데이터 베이스는 다음과 같은 정보를 포함한다.

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

2.2 음운지속시간 정규화

각 음운은 고유의 지속시간을 갖고, 이러한 고유 지속시간은 음운 발성 환경의 영향에 의해 변화한다. 따라서 각 음운의 고유길이를 최대한 배제하고 단지 음운 발성 환경의 영향에 의한 지속시간 변화만을 고려하는 정규화 지속시간 즉, 각 음운의 고유지속시간의 영향을 배제시킨 순수한 음운환경에 의한 세그먼트 지속시간 변화 요인을 분석하기 위하여 본 연구에서는 식(1)과 같은 Zscore를 사용하여 문음성DB내 세그먼트들의 지속시간을 정규화 하였다.

$$Zip = (Xip - Mp) / SDp \quad \text{------ 식 (1)}$$

Zip : 음운 p의 지속시간에 대한 i번째 세그먼트의 관측치

Mp : 음운 p의 지속시간에 대한 평균치

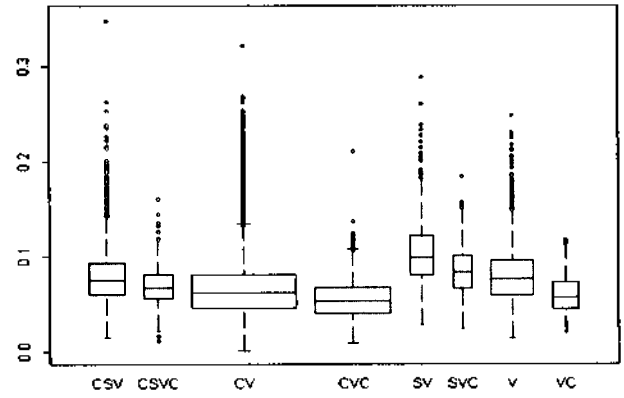
SDp : 음운 p의 지속시간에 대한 표준편차

2.3 음운지속시간 변화에 대한 통계적 분석

구축된 문음성 데이터베이스를 분석하여 세그먼트의 지속시간 변화에 크게 영향을 미치는 요인을 통계적인 방법으로 발견한다.

1) 음절 유형에 의한 영향

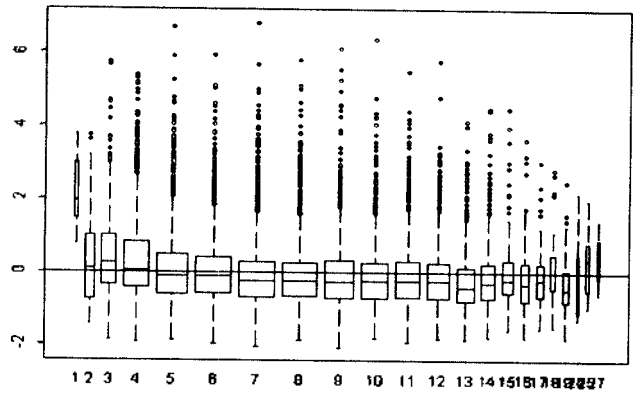
우리말의 음절유형을 다음의 8가지로 분류할 수 있다 (CV, CVC, V, VC, SV, SVC, CSV, CSVC). 음절유형에 따라 분석한 결과 음절핵은 고유의 지속시간을 갖는다. [그림 1]에 나타난 것과 같이 종성이 있는 CVC, VC 등의 음절 유형에서 음절핵의 지속시간은 비교적 짧고, 초성이 없는 음절유형에서는 초성이 있는 경우보다 비교적 길다.



[그림 1] 음절 유형별 음절핵의 지속시간 분포

2) 어절 내 음운수에 의한 영향

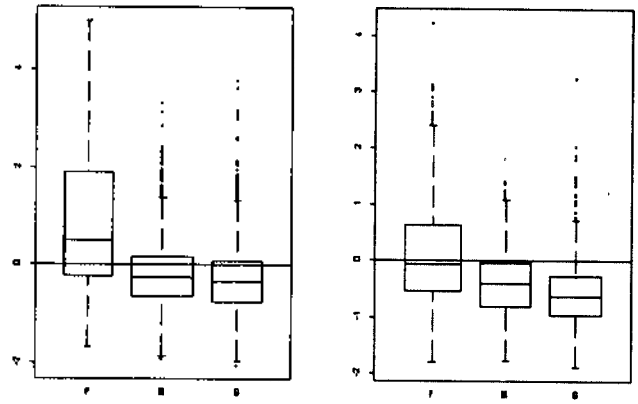
어절 내 음운수에 의한 모음의 지속시간 변화는 [그림 2]와 같이 음운의 수가 1-8개 까지 증가함에 따라 음운지속시간이 감소하고, 8-27개까지는 음운수 증가에 따라 모음의 지속시간이 거의 변화하지 않는다. 이는 각각의 음운을 조음하는데 최소한의 고유 지속시간이 유지 되어야만 하기 때문에 감소비율이 아주 작게 나타난다고 볼 수 있다.



[그림 2] 어절내 음운수에 대한 정규화 지속시간

3) 어절 내 음절 위치에 의한 영향

어절 내의 음절 위치는 문장의 한 어절 내에서의 첫 음절(S), 마지막 음절(F), 그리고 이 두 음절을 제외한 중간 음절(M) 3가지로 분류하여 분석한 결과 [그림 3]과 같이 마지막 음절에서 모음의 음운 지속시간이 길어지고, 첫 음절에서는 짧아지는 경향이 있다. 음절 유형이 CV인 경우에는 CVC인 경우보다 마지막 음절의 음운지속시간이 더 길어지는 경향이 있다.

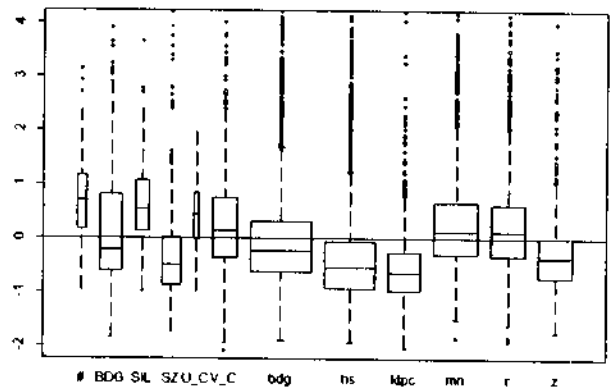


type = CV　　type = CVC

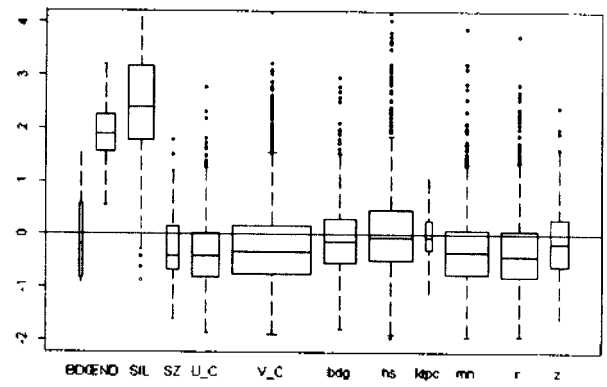
[그림 3] 어절내 음절위치별 모음의 정규화지속시간

4) 앞,뒤 인접음운에 의한 영향

음운 발성 환경 즉 고려하는 음운의 앞 음운과 뒤 음운의 영향을 분석한다. 이 때 전후 음운은 조음 양식과 위치로 분류하여 분석한다. 앞,뒤 인접음운에 의한 정규화 모음 지속시간의 분포를 비교해보면 뒤 음의 영향이 앞 음의 영향보다 크게 나타남을 알 수 있다. 따라서 뒤 음운의 영향은 음운지속시간 제어 규칙을 생성하는데 가장 중요한 특징 요소가 된다. [그림 4]와 같이 앞 음운의 영향에서는 파열, 파찰음, 마찰음 (BDG, bdg, SZ, z, hs, ktpc)의 뒤에 오는 모음의 지속시간이 짧아지는 경향이 있고, 비음 및 유음(mn, r)등의 유성음 뒤에서는 모음의 지속시간이 길어지는 경향이 있다. [그림 5]와 같이 뒤 음운의 영향에서는 마찰음, 종성, 비음, 유음(SZ, V_C, U_C, mn, r)등의 음운 앞에 오는 모음의 지속시간이 대체로 짧아지고, 또한 문말, 호흡단락(END, SIL)의 앞에 오는 모음 즉 마지막 음절의 음운 지속시간은 훨씬 길어진다.



[그림 4] 앞 음운별 모음의 정규화 지속시간



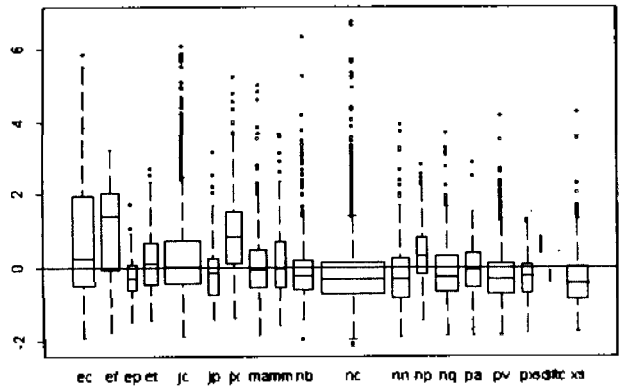
[그림 5] 뒤 음운별 모음의 정규화 지속시간

5) 문법적 요소에 의한 영향

문법적인 요인에 대한 음운지속시간을 분석하기 위하여 한국어의 품사를 다음 [표 2]과 같이 19품사로 분류한다.

[표 2] 한국어에 대한 19 품사

체언	보통명사	nc
	고유명사	nq
	의존명사	nb
	대명사	np
	수사	nn
용언	동사	pv
	형용사	pa
	보조용언	px
수식언	관형사	mm
	부사	ma
독립언	감탄사	ii
관계언	격조사	jc
	서술격조사	jp
	보조사	jx
어미	선어말어미	ep
	연결어미	ec
	전성어미	et
	종결어미	ef
접사	접미사	xs

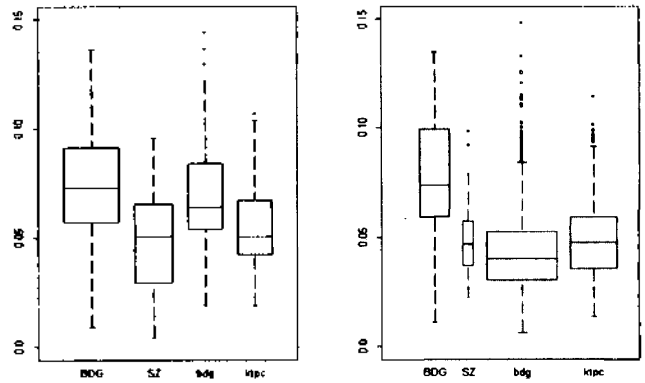


[그림 6] 품사별 모음의 정규화 지속시간

품사별 모음의 지속시간 변화를 분석한 결과[그림 6]과 같다. 품사가 ec (연결어미), ef (종결어미), jx (보조사), np (지시대명사)의 경우 음운지속 시간이 길어지는 반면, ep (선어말어미), jp (서술격조사), xs (접미사), 체언 또는 용언 등의 내용어인 경우에는 음운 지속시간이 짧아지는 경향이 있다.

6) 묵음의 지속시간

일반적으로 파열음, 파찰음 등의 앞에 묵음이 나타나는데 이러한 묵음은 조음 양식에 따라 불가피하게 나타나는 현상이다. 따라서 양질의 합성음을 위해서는 이러한 적당한 묵음의 지속시간 삽입이 필요하다. [그림 7]에서와 같이 BDG, SZ, bdg, ktpc 앞에서 발생하는 묵음은 무성음 종성 뒤에서는(왼쪽그림) 유성음 뒤에서보다(오른쪽 그림) 묵음의 지속시간이 길어진다.



[그림 7] 뒤 음운에 의한 묵음의 지속시간 분포

3. 정규화 회귀트리 모델링

3.1 정규화 회귀트리 구현

음운의 고유지속 시간의 영향을 배제시키고 순수한 음운환경에 의한 세그먼트의 지속시간을 예측하기 위하여 각 세그먼트의 지속시간을 Zscore로 정규화 한다. [표 3]은 음운 지속시간을 정규화하기 위해 구해진 자음, 모음의 일부분에 대한 평균 고유지속시간과 표준 편차이다.

[표 3] 음운의 평균지속시간과 표준편차

음운	평균지속시간 (msec)	표준편차 (msec)
B /ㅃ/	18	6
N/ㅇ/	58	22

음운	평균지속시간 (msec)	표준편차 (msec)
b/ㅂ/	27	15
g/ㄱ/	34	15
g2/ㄱ 2/	23	10
l/ㄹ/	52	24
m/ㅁ/	48	15
m2/ㅁ 2/	57	26
n/ㄴ/	37	14
n2/ㄴ 2/	70	36
a/ㅏ/	77	33
e/ㅔ/	85	36
o/ㅗ/	68	36
jo/ㅛ/	78	40
jv/ㅕ/	81	31
wa/ㅘ/	95	45

[표 3]에서는 일반적으로 모음의 표준편차가 자음의 표준편차보다 크므로 모음이 자음보다 탄성이 매우 큼을 나타낸다. 특히 유성음 종성(l /ㄹ2/, m2 /ㅁ2/, n2 /ㄴ2/, N /ㅇ/)에서는 표준편차 즉, 탄성이 크므로 지속시간의 예측이 어렵고, 음절핵에서는 e /ㅔ/, jo /ㅛ/, wa /ㅘ/에서 예측이 어렵다.

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 한다. 이 회귀트리에서 사용된 지속시간 변화 요인의 특징 요소는 다음과 같다.

- 해당음운의 조음양식
- 음절 유형
- 어절 내 음운수
- 어절 내 음절 위치
- 문장 내 어절수
- 앞 뒤 인접음운
- 품사

이러한 특징 요소를 분석한 결과 가장 큰 변화 요인은 [표 4]와 같이 뒤 음운의 영향이고 다음으로 문 음성 데이터에서는 품사가 중요한 특징 요소가 나타난다.

[표 4] 지속 시간변화의 특징 요소

순위	특징 요소
1	뒤 인접 음운
2	품사
3	어절 내 음절 위치
4	2 개 뒤 인접 음운
5	어절 내 음운 위치
6	앞 인접 음운
7	음절 유형

3.2 생성된 규칙의 평가

생성된 규칙의 타당성을 확인하기 위하여 관측치와 예측지간의 오류정도를 평가하고 오류 분석을 행한다. 이 때 예측 세그먼트 지속시간은 식(2)와 같은 방법으로 구한다. 관측치와 예측치간의 오류정도를 다중상관 계수로 평가한 결과는 [표 5]과 같다.

DURip = Mp + (Zip × SDp) ----- 식 (2)

여기에서,

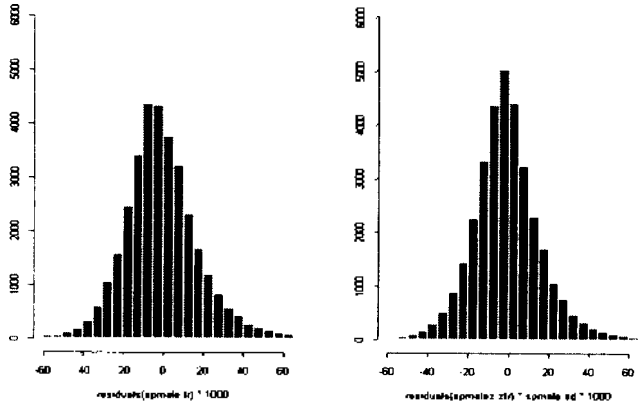
DURip : p음운의 i번째 세그맨트의 예측 지속시간
Mp : 음운p의 평균지속시간
Zip : p음운의 i번째 세그맨트의 예측 정규화 지속시간
SDp : 음운p의 지속시간의 표준편차

[표 5] 모델에 대한 평가

		지속시간	정규화
전체	다중상관 계수	0.78	0.81
	예측오차 20ms 이내	87.8%	89.4%
자음	다중상관 계수	0.80	0.82
	예측오차 20ms 이내	90.3%	91.7%
모음	다중상관 계수	0.78	0.79
	예측오차 20ms 이내	87.4%	88.1%

[표 5]에 나타난 것과 같이 모음이 자음보다 예측율이 낮은 것은 모음이 자음보다 탄성이 커서 예

측 오류율이 높기 때문이다. 정규화된 지속시간에 의한 예측과 지속시간에 의한 예측의 오류 정도를 평가한 결과 [그림 8]과 같다. 이는 정규화된 지속시간의 예측 오류율(오른쪽)이 일반 지속시간에 의한 예측 오류율(왼쪽)보다 오류 정도의 차가 더 작음을 나타낸다.



[그림 8] 정규화에 대한 예측 오류율 비교

4. 결 론

본 논문에서는 문음성 데이터 베이스(280문장)를 구축하여 음운발성 환경 및 문법적 요인을 고려한 음운지속시간을 정규화하여 통계적으로 분석하였다. 음운 지속시간 변화를 통계적으로 분석한 결과는 다음과 같다. 조음양식 및 조음위치, 음절의 유형 그리고 품사에 따른 지속시간은 고유의 분포를 갖는다. 어절 내 음운의 위치에 의해서는 마지막 음절에서 모음의 지속시간이 길어지며, 앞뒤 인접음운의 영향에 있어서는 뒤 음운의 영향을 더 크게 받는다.

음운의 고유 지속시간의 영향을 배제한 정규화 지속시간에 대해 회귀트리로 모델화 하였다. 또한 제안된 모델을 평가한 결과 예측치와 관측치간의 다중상관계수는 0.81이고 음운 지속시간 예측 오차의 89%정도가 20ms이내 이었다. 예측 오차를 좀더 줄이기 위해서는 탄성이 큰 특정 음운에 대한 조사 분석이 더욱 필요하다.

[참고 문헌]

[1] 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.

[2] Jan P.H van Santen,"Deriving text-to-speech duration from natural speech," Talking machines : Theories, Models, Designs, pp 275-285, 1992.

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

[4] 성유나, 이양희,"회귀트리에 의한 한국어 음운 지속 시간 모델", 신호처리 합동 학술대회 논문집, vol.9, Part 1, pp 53-56,1996.

[5] 이상호, 오영환 , "CART를 이용한 운율구 추출 및 음운 지속 시간 모델링",한국음향학회 학술발표, pp 135-138, 1998.

[6] 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.

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

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

* 본 연구는 한국전자통신연구소의 1998년도 수탁과제 연구지원비에 의해 연구되었습니다 *