Trainable TTS System을 위한 음운 지속시간 모델링

서지인, 이양희
동덕여자대학교 전자계산학과

An Analysis on the Phoneme Duration Modeling For the Trainable TTS System

Jiln Seo, Yanghee Lee
Dept. of Computer Science, Dongduk Women's Univ.
E-mail : petit76@dongduk.ac.kr yhlee@dongduk.ac.kr

요약

본 논문에서는 한국어 Trainable TTS System의 자연스러운 음성 합성을 위해 400문장(어절수:6,220,음운수:총43,701:자음 23,899,모음:19,802)에 대하여 단일 남성화자가 발성한 문 음성 데이터를 음운레벨세그먼트, 음운 라벨링,어절간의 띄어쓰기,어절에 대한 음운별 품사가 태깅된 문 음성 코퍼스를 사용하여 음운 환경과 품사에 의하여 음운의 지속시간이 어떻게 변화하는가에 대하여 통계적으로 분석하였다.그리고 음운 지속시간을 보다 정교하게 예측하기 위하여,각 음운에대한 고유 지속시간의 영향이 배제된 정규화 음운지속시간에 대한 회귀트리를 이용하여 정규화 지속시간에 영향을 미치는 특징요소들 간의 관계를 통계적인 방법으로 분석하였다.그 결과 문법적인 특징요소를 나타내는 요소들간에 서로 상관이 높게 나타나는 것을 알수있었다.그리고 이러한 경우 유사한 특징 요소들간에 상관이 1에 가까울 정도로 상관이 높은 요소들의 경우 예측지수가 낮은 요소들을 제거하여도 지속시간변화에 영향을 미치지 못하는 것으로 나타났다.그결과 문법적 성질이 유사한 특징요소들을 회귀트리를 통해 모델링할 경우에 요소들간의 상관정도를 분석하여 최소한의 특징요소들을 선택할수 있는 방법을 제시하였다.그리고 이를 토대로 한 정규화 회귀트리의 모델링이 지속시간 회귀트리 모델링보다 우수함을 입증하였다.

1. 서론

이 논문에서는 기존의 음운 지속시간의 모델화 방법으로 제안된 것 중,변수들이 갖는 영역을 제어요소의 의존관계에 의해 분할하는 것으로 비선형성을 표현할 수 있어 다양한 언어에 대하여 사용되고 있는 회귀트리를 사용한다.회귀트리를 사용한 한국어의 음운 지속시간 모델은 어절 데이터를 사용한 모델이기 때문에 문 음성에 대한 모델로는 불충분하고,지속시간 변화요인만을 고려하지 않고 음운의 고유 지속시간까지도 포함하여 회귀트리로 지속시간을 모델화하므로 정교한 지속시간이 예측이 불충분하다.따라서 본 논문에서는 통계적으로 처리하기에 충분히 구축된 문음성 코퍼스를 사용하여 음운의 지속시간을 변화시키는 특징요소를 splus를 사용하여 통계적으로 분석한다. 또한 이 시간 특징들 중 변화 폭이 큰 요인들을 제어요소로 각 음운의 고유길이를 최대한 배제하고 단지 음운발성 환경의 영향에 의한 지속시간 변화만을 고려하는 정규화 지속시간을 회귀트리로 모델화하고 특징요소들간의 관계를 통계적으로 분석한다.

2. 음운 지속시간 정규화와 통계적 분석

통계적인 방법으로 일반화된 규칙을 생성하기 위해서는 다양한 경우를 포함하는 많은 양의 데이터가 요구된다. 그리고 보다 일반적이며 정교한 음운 지속시간 제어 모델을 생성하기 위하여,다양한 음운 환경을 고려하는 충분히 많은 자연음성을 분석하여 음운 지속시간 변화에 영향을 미치는 요인을 추출하여야 하는데 이때 음운 지속시간을 변화시키는 요인은 크게 음운 환경적 요인(음

절의 유형,어절내 음운수,어절내 음절의위치,앞뒤 인접 음운등)과 문법적인 요인(품사)으로 나누어 생각할수 있다.이 논문에서는 구축된 문음성 데이터 베이스를 분석하여 세그먼트의 지속시간 변화에 크게 영향을 미치는 요인을 통계적인 방법으로 발견하여 예측한 후 각각의 요인들에 대한 음운 지속시간 변화에 미치는 영향을 알아보고 각각의 요인들간의 상관이 음운 지속시간 변화에 어떠한 영향을 미치는지에 대해서 분석을 하였다. 통계적으로 발견한 지속시간 변화에 영향을 미치는 특징요소는 [표 1]과 같다.

[표 1] 지속시간 변화에 영향을 미치는 요인

자음에 대해	예측지속시간	순위	모음에대해	예측지속시간
"succ"	0.7422441	1	"succ"	0.6714025
"succtag"	0.7377105	2	"succtag"	0.6646363
"rsucctag"	0.7368719	3	"rsucctag"	0.6643591
"prevtag"	0.7322104	4	"tag"	0.5337107
"rprevtag"	0.7259299	5	"prevtag"	0.5254041
"prev"	0.7105312	6	"rtag"	0.5233229
"lenloc"	0.7104035	7	"rprevtag"	0.5212083
"succ2"	0.7096529	8	"location"	0.5032852
"tag"	0.7086775	9	"succ2"	0.4883653
"stn"	0.7053369	10	"lenloc"	0.4704799
"location"	0.7050532	11	"prev"	0.4563334
"rtag"	0.7038234	12	"type"	0.4452519
"prev2"	0.7035143	13	"prev2"	0.4081083
"lencnt"	0.7018093	14	"disdirec"	0.408058
"blenloc"	0.6993012	15	"lencnt"	0.4046709
"sylcnt"	0.6991743	16	"blencnt"	0.4022898
"blencnt"	0.699162	17	"blenloc"	0.3961106
"disdirec"	0.6985432	18	"stn"	0.3912688
"wordloc"	0.697475	19	"bcnt"	0.3881877
"wordcnt"	0.6952034	20	"wordcnt"	0.38431
"direcloc"	0.6950344	21	"wordloc"	0.381112
"bcnt"	0.6955056	22	"silcnt"	0.3799033
"silloc"	0.6950183	23	"direcloc"	0.3789688
"bloc"	0.6949949	24	"sylcnt"	0.3789814
"type"	0.6949291	25	"silloc"	0.3788107
"silcnt"	0.6946057	26	"bloc"	0.3770527
"artl"	0.6939879	27	"artl"	0.3765517

[표 1]에서 예측지속시간은 아래의 [식 1]에 근거한다.

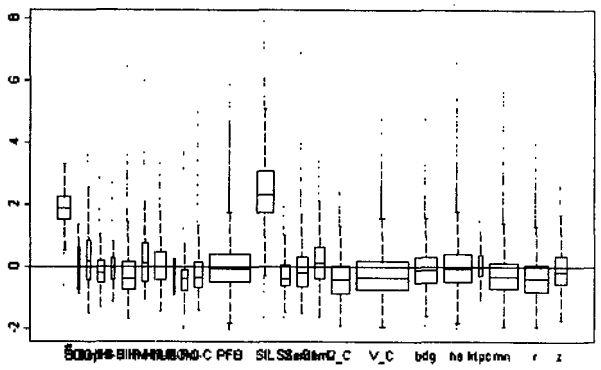
DURip = Mp + (Zip × SDp) —식 (1)	
DURip	p음운의 i번째 세그멘트의 예측 지속시간
Mp	음운p의 평균지속시간
Zip	p음운의 i번째 세그멘트의 예측 정규화 지속시간
SDp	음운p의 지속시간의 표준편차

3. 음운 지속시간을 변화시키는 요인

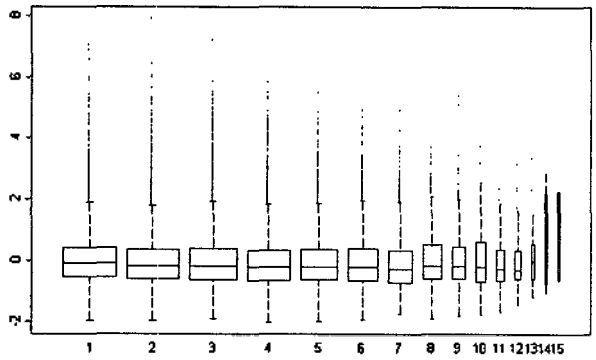
3.1 음운 환경적 요인

음운 지속시간의 변화에 영향을 미칠 요인중 문법적 요인으로 품사를 나타내는 요소인 tag,rtag,prevtag,rprevtag,succtag,rsucctag을 제외한 각각의 요소를 대상으로 음운 환경적 요인이 지속시간 변화에 미치는 영향을 [표 1]의 순위를 토대로 통계적인 방법으로 예측한결과 그 변화의 영향을 주는 정도에 대하여 예상했던 대로 [표 1]에 대한 순위대로 영향이 있음을 알수있었다.[그림 1]과 [그림 2]은 그 타당성을 뒷받침하는 예로 지속시간 변화에 가장 큰 영향을 미칠것으로 예상되는 succ와 가장 영향을 미치지 못할것으로 예상되는 bloc에 대한 분포를 나타낸다.

[그림 1]지속시간 변화에 큰 영향을 미치는 요소에 대한 정규화 지속시간 분포(모음에서succ)



[그림 2]지속시간 변화에 영향을 거의 미치지 못하는 요소에 대한 정규화 지속시간 분포(모음에서 bloc)



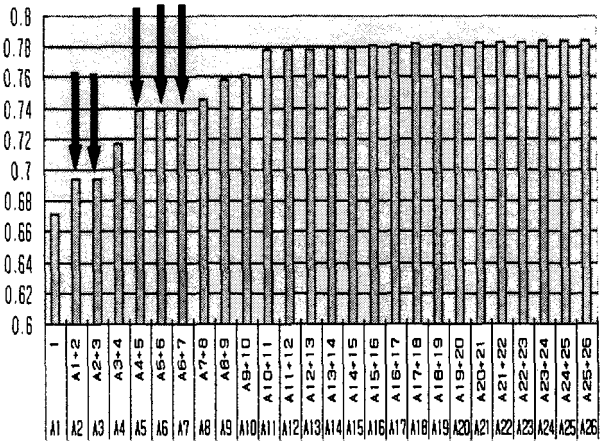
3.2 문법적 요인

문법적 요인에 대해서도 앞에서 다룬 음운 환경적 요인처럼 각각의 요소들에 대하여 통계적 방법을 사용한 결과 [표 1]의 순위가 타당하다는 입증하였다.
이제 음운환경적인 요인과 문법적인 요인들을 [표 1]의 순위대로 succ부터 bloc까지 특징요소들을 더하여 정규화 트리를 구현하여 [표 1]에서 나타난 순위대로 모든 요소들을 포함시켰을 때 지속시간 예측률이 높아지는지에 대한 정규화 회귀트리를 구현한다.

4. 회귀트리의 특징요소들간 관계분석

정규화 회귀트리 모델링에서는 음운의 고유지속 시간의 영향을 배제시키고 순수한 음운 환경에 의한 세그먼트의 지속시간을 예측하기 위하여 각 세그먼트의 지속시간을 Zscore로 정규화 한다.각 세그먼트의 정규화 지속시간은 음운이 고유지속시간을 제외한 음운 환경에만 의존하여 변화하게된다.따라서 지속시간 변화요인에 의해서만 분류되기 때문에 보다 정교하게 예측이 가능하도록 정규화 지속시간에 대해 회귀트리로 모델화한다. 이 회귀트리에서 사용된 지속시간 변화 요인의 특징요소는 [표 1]의 요소들이며 방법은 각각의 요소에 대하여 지속시간 예측의 순위가 높은 succ부터 bloc까지 요소들을 순차적으로 더하여 통계적으로 상관을 구하였다.그 결과 순위가 높은 요소들을 더할수록 정규화 회귀트리의 지속시간 변화에 따른 상관이 계속 증가할것이라는 처음의 예측에서 벗어나 [그림 3]과 같이 요인들을 계속더해도 바로 이전의 예측률과 변동이없고 오히려 순위가 더 낮은요인을 더했을 경우 예측률이 더 높아진다는 것을 발견했다. [그림 3]에서 화살표가 가리키는 것은 요소들을 더했음에도 예측률의 변동이 없는경우이다.

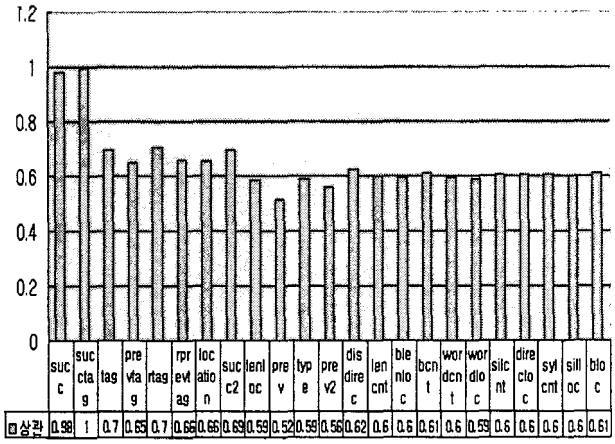
[그림 3] 정규화 회귀트리의 지속시간 변화예측률의 변동에 따른 상관계수 (모음의 경우)



[X축은 [표 1]에서 나타나는 순위대로 점차 요소들을 더해가는 것으로 오른쪽으로 갈수록 예측값이 낮은것들과의 결합이됨.Y축은 상관]

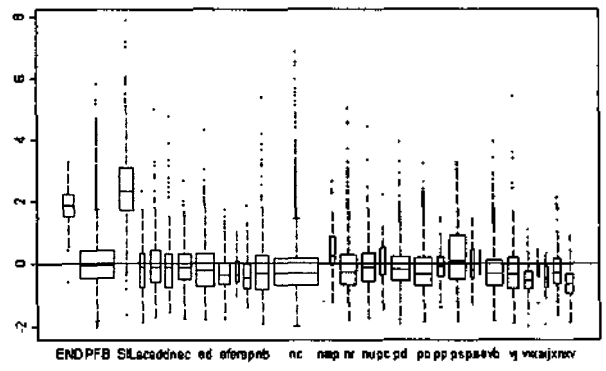
[그림 3]의 결과를 바탕으로 특징요소들을 순위에 따라 순차적으로 추가했음에도 불구하고 변동이 없는 부분에 대한 분석을 하기로한다.이 논문에서 제시하는 방법은 그림에서 A2,A3와 A5,A6,A7이 높은 순위의 결합임에도 불구하고 예측률의 변동이 없는 이유를 분석하기 위해 변동이 없는 A3의 특징요소인 rsucctag와 A1의 특징요소인 succ,A2의 요소인 succtag의 상관을 구해보고 그 외의 다른 요소들과의 상관을 구한다.그 결과는 [그림 4]와 같다.

[그림 4]rsucctag와 특징요소들과의상관

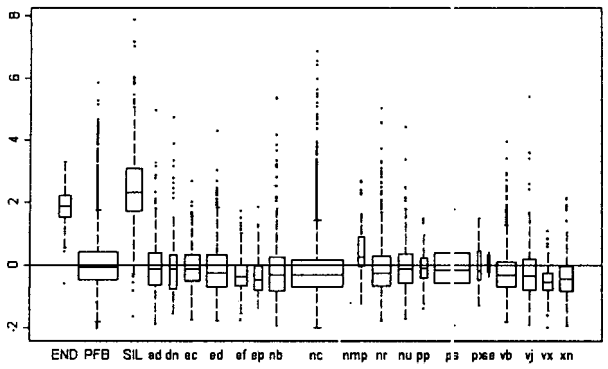


이러한 방식으로 rprevtag와 rtag도 상관을 조사한결과 rpretag는 prevtag와 상관이 매우 높고 rtag와 tag도 매우 상관이 높다는 것을 알수 있다.그러므로 이러한 경우 상관이 높은 요소와 합하여 지속시간 변화 예측률의 상관을 구할경우에는 [그림 3]의 A3,A6,A7과 같이 이미 상관이 높은 요소가 상관이 높은 요소들의 특징을 포함하는 것으로 볼수 있다.이것의 근거는 상관이 거의 1에 가까운 요소들을 합하여 다른 요소들과의 상관을 구하거나 이 둘중 순위가 높은 요소과 다른 요소들과의 상관을 구해도 지속시간 예측 상관에는 아무런 변화가 없다는 것이다.그러므로 요소들간의 상관이 높은 요인들을 이중으로 사용하여 지속시간 예측을 하는 것은 의미가 없으므로 상관이 높은 요소중 예측순위가 낮은 요소를 정규화 회귀트리 생성시 제외시킬수 있다.[그림 5]와 [그림 6]로 서로의 상관이 1에 가까운 특징요소들에 대한 정규화 지속시간 분포가 거의 유사함을 알수있다.

[그림 5]모음에서 succtag 의 정규화 지속시간 분포



[그림 6] 모음에서 succtag 의 정규화 지속시간 분포



5. 정규화 회귀트리 특징요소 선택에 대한 타당성

지속시간 회귀트리보다 정규화된 회귀트리의 타당성을 확인하기 위하여 관측치와 예측치간이 오류정도를 평가하고 오류분석을 행한다.이때 예측 세그먼트 지속시간은 [식 1]에서 구한 방식과 같다.관측치와 예측치간의 오류정도를 다중상관계수로 평가할 때 [표 1]과 같이 정규화 상관값이 높은 특징요소를 순위대로 순차적으로 모두 더하여 상관을 구한경우와 문법적요인을 가진 요소들간에 상관이 높은 것 중 상관값이 높은 하나만을 사용하여 정규화 회귀트리를 구한 상관이 같음을 알수 있었다.이를 토대로 한 정규화 회귀트리와 지속시간 회귀트리에대한 모델의 평가는 [표 2]와 같다.

[표 2] 지속시간회귀트리와 정규화된 회귀트리의 평가

예측오류율 비교	지속시간 회귀트리	정규화된 회귀트리
다중 상관계수	0.774	0.804
예측오차 25ms이내	91.30%	92.70%

6.결론

본 논문에서는 자연스러운 음성 합성을 위해 이미 구축된 400문장에대한 문음성 코퍼스를 사용하여 음운 환경과 품사 뿐만 아니라 구문구조에 의하여 음운의 지속시간이 어떻게 변화하는가에 대하여 통계적으로 분석하였다.그리고 음운 지속시간을 보다 정교하게 예측하기 위하여,각 음운에대한 고유 지속시간의 영향이 배제된 정규화 음운지속시간에 대한 회귀트리를 이용하여 정규화 지속시간에 영향을 미치는 특징요소들 간의 관계를 통계적인 방법으로 분석하였다.그 결과 문법적인 특징요소를 나타내는 요소들간에 서로 상관이 높게 나타나는 것을 알수있었다.그리고 유사한 특징 요소들간에 상관이 높을경우,각각의 특징요소별로 구한 정규화 상관지수가 낮은 요소들을 제거하여도 지속시간변화에 영향을 미치지 못하는 것으로 나타났다.이것은 문법적 성질이 유사한 특징요소들을 회귀트리를 통해 모델링할 경우에 서로간의 상관정도를 분석하여 최소한의 특징요소들을 선택할수 있음을 의미한다고 할수있다.그리고 이를 토대로 한 정규화 회귀트리의 모델링이 예측치와 관측치간의 다중 상관계수는 0.804이고 음운지속시간 예측오차의 92.7%가 25ms 이내로 지속시간 회귀트리 모델링보다 우수함을 입증되었다.

[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.

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

[8] 김상훈, 이정철, 강도규, 이영직 "대용량 운율 음성데이터를 이용한 자동합성방식 , 제 15회 음성통신 및 신호처리 워크샵 논문집 15권 1호, pp87-92, 1998.

[9] 김인영, 정지혜, 이양희, "음운지속시간의 정규화와 모델링", 제 15회 음성통신 및 신호처리 워크샵 논문집 15권 1호, pp99-104, 1998.N.

[10] ChungJihye, LeeYanghee, A Study on the Korean Concatenative Speech Synthesis System using Non-Uniform Units : ICSP 1999.