

정규화 지속시간 회귀트리를 기반으로 한 음운지속시간 모델화

A Modelling of segmental Duration based on Regression Tree of the Normalized Duration

정지혜, 김인영, 이양희

동덕여자대학교 전자계산학과

E-mail : jihye.inyoung@cs4000.dongduk.ac.kr yhlee@www.dongduk.ac.kr

요 의

본 논문에서는 자연음성으로부터 통계적인 방법으로 일반적인 음성합성 규칙을 생성하기 위해, 남녀 각각 1명이 200문장에 대해 발성한 문음성 데이터를 음운 세그먼트, 음운 라벨링, 음운별 품사 태깅, 문법 정보 태깅하여 음성 데이터베이스를 구축하였다. 이 음성 데이터베이스로부터 2가지 지속시간을 분석하여 긴 휴지와 짧은 휴지로 분류하였고, 이러한 휴지가 어느 경우에 나타나는가를 조사하였다. 음운지속시간을 보다 정교하게 예측하기 위하여, 각 음운의 고유 지속시간의 영향을 배제시킨 정규화 지속시간에 대해 2가지 class(장,단)의 휴지시간을 고려한 회귀트리로 음운지속시간을 모델화하였다. 제안된 모델의 평가 결과 예측치와 관측치 간의 다중 상관 계수는 남성은 0.82, 여성은 0.84정도로 평가되었다.

1. 서 론

자연스럽고 명료한 음성을 합성하기 위하여, 기존의 정형의 단위음성 연결 합성 방식 보다는 음운환경을 고려한 비정형 단위음성의 합성방식에 대한 연구가 활발하다 [1],[2]. 이러한 비정형 연결 합성 방식에서는 음운성 확보와 자연성 확보를 위하여 특히, 통계적인 방법에 의해 일반화된 규칙 생성을 위하여 대량의 음성 데이터가 필요하다.

우리말 음성합성에서도 비정형 단위 연결 합성 방식과 일반화된 운율 규칙 생성에 대한 연구가 활발히 진행되고 있다[3]. 이러한 연구를 수행하기 위해서는 무엇보다도 대량의 음성 데이터베이스 구축이 필요하다. 또한, 통계적

인 방법에 의해 한국어 음운지속시간 모델링에 대해 연구되었으나, 정도 높은 지속시간을 예측하기에는 아직 미흡하다[4],[5]. [4]의 경우 휴지 지속시간에 대한 고려없이 모델화하였고 [5]의 경우 운율구를 추출하여 휴지지속시간을 고려한 CART로 음운의 지속시간을 모델화하였으나 정교한 음운지속시간을 예측하기에는 불충분하다. 따라서, 본 논문에서는 전문 아나운서 남녀 각각 한명이 발성한 총 400문장(7,137어절)의 문음성 데이터에 대해 음운 세그먼트, 음운 라벨링, 품사 태깅, 띄어쓰기 정보 태깅을 행한 음성 데이터 베이스를 구축하고 한국어 음운 지속시간을 정규화 음운지속시간 회귀트리로 모델화한다. 2절에서는 통계적으로 처리하기에 충분한 문음성 데이터베이스를 구축하고, 이 문음성 데이터베이스를 사용하여 휴지 지속시간을 통계적으로 분석한다. 3절에서는 정규화 지속시간을 회귀트리로 모델화한다. 4절에서는 제안된 모델을 평가하여 타당성을 입증한다.

2. 문음성 DB구축과 휴지 지속시간 분석

2.1 문 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 자연 음성을 분석하여 음운지속시간 변화에 영향을 미치는 요인을 추출하여 규칙을 생성하여야 한다. 문헌[4],[5]에서는 남성화자 1명이 발성한 문음성 데이터베이스를 구

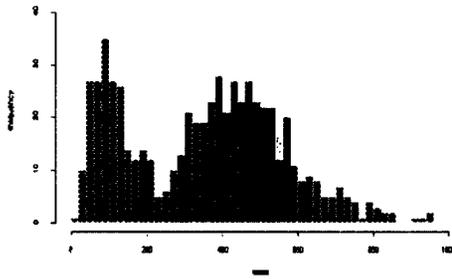
축하여 분석하였으나 여성화자에 대한 분석 또한 필요하다. 따라서 본 논문에서는 남녀 각각 1명이 발성한 표 1. 과 같은 문음성 데이터를 음운레벨로 세그먼트하고 음운 라벨링(음운기호)하였다. 그리고 음운별 품사 및 문법정보를 태깅하여 음성 데이터베이스를 구축한다.

[표 1] 음성 데이터

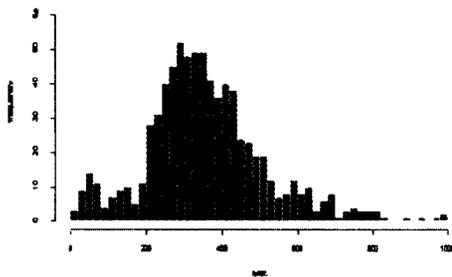
화자	남성 단일화자	여성 단일화자
데이터	200문장	200문장
어절수	3,613	3,524
음운수	25,067	25,019

2.2 휴지 지속시간 분석

음성합성의 경우 입력 텍스트로부터 휴지 지속시간의 정확한 예측은 합성음성의 음운지속시간 제어 및 핏치 제어 등에 매우 중요하므로 자연음성으로부터 휴지 지속시간을 분석하여 문장의 문법 및 문 구조와의 관계를 정확하게 모델링할 필요가 있다. 이를 위해서는 우선, 자연음성의 휴지 지속시간이 분석되어야 한다. 따라서 본 논문에서는 이들 음성데이터베이스로부터 휴지 지속 시간의 분포를 조사한다.

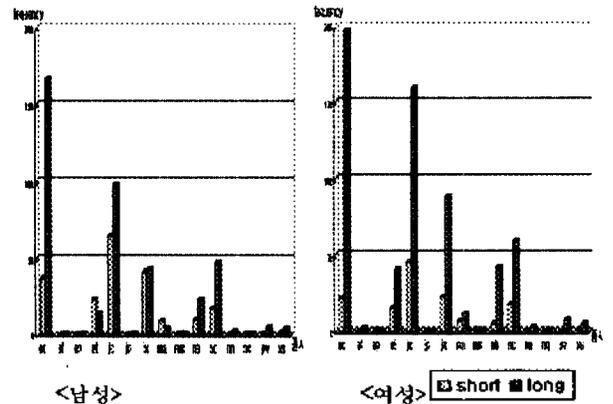


[그림 1] 남성의 휴지 지속시간 히스토그램



[그림 2] 여성의 휴지 지속시간 히스토그램

휴지 지속시간의 히스토그램은 그림 1, 2와 같다. 그림 1은 남성의 휴지 지속 시간 분포로 긴 휴지구간과 짧은 휴지구간의 차이가 뚜렷한 반면, 그림 2의 여성의 경우는 짧은 지속시간 분포가 적고 긴 지속시간의 평균 휴지 지속 시간이 남성보다 비교적 짧게 나타나고 있다. 분석된 휴지 지속시간 분포를 고려하여, 보다 간단하게 휴지 발생을 적용하기 위하여 휴지 지속시간을 남성의 경우 20ms 을 기준으로, 여성의 경우 180ms 을 기준으로 짧은 휴지와 긴 휴지로 분류한다.



[그림 3] 앞 음운의 품사에 따른 휴지기간 분포

휴지를 2개의 class (장, 단)로 분류하여 분석한 결과 그림 3과 같이 짧은 휴지는 앞 음운의 품사가 조사 (jc), 연결어미(ec), 보조사(jx)순으로, 긴 휴지는 연결어미(ec), 조사(jc), 보조사(jx), 명사(nc)순으로 남녀가 비슷하게 나타난다. 짧은 휴지는 여성의 경우 남성보다 적게 발생하고, 남성의 경우 보조사와 전성어미(et)에서 여성과는 다르게 짧은 휴지가 긴 휴지보다 많이 나타난다.

3. 정규화 지속시간에 대한 회귀트리 모델링

3.1 회귀트리 특징 요소

1) 음운지속시간의 회귀 트리

음운의 고유지속 시간의 영향을 배제시키고 순수한 음운환경에 의한 세그먼트의 지속시간을 예측하기 위하여 각 세그먼트의 지속시간을 Zscore로 정규화 한다[4].

각 세그먼트의 정규화 지속시간은 음운의 고유지속시간을 제외한 음운환경에만 의존하여 변화하게 된다. 따라서 지속시간 '변화 요인에 의해서만 분류되기 때문에 보다 정

교하게 예측이 가능하도록 정규화 지속시간에 대해 회귀 트리로 모델화 한다. 이 회귀트리에서 사용된 특징 요소는 다음과 같다.

- 해당음운의 조음양식 및 위치 :
23분류(BDG/ㄱ, ㄲ, ㅁ, SZ/ㅅ, ㅆ, Dup/-/, LO-C/ㅏ, HM-C/ㅑ, HM-B/ㅓ, HI-B/ㅕ, HI-C/ㅡ, HI-F/ㅣ, LM-F/ㅗ, HM-F/ㅛ, Sem1/j+ 단모음/, Sem2/w+ 단모음/, U_C/무성 종성/, V_C/유성종성/, bdg/ㄱ, ㄲ, ㅁ/, hs/ㅎ, ㅌ/, ktpc/ㅋ, ㅌ, ㅍ, ㅊ/, mn/ㄴ, ㄷ/, r/ㄹ/, z/ㅈ/)
- 음절 유형 :
8분류(CV, CVC, V, VC, SV, SVC, CSV, CSVC)
- 어절 내 음운 수 : (1-27)
- 한 호흡 단락 내 음운 수 : (1-87)
- 어절 내 음절 위치 :
3분류(Start, Medial, Final)
- 문장 내 어절 수
- 앞 뒤 인접음운 :
25분류 (조음양식 및 위치 23분류 + /목음/, /짧은 휴지/, /긴 휴지/, /문시작(앞)/, /문의끝(뒤)/)
- 품사와 앞 뒤 인접 음운의 품사 :
19분류(ec, ef, ep, et, jc, jp, jx, ma, mm, nb, nc, nn, np, nq, pa, pv, px, sd, xs)

2) 목음에 대한 회귀 트리 모델

일반적으로 파열음, 파찰음 등의 앞과 무성 종성 뒤에 목음이 나타나는데 이러한 목음은 조음 양식에 따라 불규칙하게 나타나는 현상이다. 이러한 목음을 일반적으로 음운의 지속시간에 포함시키나, 본 논문의 음성 DB에서는 별도로 구분하여 다루었다. 따라서 목음 지속시간을 합성하기 위해서는 적당한 목음 지속시간 삽입이 필요하다. 정확하게 예측이 가능하도록 목음 정규화 지속시간에 대해 회귀트리로 모델화 한다. 이때 사용된 특징 요소는 다음과 같다.

- 음절 유형
- 어절 내 음절 위치
- 앞 뒤 인접음운
- 품사와 앞 뒤 인접 음운의 품사

3.2 회귀트리에 대한 Pruning

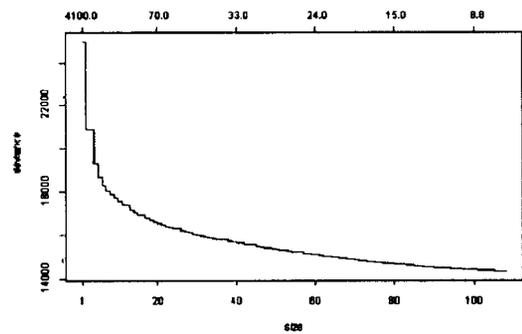
회귀트리 생성에서는 터미널 노드 수에 대한 제한을 두지 않는다. 따라서 Overfitting문제가 발생한다. 따라서 회귀트리의 Overfitting문제를 해결하기 위해 식(1)을 적용하여 회귀트리를 pruning한다.

$$D\alpha(T') = D(T') + \alpha \text{size}(T') \quad \text{--- 식 (1)}$$

$D(T')$: 부분 트리 T' 의 deviance

$\text{size}(T')$: 부분 트리 T' 의 터미널 노드수

α : cost복잡도 파라미터



[그림 4] 트리의 노드 수에 따른 deviance

그림 4에서 노드 수(size)가 60부터는 deviance가 거의 변화하지 않음을 알 수 있다. 이 실험에서는 정규화 회귀트리의 터미널 노드 수를 여성:108 → 71($\alpha=20$), 남성:115 → 64($\alpha=18$)개까지 pruning한다.

3.3 제안된 모델의 평가

생성된 규칙의 타당성을 확인하기 위하여 관측치와 예측치간의 오류정도를 평가하고 오류 분석을 행한다. 이때 예측 세그먼트 지속시간은 식(2)와 같은 방법으로 구한다. 관측치와 예측치간의 오류정도를 다중상관 계수로 평가한 결과는 표 2와 같다.

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

$DURip$: p음운의 i번째 세그먼트의 예측 지속시간

Mp : 음운p의 평균지속시간

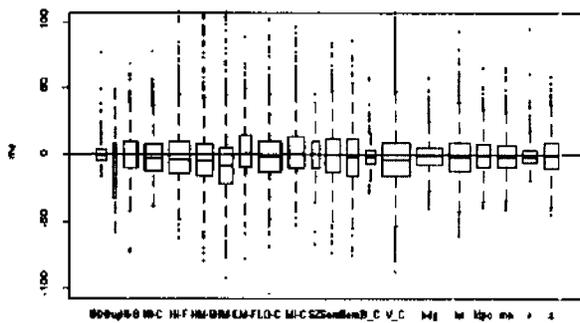
Zip : p음운의 i번째 세그먼트의 예측 정규화 지속시간

SDp : 음운p의 지속시간의 표준편차

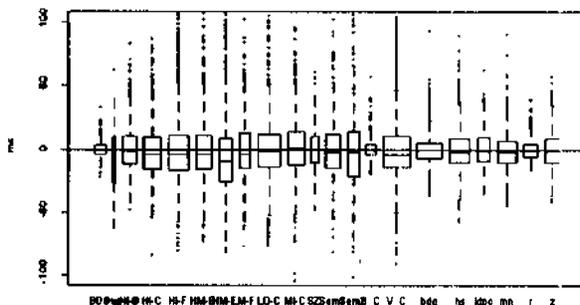
[표 2] 제안된 모델에 대한 평가

예측오류율비교		지속시간 회귀트리	정규화된 회귀트리
다중상관 계수	음	0.809 (남성)	0.821 (남성)
	운	0.828 (여성)	0.840 (여성)
	목 음	0.778 (남성) 0.812 (여성)	0.778 (남성) 0.812 (여성)
예측오차 25ms이내	음	91.6% (남성)	92.3% (남성)
	운	90.6% (여성)	91.4% (여성)
	목 음	95.3% (남성) 93.5% (여성)	95.3% (남성) 93.5% (여성)

제안된 모델에 의한 지속시간의 예측치와 관측치 간의 오류 정도를 평가하기 위하여 각 음운별 오류 분포를 그림 5, 6에 나타내었다. 그림 5, 6에서 각 클래스의 오류 분포들이 25ms내에 집중되어 있고, 남성의 오류율이 여성의 오류율보다 작음을 알 수 있다.



[그림 5] 남성의 예측 오차



[그림 6] 여성의 예측 오차

4. 결론

본 논문에서는 문음성 데이터 베이스(남성 200문장, 여

성 200문장)를 구축하였다. 이 음성 데이터베이스로부터 휴지지속시간에 대한 분포를 분석하였고 긴 휴지와 짧은 휴지로 분류하여 음운의 고유 지속시간의 영향을 배제한 정규화 지속시간을 음운의 회귀트리로 모델화하였고, 목음의 지속시간도 모델화하였다.

또한, 제안된 모델을 평가한 결과 음운의 경우 예측치와 관측치간의 다중 상관계수는 남성은 0.82, 여성은 0.84 이고 남성은 음운 지속시간 예측 오차의 92%정도가 25ms 이내 여성은 91%정도가 25ms이내이었다.

금후 연구과제로는 문음성 DB구축에 있어서 휴지구간 예측과 구분 정보를 사용하여 문음성에 대한 보다 정교한 음운지속시간 모델화가 필요하다.

[참고 문헌]

- [1] N.Campbell, A. Black, "Prosody and the selection of source units for concatenative synthesis.", Progress in Speech Synthesis. Springer Verlag, 1995.
- [2] N. Iwahashi, N. Kaiki, Y. Sagisaka, "Concatenative speech synthesis by minimum distortion criteria.", ICASSP '92, ppII-65-68, 1992.
- [3] 김상훈, 이정철, 강도규, 이영직 "대용량 운율 음성 데이터를 이용한 자동합성방식", 제 15회 음성통신 및 신호처리 워크샵 논문집 15권 1호, pp87-92, 1998.
- [4] 김인영, 정지혜, 이양희, "음운지속시간의 정규화와 모델링", 제 15회 음성통신 및 신호처리 워크샵 논문집 15권 1호, pp99-104, 1998.N.
- [5] 이상호, 오영환, "CART를 이용한 운율구 추출 및 음운 지속 시간 모델링", 한국음향학회 학술발표, pp 135-138, 1998.
- [6] M.D.Riley, "Tree-based modelling of segmental duration", Talking machines : Theories, Models, Designs, pp 265-273, 1992.
- [7] Leo Breiman, Jerome H, Friedman, Richard A, Olshen, "CLASSIFICATION AND REGRESSION TREE". Wadsworth, Inc. 1984.

* 본 연구는 한국전자통신연구소의 1998년도 수탁과제 연구지원비에 의해 연구되었습니다 *