2000년도 한국음향학회 학술발표대회 논문집 제19권 제1(s)호

TTS 시스템을 위한 휴지기간 모델링

정지혜, 이양희
동덕여자대학교 전자계산학과

The Modeling of Pause Duration For Text-To-Speech Synthesis System

Jihye Chung, Yanhee Lee
Division of Computer and Information, Dongduk Women's University

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

요약

본 논문에서는 비정형 단위를 사용한 음성 합성 시스템의 합성음에 대한 자연성을 향상 시키기 위한 휴지 구간 추출 및 휴지 지속시간 예측 모델을 제안한다. 제안된 휴지 지속시간 예측 모델은 트리 기반 모델링 기법 중 하나인 CART (Classification And Regression Trees)방법을 이용하였다. 이를 위해 남성 단일 화자가 발성한 6,220 개의 어절경계 포함하는 총 400 문장의 문 음성 데이터베이스를 구축하였고, 이 데이터베이스로부터 V-fold Cross-Validation 방법에 의해 최적의 트리를 결정하였다. 이 모델을 평가한 결과, 휴지 구간 추출 정확율은 81%로 휴지 구간 존재 추출 정확율은 83%, 휴지 구간 비존재 추출 정확율은 80%이었고, 실 휴지지속시간과 예측 휴지지속시간과의 다중상관계수는 0.84 로, 오차 범위 20ms 이내에서의 정확율은 88%이었다. 또한, 휴지지속시간을 예측하여 적용한 합성음을 청취 실험한 결과 자연 음성과 대체적으로 유사하게 나타났다.

1. 서론

자연스럽고 명료한 음성을 합성하기 위하여, 기존의 정형의 단위음성 연결 합성 방식 보다는 음운환경을 고려한 비정형 단위음성의 합성방식에 대한 연구가 활발하다 [1],[2],[7]. 이러한 비정형 연결 합성 방식에서는 음운성 확보와 자연성 확보를 위하여 대량의 음성 데이터로부터 통계적인 방법에 의해 일반화된 규칙 생성이 필요하며 음성 합성의 경우 입력 텍스트로부터 휴지지속시간의 정확한 예측은 합성음성의 음운 지속시간 제어 및 피치 제어등에 매우 중요하다. 통계적인 방법에 의해 지속시간 모델링에 대해 연구되었으나, 정도 높은 지속시간을 예측하기에는 아직 미흡하다[4]-[6]. [4]의 경우 휴지 지속시간에 대한 고려 없이 모델화하였고 [6]의 경우 운율구를 추출하여 휴지 지속시간을 고려한 CART 로 모델화하였으나 정교한 휴지 지속시간을 예측하기에는 불충분 하다.

따라서, 본 논문에서는 실험을 통해 트리를 구성하는 특징 요소를 평가해 최적의 트리를 결정하여 휴지 지속시간을 예측하였다. 2 절에서는 통계적으로 처리하기에 충분한 문 음성 데이터베이스를 구축하고, 이 문음성 데이터베이스를 사용하여 휴지 지속시간을 통계적으로 분석한다. 3 절에서는 휴지 지속시간을 회귀 트리로 모델화한 후 제안된 모델을 평가하여 타당성을 입증한다.

2. 문음성 DB 구축과 휴지 기간 분석

2.1 문 음성 DB 구축

비정형 단위 연결 합성방식과 통계적인 방법으로 일반화된 운율 규칙을 생성하기 위해서는 다양한 경우를 포함하는 많은 양의 데이터가 필요하다. 보다 일반적이며 정교한 제어 모델을 생성하기 위하여 다양한 음운 환경을 고려할 수 있도록 충분히 많은 자연 음성을 분석하여 변화에 영향을 미치는 요인을 추출하여 규칙을 생성하여야 한다. 특히 문음성의 다양한 음운환경에 의한 변화를 분석하기 위해 음운 단위 세그먼트와 휴지구간 세그먼트 되어야 하고 문

법적인 요인에 의한 변화를 분석하기 위해 품사 태깅이 필요하다. 따라서 본 논문에서는 단일 남성 화자가 발성한 400 문장 문음성 데이터를 음운 레벨 세그먼트, 음운 라벨링, 음운별 품사 및 문법정보와 지배소를 태깅한 문음성 데이터베이스를 구축한다.

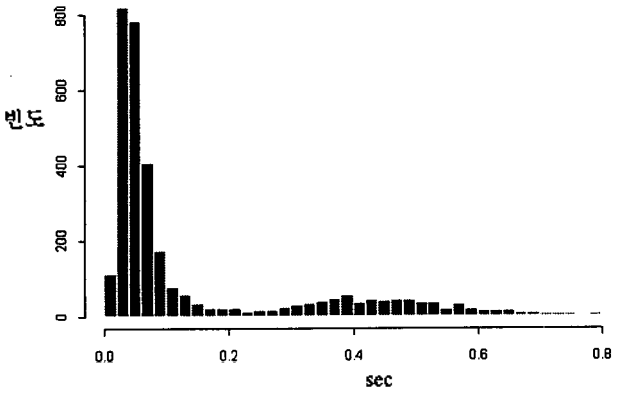
단일 남성 화자가 발성한 400 문장의 문 음성 데이터로부터 어절 경계에만 휴지 구간이 나타난다고 가정하였다. 만약 어절 사이에 휴지가 삽입되지 않았을 경우에는 휴지 기간을 0.0 msec 로 가정한 후, 나머지 휴지 구간 사이의 휴지 기간을 측정하였다. 휴지 기간을 측정한 후 어절 어절 경계들만 모아서 휴지 구간 추출 및 휴지 기간 모델링을 위한 데이터베이스를 재구성하였다.

[표 1] 문 음성 데이터 베이스

화자	남성 단일화자
데이터	400 문장
음운수	45,197
어절 경계수	6,220

2.2 휴지 구간 분석

음성 합성의 경우 입력 텍스트로부터 휴지지속시간의 정확한 예측은 합성음성의 음운 지속시간 제어 및 피치 제어 등에 매우 중요하므로 자연 음성으로부터 휴지지속 시간을 분석하여 문장의 문법 및 문 구조와의 관계를 정확하게 모델링할 필요가 있다. 따라서 본 논문에서는 이들 음성데이터베이스부터 휴지 구간의 분포를 조사한다.



[그림 1] 휴지 기간의 히스토그램

그림 1 은 측정된 6,220 개의 어절 경계 중 휴지 기간의 길이가 0.0msec 인 것을 제외한 것에 대한 히스토그램이다. 그림 1 에서 보듯이 긴 휴지 구간과 짧은 휴지 구간의 차이가 뚜렷이 관찰되며, 50 msec 이하의 휴지 기간이 상대적으로 많이 관찰됨을 알 수 있다. 이는 휴지가 삽입되지 않았거나, 짧은 휴지가 삽입된 경우도 청각적으로 휴지 구간을 자각한다는 것을 의미한다[6].

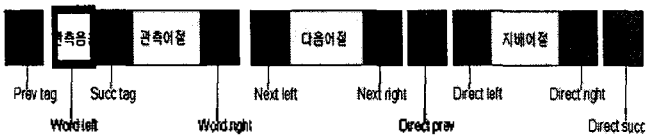
3. 휴지지속시간에 의한 회귀트리 모델링

3.1 휴지 기간 예측을 위한 특징 요소

총 28 개의 특징 요소들을 이용하여 결정 트리 및 회귀트리를 생성한다. 사용된 특징 요소들은 다음과 같다.

- 앞,뒤 음운의 조음양식 및 위치
  : 총 23 분류(BDG/ㄱ,ㄸ,ㅃ/, SZ/ㅆ,ㅉ/, Dup/ㅢ/, LO-C/ㅏ/, HM-C/ㅓ/, HM-B/ㅗ/, HI-B/ㅜ/, HI-C/ㅡ/, HI-F/ㅣ/, LM-F/ㅐ/, HM-F/ㅔ/, Sem1/j+ 단모음/, Sem2/w+ 단모음/, U_C/무성 종성/, V_C/유성종성/, bdg/ㄱ,ㄷ,ㅂ/, hs/ㅎ, ㅅ/, ktpc/ㅋ,ㅌ,ㅍ,ㅊ/, mn/ㅁ,ㄴ/, r/ㄹ/, z/ㅈ/)
- 어절 내 음절의 위치,개수
- 어절 내 음운의 위치,개수
- 어절 내 음운의 상대적 위치
  : 3 분류(S, M, F)
- 문장 내 어절의 위치,개수
- 문장 내 어절의 상대적 위치
  : 3 분류(S,M,F), 5 분류(S, SM, M, Mf, F)
- 지배어절의 위치, 관측어절과 지배어절과의 어절거리
- 관측어절과 지배어절과의 음운거리
- 관측 음운의 품사
  : 총 20 분류(숫자/nb/, 보통명사/nc/, 고유명사/nr/, 단위성 의존명사/nu/, 대명사/np/, 수사/nn/, 접미사/xn/, 연결어미/ec/, 선어말어미/ep/, 어말어미/ef/, 동사/vb/, 형용사/vj/, 보조용언/vx/, 관형사/dn/, 부사/ad/, 감탄사/it/, 주격조사/sp/, 서술격조사/pp/, 보조사/px/, 관형사형 연결어미/ed/)
- 앞,뒤 음운의 품사 (prev tag, succ tag)
- 관측 어절의 좌,우 품사 (word left, right)
- 관측 어절 다음의 좌,우 품사 (next left, right)
- 지배어절의 앞,뒤 품사(direc prev, succ)
  - 지배어절의 좌,우 품사(direc left, right)

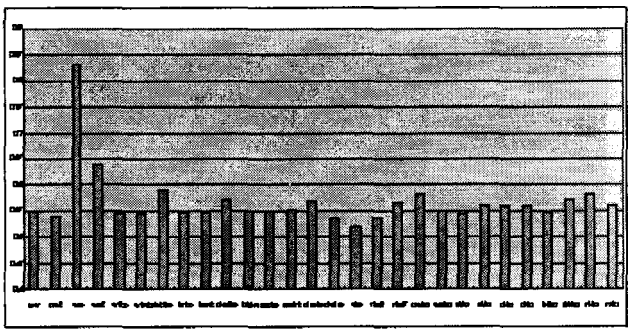
몇몇 품사에 관련된 변수의 도식도는 그림 2 에 예를 들어 도식화한다.



[그림 2] 품사 관련 특징 요소

3.2 휴지 구간 추출을 위한 결정 트리 생성

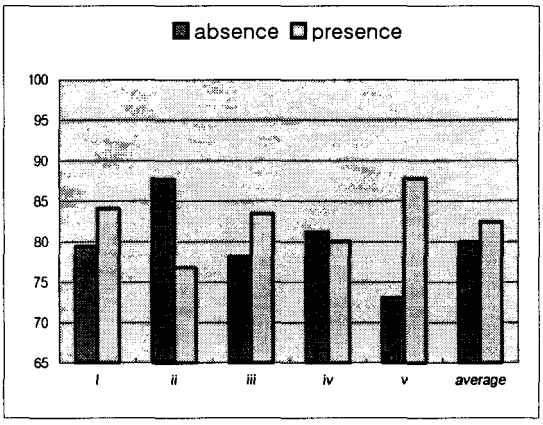
각각의 특징 요소별 휴지 구간 예측율은 그림 3 과 같다.



[그림 3] 특징 요소별 휴지 구간 유무 예측율 그래프

특징 요소별 운율구 유무 예측율을 보면 주로 어절 경계가 나타나는 뒤 음운의 영향이 아주 크게 나타난다. 마찬가지로 관측 어절 경계의 뒤 음운의 품사가 많은 영향을 주고 있다. 위의 특징 요소들과 예측 변수들을 V-fold cross-validation 방법에 의해 최적 트리를 결정하였다.

그림 4 는 원 데이터가 휴지 구간이 없을 경우에 휴지 구간이 없다고 예측한 경우(absence), 휴지 구간이 존재할 경우 휴지 구간이 있다고 예측한 경우(presence)의 예측율에 관한 그래프이다. 각각의 폴더(I~V)에 따라 예측율의 기복이 있지만 전체적으로 볼 때 휴지 구간이 존재할 경우 휴지 구간이 있다고 예측하는 것이 확률이 그 반대의 경우보다 더 크다는 것을 알 수 있다.

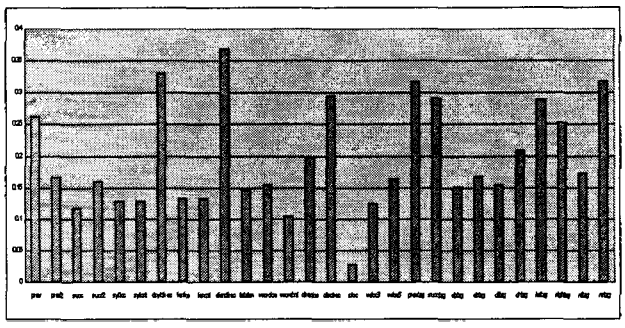


[그림 4] 휴지 구간 예측율 (폴드별, 평균)

3.3 휴지 기간 예측을 위한 회귀 트리 생성

특징 요소별 휴지 기간 상관 계수를 살펴보면 휴지 구간 유무 예측율과는 다르게 주로 어절 경계가 나타나는 앞 음운의 영향이 큰 것을 알 수 있다. 즉, 어떤 음운으로 어절이 끝나느냐에 따라 휴지지속시간에 영향을 미친다. 또한, 품사 관련 특징 요소와 지배소 관련 특징 요소의 상관 계수가 높은 것을 알 수 있다. 관측 음운의 앞뒤 품사나 관측 어절의 좌우 품사, 관측 어절 다음 어절의 좌우 품사 모두 상관 계수가 높게 나타났다. 이와 함께 지배소와 관측 어절(피지배소)간의 거리를 나타 내는 세 특징 요소(지배소까지의 음운 거리, 음절 거리, 어절 거리) 모두 높은 상관 계수로 나타난다.

각각의 특징 요소별 휴지 기간 예측율은 그림 5 와 같다.



[그림 5] 각 특징 요소별 휴지 기간 예측율 그래프

3.4 제안된 모델의 평가

생성된 규칙의 타당성을 확인하기 위하여 관측치와 예측치 간의 오류정도를 평가하고 오류 분석을 행한다.

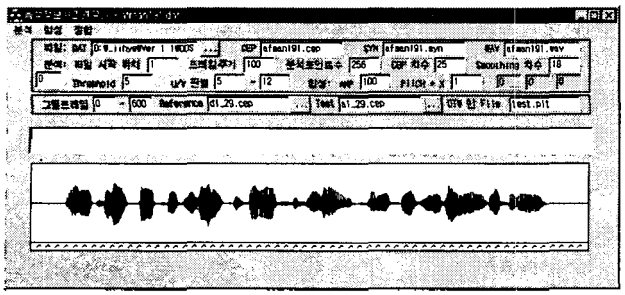
[표 2] 제안된 모델에 대한 평가

	휴지 구간 예측율 (평균)	
휴지구간 존재	무(2,990)	유(3,230)
무(absence)	2,390(79.9%)	566(17.4%)
유(presence)	600(20%)	2,664(82.5%)

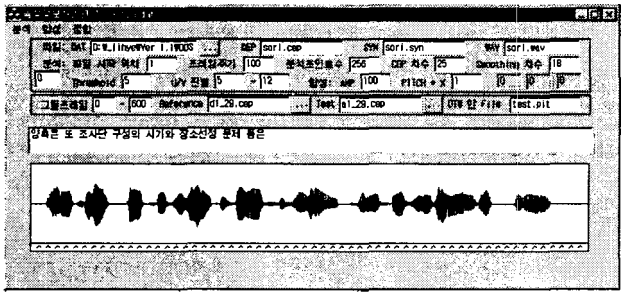
	휴지 지속시간 회귀트리
다중상관계수	0.84
예측오차 20ms 이내	88%

3.5 합성음 생성

결정된 트리를 기존의 트리 코드 생성기를 통하여 모델화하여 합성음의 자연성을 위해 휴지 구간을 예측하고 휴지가 발생하였을 경우 휴지 지속시간을 예측하여 합성음에 삽입한다. 그림 8은 원음의 파형이며 그림 9는 휴지 구간을 삽입한 합성음의 파형이다. 그림 9에서 보여지듯 휴지 구간을 삽입하였을 경우 좀 더 자연스러운 합성음이 생성됨을 알 수 있다.



[그림 8] 원음의 파형



[그림 9] 휴지 구간을 삽입한 합성음의 파형

4. 결론

본 논문에서는 남성 단일 화자가 발성한 문음성 데이터베이스(400 문장, 45,197 음운)으로부터 휴지 지속시간을 예측하기 위하여 휴지 구간만을 추출한 데이터 베이스(총 6,220)를 구축하였다. 이 데이터 베이스로부터 휴지 구간에 대한 분포를 분석하여 휴지 구간 추출을 위한 결정 트리를 생성하였다. 또, CART를 이용하여 V-fold cross-validation 방법에 의해 최적의 휴지 지속 시간을 회귀 트리로 모델링 하였다. 그 결과 휴지 기간 예측 상관 계수는 0.84로 나타났으며 정확율 20ms 이내에서 88%로 나타났다.

결정된 트리로 합성음의 자연성을 위해 휴지 기간을 예측하고 휴지 구간을 삽입하였다. 휴지 구간의 추출은 대체적으로 원음의 휴지 구간과 일치하였고, 휴지지속시간의 예측 후 합성음에 삽입 하였을 경우 휴지지속시간이 길게 나타나는 구간도 있었지만 대체적으로 원음과 유사하게 나타나 휴지 구간을 삽입 하지 않았을 경우보다 매우 자연스러웠다.

금후 연구과제로는 아직 자연스러운 운율을 기대하기에는 좀 더 높은 정확율이 요구되므로 최근에 제안되어지고 있는 여러 방법을 통하여 정확율를 높이도록 해야겠다.

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

[4] 정지혜, 이양희, "정규화 지속시간 회귀트리를 기반으로 한 음운 지속시간 모델화", 한국음향학회 학술대회발표, PP 278-281 1998

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

[6] JH Chung, YH Lee, "A Study on Korean concatenative speech synthesis using Non-uniform units", ICSP99, pp 167-172, 1999.