POSTTS : 자연어 분석을 통한 코퍼스 기반 한국어 TTS

*하주홍°, *정 옥, **김병창, *이근배
*포항공과대학교 컴퓨터공학과
**위덕대학교 컴퓨터멀티미디어공학부

POSTTS : Corpus Based Korean TTS based on Natural Language Analysis

*Ju-Hong Ha, *Yu Zheng, **Byeongchang Kim, *Gary Geunbae Lee
*Dept. of CSE, POSTECH
**Division of Computer and Multimedia Engineering, UIDUK University
E-mail : {miracle°, zhengyu, bckim, gblee}@postech.ac.kr

Abstract

In order to produce high quality synthesized speech, it is very important to get an accurate grapheme-to-phoneme conversion and prosody model from texts using natural language processing. Robust preprocessing for non-Korean characters should also be required. In this paper, we analyzed Korean texts using a morphological analyzer, part-of-speech tagger and syntactic chunker. We present a new grapheme-to-phoneme conversion method, i.e. a dictionary-based and rule-based hybrid method, for unlimited vocabulary Korean TTS. We constructed a prosody model using a probabilistic method and decision tree-based method.

I. 서론

음성 합성에 대한 연구가 18세기부터 시작되었지만 음성 합성을 위한 텍스트 분석은 20세기 후반에야 이루어졌다. 최초의 완전한 TTS (Text-to-Speech)는 영어를 대상으로 하는 MITalk으로 일반 단어뿐만 아니라 특수기호나 숫자, 약어 등도 음성 합성이 가능하였다. 그 뿐만 아니라 형태소 분석이나 문자 발음 변환(grapheme-to-phoneme)에 사용된 기법들도 정교하여 텍스트 분석의 필요성에 대한 선례를 남겼다. 그 이후로 TTS에 사용되는 언어학적 지식은 거의 필수적이라 여겨지고 있다[1]. 현재 선진국들의 음성합성 기술은 어느 정도 이상의 수준에 이르러, 주로 텍스트를 분석하여 그 결과에 따른 합성음의 변형, 즉 합성음의 고저를 조절한다든지, 합성음의 속도를 조절하는 등의 연구를 수행하고 있다.

국내의 TTS에 대한 연구는 이미 대기업 연구소를 중심으로 상품화 수준으로 행해지고 있다[2]. 그 중 일부에서는 기본적인 텍스트 분석을 통한 운소의 추출과 적용을 연구하고 있으나 아직 수준 있는 한국어 텍스트 분석 기술의 미보유와 한국어 음운 현상에 대한 체계적인 정리가 전반적으로 미흡한 실정이다[3]. 그 이유는 국내의 음성합성 연구자들이 신호처리를 주 연구 대상으로 하여 자연어처리 분야에 제한된 지식만을 가지고 있기 때문인데, 이것을 극복하기 위해서는 본 논문에서처럼 자연어 처리 기술과 음성 합성 기술과의 접목이 시급히 이루어 져야 한다.

본 논문에서는 형태소 분석, 품사 태깅, 구문 단위화(chunking)를 통해 효율적인 자연어 처리를 하고, 음운 접속 정보에 의한 음운 현상 처리, 단위화와 기계학습 결과에 의한 휴지구간(phrase break) 추출 등에 기반한 보다 자연스러운 합성음을 생성하는 POSTTS[1] 시

1) POSTECH Text-to-Speech system
데모사이트: http://nlp.postech.ac.kr/~project/Research/POSTTS/demoframe.html

스템을 구축하였다. 본 논문은 새로운 이론적 사실의 발견보다는 이러한 POSTTS 시스템의 전반적인 소개를 목적으로 한다.

따라서 본 논문의 구성은 다음과 같다. 2장에서는 형태소 분석과 품사 태깅, 단위화의 자연어 처리 모듈에 대해 기술하고, 3장에서는 띄어읽기 단위인 언절 경계 추정에 대해 살펴보고, 4장에서는 음운 현상을 고려한 문자 발음 변환 방법에 대해 다룬다. 5장에서는 보다 자연스러운 합성음 생성을 위한 운율 추정 모듈에 대해 살펴보고, 6장에서는 문자 발음 변환에 대한 실험 및 분석을 하며, 7장에서 결론 및 고찰을 기술한다.

II. 텍스트의 자연어 처리와 분석

(1) 형태소 분석 및 품사 태깅

TTS 시스템에서 정확한 발음열 변환 및 자연스러운 합성음 생성의 기초 핵심 작업으로 그림 1과 같이 한국어 텍스트 분석을 수행한다.

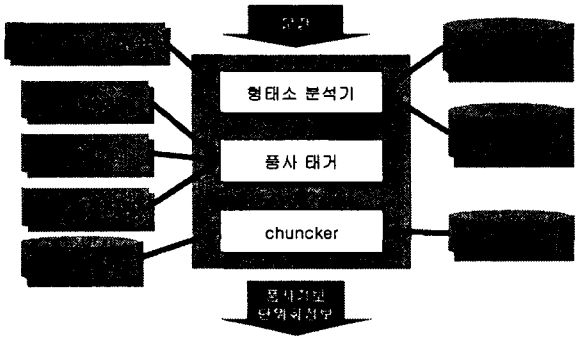

그림 1. 형태소 분석기, 품사 태거, chunker의 통합

형태소 분석기는 Character-synchronous Dynamic Programming Algorithm과 Trie 사전을 기반으로 구현되어 있으며, 그림 1에서와 같이 자소열-형태소 사전, 자소열-형태소 패턴 사전을 가지고 있다. 자소열 형태소 사전에는 여러 어절에 걸친 형태소를 분석하기 위한 다어절어 사전을 포함하고 있다. 그리고 자소열-형태소 패턴 사전을 이용하면 분석 실패 후 추가분석을 하는 기존의 방법과는 달리 형태소 위치와 개수에 무관하게 미등록어 추정이 가능하다. 그리고 패턴사전은 등록어와 같은 방법으로 미등록어의 품사를 추정할 수 있도록 하여 형태소 분석기의 구조를 간단하게 하는 효과가 있다. 품사 태거에서는 형태소 분석과 동시에 Viterbi 검색 방법을 이용한 태깅을 수행한다. 후에 오류 수정 규칙을 이용하여 확률 태거의 오류를 수정하면 가장 최적의 품사열이 구해진다[4].

(2) 단위화 (chunking)

자연어 문장의 구문분석에 있어서 가장 문제가 되는 것은 다양한 분석 결과들이 존재한다는 것이다. 다양한 결과들을 도출, 처리, 분석하는데 지나치게 많은 시간이 소요될 수 있으며 잘못된 분석 결과를 얻어낼 확률도 그만큼 증가하게 되는 등의 문제가 있다. 본 논문에서는 위에서 언급한 문제점을 해결하기 위해 문장에 대해 명사구/ 동사구 패턴을 적용하여 문장을 단위화하였다. 단위화는 구문분석에 비해 감소된 분석 시간과 복잡도를 가지고 언절 경계 추정에 필요한 정보들을 생성한다.

III. 언절 경계 추정

(1) 결정 트리(decision tree)를 이용한 언절 경계 추정

언절 경계 추정을 위한 통계적 방법은 일반적으로 많은 양의 훈련 데이터를 필요로 한다. 이러한 훈련데이터 부족 현상을 해소하기 위해 smoothing 기법을 사용하지만 그 현상을 완벽하게 해소할 수는 없다. 본 연구에서는 이러한 데이터 부족 현상을 극복하기 위해 확률적 추정과 동시에 결정 트리를 언절 경계 추정에 사용하였다.[5]

형태소 분석 결과를 바탕으로 언절 경계 후보 (띄어쓰기)에 대해서 확률적 추정을 바탕으로 결정 트리를 사용하여 언절 경계를 수정한다. 언절 경계 후보를 기준으로 좌우 각각 7개 어절의 마지막 형태소의 품사 태그를 수정 시 자질(feature)로 사용한다. 7개 미만일 경우 "z"라는 표기로 어절이 존재하지 않음을 표시한다. 언절 경계 후보 바로 앞 어절의 위치정보, 현재 문장의 길이 등을 어절단위로 표현하여 형태소 품사 태그열과 함께 자질로 사용한다.

형태소 분석기에 의한 형태소 품사 태그열과 정답 언절 경계를 포함하는 형태소 품사 태그열을 비교하여 결정 트리를 C4.5에 의해 구축하고, 형태소 분석 결과와 현재 언절 경계 후보의 선행 어절에 대한 위치 정보를 사용해서 언절 경계를 추정하게 된다.

(2) 단위화 정보를 사용한 언절 경계의 보정

단위화 정보는 문장의 구조를 파악하고, 문장 내에서 역할이 같은 몇 개의 어절 단위로 묶는 것을 말한다. 이러한 단위화 정보는 구조상 그 자체만으로도 언절 경계가 될 수 있으나, 보다 안전하고 강건한 언절 경계 추정을 위해서 결정 트리와 상호 보완할 수 있는 방법을 고안하였다. 즉, 결정 트리의 결과와 단위화 결과가 일치할 경우 추정된 언절 경계의 강도를 심화시키고, 일치하지 않을 경우 추정된 언절 경계의 강도를

약화시켰다. 따라서 낮은 빈도로 나타나는 짧은 휴지기의 빈도를 높여 줄 수 있으며, 극단적인 에러의 발생을 감소시킬 수 있다.

Ⅳ. 문자 발음 변환

한국어의 음운 변이는 주로 품사에 의해 많이 영향을 받는다. 그림 2는 형태소 분석기와 품사 태거를 기반으로 하는 자소열-음소열 변환기의 구조를 나타내고 있다.

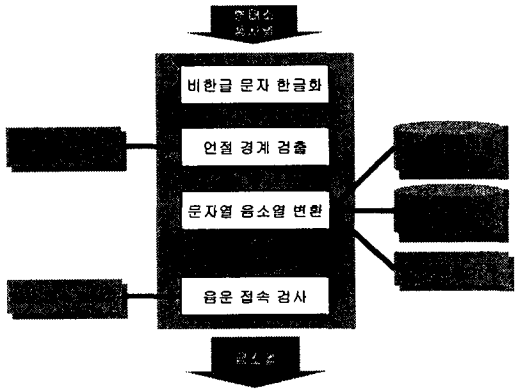

그림 2. 음운 현상을 고려한 발음 생성

미등록어 추정이 가능한 품사 태거의 분석 결과를 사용하여 각 형태소의 분류와 이형태 정보로 형태소-음소열 사전을 검색하면 각 형태소의 가능한 발음들을 얻을 수 있다. 음소열 미등록어인 경우에는 형태소-음소열 패턴 사전으로부터 경계 자소의 가능한 음운변이에 따라 음소열 패턴들이 검색된다. 음소열 미등록어의 경계 자소를 제외한 내부 자소들은 종성, 초성, 중성을 한 단위로 하는 CCV 변환 규칙에 따라 음소열로 변환한다. 각 형태소의 음소열은 음운 접속 정보에 따라 문장의 음소열을 형성한다[6],[7].

한편, 입력 문장에는 순수한 한글 뿐만 아니라 많은 숫자와 영문자, 기호 등이 포함될 수 있다. 이러한 비한글 문자들도 일반적인 한글과 함께 음성으로 변환될 수 있도록 올바른 음소열로 변환하여야 한다. 본 연구에서는 숫자는 유한 오토마타를 사용하고, 날짜, 점수, 수식 등은 패턴을 분석하여 해당 발음으로 변환하였다.

Ⅴ. 운율 추정

휴지 구간(phrase break)을 제외하고 운율은 음의 피치(pitch)와 길이(duration), 세기(intensity) 세 요소로 볼 수 있는데, 음의 세기의 경우 경험적으로 피치나 길이에 비해 그 역할이 미미함으로 평균 세기를 미리 구해 추정 시에 사용하고, 본 장에서는 피치와 길이에 대한 학습과 추정을 다룬다.

합성해내고자 하는 음을 위한 운율은 크게 학습과 추정 두 단계를 걸쳐 추정하게 된다. 즉, 음성 DB를 기준으로 운율을 오프라인으로 학습하여 운율 모델을 마련한 뒤, 온라인 합성 시에는 그 모델에 의해 운율을 추정하는 방식이다.

(1) 언어 특징 및 음성 특징 추출

운율 학습에 쓰이는 특징 벡터는 표 1과 같다.

표 1. 특징 벡터

① 현 어절의 어절 태그	⑦ 현 어절 문장 내 상대 위치
② 왼쪽 어절의 어절 태그	⑧ 현 어절의 언절 내 상대 위치
③ 오른쪽 어절의 어절 태그	⑨ 현 어절 첫 음절의 문장 내 상대 위치
④ 현 어절의 음절 수	⑩ 현 어절 첫 음절의 언절 내 상대 위치
⑤ 현 언절 내 어절 수	⑪ 현 어절 첫/중간/끝에서 두 번째/ 끝 모음의 피치/길이
⑥ 현 언절 내 음절 수	

처음 10개는 텍스트 표현을 분석하여 구할 수 있는 언어 특징들이고, 마지막 특징은 문장의 음성 표현을 근거로 하여 얻어지는 음성 특징이다. ①~④는 어절 태깅 과정에서 얻을 수 있다. ⑤~⑩은 정보 추출 범위를 언절 또는 문장 단위까지 확대하여 획득한 특징들로서, 운율 생성에 영향을 줄 수 있는 광역의 언어 정보도 이 특징들을 통해 함께 고려할 수 있다.

(2) 음성 특징 추출

학습은 11번째 음성 특징을 모델로 추상화하는 것으로, 나중에 운율 생성 시에 모델로부터 이 특징을 추정한다. 피치 특징은 각 학습 문장의 음소 분할 정보와 피치 정보를 근거로 하여, 길이 특징은 음소 분할 정보만을 근거로 하여 각각 추출한다. 한 어절의 수가 1인 경우 마지막 음절로, 2인 경우에는 첫 음절과 마지막 음절로, 3인 경우에는 첫 음절, 끝에서 두 번째 음절, 마지막 음절로, 4 이상의 경우에는 첫 음절, 중간 음절, 끝에서 두 번째 음절, 마지막 음절로부터 네 지점을 위한 음성 특징 네 개를 동일한 방식으로 추출한다. 이런 방식으로 학습 데이터를 피치, 길이 각각 4개씩 8개 파일을 생성한다. 학습은 기계 학습 툴 중 하나인 Weka[2]로 이루어졌다.

(3) 운율 추정

운율 추정은 합성해내고자 하는 각 음소의 운율, 즉 피치와 길이를 추정하는 것으로 학습된 모델을 근거로 하고, 모델 학습에 쓰였던 특징 벡터의 마지막 원소인 음성 특징을 추정하는 것이다. 예를 들어, 어떤 어절의

2) 무료로 공개된 소프트웨어로서, 다운로드를 위한 웹 페이지는 http://www.cs.waikato.ac.nz/~ml/weka/index.html 이다.

마지막 모음에 관한 피치를 추정하고자 한다면 그 어절로부터 10차원의 언어 특징 벡터를 구성한 뒤에 어절 내 마지막 모음의 피치 추정을 위한 모델에 의거하여 그 모음을 위한 피치를 유도해낸다.

다음으로 음절수가 5 이상인 어절일 경우 그림 3과 같이 거점 모음들 외의 모음들은 거점 모음들의 단순 내삽(interpolation) 기법을 적용해 피치, 길이 곡선을 그리고 각 모음들 위치에 해당하는 곡선 위의 점을 해당 모음의 피치와 길이 값으로 사용한다.

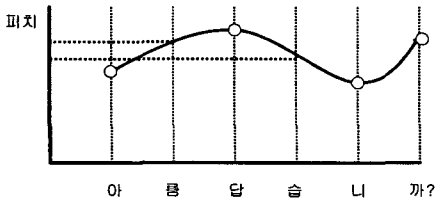

그림 3. 내삽에 의한 모음 피치 추정

한편 자음의 피치는 현재 모음의 피치에서 직전 모음의 피치를 뺀 값의 반을 구한 뒤에 현재 모음의 피치에서 그 반값을 뺀 값을 초성 자음을 위한 피치로, 그 반값을 더한 값을 종성 자음의 피치로 각각 부여한다. 자음의 길이는 음성 DB로부터 미리 구해 놓은 각 자음의 평균 길이를 그대로 가져다 쓴다.

Ⅵ. 실험 및 분석

언절 경계 검출을 위하여 MBCNEWSDB[3]를 사용했다. 훈련 데이터 부족을 보완하기 위해 앞에서 언급했듯이 통계적 방법으로 언절 경계를 추출하고, 결정 트리를 이용하여 후처리를 수행하였다. 통계적 학습, 결정 트리 후처리 학습, 평가 데이터를 4:5:1 비율로 했을 때 언절 경계 정확도 62.74%, 띄어쓰기 정확도 85.57%로 통계적 방법만 사용한 것보다 각각 10%, 4%의 성능 향상을 보였다.

문자 발음 변환에서는 Sent3000DB[4] 중 9,773문장을 사용하여 2,030여 개의 CCV 변환 규칙을 생성하였다. 나머지 4,973 문장을 변환기의 입력으로 사용해 출력된 음소열과 형태소-음소열 말뭉치를 비교하여 변환기의 성능을 평가하였다. 4,973 문장 중 4,853 문장이 정확하게 음소열로 변환되어 97.5%의 변환율을 보였고, 자소 218,687개 중 218,514개가 정확하게 변환되어 자소 변환율은 99.9%가 되었다. 모든 문장이 정확하게 변환되지 못한 주요 요인은 음소열 미등록어에 대해서 같은 기본 자소단위가 다른 두 발음으로 변하는 경우가 대부분이었다. 그리고 품사 태깅의 오류에 의해서 음소열이 바르지 못하게 변환된 것도 있었다. MBCNEWSDB의 210문장에 대한 자소열-음소열 변환 실험에서 85.5%의 문장과 99.0%의 자소가 정확하게 음소로 변환되었으며, 언절 경계 검출 후 자소열-음소열 변환은 90.8%와 99.3%로 향상되었다.

3) 본 실험실에서 MBC NEWDESK 앵커의 발화를 녹음, 기록, 발음 변환하여 수집한 코퍼스 (10,000여 문장의 음성)

4) 한국과학기술원 통신연구실의 무역상담 도메인 음성 DB (3,000여 개 단어, 14,746 문장의 음성)

Ⅶ. 결론

본 논문에서는 보다 자연스러운 TTS 합성음을 위한 운소의 첨가와 음운 현상의 모델링에 필요한 자연어 처리 기법을 중심으로 코퍼스 기반 한국어 TTS 시스템에 대해 살펴보았다. 이를 통해 형태소 품사 정보와 음운 변이와의 관계, 구묶음과 언절 경계와의 관계를 파악하고, 한국어의 운율 생성에 필요한 언어 정보를 얻을 수 있었다.

참고문헌

[1] Shih, Chilin and Sproat, Richard, "Issues in Text-to-Speech Conversion for Mandarin", *Computational Linguistics and Chinese Language Processing,* Vol.1, no14, pp37-86, 1994

[2] 이준우, 김세린, 김상수, 이종석, 김민성, "수정된 음절을 이용한 한국어 문장-음성 변환 시스템", *제 13회 음성 통신 및 신호처리 워크샵 논문집,* pp.237-240, 1996

[3] 이양희, "음성 합성 기술 개발의 현황과 과제", *제 1회 음성학 학술대회 자료집,* pp145-155, 1994

[4] 차정원, "일반화된 미등록어 처리를 이용한 혼합형 품사 태거", *포항공과대학교 컴퓨터공학과 석사 학위 논문,* 1999

[5] Eric Sanders, "Using Probabilistic Methods to Predict Phrase Boundaries for a Text-to-Speech System", *MS Thesis of University of Mijmegen,* 1995

[6] 박수현, 권혁철, "한국어 Text-to-Speech 변환을 위한 음운 변동 시스템에 관현 연구", *'95 한글 및 한국어 정보처리 학술 발표 논문집,* pp.35-38, 1995

[7] 위선희, 정민화, "음운변화 규칙을 이용한 음성 사전 생성", *HCI '97 학술대회 발표 논문집,* pp.308-313, 1997