

한국어 운율규칙을 이용한 음성합성시스템

Speech synthesis system using Korean prosodic rules

이기영, 송민석*

관동대학교 전자통신공학과, *관동대학교 영어영문학과
kylee@kdcgs.kwandong.ac.kr, *mssong@kdcgs.kwandong.ac.kr

Abstract

This paper proposes the speech synthesis method using Korean prosodic rules as an important technique for Korean speech synthesis. The prosodic model for speech synthesis is composed of accentual phrases and intonational phrases which are derived from hierarchical structure of prosody. This prosodic model controls duration time, intonation and pause of synthesized speech. Synthesis units constitute of demi-syllables and VCV-triphones which can make unlimited vocabularies, and TD-PSOLA is used as the synthesis method.

요 약

본 연구에서는 한국어 음성합성을 위한 핵심 기술로서 한국어 운율규칙을 이용한 음성합성 방법을 제안하고 음성합성시스템을 구현한다. 여기서 한국어 운율구는 운율의 계층구조로부터 유도하여 정의된 억양구와 강세구를 사용한다. 또한 강세구와 억양구를 이용하여 생성된 운율규칙이 고품질의 음성합성에 필요한 지속 시간, 억양 및 휴지기를 제어한다. 합성단위는 무제한 어휘가 가능한 반음절과 triphone(VCV)의 조합이며 합성방식은 TD-PSOLA를 사용한다.

1. 서 론

음성합성은 기계가 인간의 음성을 합성하여 내도록 하는 기술이며, 기계에 의한 합성음성은 영묘하고 자연스러워야 한다. 최근 음성신호처리기술의 발달과 함께 맨-머신 인터페이스에서 처리결과를 무제한 음성 메세지로 비교적 명확하고 자연스러운 목소리로 들려주기 위한 합성방식으로 활발한 연구 및 개발이 진행되고 있다. 그러나 아직까지 규칙합성에 의한 합성음의 음질 수준이 자연음성에 비하여 매우 큰 열세를 보이고 있는 것은 자연스러운 음성의 합성을 위한 기본적 기술인 언어처리 기술, 운율제어기술 및 음성생성기술 모두가 현재 충분한 수준에 도달하지 못하였기 때문이다.

문장단위 이상의 연속음성은 단독 발생된 고립단어 음성과 달리 지속시간, 억양, 크기 및 휴지기 등의 운율이라는 언어적 정보를 가지고 있다. 대화를 하거나 낭독을 하는 화자는 이 운율을 이용하여 문장의 의미, 감정, 의도 및 마음가짐 등을 전하며, 청취자는 화자가 발성한 음성으로부터 운율을 감지하여 전체적인 의미를 파악한다.

일부 음성합성 연구에서는 일본에서 제안한 합성음성의 운율제어에 사용하고 있는 Fuzisaki 모델을 이용하고 있으니 이모델은 주로 액센트언어인 일본어와 영어에 적합하므로 한국어

에 적용할 경우 합성음성이 어색하며 모델에 사용하는 파라메타가 한국어에 맞도록 변형되어야 할 부분이 많다. 이러한 운율 특성은 언어마다 고유한 것으로, 외국의 합성시스템을 그대로 모방함으로써 한국어에 적합한 합성시스템의 구현이 이루어 질 수 없다.

본 연구에서는 한국어 음성합성을 위한 핵심 기술로서 한국어 운율구를 이용한 음성합성 방법을 제안하고 한국어 음성합성시스템을 구현한다. 여기서 한국어 운율구는 각 국의 언어에 공통적으로 적용할 수 있는 운율의 계층구조로부터 유도하여 정의된 억양구와 강세구를 사용한다. 억양구는 주로 휴지기를 동반하며 문장중간의 것과 문장말의 것으로 나뉘어 진다. 강세구는 문법적 어절의 경계와 거의 일치하나 100% 일치하지는 않는다. 이러한 한국어 운율구의 두 요소를 한국어 음성합성의 운율제어에 적용하였으며 반음절단위와 VCV 단위의 한국어 음성합성시스템을 구현하였다. 평가를 위하여 운율제어 없는 합성음성과 청취비교한 결과 한국어 운율구를 적용한 합성음성이 우수함을 확인하였다.

2. 한국어 운율규칙

운율구조는 현재까지의 음성성언어처리 시스템이 어휘가 적거나 응용분야가 한정된 경우의 문장들을 대상으로 하는 한계를 벗어나기 위한 가장 좋은 방법으로 주목되어, 유럽, 미국 및 일본등이 이에 대한 많은 연구가 시작되었다. 이

러한 연구들에서 주로 사용하고 있는 각국의 음성언어에 적용될 수 있는 범용적 운율 계층구조는 1986년 Nespor와 Vogel이 제안한 계층적 운율구조(hierarchical prosodic phrase)이다.

(1) 한국어 운율구조

전 세계의 언어를 총괄적으로 생각해 보면 Nespor와 Vogel이 제안한 계층적 운율구조가 모두 나타나고 있으니, 각국의 언어마다에서 이들의 계층구조가 그대로 적용되는 것은 아니다. 언어학자 전선아(3)는 한국어의 경우 강세구(accentual phrase)와 억양구(intonational phrase)가 한국어의 기본적 운율구조를 구성하고 있다고 제안했다. 본 연구에서는 전선아의 운율구조 이론을 받아들여 한국어 음성합성을 위한 운율구 단위의 언어학적 기반으로 삼았다. 여기서 억양구(intonational phrase)는 Nespor와 Vogel이 제안한 보편적 계층구조와 일치하는 개념으로서, 한국어의 경우 톤레벨(tone level)이 H(igh)와 L(ow)로 구성되는 억양격체에 의해 정의된다. 여기서 억양격체는 피치엑센트(pitch accent)와 어절톤(phrase tone)으로 구성되며, 피치엑센트란 어휘 고유의 엑센트(lexical stress)이며 어절톤은 어절의 엑센트와 경계톤(boundary tone)으로 구성된다. 어절의 엑센트는 억양구보다 작은 단위이며, 경계톤은 억양구 단위를 구분하는 부분이다. 이와 같이 운율구 단위를 억양구와 그보다 작은 단위로 나눌 수 있으며 억양구 단위보다 작은 단위를 강세구(accentual phrase)

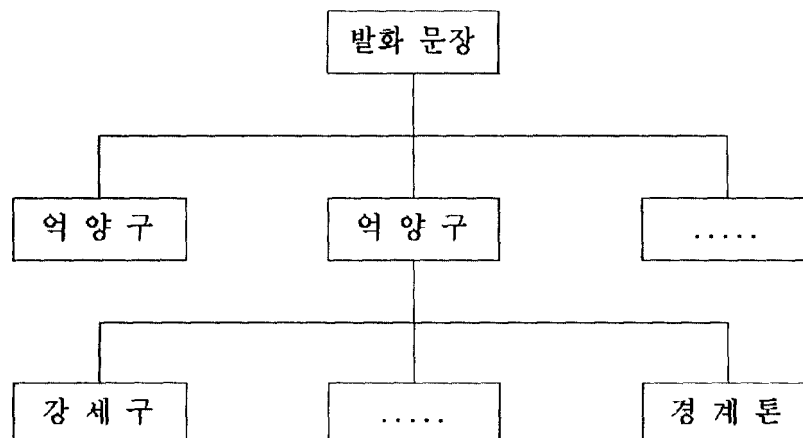


그림 1. 한국어 운율구조

한국어 운율규칙을 이용한 음성합성시스템

라 한다. 언어학적으로 제시된 억양구(intonational phrase)와 강세구(accentual phrase)를 이용하여 한국어 발화문장(utterance)에 대한 운율구조는 그림 1 과 같다.

(2) 한국어 운율분석과 규칙

한국어 운율구조를 음성합성에 적용하기 위하여 입력된 텍스트로부터 지속시간, 억양 및 휴지기를 구하기 위한 운율의 분석과정은 다음과 같다.

첫째, 지속시간은 텍스트 분석과정에서 부여된 음절별 지속시간을 그대로 이용한다.

둘째, 텍스트로부터 억양을 분석해 내기 위한 문장의 각 어절을 강세구로 하되, 침표(.) 또는 마침표(.)로 끝나는 마지막 서술어절을 경계톤으로 하며, 억양구는 시작하는 강세구부터 경계톤까지로 한다.

셋째, 휴지기는 이미 분석된 억양구마다 일정시간을 삽입한다.

다음 그림 2 는 입력된 텍스트 “자연어는 학교에 갑니다.”를 운율분석하여 생성된 운율규칙이다.

3. 한국어 음성합성 시스템 구현

(1) 데이터베이스의 구축

고품질의 무제한 어휘에 대한 음성합성 시스템을 구현하기 위하여 이용하는 합성방식은 TD-PSOLA 방식이며 데이터베이스의 합성단위는 반음절과 triphone (VCV)의 조합이다. 데이터베이스의 용량은 약 7Mbyte 이지만 pitch-synchronous한 원음성의 파형이므로 실시간 합성에 용이하다.

본연구의 음성합성 시스템을 구현하기 위하여 사용하는 데이터베이스의 합성단위는 반음절과 triphone(VCV)의 조합이며 합성방식은 운율제어에 뛰어난 PSOLA 합성법을 이용한다.

데이터베이스는 이상의 채취 기준에 의한 반음절단위 및 VCV 단위의 음성으로부터 각각의 PSOLA 처리된 파형으로 구성한다. 이 데이터베이스 구축을 위한 해식조건은 표 1 과 같다.

표 1. 해식 조건

A/D data	10kHz sampling, 16-bit
window function	Hamming
window length	25.6 ms
window shift	6.4 ms
PSOLA signal	pitch-synchronous analysis & Hanning window

(2) 합성결과

실제로 입력된 문장이 “자연어는 학교에 갑니다.”라고 하면 우선 입력된 문장에 음운규칙을 적용하여 소리나는 대로 문자를 변환해 주며 구문분석과정을 통하여 합성음성에 해당하는 운율규칙의 기본형을 작성한다. 다음 그림 3(a)는 “자연어는 학교에 갑니다.”라는 입력문장에 대해 운율규칙을 적용하지 않고 합성한 결과이며 그림 3(b)는 한국형 운율규칙의 기본형을 적용하여 합성한 결과이다.

4. 결론

본 연구에서는 한국어 음성합성을 위한 핵심 기술로서 한국어 운율규칙을 이용한 음성합성 방법을 제안하고 음성합성시스템을 구현하였다. 여기서 한국어 운율구는 억양구와 강세구를 사용하였다. 이러한 한국어 운율구의 두 요소를 한국어 음성합성의 운율제어에 적용하였으며 반음절단위와 VCV 단위의 한국어 음성합성시스템을 구현하였다. 평가를 위하여 운율제어 없는 합성음성과 청취비교한 결과 한국어 운율구를 적용한 합성음성이 우수함을 확인하였다.

* 참고 문헌

- [1] Colin W. Wightman and M. Ostendorf, "Automatic Labeling of Prosodic Patterns," IEEF Trans., Speech, Audio Processing, vol. 2, no. 4, pp. 1994.

- [2] 전선어, The Phonetics and Phonology of Korean Prosody, Ph.D. Dissertation, Ohio State University, 1993.
- [3] H. Valbret, E. Moulines, J. P. Tubach, "Voice transformation using PSOLA Technique," EUROSPEECH 91, pp. 345-348, 1991.
- [4] F. J. Charpentier, M. G. Stella, "Diphone synthesis using overlap-add technique for speech waveforms concatenation," ICASSP 86, pp.2015-2018, 1986
- [5] 이기영 외 3인, "타화자의 어기신호를 이용한 억양 변환," 한국음향학회지 제 14 권 제 4 호, pp. 21-28, 1995.
- [6] 이기영, 송민석, 최창석, "한국어 중의성문장을 구분하기 위한 억양패턴의 자동 인식," 한국음향학회 학술대회 논문집, 131-134, 1995.

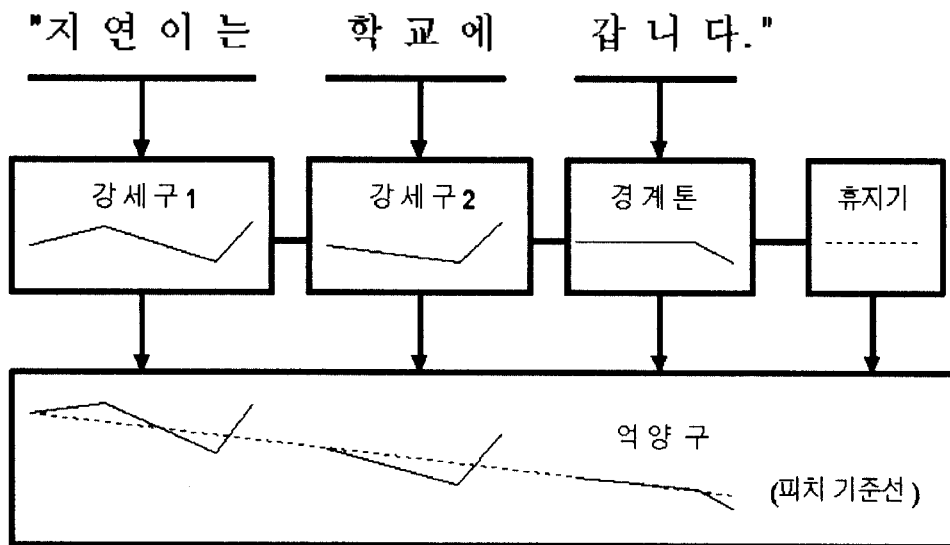


그림 2. 한국어 운율규칙