

# 한국어 Text-to-Speech 변환을 위한 음운 변동 시스템에 관한 연구

박 수현, 권 혁철  
부산대학교 전자계산학과

## Implementation to phonological alteration module for a korean text-to-speech

Su-Hyun Park, Hyuk-Chul Kwon  
Dept. of Computer Science, Pusan National University

### 요 약

Text-to-speech 시스템은 텍스트를 입력으로 받아 텍스트와 일치하는 음성을 출력하는 시스템으로, 인간이 자신의 모국어로 텍스트를 읽는 것과 비슷한 수준의 음성을 출력하는 데 목적이 있다. 한국어의 각 단어들은 한 단어 내에 있는 형태소들 사이에 음운 변동 현상을 일으켜 쓰여진 형태와 다르게 발음된다. 그러므로 한국어 텍스트를 자연스럽게 발음하기 위해서는 음운 변동 현상을 효율적으로 처리할 수 있어야 한다. 한국어에서 음운 변동을 일으키는 규칙은 여러 가지이고, 정확한 발음을 위해서는 이러한 규칙들이 차례대로 적용되어야 한다. 따라서 본 논문에서는 이러한 한국어의 발음상의 특성을 고려하여 two-level 모델에 기반한 음운 변동 시스템을 구현한다.

### 1. 서 론

Text-to-speech 변환 시스템은 자료 혹은 텍스트로 된 언어 정보(linguistic information)를 알맞은 음성으로 변환하는 시스템으로, 인간이 자신의 모국어로 텍스트를 읽는 것과 비슷한 수준의 음성을 출력하는데 그 목적이 있다. Text-to-speech 변환 시스템은 사용자에게 더욱 친근한 인터페이스(interface)를 제공하기 위하여 발전되어 왔다. 즉, 소경(blind people)들을 위한 음성 독서 장치(audio reading device), 전자 우편(electronic mail)을 음성 우편(voice mail)으로 변환하는 시스템 등에 응용될 수 있다.

기존의 음성 합성 기술(speech synthesis technology)들은 제한된 수의 문장으로 고정된 텍스트만을 합성하는 수준이었고, 이러한 능력을 가진 장치들은 상업적으로 유용하게 사용되었다. 이러한 장치를 위해 사용된 기술들은 기본적으로 실현하려고 하는 목표 언어의 특성과는 무관하다. 그러나 제약 없이 가하지 않은 텍스트를 입력으로 받아, 인간이

자신의 모국어로 텍스트를 읽는 것과 비슷한 수준의 음성을 출력하는 Text-to-speech 변환 시스템을 개발하기 위한 기술은 일반적으로 목표 언어의 특성에 종속적이다. 즉, 한국어 Text-to-speech 변환 시스템은 한국어 음운 규칙(phonological rule)의 특성을 고려하여 설계되어야 한다.

한국어의 음운(phonology)은 음소(phoneme)인 자음과 모음, 운소(prosodeme)인 소리의 길이(length)로 구성된다. 또한 한국어에서는 한 형태소의 음소가 그 놓이는 환경에 따라 여러 가지 변이음을 가진다. 예를 들어 [값]은 한 형태소이다. 그런데 이 형태소는 뒤에 오는 조사에 따라 [값이]는 [갑시]로, [값도]는 [갑도]등의 여러 가지 형태(morph)로 변환하게 된다. 이처럼 음소가 놓이는 환경에 따라 다른 음소로 바뀌는 현상을 음운 변동 규칙(phonological alternation rule)이라 하고, 한 형태소의 음소가 다른 변이음으로 발음되는 데는 이러한 음운 변동 규칙들이 차례를 지켜 적용되어 진다.

본 논문에서는 한국어 Text-to-speech 변환 시스

템을 위하여 two-level 모델에 기반하여 음운 변동 규칙을 구현하였다. 본 시스템에서는 한국어의 음운 변동 현상을 고려하여 여러 가지 현상들을 5단계의 규칙으로 정의함으로써 규칙의 수를 최대한 줄였고, 또한 각 어절이 변형되는데 적용되어지는 순서를 고려하여 composition함으로써 특정한 입력에 대해서 특정 유형의 규칙이 적용되게 하거나 특정 유형의 규칙은 적용 대상에서 제외되도록 하였다.

본 논문에서 구현한 한국어 음운 변동 시스템은 "twolc"와 "ifsm"을 이용하여 구현하였다.

## 2. Two-level 모델에 기반한 한국어 음운 변동 시스템

### 2.1 한국어 음운 변동 현상의 분석

한국어를 발음하는데 나타나는 음운 변동 현상은 한 형태소의 음소가 그 놓이는 환경에 따라 여러 가지 형태이다. 예를 들어 형태소 [발]은 뒤에 따라 오는 조사에 따라 다음과 같이 서로 다른 형태로 발음된다. 이와 같이 음소의 형태를 변화시키는 데는 여러 가지 변동 규칙들이 있고, 이들 규칙들은 각 어절에 정해진 순서대로 적용되어 진다.

(1)의 가)의 [발음]은 종성 'ㅌ'이 연음되어 [바들]로 발음된다. 그러나 만약 한국어에서 음절의 끝소리가 일곱 가지뿐이라는 규칙이 먼저 적용되고 다음에 연음 된다면, 형태소 [발]의 종성은 'ㄷ'으로 바뀌고 'ㄷ'이 다음 음절로 연음되어 [바들]로 발음나게 된다. 나)는 끝소리 규칙과 경음화 규칙이 차례 대로 적용되어 [발또]로, 다)는 대표음 규칙과 자음동화 규칙이 적용되어 [반만]으로 발음되고, 라)는 구개음화로 인하여 [바치]로 발음된다. 나), 다), 라)에서도 마찬가지로 규칙의 적용 순서가 바뀌게 되면 잘못 발음되게 된다.

(1)	가) -을	-> 바들
	나) -도	-> 발또
[발]	다) -만	-> 반만
	라) -이	-> 바치

위와 같이 뒤에 오는 형태소에 따라 한쪽 혹은 양쪽의 발음의 변화를 일으키는 음운 변동 현상의 종류를 살펴보면 다음과 같다.

가) 복자음 줄이기: 복자음의 둘째 자음은 끝소리 자리에서 줄어진다.

나) 일곱 끝소리 되기: 음성으로 표현되는 음절의 종성은 'ㅂ, ㄷ, ㄱ, ㅁ, ㄴ, ㅇ, ㄹ'의 일곱 가지뿐이다. 음절의 종성 'ㅍ' -> 'ㅂ', 'ㅅ, ㅆ, ㅈ, ㅊ, ㅌ' -> 'ㄷ', 'ㅋ, ㆁ' -> 'ㄱ'으로 발음된다.

다) 자음동화: 음절의 종성이 다음 음절의 자음과 만나면 서로 변화하는 현상이다.

① 'ㅂ, ㄷ, ㄱ'은 비음 'ㅁ, ㄴ' 앞에서 각각 'ㅁ, ㄴ, ㅇ'이 된다.

② 'ㅂ, ㄷ, ㄱ' 뒤의 'ㄹ'은 'ㄴ'이 되고, 'ㅁ, ㄴ, ㄱ'은 각각 'ㅁ, ㄴ, ㅇ'이 된다.

③ 비음 'ㅁ, ㅇ'과 'ㄹ'이 만나면 'ㄹ'이 'ㄴ'으로 된다.

④ 'ㄹ' 앞뒤의 'ㄴ'은 'ㄹ'로 된다.

라) 구개음화: 종성이 'ㄷ', 'ㄷ' 인 형태소가 모음 'ㅣ' 나 반모음 'ㅣ' 가 되는 형태소와 만나면 'ㅈ, ㅊ'으로 민화된다.

마) 적음화: 거센소리 짝이 있는 약한 소리가 'ㅎ'과 이어 나면 두 개의 소리가 하나의 거센소리로 줄어 발음된다.

바) 경음화: 두 개의 안울림소리가 서로 만날 때, 뒤의 소리가 된소리로 발음되는 현상이다.

사) 연음: 앞 음절의 종성에 모음으로 시작되는 형식 형태소가 이어지면 앞 음절 종성이 다음 음절의 초성으로 발음되는 현상이다.

### 2.2 Two-level 모델에 기반한 음운 변동 시스템 개요

본 논문에서 제안한 시스템은 한국어로된 텍스트를 자연스럽게 발음을 하기 위해 한국어의 발음상의 특성인 음운 변동 현상을 two-level 규칙으로 나타내었다. 먼저, 본 시스템에서는 입력된 텍스트를 어절 단위로 잘라낸 다음, 각각의 어절은 two-level 규칙으로 정의된 음운 변동 시스템에 적용되어 한국어 발음 규칙에 맞는 형태의 음운으로 민환된다.

2.1에서 정의한 음운 변동 규칙을 two-level 모델로 설명해 보면 다음과 같다.

(2)가) 구개음화 현상은 뒤에 모음 'ㅣ'나 반모음

‘l’가 올 때, 앞 음절의 ‘ㄷ, ㅌ’은 ‘ㅅ, ㅆ’으로 바뀌도록 하는 규칙으로 나타낼 수 있다.

나) 연음 현상은 음절의 종성인 자음과 다음 음절의 모음 사이에 존재하는 음가가 없는 자음인 ‘ㅇ’이 앞 음절의 종성으로 바뀌도록 하는 규칙과 뒤 음절의 초성인 ‘ㅇ’이 바뀔 때 앞 음절의 종성은 없어지도록 하는 규칙으로 나타낼 수 있다.

다) 음절 끝소리의 대표음을 얻기 위해서는 다음과 같은 규칙들로 표현할 수 있다.

- ① 음절의 종성이 ‘ㅇ’이면 ‘ㅁ’으로 바꾸는 규칙
- ② 음절의 종성이 ‘ㅅ, ㅆ, ㅈ, ㅊ, ㅌ’이면 ‘ㄷ’으로 바꾸는 규칙,
- ③ 음절의 종성이 ‘ㄱ, ㅋ’이면 ‘ㄱ’으로 바꾸는 규칙
- ④ 음절의 종성이 복자음인 경우에 ‘ㅃ, ㅍ, ㅈ, ㅊ, ㅌ’을 첫번째 자음으로 바꾸는 규칙

라) 경음화 현상과 격음화 현상은 서로 독립적이므로 한 개 규칙으로 표현할 수 있다.

- ① 격음화 현상은 모음과 자음 ‘ㅎ’ 사이에 오는 거센소리를 내는 짝이 있는 자음 ‘ㄱ, ㄷ, ㅂ, ㅅ’은 각기 자신의 거센소리 짝으로 바꾸는 규칙과 종성의 자음이 거센소리 짝으로 바뀔 때 뒤에 오는 ‘ㅎ’을 없애는 규칙으로 나타낼 수 있다.
- ② 경음화 현상은 자음에서 비음과 유음을 제외한 자음이 종성인 음절과 다음 음절의 모음 사이에 오는 자음이 된소리 내는 짝으로 바꾸는 규칙으로 나타낼 수 있다.

마) 자음동화 현상은 다음과 같은 규칙들로 표현할 수 있다.

- ① ‘ㄹ’ 앞에 오는 종성 ‘ㅂ, ㄷ, ㄱ’을 각기 ‘ㅃ, ㄸ, ㅇ’으로 바꾸는 규칙
- ② ‘ㄹ’ 뒤에 오는 다음 음절의 초성 ‘ㄱ’을 ‘ㄴ’으로 바꾸는 규칙
- ③ ‘ㄹ’ 앞의 종성이나 뒤의 초성에 ‘ㄴ’이 오면 이를 ‘ㄹ’로 바꾸는 규칙

위의 규칙들 각각은 “twolc”를 이용하여 각각 FST(finite state transducer)로 컴파일되어지고, 또

한 적용되는 순서의 아무런 제약이 없는 FST들(즉, (2)의 각 항에 속한 규칙들)은 한 개의 transducer로 intersect될 수 있다. 그러나 음운 변동 시스템을 구성하는 규칙들은 복잡하기 때문에, 모든 규칙을 intersect를 통해 통합할 수는 없고, composition을 통하여 규칙 적용 순서를 적절하게 정해주어야만 한국어의 음운현상을 모두 표현할 수 있다.

### 3. 한국어 음운 변동 시스템의 구현

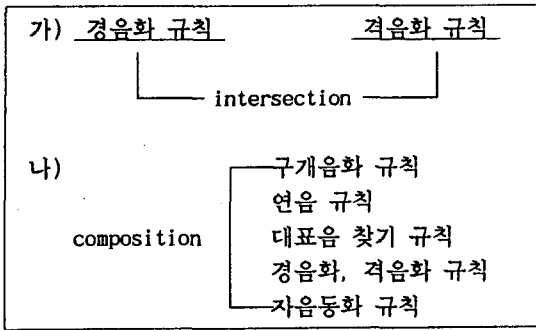
2.1에서 살펴본 바와 같이 한국어에서의 음운은 한 형태소에 몇 가지의 다른 변동 규칙들이 적용되어 변이된 형태로 발음되게 되는데, 이러한 규칙들은 차례를 지켜 적용되어야 한다. 예를 들어 (1)의 다)의 형태소 [발]이 조사 [만]과 결합하여 [만만]으로 정확히 발음되기 위해서는 음절의 끝소리가 일곱 가지뿐이라는 규칙과 자음 동화 규칙이 차례대로 적용되어야 한다.

만약 두 규칙의 적용 순서가 바뀌면 [만만]과 같이 잘못된 결과를 얻게 된다. 또한 (1)의 [발음]은 음절의 끝소리가 일곱 가지뿐이라는 규칙이 적용되기 전에 연음 규칙이 적용되어야만 [바들]로 발음된다. 만약 두 규칙의 적용 순서가 바뀌게 되면 [바들]로 잘못된 결과를 얻게 된다. 마찬가지로 [발이]의 경우에도 음절의 끝소리가 일곱 가지뿐이라는 규칙이 적용되기 전에 구개음화 규칙이 적용되어야만 [바치]로 발음 나게 되고 그렇지 않은 경우에는 [바지]로 발음 나게 된다. [발도]는 음절의 끝소리가 일곱 가지뿐이라는 규칙과 경음화 규칙이 차례대로 적용되어야 [반또]로 발음나게 된다.

- (3) 가) [발만]  
 [만만] <- 대표음 찾는 규칙  
 [만만] <- 자음 동화 규칙

- 나) [발만]  
 [발만] <- 자음 동화 규칙  
 [만만] <- 대표음 찾는 규칙

경음화 규칙과 격음화 규칙의 적용 순서는 발음의 형태를 변화시키지 않으므로 각기 다른 규칙으로 구현하더라도 <그림1>의 가)처럼 한 개로 합쳐진다. 그러나 그외의 규칙들은 일정한 순서에 따라 적용되어야 하므로 <그림1>의 나)에서처럼 composition을 통해 한 개의 FST를 구성한다.



<그림1> 규칙의 intersection과 composition

#### 4. 결론 및 향후 연구 과제

본 논문에서는 한국어 Text-to-speech 변환 시스템을 위한 음운 변동 시스템을 two-level 모델에 기반하여 구현하였다. 음운 변동 시스템은 한국어의 음운 변동 현상을 각기 독립적으로 구현하여 규칙을 간단히 하였고, 규칙들간의 상호 관계를 줄임으로써 프로그램의 모듈화를 이룰 수 있도록 하였다. 또한 상호 독립적인 규칙들은 한 개의 규칙으로 합하므로 규칙의 수를 줄였고, 입력 어절에 따라 특정 유형의 규칙 적용을 막는 것을 방지하기 위하여 각 규칙들간의 적용 순서를 한국어의 음운 현상을 충분히 고려하여 통합하였다.

그러나 보다 자연스러운 한국어 Text-to-speech 변환 시스템을 구현하기 위해서는 앞으로는 한국어의 뜻을 분화시킬 수 있는 음운인 소리의 길이에 대한 처리가 수행되어야 한다. 즉, 한국어에서 눈(사랍의 눈)과 눈:(하늘에서 내리는 눈)의 음운 같지만 소리의 장단에 의해서 그 뜻이 달라진다. 그러므로 입력된 텍스트를 정확하게 발음하기 위해서는 소리의 장단음을 처리할 수 있어야 한다.

장단음을 처리하는 한국어 Text-to-speech 시스템을 위해서는 형태소 분석기의 구현이 꼭 필요하다. 그러므로 향후 장단음을 처리할 수 있는 형태소 분석기에 대한 연구를 진행하여 이 결과를 통해서 한국어 Text-to-speech 변환 시스템의 효율을 향상시킬 수 있도록 하여야 한다.

#### [참고문헌]

- [1] Dennis H. Klatt, "Review of text-to-speech conversion for English," J. Acoustical Society America Vol. 82, No. 3, September 1987.
- [2] Karttunen, Lauri, Kaplan, Ronald M., and Zaenen, "Two-Level Morphology with Composition," Proceedings of the 15th International Conference on Computational Linguistics, Vol. 1, pp. 141-148, 1992
- [3] Karttunen, Lauri and Beesley, Kenneth R. "Two-level rule compiler," Technical Report, ISTL-92-2, Xerox Palo Alto Research Center, October 1992.
- [4] Kwon, H. and Lauri, "Incremental Construction of a Lexical Transducer for Korean," Proceedings of the 17th International Conference on Computational Linguistics, Vol.1, No. 18, pp. 41-59. March 1994.
- [5] Lin-Shan Lee, Chu-Yu Tsebg, and Ming Ouh-Young, "The synthesis Rules in a Chinese Text-to-Speech System," IEEE Trans. On Accoustics, speech, and signal processing, Vol. 37, No. 9, pp. 1309-1320, September 1989.
- [6] Michael h. O'Malley, "Text-To-Speech Conversion Technology," IEEE Computer, pp. 17-23, August 1990.
- [7] Sproat R. *Morphology and Computation*, MIT Press, 1992.
- [8] 강승식, "다중 형태론과 한국어 형태소 분석 모델", 한글 및 한국어 정보처리 학술발표 논문집, pp. 140-145. 1994.
- [9] 허용, *국어학 -우리말의 오늘·이제-*, 샘문화사, 1994.