김일성종합대학학보
정보과학
주체110(2021)년
제67권 제3호

조선어음성합성을 위한 소리뭉침경계예측의 한가지 방법

한 철 진

경애하는 **김정은**동지께서는 다음과 같이 말씀하시였다.

《과학기술과 경제의 일체화를 다그치고 나라의 경제를 현대화, 정보화하는데서 과학기술부문이 주도적인 역할을 하도록 하여야 합니다.》

통계적음성합성[2, 3]은 본문해석, 음향모형, 파형생성의 3단계로 구성되며 여기서 첫 단계인 본문해석성능이 합성음의 자연성에 미치는 영향이 크다.

본문해석에서는 입력된 본문에 대하여 전처리와 본문정규화, 발음변환과 같은 처리를 진행하고 음성합성에서 쓰이는 어학적특징들을 추출하여 음향모형화를 위한 문맥특징으로 리용한다.

조선어억양구조에서 하나의 발성(문장)은 하나 혹은 그 이상의 소리매듭(억양구)으로, 하나의 소리매듭은 하나 혹은 그 이상의 소리뭉침으로, 하나의 소리뭉침은 하나 혹은 그 이상의 소리마디로 구성된다. 소리뭉침[1]은 조선어억양구조의 고유한 률동단위이며 조선어음성합성에서 소리뭉침과 관련한 어학적특징들이 합성음의 자연성에 일정한 영향을 준다. 이로부터 자연스러운 합성음의 생성을 위하여 소리뭉침의 경계를 정확히 예측하는것이 필요하다.

론문에서는 조선어소리뭉침의 특성에 대한 연구에 기초하여 소리뭉침경계예측을 위한 방법을 제안하였다.

1. 소리뭉침의 특성

조선어억양구조의 고유한 단위인 소리뭉침은 하나 혹은 그 이상의 형태부로 이루어지며 형태부경계에서만 소리뭉침경계가 이루어진다. 또한 하나의 소리뭉침은 보통속도의 랑독에서 1～6음절로 이루어지며 여기서도 특히 3～4개의 음절로 이루어진 경우가 보다 일반적이다.

실례를 들면 다음과 같다.

이 프로그람은/ 조선어/음성합성/프로그람입니다.

나는/ 학교에서/ 공부하고/있었습니다.

이분은/ 나의/ 아버지입니다.

거기에는/ 소,/ 염소,/ 양,/ 돼지가/ 있었다.

실례에서 보여준것처럼 소리뭉침은 보통 띄여쓰기로 구분되는 단어(이것을 단어라고 표기함.) 1개로 이루어지며 그외에 하나의 단어가 둘 혹은 그 이상의 소리뭉침으로 갈라지는 경우와 여러개의 단어가 합쳐져 하나의 소리뭉침을 이루는 경우도 있다.

소리뭉침경계는 형태부경계에서 실현되며 앞, 뒤형태부에 따라 결정된다.

실례로《이, 저, 그》와 같은 지시대명사나《더》와 같은 일부 부사뒤에서는 일반적으로 소리뭉침경계가 실현되지 않고 토앞에서와 말뿌리, 뒤붙이사이의 경계에서 실현되는 경우가 없다.

실례로 문장《나는 학교에서 공부하고있었다.》의 소리뭉침경계는《나는/ 학교에서/ 공부하고/있었다.》와 같이 될수는 있지만《는》,《에서》,《고》,《었다》의 앞에서 소리뭉침경계가 이루어질수 없으며《구두쟁이》와 같은 단어에서도《구두》와《쟁이》사이에서 소리뭉침경계가 이루어질수 없다.

여러개의 명사들이 결합된 비교적 큰 소리토막들에서는 소리뭉침경계가 어휘화정도에 따라 어휘화정도가 높은 단어들끼리 하나의 소리뭉침을 형성하게 된다. 이외에 발성속도에 따라 소리뭉침을 이루는 소리마디개수가 달라지며 뜻마루에 따라서도 소리뭉침경계가 달라진다.

2. 소리뭉침경계예측

소리뭉침의 특징을 종합해보면 형태부의 길이와 형태부의 의미, 발성자의 의도와 발성속도에 따라 소리뭉침경계가 결정된다고 볼수 있다. 음성합성의 견지에서는 평문랑독과 보통의 발성속도를 생각하는것으로 하여 여기서는 발성자의 의도나 발성속도에 따르는 소리뭉침경계에 대해서는 론의하지 않는다. 이로부터 형태부의 길이와 형태부의 의미에 따라 소리뭉침경계를 예측하기 위한 방법에 대하여 보기로 한다.

하나의 단어가 2개이상의 소리뭉침으로 갈라지는 경우와 여러 단어가 합쳐져 하나의 소리뭉침을 이루는 경우가 있으며 특히 어떤 경우에는 앞단어가 뒤단어의 전체가 아니라 일부와 합쳐져 하나의 소리뭉침을 이루는 경우도 있는것으로 하여 분리와 결합의 두 단계로 나누어 소리뭉침경계예측을 진행한다.

1) 분리단계

분리는 길이가 k이상인 단어들에 대하여 적용된다.

단어 w가 n개의 형태부 $m_1, m_2, \cdots, m_n$으로 이루어졌다고 할 때 매 형태부경계위치에서 다음과 같은 확률 P_i를 계산한다.

$$P_i = \frac{1}{2}(P_a(m_i) + P_b(m_{i+1})) \tag{1}$$

여기서 $P_a(m_i)$와 $P_b(m_i)$는 각각 형태부 m_i의 뒤, 앞에서 소리뭉침경계가 실현될 확률로서 소리뭉침경계가 수동적으로 표기된 학습자료로부터 얻은 확률값들을 리용한다.

$1 \le i \le n-1$에 대하여 계산된 P_i중에서 값이 최대로 되는 위치

$$s = \arg\max_i P_i \tag{2}$$

를 찾고 확률 P_s가 어떤 턱값 τ보다 크면 단어 w를 두 단어 w_1과 w_2로 분할한다.

$$w_1 = m_1 \cdots m_s, \quad w_2 = m_{s+1} \cdots m_n$$

두 단어 w_1, w_2중에서 길이가 k이상인 단어에 대하여 우와 같은 공정을 반복한다. 이 반복과정은 길이가 k이상인 토막이 없거나 확률값이 모두 τ보다 작을 때까지 진행한다.

2) 결합단계

결합은 문장에서 공백(띄여쓰기)위치에서만 진행되며 규칙에 의한 방법으로 진행한다.

소리뭉침결합규칙의 실례를 표 1에 보여주었다.

표 1. 소리뭉침결합규칙의 실례

규칙	실례
대명사 + 명사	우리 나라, 내 나라, 우리 집, 이분
부사《더》+ 형용사	더 큰, 더 많은
토《여야/어야/아야》+ 동사《하(다.)》	하여야 한다.
토《게/도록》+ 동사《하/되(다.)》	하게 된다., 하도록 하다.
…	…

공백앞, 뒤의 소리뭉침후보들에 대하여 그 길이의 합이 k보다 작은 경우 표 1에서 보여준것과 같은 규칙에 맞으면 두 단어를 결합하여 하나의 소리뭉침으로 설정한다.

3. 실험 및 평가

소리뭉침경계의 예측은 학습단계에서는 리용되지 않고 합성음생성을 위한 본문해석단계에서만 리용되며 그 경계위치들이 발성자마다 일정하게 다르고 또 긴 단어속에서 소리뭉침경계가 유일하게 결정되지 않는것으로 하여 경계예측의 정확성평가가 어렵다.

실험에서는 수동적으로 표기된 소리뭉침경계를 리용한 음성합성체계(체계 1)와 론문에서 제안한 방법으로 소리뭉침경계를 예측하여 리용한 체계(체계 2)를 구축하고 그것들사이의 합성음질을 비교하는 방법으로 체계성능을 평가하였다. 길이턱값을 $k=5$, 분리확률턱값을 $\tau=0.3$으로 설정하였다. 그것은 길이가 긴 단어에 대해서는 앞단어나 뒤단어중 어느 한 단어의 확률이라도 높으면 분리하는것이 보다 효과적이라고 보았기때문이다.

조선어음성합성을 위한 문맥특징으로 음소, 음절, 소리뭉침, 억양구, 문장과 관련한 특징들을 추출하여 리용한다.

조선어음성합성을 위하여 리용한 문맥특징들을 표 2에 보여주었다.

표 2. 조선어음성합성을 위하여 리용한 문맥특징

준위	특징
음소	현재음소와 앞, 뒤 각각 2개 음소의 이름
	음절에서 현재음소의 위치
음절	이전, 현재, 다음음절에 력점이 있는가
	이전, 현재, 다음음절의 음소개수
	소리뭉침, 억양구안에서의 음절위치(앞, 뒤)
	소리뭉침안에서 력점위치로부터의 거리
	현재음절의 모음이름
소리뭉침	이전, 현재, 다음소리뭉침의 품사정보
	이전, 현재, 다음소리뭉침의 음절수
	억양구안에서의 소리뭉침위치(앞, 뒤)
억양구	이전, 현재, 다음억양구의 음절개수
	이전, 현재, 다음억양구의 소리뭉침개수
	문장에서의 억양구위치(앞, 뒤)
	억양구의 이음억양
문장	문장의 음절개수, 소리뭉침개수, 억양구개수
	문장의 구두점정보

음성합성의 자연성평가실험은 신경망에 기초한 음성합성체계구축도구인 Merlin을 리용하여 진행하였다. 학습자료로는 30대의 남성(문화어소유자)이 표준속도로 발성한 록음자료(표본화주파수 16kHz, 량자화비트수 16bit, mono) 1 000문장(3h 20min)을 리용한다. 음성합성을 위한 지속 및 음향모형은 모두 숨은층 4개, 매 층에 512개의 세포를 가지는 전결합신경망으로 구성한다. 신경망훈련은 소묶음방식으로 진행하였으며 훈련을 위한 초파라메터로 초기학습률은 0.01, 훈련반복수는 200, 묶음크기는 64로 설정하였다.

학습자료에 포함되지 않은 임의의 50개 문장에 대하여 앞에서 론의한 두가지 체계로 합성음을 생성하고 선택점수(Preference Score)를 리용하여 평가를 진행하였다. 즉 청취자들이 두가지 합성음을 듣고 보다 자연스러운 합성음을 선택하며 그들이 선택한 개수를 종합하여 합성음의 질을 평가한다.

평가실험결과를 표 3에 보여주었다.

표 3. 평가실험결과

체계 1/%	체계 2/%	중립/%
21.2	19.3	59.5

표 3의 평가실험결과에서 보여준것처럼 두 체계의 자연성이 거의 같다. 이로부터 론문에서 제안한 소리뭉침경계예측방법이 조선어음성합성의 자연성을 높이는데서 효과적이라는것을 알수 있다.

맺 는 말

조선어억양구조의 고유한 단위인 소리뭉침의 경계예측을 위한 한가지 방법을 제안하고 소리뭉침특징을 반영한 조선어음성합성의 자연성평가를 통하여 그 성능을 실험적으로 검증하였다.

참 고 문 헌

[1] 김동철; 조선어억양구조연구, 사회과학출판사, 55～104, 주체108(2019).

[2] H. Zen et al.; Speech Communication, 51, 11, 1039, 2009.

[3] H. Zen et al.; IEEE ICASSP 2013, 7962, 2013.

주체110(2021)년 5월 5일 원고접수

A Method of Accentual Phrase Boundary Prediction for Korean Speech Synthesis

Han Chol Jin

In this paper, we investigated the characteristics of the Korean accentual phrase, proposed a method of its boundary prediction, and verified the efficiency empirically.

Keywords: Korean speech synthesis, accentual phrase, deep neural network