일본어 TTS의 가변 Break를 이용한 합성단위 선택 방법

*나덕수, **배명진
(주)보이스웨어, 숭실대학교 정보통신공학과
e-mail : dsna@voiceware.co.kr, mjbae@ssu.ac.kr

A Unit Selection Methods using Variable Break in a Japanese TTS

*Deok-Su Na, **Myung-Jin Bae
Voiceware Co. Ltd.
**Information and Communication Engineering
Soongsil University

Abstract

This paper proposes a variable break that can offset prediction error as well as a pre-selection methods, based on the variable break, for enhanced unit selection. In Japanese, a sentence consists of several APs (Accentual phrases) and MPs (Major phrases), and the breaks between these phrases must predicted to realize text-to-speech systems. An MP also consists of several APs and plays a decisive role in making synthetic speech natural and understandable because short pauses appear at its boundary. The variable break is defined as a break that is able to change easily from an AP to an MP boundary, or from an MP to an AP boundary. Using CART (Classification and Regression Trees), the variable break is modeled stochastically, and then we pre-select candidate units in the unit-selection process. As the experimental results show, it was possible to complement a break prediction error and improve the naturalness of synthetic speech.

I. 서론

음성합성 시스템의 합성단위 선택 과정은 문맥 정보와 운율 파라미터에 의해 결정되는데 보다 자연스러운 합성음을 얻기 위해서는 정확한 운율 모델링이 필수적이다. 운율구의 구조는 언어마다의 독특한 차이로 인해 동일하지는 않지만 일반적으로 “utterance” > “intonational phrase” > “major phrase” > “minor phrase” > “word”와 같다.

1996년 Cambell의 연구[1]를 살펴보면, 자동으로 예측한 BI(Break Index)와 사람이 레이블링한 BI가 일치하는 정확도는 69%정도이나 예측 값을 +/- 1로 조정한다면 90%로 올라가는 것을 알 수 있다. 이러한 결과는 break index 사이의 불명확성을 나타내는 것으로 BI 예측을 어렵게 하는 요인이지만 합성기에서 이러한 특징을 이용한다면 보다 자연스러운 합성음을 얻을 수 있다. 본 논문에서는 이러한 특징을 이용하기 위해 가변 break를 정의하고 통계적 방법으로 예측한 break의 확률을 얻고 이것을 이용하여 합성단위 선택을 수행하는 방법을 제안한다.

II. 제안하는 합성단위 선택 방법

일본어 합성기에서 단어 및 AP(Accentual Phrase) 경계는 발음변환 결과로 얻어지는 단어, 발음열, 악센

트 등으로 예측할 수 있다. 합성기의 break 예측기는 AP 경계 중 MP 경계인 것을 찾는 것이 중요하다. 본 논문에서 제안하는 가변 break는 AP 경계 중 MP 경계도 될 수 있는 것과 MP 경계 중 AP 경계도 될 수도 있는 break를 의미하고 CART의 회귀 트리(regressiion tree)로 모델링 하였다. 합성기 개발을 위해 구축된 코퍼스 중 8,586문장을 선택하여 5,769문장으로 학습하고 2,718문장으로 성능 평가하였고 각 문장은 녹음된 음성 데이터를 청취하여 Break를 레이블링을 수행 하였다.

표 1. CART 모델링에 사용한 특징

No	Factors
1	Number of mora in (Wk-2,Wk-1,Wk,Wk+1,Wk+2)
2	Number of mora before/after Wk within P
3	Part-of-speech type of (Wk-2,Wk-1,Wk,Wk+1,Wk+2)
4	Tone pattern of consecutive five mora before Mn
5	Tone pattern of consecutive five mora after Mn
6	Phoneme of mora in (Wk-2,Wk-1,Wk,Wk+1,Wk+2)
7	Kind of morph (Wk-2,Wk-1,Wk,Wk+1,Wk+2)
8	Break of (Wk-2,Wk-1,Wk)

표 1은 CART 모델링에 사용한 특징들이다. Wk는 break를 예측하려는 경계의 단어이고, P는 IP이고, Mn은 Wk의 마지막 mora이다.

$$\widehat{P}_i = [P_{AP}(i), P_{MP}(i)] \quad \text{(식 1)}$$

$$P_{AP}(i) = 1 - P_{MP}(i) \quad \text{(식 2)}$$

$\widehat{P}_i$는 AP/MP 예측 트리의 출력 형태를 나타낸 것으로 예측하려는 단어 경계 i가 AP 경계일 확률 $P_{AP}(i)$와 MP 경계일 확률 $P_{MP}(i)$의 벡터형태이다.

$$\delta_{P_{AP}}(i) = \begin{cases} 1 & i \in AP,\ P_{AP}(i) > P_{MP}(i) \\ 0 & otherwise \end{cases} \quad \text{(식 3)}$$

$$\overline{P_{AP}} = \frac{1}{N_{AP}} \sum^{N_{AP}} P_{AP}(i)\delta_{AP}(i) \quad \text{(식 4)}$$

$$\delta_{P_{MP}}(i) = \begin{cases} 1 & i \in MP,\ P_{MP}(i) > P_{AP}(i) \\ 0 & otherwise \end{cases} \quad \text{(식 5)}$$

$$\overline{P_{MP}} = \frac{1}{N_{MP}} \sum^{N_{MP}} P_{MP}(i)\delta_{MP}(i) \quad \text{(식 6)}$$

$\overline{P_{AP}}$는 AP 경계의 예측이 정확한 경우의 예측확률, $P_{AP}(i)$의 평균이고, $\overline{P_{MP}}$는 MP 경계의 예측이 정확한 경우의 $P_{MP}(i)$의 평균이다.

$$\widetilde{P}_i = \begin{cases} [1,0] & P_{AP}(i) \geqq \overline{P_{AP}} \\ [0,1] & P_{MP}(i) \geqq \overline{P_{MP}} \\ \widehat{P}_i & otherwise \end{cases} \quad \text{(식 7)}$$

가변 break는 $P_{AP}(i)$와 $P_{MP}(i)$ 모두 각각의 예측확률의 평균 보다 작은 경우이고 그렇지 않은 경우 각각의 확률을 1 또는 0으로 조정한다.

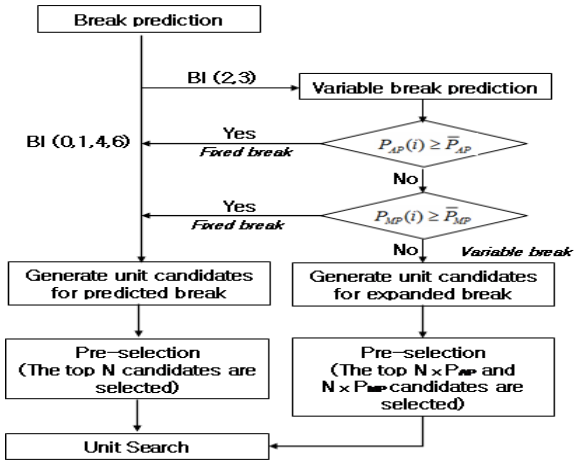

그림 1. 제안하는 합성단위 선택방법

III. 실험및 결론

시스템의 성능을 평가하기 위해 합성음의 MOS 테스트를 수행하였다. 테스트는 일본인 여성 5명이 참가하였고, 테스트 문장은 JEITA 종합평가문장[2] 중 127문장을 선택하여 실험하였다. MOS 테스트는 원음 127개와 유동 break를 사용한 시스템 1과 사용하지 않은 시스템 2로 생성한 합성음 254개를 섞어 불규칙한 순서로 청취하고 5개의 레벨(1~5, Bad, Poor, Fair, Good, Excellent) 중 하나를 선택하도록 하였다.

표 2. 음성 코퍼스

성별	녹음시간	개수			
		문장	IP	AP	음소
여성	41.04	17230	35871	142061	1104450

표 3. MOS 테스트

원음	시스템 1	시스템 2
4.99	4.20	4.01

논문에서는 합성음의 자연성을 향상시키기 위해 코퍼스 기반 일본어 합성기에서 생성된 운율을 보다 효율적으로 이용하여 합성단위를 선택하는 방법으로 운율정보의 하나인 break에 가변 break 개념을 도입하였다. 이것은 음성 DB의 각 세그먼트가 가지는 다양한 운율정보를 이용할 수 있는 합성단위 검색을 가능하게 하였다.

참고문헌

[1] J. Venditti, "Japanese ToBI labeling guidelines," OSU Working Papers in Linguistics, pp. 127-162, 1997.

[2] Technical Standardization Committee on Speech Input/Output Systems, "Speech Synthesis System Performance Evaluation Methods," JEITA IT-4001, pp.42-45, 2003.