

# 성대특성 보간에 의한 합성음의 음질 향상

- 음성코퍼스 내 개구간 비 보간을 위한 기초연구 -

배재현\* 오영환\*

\* 한국과학기술원 전자전산학과 전산학전공 음성인터페이스 연구실

## Synthetic Speech Quality Improvement

### By Glottal parameter Interpolation

- Preliminary study on open quotient interpolation in the speech corpus -

Jae-Hyun Bae\*, Yung-Hwan Oh\*

\* Voice Interface Lab., Div. of Computer Science, Dept. of EECS, KAIST

jhbae@bulsai.kaist.ac.kr, yhoh@bulsai.kaist.ac.kr

## Abstract

For the Large Corpus based TTS, the consistency of the speech corpus is very important. It is because the inconsistency of the speech quality in the corpus may result in a distortion at the concatenation point. And because of this inconsistency, large corpus must be tuned repeatedly. One of the reasons for the inconsistency of the speech corpus is the different glottal characteristics of the speech sentence in the corpus. In this paper, we adjusted the glottal characteristics of the speech in the corpus to prevent this distortion. And the experimental results are showed.

## I. 서론

여러 음소유닛이 접합된 합성음에서는 주로 접합부에서 음성왜곡이 일어난다. 이는 접합된 두 음소유닛 사이의 음향학적인 특성이 다르기 때문이며, 이로 인한 왜곡을 줄이기 위해 많은 연구가 수행되었다. 이러한 기존 연구로는 음향학적 특성들을 다항식으로 보간하거나, 중첩 윈도우를 사용하거나, 대용량 코퍼스에

서 가장 적당한 후보음소유닛을 선택하는 방법 등이 있다.[1][2][3] 그 결과 접합부에서의 음성 왜곡이 많이 줄어든 합성음을 얻을 수 있었다. 그러나 대부분의 이러한 연구들은 성도 특성을 보간하는 데 초점이 맞추어져 있다. 본 논문에서는 음성의 성대특성에 초점을 맞추어 연구하고자 한다; 만약 합성음의 성대특성이 일관된 특징을 보인다면, 보다 자연스러운 합성음을 생성할 수 있을 것이다. 본 논문에서는 이를 위한 기초연구로서 접합되는 유닛 간의 개구간 비(OQ)에 대한 보간을 수행하고, 그 결과를 보인다. 개구간 비는 접합부 주변 여러 프레임에 걸쳐 선형보간하였다. 2장에서는 성대과의 다양한 특성에 대하여 설명하고, 3장에서는 개구간 비의 선형보간에 대하여 설명하고 이에 대한 기초실험을 보이며 4장에서 결론을 맺는다.

## II. 성대 특성

음성의 여러 가지 특성 중 음색은 성대특성에 의해 영향을 많이 받는 특성으로, 본 논문에서는 이에 관한 연구를 수행한다. 본 논문에서는 성대특성을 모델링하기 위하여 LF모델을 사용하였다.[4] LF모델에서는 성대특성을 2부분으로 나누어 표현하고 있으며, 이는 식 (1)과 같다.

$$g(t) = \begin{cases} E_0 e^{at} \sin \omega_y t & , t \leq T_3 \\ -\frac{E_e}{\epsilon T_a} [e^{e(t-T_3)} - e^{-e(T_0-T_3)}] & , T_3 \leq t \leq T_0 \end{cases} \quad (1)$$

식 (1)에 사용된 파라미터는 그림 1에 표현되어 있다. 그림 1은 성대파와 그 미분값을 표현한 그림으로, 식 (1)에서와 같이 EE값을 기준으로 두가지 식으로 표현할 수 있음을 알 수 있다.

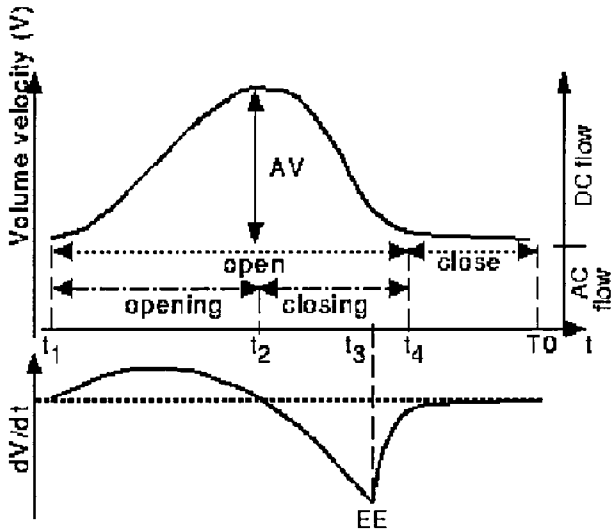


그림 1 음성 성대파의 모델링

음성의 성대특성은 그림 1과 같이 여러 종류가 있으며 이 중 EE는 여기강도(excitation strength)를 나타내며 이는 음성의 강도에 관계된 특성으로 EE의 절대값이 작고 명확하지 않으면 숨소리가 섞인 음성(breathy voice)이 나타나고 그 값이 크고 명확할수록 강한 음성이 된다. 개구간 비(OQ)는  $t_1/T_0$ 로 나타낸다.[5] 개구간 비는 음성 스펙트럼의 낮은 차수의 하모닉의 크기에 영향을 주며, 음성의 F0에 따른 자연성에 관계된 파라미터이다.[5]

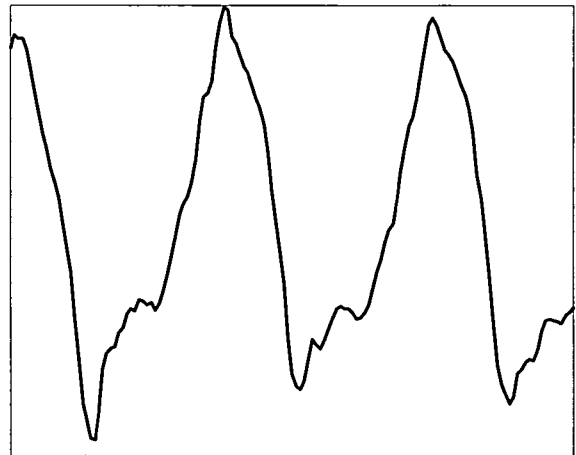
합성된 문장을 살펴보면 사람이 발성한 것과는 다른 부자연스러움을 관찰할 수 있는데, 이는 운율 모델링시의 오류, 신호조작 및 접합시의 음성의 음향학적 특성 등의 차이에 기인한다. 이러한 왜곡 중 합성된 문장의 음색이 변하는 것은 그림 2와 같이 접합유닛 사이에 음성의 성대특성이 차이가 발생하기 때문이라고 할 수 있다.

그림 2는 /t/ 음성에 대한 성대파를 표현한 것으로, 이 두 유닛을 접합할 때에는 서로 다른 성대특성

으로 인해 음색의 갑작스런 변화가 관찰된다. 본 논문에서는 합성음의 음색을 결정하는 여러 상대 특성 중 개구간 비에 대한 실험을 수행하였다.



(a)



(b)

그림 2 /t/ 발음의 성대파

3장에서는 접합 유닛간의 음색왜곡을 줄이기 위해, 각 유닛에서 접합부에서의 개구간 비(OQ)에 대한 선형보간법에 대해 설명한다.

### III. 개구간 비(OQ) 보간

본 장에서는 이전 음소유닛과 현재 음소유닛의 개구간 비를 보간하기 위한 방법을 설명하고, 이에 대한 결과를 보인다. 한 음소유닛 내에서는 음색의 변화가 없다는 가정을 바탕으로, 접합하는 두 음소유닛의 중심 사이에서 개구간 비가 천천히 변화하도록 보간함수를 설계하였다.

음소유닛 접합시 이전 음소유닛과 현재 음소유닛의

중심에서의 개구간 비를 집합점을 기준으로 주위 여러 프레임에 걸쳐 선형보간한다. 보간은 순방향, 역방향의 두 방향에 대하여 개구간 비율을 구하여 수행된다. 이전 음소유닛의 개구간 비율은, 이전 음소유닛에서는 본래의 개구간 비율을 갖도록 하고, 현재 음소유닛의 중심에서 그 영향이 가장 작아지도록 하였다. 현재 음소유닛 중심에서의 개구간 비율 또한 이전 음소유닛의 중심에서 가장 작은 영향을 주며, 현재 음소유닛의 중심에서는 현재 음소유닛 본래의 개구간 비율을 갖도록 보간 알고리즘을 고안하였다. 음소유닛의 집합부에서는 음색이 갑자기 변하지 않도록 각 음소유닛의 중심부 개구간 비율의 영향을 적절히 받도록 한다. 보간시에는 식 (2), (3)을 사용한다.

$$OQ_t = wp_t OQ_p + wc_t OQ_c \quad (2)$$

식 (2)에서  $OQ_p$ 와  $OQ_c$ 는 각각 이전 음소유닛과 현재 음소유닛의 중심에서의 개구간 비를 나타낸다. 보간가중치  $wp_t$ 와  $wc_t$ 는 식 (3)에서와 같다.

$$\begin{cases} wp_t = \frac{T-t}{T} \\ wc_t = \frac{t}{T} \end{cases}, 0 \leq t \leq T \quad (3)$$

식 (3)에서  $T$ 는 현재 음소유닛과 이전 음소유닛 각각의 중심사이의 거리를 나타낸다. 식 (2), (3)에서는 이전 음소유닛 중심에서의 시간 인덱스를 0으로 가정한다. 보간 과정은 그림 3으로 도시하였다. 그림 3에서와 같이 이전 음소유닛과 현재음소유닛의 집합부를 가운데로 하여 이전 음소유닛의 중심에서부터 시작하여, 현재 음소유닛의 중심에 이르기까지 점차적으로 개구간 비를 조절한다. 개구간 비를 집합 유닛간에 천천히 변하게 함으로써, 유닛 집합시 갑작스런 음색의 변화를 방지하고, 보다 자연스러운 성대파형의 변화를 가져올 수 있다.

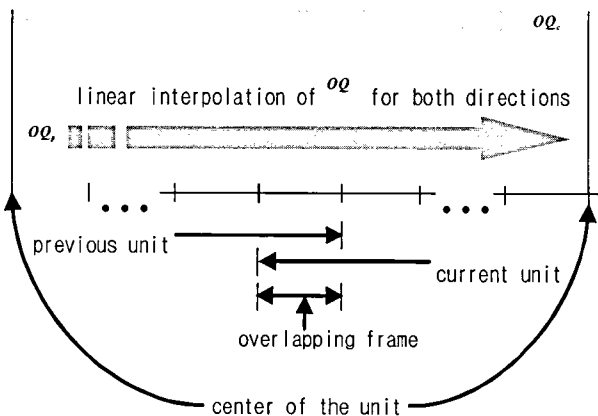


그림 3 개구간 비의 선형보간 과정

실험은 약 5000문장으로 이루어진 대용량 코퍼스기

반 합성기의 개선을 목적으로 행해진 기초실험으로, 제안한 알고리즘을 기반으로 수동으로 보간실험을 수행하였다.

그림 4는 본 논문에서 설명한 개구간 비 보간법을 사용하여 그림 2의 /t/ 음소를 보간한 결과를 나타낸

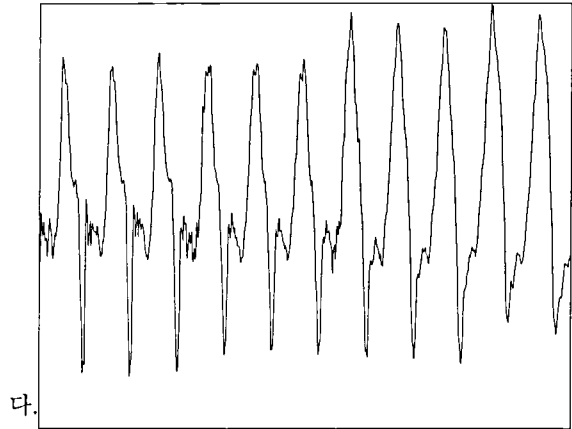


그림 4 /t/음소의 개구간 보간 결과

다. 실험결과 집합된 음소유닛간의 갑작스런 음색변화를 줄일 수 있었으며, 부드럽게 연결된 성도파를 얻을 수 있었다.

#### IV. 결론

대용량 코퍼스 기반 합성방식에 사용되는 음성코퍼스는 그 양이 방대하여, 코퍼스 전체에 대하여 동일한 음색을 유지하는 것은 매우 어렵다. 또한 집합되는 두 유닛 사이에서 동일한 음색을 유지하는 것 또한 많은 노력과 시간이 필요한 작업이다. 그리고, 음소유닛 집합 방식을 사용하는 기존 합성방법에서는 성도특성에 대한 보간을 주로 다루었다. 본 논문에서는 성도특성이 음색에 많은 영향을 주는 것에 착안하여 합성음 생성시 성도 특성을 보간하는 알고리즘을 고안하고, 이에 대한 기초실험을 수행하였다. 본 논문에서는 개구간 비에 대한 실험을 수행하였고, 부드럽게 연결된 성도파형을 얻을 수 있었다. 향후 여러 가지 성도특성을 이용하여 합성음의 음질을 보다 자연스럽게 하는 연구가 필요하며, 이를 더 발전시켜 풍부한 음색의 합성음을 생성하는 TTS 시스템, 소용량 고품질의 TTS시스템을 개발하고자 한다.

#### 참고문헌

[1] T. F. Quatieri, R. J. McAulay, "Shape Invariant

- Time-Scale and Pitch Modification of Speech, " *IEEE Trans. on Signal Processing*, vol. 40, no. 3, pp.497-510, Mar. 1992
- [2] R. J. McAulay and T. F. Quatieri, "Speech Analysis/Synthesis Based on a Sinusoidal Representation, " *IEEE Trans. on ASSP*, vol. 34, pp.744-753, Aug. 1986
- [3] Hunt, A. J., Black, A. W. , "Unit Selection in a concatenative speech synthesis system using a large speech database, " *IEEE ICASSP* vol.1, pp 373-376, 1996.
- [4] Fant, G., Liljencrants, J., and Lin, Q.," A four-parameter model of glottal flow," *STL-QPSR*, Vol. 4, pp. 1-13, 1985
- [5] Toohar, M. and McKenna, J.G., "Prediction of Glottal LF Parameters using Regressin Trees," *ICSLP*, pp1089-1093 2004.