

포만트 합성기용 반음절 세트의 구축에 관한 연구

이정석*, 오영환

한국과학기술원 전산학과

Implementation of Demissyllable database for formant synthesizer

Jung-Suk Lee, Yung-Hwan Oh

Computer Science Dept. KAIST

요약

포만트형 합성기에 사용될 반음절 데이터 베이스의 구성과 필요한 파라미터의 추출 과정에 대하여 논한다. 포만트 합성기는 많은 구동 파라미터를 필요로 하기 때문에 저장 장소를 절약하기 위해서 적절한 합성단위의 선택과 합성단위의 효율적인 표현이 필요하다. 본 연구에서는 포만트 합성기에 있어서 합성음의 음질에 큰 영향을 미치는 포만트궤적의 추출과 데이터베이스의 구성에 대하여 기술한다.

1. 서론

문자-음성변환시스템(text-to-speech system)은 작은 음운단위를 기본단위로 하여, 언어학적 지식에 근거한 규칙들을 적용함으로써 임의의 문장을 음성으로 변환시키는 기능을 행하게 된다. 이러한 문자-음성변환 시스템은 사용하는 음운단위와 합성기의 종류에 따라 여러방식으로 구현될 수 있다. 본 연구에서는 포만트형 합성기(그림 1)를 사용한 문자-음성변환시스템의 구현시에 필요한 반음절 데이터 세트의 구축에 대하여 논한다. 포만트형 합성기는 LPC 계열의 합성기보다 많은 구동 파라미터를 필요로 하여 구현하기가 곤란하다는 단점이 있으나, 사용되는 파라미터가 사람이 직관적으로 이해할 수 있는 형태이기 때문에 시스템의 조율이 용이하다는 이점이 있다. 포만트합성기에서는 다양한 조음환경에서의 음운의 변화를 규칙화하여 임의의 문장을 합성하는데 이러한 방식은 필요한 저장장치 양은 매우 적으나 규칙의 개발에 많은 음성 데이터의 분석을 필요로 하여 구현이 용이하지 않다. 따라서 본 연구에서는 비교적 합성 단위의 수가 적으면서 다양한 조음결합 현상을 포함할 수 있는 반음절(demissyllable)을 기본 단위로 채택하였다. 반음절을

합성단위로 사용할 경우에는 위에서 언급한 조음환경에 따른 음운의 변동이 반음절내부에 포함되므로 별도의 처리규칙을 요하지 않는다. 포만트합성기에 사용되는 파라미터중의 하나인 포만트 주파수는 LPC 또는 Cepstrum 분석으로 구하게 되나 이러한 분석방법은 비음화되지 않은 모음구간을 제외하고는 정확한 포만트 궤적을 구하기 어렵다. 포만트의 불규칙한 변화는 음질 저하의 원인이 되므로 포만트 궤적이 잘 추출되지 않은 구간에서는 별도의 처리를 해주어야 한다. 본 연구에서는 반음절 데이터 세트의 구성 및 포만트 합성기에 필요한 파라미터의 추출과정에 관하여 기술한다.

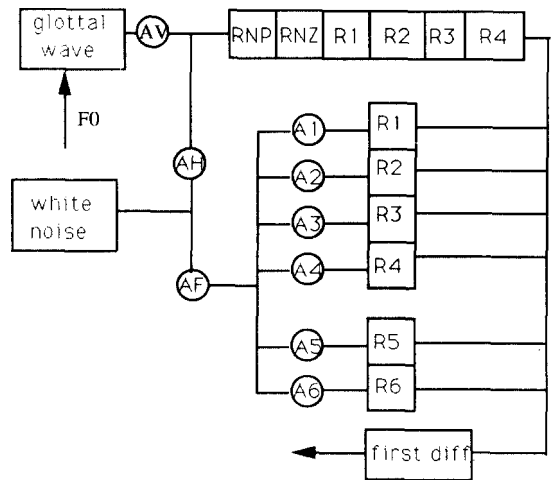


그림 1. 포만트 합성기

기음 + 모음

우리말의 음소중에서 기음에 해당하는 것은 "ㅎ"인데, 기음구간의 포먼트 값은 뒤이어 오는 모음이 시작될 때의 포먼트 값을 사용하면 된다. (그림 4)

모음 + 비음

이런 형태의 반음절에서는 모음의 비음화가 시작되는 구간부터 제 2 포먼트가 사라지거나 거의 추출이 안되는 어려움이 있다. 이런 경우에는 제 1포먼트나 제 3포먼트 궤적을 관찰하여 비음의 안정화 구간을 결정한 다음 그 시점에 나타나는 제 2 포먼트 값과 선형보간하여 천이구간에서의 제 2 포먼트 값을 결정한다. (그림 5)

모음 + 폐쇄음 :

폐쇄음 구간에서 포먼트의 변화가 불규칙하게 나타난다. 폐쇄음 구간의 특징이 나타나는 프레임의 포먼트 값을 다른 프레임에 사용한다.

유음 + 모음, 모음 + 유음 형태의 반음절은 포먼트 궤적의 변화가 완만하여 별도의 처리를 필요로 하지 않는다

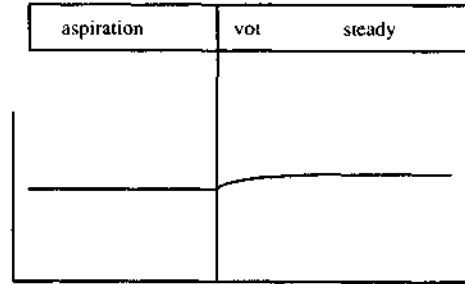


그림 4. 기음 + 모음

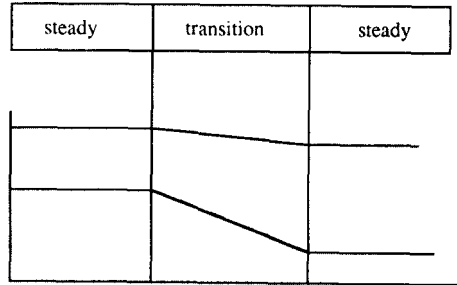


그림 5. 모음 + 비음

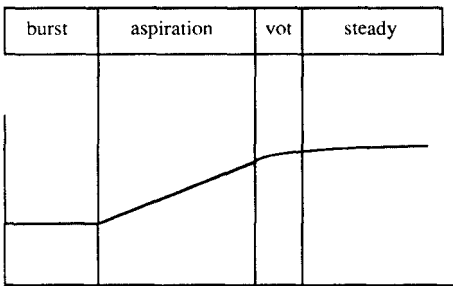


그림 2. 무성파열음 + 모음

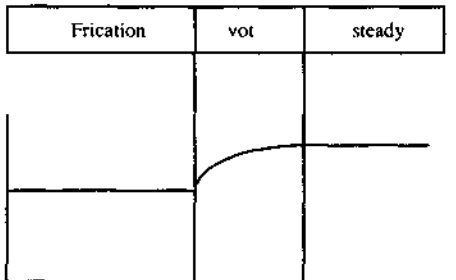


그림 3. [유성파열음, 마찰음] + 모음

time	acoustic-event
0.1312	Steady
0.1879	Fricative
0.2660	End

표 1. 반음절 "악"의 제어화일

III. 결론

본 연구에서는 포먼트 합성기에 사용될 수 있는 반음절 데이터 세트의 구현에 관하여 기술하였다. 400 여개의 반음절을 합성단위로 사용함으로써 합성단위 세트의 크기를 적정선에서 유지하면서, 다양한 음운환경에서의 조음결합현상을 포함하여 문자-음성 변환장치시스템의 구현이 용이하도록 하였다. 각각의 반음절은 음원의 세기, 포먼트 및 대역폭, 기본주파수의 시간적 변화를 프레임 단위로 저장한 파라미터 화일과 각 반음절에서 음원의 변화 정보 및 접속 및 길이 조절을 위한 천이정보를 가지고 있는 제어화일로 구성하였다. 제어화일은 <시간, 음향학적 사건>의 열로 구성하여 문자-음성변환장치의 구현시에

는 각각의 음향학적 사건이 일어난 시점만의 파라미터 값만 저장하고 있어서 가능하도록 하였다. 포먼트 합성기에 있어서 음질에 큰 영향을 미치는 요소중의 하나인 포먼트 궤적을 음절의 종류에 따라 추측, 처리하는 방법을 기술하였다. 제안된 포먼트 궤적의 처리방법에 따라 포먼트궤적의 처리가 곤란한 종성폐쇄음에 대하여 실험한 결과 비교적 좋은 결과를 얻었다. 무성음 구간에서의 포먼트의 전폭의 궤적을 효과적으로 근사할 수 있는 방법과 음절의 표준값이에 대한 연구가 진행되어야 할 것으로 판단된다..

참고 문헌

- [1] Special Issue on Man-Machine communication by Speech, Proc. IEEE, vol. 73, no. 11, Nov. 1985
- [2] D. H. Klatt, "Review of text-to-speech conversion for English," J. Acoust. Soc. Am, vol.82, no. 3, pp737 - 793, 1987
- [3] D. H. Klatt, " Software for a cascade/parallel formant synthesizer", J. Acoust. Soc. Am, vol. 82, no. 3, pp. 971 - 995, 1980
- [4] 김 용갑, " A Rule-Based Text-to-Speech System for Korean", Proc. of Korea-Japan Joint WorkShop on Advanced Tech. of Speech Recognition & Synthesis, pp. 109 - 114, 1991