

우리말 규칙합성에 관한 연구 (II)  
 - 반음절 단위의 음성합성

천 강식 이 성준 이 재홍  
 서울대학교 전자공학과

Synthesis-by-rule of Korean: Part II  
 - Speech Synthesis Using the Units of Demisyllables

Kang Sik Cheon Sung Jun Lee Jae Hong Lee  
 Department of Electronics Engineering  
 Seoul National University

Abstract

A new set of the units of demi-syllables is presented for Korean speech synthesis. The performance of the set of demi-syllable units is compared with that of the set of syllable units in the aspects of the quality of synthesized speech using each set of the units and the size of the computer memory which each set of units occupies. The set of demi-syllable units achieves comparable speech quality and occupies smaller memory size than the set of syllable units.

추출하여 음성합성 단위 세트를 구성하고 기억장치에 저장해 두었다가 합성하고자 하는 문장에 따라 해당되는 부분을 검색하여 차례대로 연결함으로써 이루어진다. 음성합성은 합성 가능한 어휘 수에 따라 분류하면 제한 단어 음성합성과 규칙합성 (또는 무제한 단어 음성합성)으로 나누어진다 [1]. 본 논문에서는 한국어 규칙합성을 위해서 반음절 (demi-syllable)을 합성의 기본단위로 하여 구성한 음성합성 단위 세트를 제안하고 음성합성 단위 세트의 성능 검토를 위해 제작한 음성합성기를 소개한다.

I. 서론

지금까지는 인간과 컴퓨터간의 정보 전달의 수단으로써 키보드와 모니터를 이용하는 시각적 방법이 사용되어 왔다. 정보의 종류가 다양해짐에 따라 시각적 방법만으로는 불충분하게 되어 청각적인 방법을 도입하게 되었다. 이를 위해 음성인식과 음성합성에 관한 연구가 최근 활발히 진행되고 있다. 이 중 음성합성은 컴퓨터가 인간에게 음성의 형태로 정보를 전달하게 하는 수단이다. 음성합성은 기본적으로 인간의 음성을 단어, 음절, 음소 등의 단위로

II. 음성합성 시스템의 개요

합성할 한국어 문장이 입력되면 먼저 음운변동 처리를 통해서 발음나는 대로의 문장으로 변환된다 [2]. 이 변환된 문장에 따라 제어 프로그램은 합성에 필요한 합성단위를 음성합성 단위 세트로부터 가져와서 순서대로 합성기로 출력한다. 음성합성의 흐름도는 그림 1과 같다.

III. 한국어 음성합성의 기본단위

한국어 음성합성의 단위로는 문장, 단어, 음절, diphone, 이음(allophone)등이 있다. 이 중에서 문장이나 단어는 제한단어 합성에는 적합한

합성단위이나, 규칙합성에는 필요한 합성단위의 갯수가 매우 많아서 규칙합성의 합성단위로는 적합하지 않다 [1]. 음절은 규칙합성의 합성단위로 사용되면 좋은 음질을 낼 수 있으나 필요한 합성단위의 갯수가 많아서 PC를 이용한 음성합성 시스템에서는 적합하지 않다 [3]. diphone이나 이음을 합성단위로 사용하면 필요한 합성단위의 갯수는 음절을 합성단위로 사용한 경우에 비하여 크게 줄일 수 있다 [4], [5]. 그러나 아직 한국어 diphone 및 이음의 음성학적 정량적 특성에 관한 철저한 분석이 부족한 까닭으로 diphone과 이음을 합성단위로 사용하여 합성한 음성은 그 음질이 음절을 합성단위로 사용한 경우에 비해 상당히 떨어진다 [6]. 이에 대한 대안으로서 본 논문에서는 반음절 (demi-syllable)을 한국어 규칙합성의 기본단위로 사용할 것을 제안한다. 반음절을 합성단위로 사용하면 합성단위의 갯수를 음절을 사용한 경우에 비하여 약 1/2로 줄일 수 있다.

#### IV. 한국어 음성합성을 위한 음성합성 단위 세트의 구성

한국어의 음절은 초성, 중성, 종성으로 구성되는 구조적 특성을 가지고 있다. 이러한 구조적 특성에 착안하여 반음절로 음성합성 단위 세트를 구성할 수 있는데 그 방법에는 다음 두 가지가 있다. 첫째, 모든 초성, 중성 및 종성에 대하여 '초성 + 중성'과 '중성의 일부 + 중성'을 합성단위로 하여 음성합성 단위 세트를 구성한다. 둘째, 모든 초성 중성 종성에 대하여 '초성 + 중성의 일부'와 '중성 + 종성'을 합성단위로 하여 음성합성 단위 세트를 구성한다. 둘째 방법으로 음성합성 단위 세트를 구성하는 합성단위는 표 1과 같다.

#### V. 음성합성기

음성합성 단위 세트의 성능 검토에 사용된 음성합성기의 구성도는 그림 2와 같다. 이 음성합성기는 PC/XT 호환기종에 연결하여 사용한다 [7]. PC/XT로부터 음성합성에 필요한 데이터는 handshaking 방식을 통하여 PPI (programmable peripheral interface)에 입력된다. PPI는 이 데이터를 음성합성 프로세서(VSP: voice synthesis processor)에 전달한다. 음성합성 프로세서는 펄스 열(pulse train)과 불규칙 잡음원을 조정함으로써 입력된 데이터에 각각 대응되는 유성음과 무성음을 LPC기법을 사용하여 합성한다 [8]. 합성된 음성신호는 출력부에서 변환 증폭되어 스피커를 통하여 음성으로 출력된다.

#### VI. 실험 결과

한국어를 규칙합성하는데 있어서 IV절에서 제안한 반음절을 합성단위로 하는 두 가지 음성합성 단위 세트를 구성하고, 각각의 음성합성 단위 세트를 사용하여 합성음을 만들어 그 음질을 비교하였다. 그 결과 두 방법 중 후자('초성 + 중성의 일부'와 '중성 + 종성')에 의한 합성음이 전자('초성 + 중성'과 '중성의 일부 + 중성')보다 우수하였다. 두 가지 반음절 합성단위 중 보다 우수한 후자를 합성단위를 사용하여 합성한 음성과 음절을 합성단위로 사용하여 합성한 음성의 음질을 비교한 결과 거의 차이가 없었다. 한편, 발음질을 합성의 기본단위로 하여 음성합성 단위 세트를 구성하면 음절을 합성단위로 하여 음성합성 단위 세트를 구성할 때보다 음성합성 단위 세트가 점유하는 컴퓨터 기억장치(PC의 경우 RAM)가 약 1/3 정도로 줄어 든다.

#### VII. 결론

한국어 규칙합성을 위한 기본 합성단위로서 반음절을 제안하였다. 반음절을 합성의

기본단위로 사용하면 음성합성 단위 세트의 크기가 음절을 합성의 기본단위로 사용한 경우에 비하여 약 1/3 정도로 줄어 든다. 또한 반음절을 합성단위로 사용하여 합성한 음성의 음절은 음절을 합성단위로 사용하여 합성한 음성의 음절과 거의 차이가 없다. 합성단위에 따른 합성음의 음절을 비교 검토하는데 사용된 합성기는 음성합성 프로세서들 사용하여 구성한 것으로 그 성능이 제한적이었다. 음성합성 프로세서 대신에 범용 디지털 신호처리 IC를 사용한다면 보다 우수한 합성 음절을 얻을 수 있을 것으로 기대된다.

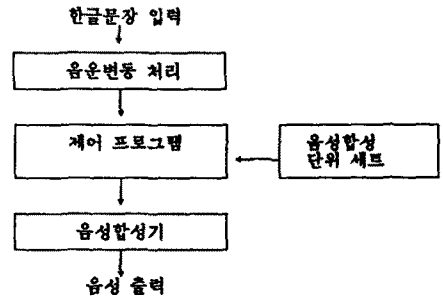


그림 1. 음성합성의 흐름도  
Fig. 1. Flow chart of speech synthesis.

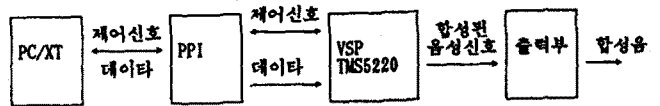


그림 2. 음성 합성기 구성도  
Fig. 2. Block diagram of a speech synthesizer.

표 1. 음성합성 단위 세트의 구조  
Table 1. Set of units for speech synthesis.

a) 초성 + 중성의 일부

가	가	거	저	고	교	구	규	그	기	개	개	개	계	과	과	피	권	웨	귀
까	까	꺼	저	꼬	교	꾸	규	끄	끼	깨	깨	깨	계	과	과	피	권	웨	귀
나	나	너	너	노	노	누	뉴	느	니	내	내	내	내	와	와	외	외	외	외
다	다	더	더	도	도	두	듀	드	디	대	대	대	대	와	와	외	외	외	외
따	따	더	더	도	도	두	듀	드	디	대	대	대	대	와	와	외	외	외	외
라	라	러	러	로	로	루	류	르	리	래	래	래	래	와	와	외	외	외	외
마	마	머	머	모	모	무	무	므	미	매	매	매	매	와	와	외	외	외	외
바	바	버	버	보	보	부	부	브	비	베	베	베	베	와	와	외	외	외	외
사	사	서	서	소	소	수	수	스	시	새	새	새	새	와	와	외	외	외	외
아	아	어	어	오	오	우	유	으	이	에	에	에	에	와	와	외	외	외	외
자	자	저	저	조	조	주	주	즈	지	재	재	재	재	와	와	외	외	외	외
차	차	저	저	조	조	추	추	츠	치	채	채	채	채	와	와	외	외	외	외
카	카	커	커	코	코	쿠	쿠	크	키	카이	카이	카이	카이	와	와	외	외	외	외
타	타	터	터	토	토	투	투	트	티	태	태	태	태	와	와	외	외	외	외
파	파	퍼	퍼	포	포	푸	푸	프	피	패	패	패	패	와	와	외	외	외	외
피	피	퍼	퍼	포	포	푸	푸	프	피	패	패	패	패	와	와	외	외	외	외
하	하	허	허	호	호	후	후	후	히	해	해	해	해	와	와	외	외	외	외

단위의 갯수 = 351

b) 중성 + 중성

Prediction of Speech. Springer-Verlag,  
1976.

아	악	안	안	알	알	알	알
야	약	얀	얀	랄	랄	랄	랄
이	익	인	인	일	일	일	일
의	의	연	연	염	염	염	염
오	옥	온	온	울	울	울	울
요	육	운	운	을	을	을	을
우	육	운	운	음	음	음	음
유	육	운	운	음	음	음	음
으	육	인	인	일	일	일	일
이	익	연	연	염	염	염	염
애	익	연	연	염	염	염	염
애	익	연	연	염	염	염	염
애	익	연	연	염	염	염	염
와	학	환	환	환	환	환	환
왜	학	환	환	환	환	환	환
의	학	환	환	환	환	환	환
위	학	환	환	환	환	환	환
위	학	환	환	환	환	환	환
의	학	환	환	환	환	환	환

단위의 갯수 = 168

참고 문헌

[1] G. Bristow, Electronic Speech Synthesis. McGraw-Hill, 1984.

[2] 임 성빈, 이 재홍, 최 계근, "우리말 규칙합성에 관한 연구 (I) - 음운변동 처리 시스템," 대한전자공학회 추계 종합학술대회는논문지, 1987.

[3] 은 중판, "음성에 의한 Man-Machine Communication 기술의 현황," 대한전자공학회 전자공학회지, vol. 15, no. 2, April 1988.

[4] 김동성, Diphone를 이용한 한국어 음성합성에 관한 연구. 서울대학교 석사학위 논문, 1988.

[5] 이 정철, 음소를 이용한 음절 단위 한국어 규칙합성에 관한 연구. 서울대학교 석사학위 논문, 1988.

[6] 서울대학교 생산기술연구소, 한국어 규칙합성에 관한 연구. 서울대학교 생산기술연구소, 1987.

[7] J. Coffron, The IBM PC Connection. Sybex, 1984.

[8] J. Markel and A. Gray, Jr., Linear