

단음절 합성단위음을 사용한 시간영역에서의 한국어 다음절어 규칙합성을 위한 음절간 접속구간에서의 에너지 흐름 제어에 관한 연구

정회원 강 찬 희*, 김 윤 석*

On the Control of Energy Flow between the Connection Parts of Syllables for the Korean Multi-Syllabic Speech Synthesis in the Time Domain Using Mono-syllables as a Synthesis Unit

Chan-hee Kang* Regular Member, Yun-seok Kim* Regular Members

요 약

본 논문은 시간영역 상에서의 단음절 단위합성음을 사용한 다음절어 합성에 관한 연구이다. 특히, 파형 연결시 접속구간에서의 에너지 흐름의 형태를 제어하기 위한 연구이다. 이를 위하여 시간영역 상에서 추출한 운율요소 제어용 매개변수¹⁾를 사용하여 제어하였으며, 음절간 파형 형태의 접속규칙을 도출하여 합성시킴으로써 에너지 흐름의 형태를 시간영역 상에서 제어시킨 결과를 제시하였다. 실험결과, 단음절 단위의 저장된 파형을 연결시킴으로써 발생하는 에너지 흐름의 불연속성을 제거할 수 있었으며, 또한 합성음의 음질 및 자연성이 향상되었다.

ABSTRACT

This paper is to synthesize korean multi-syllabic speeches in the time domain using mono-syllables as a synthesis unit. Specially it is to control the shape forms of speech energy flows between the connection parts of syllables in the case of concatenating mono-syllables. For this it is controlled with the prosody parameters¹⁾ extracted from speech waveforms in the time domains and presented the experimental results controlled the energy flows by using the induced concatenation rules from the korean syllable shapeforms in connection parts of syllables. In the results of experiments, it is removed the discontinuities of energy flows in the connection parts produced by concatenating the mono-syllables in the time domain and also improved the qualities and naturalities of synthesized speeches.

I. 서론

본 연구의 궁극적인 목적은 텍스트를 음성으로 변환하기 위함이다. 음성합성에 있어서 그 합성방식 또한 매우 다양하며, 이들 합성 방식이 지니고 있는 문제점 및 특성 또한 널리 알려져 있다. 음성합성분야는 광의적으로 해석하면, 기초요소 기술로서 그 용

용분야는 매우 광범위하다. 따라서 본 논문에서의 주 연구대상은 자연스럽고 음질이 우수한 새로운 합성방식의 개발이다. 본 논문에서의 연구개발의 전 반적인 관점은 첫째로 인간이 발성한 음질의 수준에 비견되는 합성음을 얻기 위한 시간영역에서의 합성 방식의 개발이다. 그 이유는 주파수 영역에서의 합성음이 지니고 있는 한계점 때문이다. 주파수

* 상지영서대학 전자과(chkang@sangji-c.ac.kr)

논문번호 : 99249-0628, 접수일자 : 1999년 6월 28일

※ 이 논문은 한국학술진흥재단의 '97 자유공모과제(1997-004-E00224) 연구비에 의하여 연구되었음.

영역에서의 합성 모델은 소스 여기원과 성도모델을 추정하여 합성음을 생성시키는 방식이므로 이들을 각각 추정할 때 이미 어떤 방식이든지 정확한 추정이 어려워 오차가 발생하므로 합성 음질 차원에서는 인간 육성에 의한 시간영역에서의 합성방식에 의한 음질에는 미치지 못하기 때문이다. 둘째로 시간영역에서의 합성 방식이 지니고 있는 한계점의 극복이다. 즉, 시간영역상에서의 합성방식은 주파수 영역에서의 합성 방식과는 달리 저장된 단위 합성음을 그대로 이용하므로 연속음 합성시, 고정된 파형의 형태를 변형시키기가 어려워 자연성이 매우 열악하다는 점이다. 따라서 주파수 영역에서의 합성방식이 지니고 있는 운용요소 제어 기법을 이용하여(즉, 주파수 영역에서의 운용요소 제어 원리는 소스 여기원의 한 예로 임펄스 여기원을 들면, 역양성분을 제어하기 위하여 임펄스 열의 간격으로 피치주기를 제어하고, 임펄스 열의 개수로 장단요소를 제어하고, 강약요소는 앰프 특성과 같이 임펄스의 진폭성분으로 제어하듯이, 시간영역에서 저장된 고립어 형태의 단위합성음으로부터 임펄스 열의 간격에 해당하는 단위합성음의 매 단위피치주기들을 추출하고, 또한 임펄스의 크기에 해당하는 진폭제어성분을 단위 피치주기로부터 추출하여 매개변수화 하여 장단강약고저 성분을 제어함) 시간영역상에서 이미 저장된 파형으로부터 장단강약고저 성분을 제어하기 위한 파라미터를 추출하여 운용요소의 제어가 가능한 자연성과 음질이 양호한 무제한 연속음에 대한 합성방식을 개발하고자 한다.¹⁾²⁾ 셋째로, 고립어 단위합성음의 사용이다. 그 이유는 텍스트를 음성으로 변환할 때 한글이 지니고 있는 특징이 고립어 단위로 창제되어서 1: 1의 대응이 가능하여져 오디오텍스의 구현 및 데이터 통신시 2 바이트 단위의 한글 등을 음성 데이터로 변환하여 전송할 수 있어서 ARS 시스템에도 활용될 수 있기 때문이다. 또한 다이폰 단위음이나 반음절어를 사용할 때 그 경계구간이 애매하여 합성단위를 추출 알고리즘에 따라 단위합성음 경계구간에서의 불협화음으로 인한 에코음이 발생하는 현상을 극복할 수 있는 장점이 있다. 그러나 한국어 합성시, 합성에 사용되는 단위음으로는 다이폰을 사용하여 음절과 음절 사이에 발생하는 조음변화로 인한 변이음 문제를 해결하는 연구가 주로 이루어지고 있다. 특히 한국어가 음절 단위의 음성임에도 불구하고 단음절 단위음을 사용하지 못하고 있는 주된 이유 중의 하나가 한정된 1,024개 정도의 단위음으로 음절간에서 발생하는 변

이음에 대처하기 어려운 문제점을 내재하고 있기 때문이나 이를 극복하고자 한다. 넷째로, 파형연접에 의한 시간영역 상에서의 독자적인 합성방식 개발이다. TD-PSOLA³⁾⁴⁾ 방식에서 사용하는 방식은 합성단위음으로 CDU라는 합성단위음을 저장하여 운용요소를 제어하기 위하여 가중합수를 가하여 피치를 변경시키거나 이는 저장된 원음질을 변형시키는 또 다른 요인의 하나이므로 가끔적이면 저장된 원음성 자체를 이용하여 원음이 지닌 음질 자체를 유지시키는 방법을 개발하고자 하였다.

그러나 위에서 열거한 합성방식¹⁾²⁾에 의한 연속음 합성시에는 수많은 문제점들이 내재하고 있다. 그 중 상당한 문제점들이 수년간에 걸친 개발의 결과로 많이 개선되어 본 방식이 지닌 특유의 자연성과 음질을 확보할 수 있게 되었다. 이를 문장단위의 합성음이나 다음절어 단위의 연속음 합성에 적용시키기 위하여는 이 방식에 적합한 에너지 컨투어 및 피치 컨투어 생성이 이루어져야 한다. 본 연구에서는 연속음 합성을 위한 에너지 컨투어 생성시 발생하는 고립어 단위의 파형 연접 구간에서의 음질 저하 현상을 개선시킨 다음절어 합성에 관한 연구이다. 즉, 다음절 단위음을 사용한 시간영역에서의 합성방식 중에서 연속음 합성을 위한 여러가지 문제점 중에서 파형 연접 구간에서의 음질저하현상을 개선하기 위한 음성파형 포락선 형태 제어에 관하여 주로 논하였다.

일반적으로 파형 연접에 의한 합성 방법은 위상 왜곡으로 인한 잡음이라든가, 접속 점에서의 불일치로 인한 스파이크성 잡음이 발생되어 음질이 저하되고 운용요소의 제어가 용이하지 못하다. 따라서 제한한 방식에서는 이러한 점들을 극복하기 위하여 운용요소를 시간영역 상에서 효율적으로 제어하기 위한 몇 가지 매개변수의 추출 및 음절간 접속구간에서의 파형 포락선 제어 알고리즘(Ⅱ장)과 연속음 합성을 위한 음절간 파형 접속규칙(Ⅲ장)을 도출하여 다음절어를 합성시킨 결과를 V장에 제시한다.

Ⅱ. 규칙합성용 매개변수 추출 및 파형 포락선 제어 알고리즘

저장된 고립어 단위의 단위합성음을 단순히 텍스트에 따라 파형연접시켜 합성음을 발생시킬 수는 없다. 인간이 발생한 음성은 에너지로서 장단성분이 라든가, 강약 및 고저 성분들이 어우러져 서로의 상관관계에 의하여 유체역학적인 자연스런 에너지

flow를 보이고 있기 때문이다. 즉, 단순 파형 연결은 음절단위의 합성음 음절은 좋다할지라도 음절과 음절이 모여 화자의 감정상태나 의미의 전달등을 나타내는 운율요소의 모든 성분들이 제어가 안되어서 매우 부자연스러울 뿐만아니라 인간이 인식하기 어려운 수준의 합성음의 결과가 되기 때문이다. 본장은 수많은 실험의 시행착오 끝에 하나씩 잘못된 부분을 개선해 나가는 방식으로 이루어진 것으로서, 시론에서 언급한 바와 같이 연속음 합성을 위한 연구이다. 이장에서는 연속음 합성을 위해서는 수많은 과제들이 산재해 있지만 우선, 저장된 파형의 형태를 연속음 형태로 시간 영역 상에서 변형시켜 파형 연결 구간에서의 이질음화로 인한 부자연성을 개선하기 위하여 그림1에서와 같이 저장된 단음절단위 음으로부터 추출된 파라미터를 이용하여 합성음의 에너지 flow를 제어하고, 음절간 접속구간에서의 파형접속규칙을 도출된 결과(III장에서 제시)를 이용하여 연속적인 파형 포락선 형태를 제어하기 위한 문제에 관하여 논하였다.

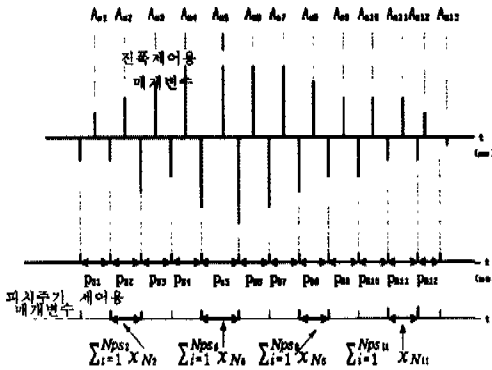


그림 1. 음성 파형 분석에 의한 운율 제어용 매개변수 추출

1. 시간영역에서의 규칙 합성음 매개변수 추출

시론에서와 같이 고립어 단위의 음성 파형을 입력시켜 파형을 분석한 후, 규칙 합성음 매개변수를 추출하여 지속시간, 강약 및 억양(피치 주기)등과 같은 운율요소를 시간 영역 상에서 제어하여 합성하였다. 주파수영역에서의 합성 시, 억양성분(피치 주기)은 임펄스 열 간격, 강약성분은 임펄스 세기, 장단성분은 프레임 개수 등을 조절하여 운율요소를 제어하므로 본 절에서도 파형을 분석하여 이에 대응하는 운율 제어용 매개변수를 시간영역 상에서 추출하여 운율요소를 제어하는 방법을 사용하였다.

① 피치주기 제어용 매개변수 : 1PERIOD(고립어내 1 피치 주기 프레임 당 음성 데이터)

고립어 내 유성음 구간에서의 단위 피치 주기 프레임 총 갯수(변수:Np)와 단위 피치 주기 프레임 내 음성 데이터(변수:1PERIOD)에 3차 Lagrange 보간법을 적용시켜 변경시키고자 하는 새로운 피치 주기의 음성 데이터 열로 바꾸어 피치 주기를 제어함.

② 진폭 포락선 제어용 매개변수 : RATIO

식 2)에서와 같이 고립어 내 단위 피치 주기 프레임 간격 당 최대·최소 진폭 열을 추출하여 포락선형태를 제어함.

③ 지속시간 제어용 매개변수 : Np

피치 주기 프레임의 총 갯수(변수:Np)를 조절하여 지속시간을 제어함.

④ 강약 제어용 매개변수 : MAX, MIN

식 2)에서와 같이 단위 피치 주기 프레임 간격 내에서의 최대·최소 진폭 비율을 조절하여 제어함.

⑤ 쉼 제어용 매개변수 : ZERO

구문 성격에 따라 0.2초, 0.3 초, 0.5초, 1초 등 적당한 쉼 구간 데이터를 부여하여 제어함.

먼저, 음성 파형으로부터 규칙 합성음 매개변수를 추출하기 위하여, 임의의 음성 데이터를 $x(n)$, 고립어의 음성 데이터 총 갯수를 N, 무성음부의 음성 데이터 갯수를 N_c , 유성음부의 음성 데이터 갯수를 N_v , 고립어 유성음부 내에서의 단위 피치 주기 프레임 총 갯수를 N_p 로 각각 정의한다. 그러면, 고립어 내의 음성 데이터 열은 $\sum_{n=1}^N x(n)$ 으로 표기된다. 이 때, 각각의 단위 피치주기 프레임 당 구간의 경계를 $P_{s1}, P_{s2}, P_{s3}, \dots$ 등으로 나타내고, 각 단위 피치주기 프레임 구간에서의 음성 데이터 갯수를 $N_{ps1}, N_{ps2}, N_{ps3}, \dots$ 등을 배열 $N_{ps}()$ 로 표기하면, N 개의 음성 데이터 열 $\sum_{n=1}^N x(n)$ 을 1차원 배열인 단위 피치 주기 프레임의 N_p 개 소블록의 합으로 표기 가능하다. 따라서 이를 2차원 배열로 표시하면,

$$\sum_{n=1}^N x(n) = \sum_{i=1}^{N_c} x(i) + \sum_{i=1}^{N_v} x(i) \quad (1)$$

(단, $N = N_c + N_v$ 임)

$$\sum_{j=1}^{N_1} x(j) = \sum_{n1=1}^{N_p} \sum_{n2=1}^{N_{ps}(n1)} x(n1, n2) \quad (2)$$

이 된다. 여기서, $x(5,10)$ 는 5번째 단위 피치주기 프레임 구간의 10번째 데이터를 의미한다. 즉, $x(n1,n2)$ 는 x (고립어 내 단위 피치주기 프레임 번호, 단위 피치 주기 프레임 구간 내 음성 데이터 열 번호)를 의미한다. 또한, 단위 피치 주기 프레임 구간 내에서의 음성 데이터 열의 최대 진폭의 절대치를 각각 $A_{m1}, A_{m2}, A_{m3}, \dots$ 등으로 정의하고, 각각의 피치 주기 프레임 구간내의 데이터 열을 일정한 크기로 정규화시킨 임의의 음성 데이터를 $x_N(n)$ 으로 정의하면, 2 차원 블록화 배열로 표시된 음성 데이터 $\sum_{n1=1}^{N_p} \sum_{n2=1}^{N_{ps}(n1)} x(n1, n2)$ 은

$$\sum_{n1=1}^{N_p} \sum_{n2=1}^{N_{ps}(n1)} x(n1, n2) \approx \sum_{n1=1}^{N_p} \sum_{n2=1}^{N_{ps}(n1)} A_m(n1) \cdot x_N(n1, n2) \quad (3)$$

이 된다.

위 식들로부터 추출·저장하여 고립어 단위 음성 DB에 작성된 주요 매개변수는

- 1) 고립어 내 전체 데이터 갯수 정보 : $N(2$ 바이트)
- 2) 고립어 내 단위 피치 주기 프레임 갯수 정보 : $N_p(1$ 바이트)
- 3) 각 피치주기 프레임 구간 내 음성 데이터 열 갯수 정보 : $\sum_{i=1}^{N_p} N_{ps}(N_p$ 바이트)
- 4) 고립어 내 단위 피치 주기 프레임 당 최대진폭 정보 : $\sum_{i=1}^{N_p} A_m(i) (N_p$ 바이트)
- 5) 단위 피치 주기별로 정규화된 음성 데이터 정보 : $\sum_{n1=1}^{N_p} \sum_{n2=1}^{N_{ps}(n1)} x_N(n1, n2) (N$ 바이트)

등이다.

2. 음절간 접속구간에서의 파형 포락선 형태 제어
한국어 음성에 있어서 연속음 파형의 기본 진폭 포락선 형태를 4가지 유형으로 분류(그림3)하여 합성에 이용하였다. 단음절어 합성시에는 그림3(a)의 포락선 형태로, 2 음절어 이상의 연속음을 합성할 경우에는 III장에서 제시한 음절유형간 진폭 포락선 결합규칙에 따라서 도입부의 진폭 포락선을 제어시키는 경우(그림3(b)와 쇠퇴부를 제어하는 경우(그림 3(c)) 및 도입부와 쇠퇴부를 동시에 제어하는 경우 (그림3(d))로 세분화시켜 파형의 포락선 형태를 제

어하였다.

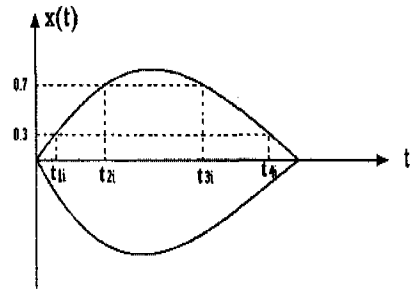


그림 2. 단음절 단위 음성의 진폭 포락선 제어 설명도

그림4는 3음절어의 “우리는”의 연속적인 진폭 포락선의 형태를 생성하기 위한 설명도이다. 이 그림에서와 같이 각각의 음절에 대한 진폭 포락선 형태가 진폭 포락선 결합규칙에 의하여 결정되면 진폭 포락선 제어에 필요한 4개 지점들을 구하여야 한다. 즉, 연속음 중에서 “우”에 해당되는 합성음을 생성시키기 위하여는 먼저 t_{31} 와 t_{41} 을 구하여야 한다. 그 다음에 t_{41} 지점에서 단음절어의 파형이 종료되는 지점까지의 파형은 합성에 이용하지 않으므로 절취하여 버리며, 파형이 시작되는 부분부터 t_{31} 지점까지의 파형 포락선은 원형 그대로 유지시키며 t_{31} 에서 t_{41} 까지의 부분은 진폭포락선을 일정한 진폭비를 유지하도록 제어하여 합성음을 생성하였다.

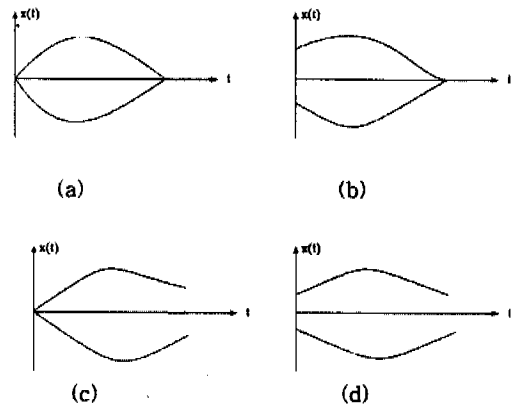


그림 3. 진폭 포락선 4 가지 유형도

이와같은 방법으로 두번째 음절 “리”는 $t_{12}, t_{22}, t_{32}, t_{42}$ 를 각각 구한 후, t_{12} 이전에 이르는 부분의 파형과 t_{42} 이후의 파형은 절취하여 버린 후 t_{12} 에서 t_{22} 까지에 이르는 파형과 t_{32} 에서 t_{42} 까지에 이르는

파형은 진폭 포락선을 제어시켜 앞 음절 “우”와 다음에 이어지는 음절 “는”의 파형 진폭 포락선을 제어시킨 후 각각의 음절을 연결시켜 연속음의 파형을 생성시키는 방법을 설명한 것이다. 그림 5는 이러한 방법에 의하여 규칙합성된 합성음 “우리는”의 파형도를 표시한 것이다.

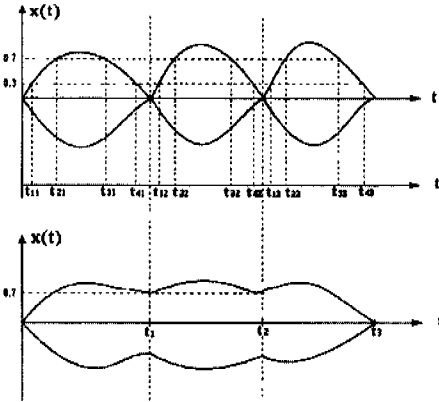


그림 4. 음성파형의 진폭 포락선 제어 예

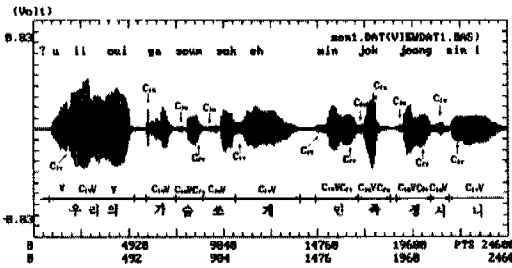


그림 5. 음절간 진폭 포락선 파형분석도(원음: “우리의 가슴 속에는 민족 정신이”)

표 1. 규칙합성에 사용된 한국어 자음 형태별 분류

No.	자음 유형	종류	갯수
1	C_{lu}	ㄱ, ㅋ, ㆁ, ㄷ, ㅌ, ㅃ, ㅍ, ㅅ, ㅆ, ㅈ, ㅊ, ㅊ, ㅅ, ㅆ, ㅈ, ㅊ, ㅊ, ㅅ, ㅆ, ㅈ, ㅊ	15
	C_{lv}	ㄴ, ㄹ, ㅁ	3
2	C_{fv}	ㄴ, ㄹ, ㅁ, ㅇ	4
	C_{fu}	ㄱ, ㆁ, ㅃ	3

III. 음절간 파형 포락선 형태 접속규칙

이 장에서 논하는 파형 포락선 형태 접속 규칙이 란 시간영역 상에서 파형 연결시 발생하는 조음변

화 현상과 음절간 이질음으로 인한 불협화음이 발생하는 현상을 개선하기 위한 과정으로 한국어의 음성학적인 면에서 고찰한 결과를 이용하여 음절간 접속구간에서의 파형 포락선 형태에 대한 접속 규칙을 도출하여 적용시킨 결과를 IV장에 제시하였다. 예를들면, 그림 6과 표2는 “우리는”을 규칙합성시킬 때 사용되는 음운기호를그림으로 표시한 것이며, 이를 음절간 접속구간에서의 진폭 포락선 결합 규칙과 연계지어 살펴보면, 우선 “우리는”의 음절 유형별 비트(b;b₀)는 “00-00-00”를 나타내며, 이는 진폭 포락선 결합 규칙3-1)이 적용되어 각각의 음절 진폭 포락선형태가 type 1+type 3+type 2로 되고, 파형 포락선의 비트b₁b₀은 “01-11-10”로 결정되어 합성음“우”는 종성구간(파형의 쇠퇴부), “리”는 초성과 종성구간(파형의 도입부와 쇠퇴부), “는”은 초성구간(파형의 도입부)의 파형 진폭 포락선을 제어하여 연속음 형태의 진폭 포락선을 지닌 합성음을 생성시킬 수 있도록 음운기호를 설계하여 규칙합성에 이용하였다. 그림7은 규칙합성과정을 단계별로 나타낸 것으로써, 음운기호가 생성되면 단음절 단위 음성 DB에 등록된 운율제어용 파라미터와 음성데이터를 읽어들이 음운기호에 해당되는 합성음을 생성하는 과정을 단계별로 설명한 것이다. “우리는”이란 합성음을 규칙합성하기 위하여는 음운기호에 의거하여, 정해진 규칙을 적용시켜 합성음을 생성하게 된다. 즉, 합성시 “우”에 해당되는 강약장단진폭 정보는“010101”을 나타내므로 이는 “보통의 강약에 보통의 발생속도(초당 5음절어 정도 : 200 msec 정도)를 지닌 합성음으로써 다음 음절의 진폭 포락선과는 연속적으로 이어진다”라는 정보로, “리”에 해당되는 음운기호내 운율정보는 “010111”로, “는”에 해당되는 음운기호내 운율정보는 “010110”로 각각 해독된다. 그 다음에 단음절 단위 음성 DB에 저장된 음절의 지속시간이 300ms 정도이면 음절내의단위피치 갯수정보를 읽어들이 200msec정도로 단위피치 주기 간격으로 음성데이터를 decimation 시켰으며, 음절의 지속시간이 200msec 이하이면 단위피치 주기 간격의 음성데이터를 일정한 비율로 보간시켜 지속시간을 제어하였다. 이러한 과정을 거친 후, 두 번째 운율정보 성분에 해당되는 강약 성분을 음운기호로부터 읽어들이 강약 제어의 과정을 수행한다. 그 다음에 진폭 포락선에 해당하는 비트값을 해독하여 II장 2절에서 제시한 합성음의 진폭 포락선의 기본 4 가지 유형 즉,

1)type 0:파형진폭의 변화가 없음.

2)type 1:파형의 종료부(즉, 쇠퇴부로 명하였음)를 진폭제어함.-“우”

3)type 2:파형의 시작부를 진폭제어함.-“는”

4)type 3:파형의 시작부와 종료부를 진폭제어함.-“리”에 따라서 다음절어 합성음을 생성하였다. 이때, 적용된 규칙은 아래와 같으며, type0 등은 음절간 접속시 파형 포락선의 기본 형태를 의미하며, 도출된 다음절어 포락선의 기본 형태가 결정되면 II 장 1절에서 추출한 파라미터를 이용 하여 다음절어의 장단강약고저 성분을 제어하여 합성음을 생성한다.

규칙 1) 단음절어 발생시에는 type 0의 규칙 적용.

규칙 2) 2 음절어 발생시에는

Byte 2						Byte 1		Byte 0
b7	b6	b5	b4	b3	b2	b1	b0	음가코드
2 바이트 조합형								
(b7b6:음절유형, b5b4:강약, b3b2:지속시간, b1b0:진폭)								

그림 6. 규칙합성용 음운기호

표 2. 규칙합성용 음운기호내 운율정보제어용 코드

b7b6b5b4b3b2b1b0	음절유형 b7b6	악센트 b5b4 (강약)	지속시간 b3b2 (장단)	진폭 b1b0 (윤곽선)
00	(C _{iv})V(C _{fv})	약음	장음	단음절어
01	C _{iu} V(C _{fv})	평상음	평상음	종성구간 변이음
10	(C _{iv})VC _{fu}	강음	단음	초성구간 변이음
11	C _{iu} VC _{fu}			초,종성 "

2-1) (C)V(C_{fv})+(C_{iv})V(C) : type 1+type 2의 규칙 적용.

2-2) (C)V(C)+C_{iu}V(C) : type 0+type 0의 규칙 적용.

2-3) (C)VC_{fu}+V(C) : type 0+type 0의 규칙 적용

규칙 3) 다음절어 발생시에는

3-1) (C)V(C_{fv})+(C_{iv})V(C_{fv})+(C_{iv})V(C) : type 1+type 3+type 2의 규칙 적용. (예:“우리는”:“우리는”).

3-2) (C)V(C_{fv})+(C_{iv})V(C_{fv})+C_{iu}V(C) : type 1+type 2+type 0의 규칙 적용. (예:“알맞다”:“알따”).

3-3) (C)VC_{fu}+(C)V(C_{fv})+(C_{iv})V(C) : type 0+type 1+type 2의 규칙 적용. (예:“앞가림”:“앞가림”).

3-4) (C)VC_{fu}+(C)VC_{fu}+(C)V(C) : type 0+type 0+type 0의 규칙 적용 (예:“학적부”:“학적부”). 와 같다.

IV. 다음절어 규칙합성

규칙합성과정은 그림 7에서와 같이, 저장된 음성 데이터로부터 파형분석과정을 거쳐 운율제어용 파라미터를 추출하여 음성합성용 데이터 포맷사전에 등록시킨다. 한국어 무제한 규칙합성시에는 입력되어진 문장단위의 기호열로부터 생성된 음가코드(그림6, 표2)를 발생하여 데이터 포맷사전에 등록된 음절단위의 음성데이터 및 운율제어용 파라미터를 읽어 들여 도출된 규칙합성에 필요한 규칙을 적용시켜 그림 7에서와 같이 단계별로 합성음을 생성시킨다. 이때, 음질이 저조한 합성음은 등록된 데이터 포맷사전을 수정하여 양질의 합성음을 생성시킬 수 있는 착오정정 기능을 지니고 있다. 이러한 과정을 요약하여 단계별로 설명하면,

[단계1]: 단음절어 단위의 음성파형을 10kHz 샘플링 비로 A/D변환하여 메모리에 저장시킨 후 파형을 분석하여 자음의 유무에 따라 다음절어 합성을 위한 단음절어의 기본형태소를 결정한다

[단계2]: 자음과 모음의 경계구간을 구한 후, 자음소 개수를 데이터 포맷 사전에 등록시킨다.

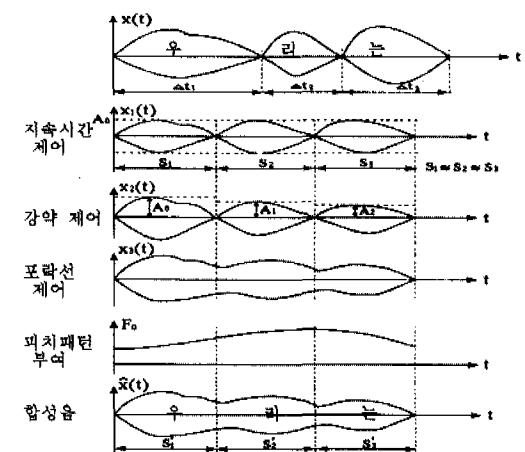


그림 7. 시간영역에서의 규칙합성 과정도

[단계3]: 규칙합성에 필요한 파라미터를 구하여 데이터 포맷 사전을 구축한다.

[단계4]: 데이터 포맷 사전으로 부터 합성시키고자 하는 음절의 화일을 읽어 생성된 음가 코드표에 의하여 장단강약고저음의 다음절어의 규칙합성음을 생성시킨다.

[단계5]: 규칙합성음을 청취 평가한 후, 저음질의 음성파형은 데이터 포맷 사전에 등록된 부분에 잘못된 부분을 정정하여 합성음질을 보정시켜 합성음의 음질을 개선시킨다.

[단계6]: 한국어의 언어학적, 음향학적 특성을 고찰하여 규칙합성 알고리즘을 작성하여 규칙합성에 이용한다.

와 같다. 그림 8은 이러한 과정에 의하여 생성된 다음절어 합성음의 한 예이며, V장에서 진폭, 지속시간, 피치 및 음절간 접속구간에서 합성음 파형의 포락선이 제어된 결과를 제시하였다.

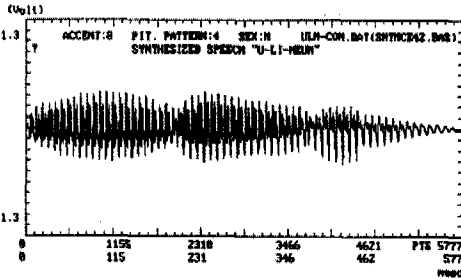


그림 8. 그림7에 의한 합성결과(합성음 “우리는”)

V. 음절간 접속구간에서의 파형 포락선 형태 제어결과

저장된 고품어 단위의 음성 데이터를 연결하였을 경우의 파형에서는 고품어 마다 각기 다른 고저, 세기, 장단을 지니고 있어서 청취하면 매우 부자연스럽다. 연속적으로 발생한 원 음성 “나는 강 찬회입니다”의 파형에서는 파형진폭의 흐름이 거의 영으로 떨어지는 일이 없어 연속적으로 이어지나, 고품어 단위에 의한 합성시에는 음절간 파형 접합점에서 영으로 떨어져 연결되어, 합성음 청취시 음절 사이가 끊겨져 청취되어 매우 부자연스러워진다. 본 논문에서는 이러한 문제를 해결하기 위하여 음성분석에서 추출된 임펄스 진폭 비 매개변수¹⁾를 이용하여 파형 접합면에서의 에너지 윤곽선의 흐름을 제어하였다(그림4,7,8,10). 연속어의 합성 시, 파형 진폭의 연속성은 에너지 윤곽선의 연속성과 일치하게 되므로, 본 논문에서는 음절간 파형 접합점에서의 진폭

포락선의 형태를 연속적으로 제어하였다(그림8,10). 이를 위하여 에너지 윤곽선의 시작점과 끝점을 합성음의 발생 속도 등에 따라 결정한 후, 저장된 임펄스 열의 진폭 비를 조절하여 파형 포락선의 형태를 여러 가지 규칙에 따라 지속시간, 피치 및 에너지의 흐름이 연속적이 되도록 생성하였다(그림9, 그림10). 그림 9와 그림10은 본 논문에서 제안한 시간 영역에서 합성시킨 연속음의 합성음 결과이다.

VI. 결론

본 논문은 음절간 접속구간에서의 파형 에너지 흐름 제어에 관한 연구로서 음절간 접속규칙을 도출하여 시간 영역에서 합성시키는 방식을 제안하였다. 그 가능성에 대한 타당성 검토로 V장에서 실험 결과를 제시하였다. 합성시 매개변수를 수식화하여 운용요소(장단강약고저)를 제어함으로써 통합적인 운용요소의 제어가 가능하여져 자연성이

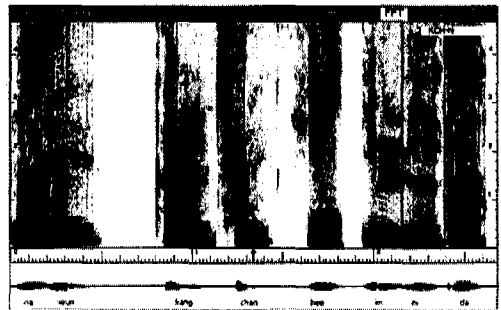


그림 9. 연속음에 대한 합성결과(나는 강찬회입니다)

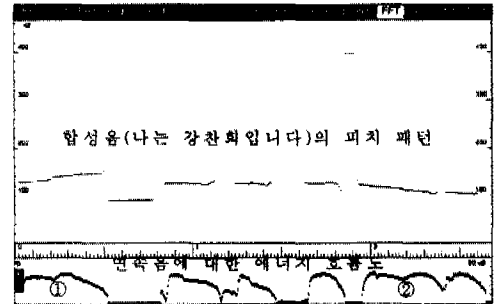


그림 10. 연속음에 대한 합성결과(나는 강찬회입니다)

향상되었다. 음절 또한 매우 자연스럽고 양질의 음질을 지닌 합성음으로 생성 가능하였다. 연속음 합성 시 음절간 접속구간에서의 파형 윤곽선 제어 및 변이음 처리가 개선되었다. 본 논문은 시간 영역에

서의 다음절어 합성에 관한 연구로서 저장된 음성 데이터로부터 매개변수를 추출하여 제어시키는 방식을 이용하였다. 합성시 II장 1절에서와 같이 매개변수를 수식화하여 저장된 고립어 단위의 단위합성음으로 부터 운율요소(장단강약고저)를 시간영역에서 제어함으로써 통합적인 운율요소의 제어가 가능하여져 자연성이 향상되다(그림9, 그림10). 실험결과, 시간영역에서의 단위절 단위합성음을 사용한 합성시 문제가 되는 변이음 처리시 음절간 접속구간에서의 음절간 파형 접속 규칙을 도출하여 파형 윤곽선을 제어하여 합성시킨 결과 음절간 접속구간에서의 음질과 자연성이 개선되었다. 다음절어 및 문장단위의 합성을 위하여는 한국어에 대한 언어학적인 연구 즉, 한국어 구문 및 문장 단위에서의 음절별 지속시간, 진폭, 피치 패턴 등과 같은 수많은 연구가 지속적으로 이루어져야 보다 자연스런 문장단위의 합성이 가능해질 것이다.

참고 문헌

[1] 강찬희, "시간영역에서의 무제한 고립어 합성을 위한 운율 요소 제어용 알고리즘 개발," 전자공학회 논문지, 제35권 C편 제7호, 1998.7

[2] 강찬희, 진용욱, "한국어 문어변환 시스템 내에서의 음성합성기 개발," 한국음향학회 논문지, 1993.2

[3] T. Dutoit, H. Leich, "Improving the TD-PSOLA Text-To-Speech Synthesizer with a Specially Designed MBE re-Synthesis of the Segments Database", ICASSP 92, vol. 1, pp.343-346.

[4] T. Dutoit, "High Quality Text-To-Speech Synthesis: a Comparison of Four Candidate Algorithms", Proc. ICCASSP 94, vol. 1, pp.565-568.

[5] Jonathan Allen, M. Sharon Hunnicutt and Dennis Klatt, From Text to Speech : The MITalk system, Cambridge Univ. Press, 1987

[6] Shuzo Saito, Fundamentals of Speech Signal Processing, Academic Press, 1981

[7] G. Rigoll, "The DECTalk system for German : A study of the modification of a text-to-speech converter for a foreign language," IEEE Proc. ICASSP '87, 1987

[8] 이현복, "현대 한국어의 악센트" 서울대학교 문

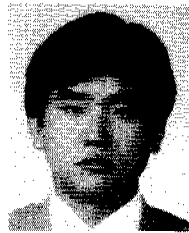
리대학보 19권 합병호(통권28호), 1973

[9] 성철재, "표준 한국어 악센트의 실험 음성적 연구 - 청취 테스트 및 음향분석," 서울대학교 대학원 석사학위 논문, 1991

[10] 이현복, 한국어의 표준발음, 교육과학사, 1989

강 찬 희(Chan Hee Kang)

정회원

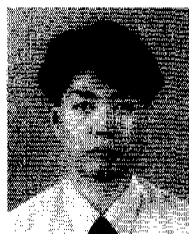


1980년 2월: 경희대학교 전자공학과 졸업(공학사).
 1982년 2월: 경희대학교 대학원 전자공학과 졸업 (공학석사).

1983년 7월~1985년 7월: 해군사관학교 교수부 전자과 전임강사.
 1985년 8월~1986년 8월: 삼성 통신 연구소 근무 (연구원).
 1986년 9월~1994년 8월: 경희대학교 대학원 전자공학과 졸업(공학박사).
 1989년 3월~현재: 상지영서대학 전자과 부교수.
 <주관심 분야> 한국어 TTS 시스템내 음성합성 알고리즘 개발, 구문분석기 설계, ARS 시스템 구현, 음성인식 및 음성 신호 처리 등임.

김 윤 석(Yun Seok Kim)

정회원



1989년 2월: 경원대학교 전자공학과 졸업(공학사).
 1991년 8월: 경희대학교 대학원 전자공학과 졸업 (공학석사).
 1996년 2월: 경희대학교 대학원 전자공학과 박사과정수료

1993년 10월~현재: 상지영서대학 전자과 조교수.
 <주관심 분야> 통신망, ATM 통화 량 제어, 신경회로망, 음성신호처리 등임.