

## 한국어 음성의 규칙 합성

鄭麟淙\*, 景淵靖\*\*, 李亮熙\*\*

\* 漢陽大學校 電子計算學科

\*\* 同德女子大學校 電子計算學科

### I. 서론

최근 음성언어에 의한 Man-Machine 인터페이스의 실현은 많은 관심을 모으고 있다. 이를 위한 음성 생성 방법은 크게 녹음 편집형, 분석합성형과 규칙 합성형으로 나눌 수 있다.<sup>[1]</sup> 이 가운데 규칙합성은 언어 정보와 음운 규칙으로부터 음성을 생성해내는 방식으로서, 규칙합성 기술은 근년 반도체 집적 회로의 진보에 따라 선진 각국을 중심으로 활발히 연구되어 일부 실용화 되었고, 음질의 개선을 위한 연구도 계속 진행되고 있다.<sup>[2], [6]</sup>

근래에 국내에서도 한국어 음성의 규칙 합성에 대한 연구가 활발히 진행되고 있다.<sup>[7], [14]</sup> 그러나 아직 실용화하기에는 음질면에서 미흡한 실정이므로 좋은 규칙 합성 시스템을 만들기 위해서는 우리말에 대한 많은 분석을 토대로 하여 규칙 합성 시스템을 이루고 있는 부분 시스템들을 확립하여야 한다.

본 고에서는 한국어 규칙합성 시스템의 개요 및 기술과 규칙 합성 시스템의 부분 시스템인 언어 처리 기능과 음성 합성 기능을 소개한다.

### II. 음성 생성 과정의 모델링

인간이 어떠한 정보를 음성으로 표현하는 과정을 몇 가지로 나누어 모델화하면 그림 1과 같이 나타낼 수 있다. 사과의 내용, 의도를 말로서 표현하는 인간과 같은 정도의 표현능력을 갖기 위한 “개념으로부터의 음성합성”<sup>[15], [16]</sup>을 생각하면, 그림1의 첫째 과정

인 개념의 표현과 개념의 언어화가 필요하다. 현 시점에서는 실현하기 어려운 문제이나, “개념으로부터의 음성합성”은 인간과 기계, 또는 인간과 인간의 원활한 대화를 실현하는 데 하나의 목적이 있다. 그러나 현단계에서는 문장의 기저구조를 음성으로 변환한다는 시도에 그쳐있다. 언어화 과정을 생략한 나머지 부분은 통상 문자열로 나타내어진 문장을 음성으로 변환하는 것으로서 텍스트 음성변환(Text-To-Speech, TTS)이라 하며, 규칙 합성(Synthesis-by-rule)으로 이를 실현할 수 있다. 음성신호는 조음 파라미터 계열과 음원파라미터 계열로 나타내어지고, 이 파라미터 계열들로 부터 음성신호로의 변환은 이전부터 성도 아날로그(구조 아날로그)방식, 포만트(formant)합성(터미널 아날로그, 스펙트럼 아날로그

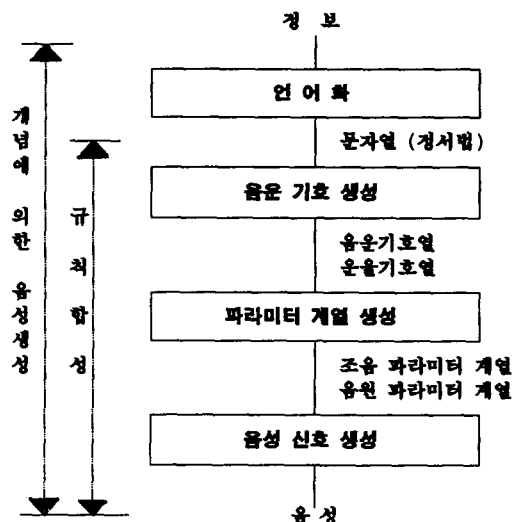


그림 1. 음성 생성과정의 모델

그)방식이 이용되고 있다. 최근에는 분석합성 방식으로서 유용한 PARCOR(PARTIAL CORrelation)방식, LSP(Line Spectrum Pair)방식과 첵스트럼(cepstrum)방식등과, 음성 파형의 단위 음성을 접속하여 음성을 생성하는 PSOLA(Pitch - Synchronous OverLap Add)방식도 이용되고 있다. [17], [19]

파라미터-음성변환 방식의 선택 혹은 새로운 방식의 확립은 요구조건이 엄격하기 때문에 쉽지 않다. 음성을 규칙합성하기 위한 시스템을 구성하기 위해서는, 먼저 언어처리 기능에 의해 입력문장인 문자 계열로부터 음운기호 계열과 운율기호 계열이 생성되고, 또 이들 음운기호열과 운율기호열로부터 파라미터계열이 생성된다. 이 계열들은 각각의 규칙들을 사용하여 생성할 수 있지만, 이 규칙들은 파라미터-음성변환 방식에 따른다. 따라서, 규칙합성시스템에 있어서 파라미터-음성변환 방식의 선택은 시스템 전체의 각 부분 시스템과 밀접한 관계가 있다.

### Ⅲ. 규칙 합성계

텍스트 음성변환 시스템은 언어에 따라 다른데, 한국어의 텍스트 음성변환 시스템은 그림 2와 같이 구성할 수 있다. 이 시스템은 크게 언어처리 기능과 음성합성

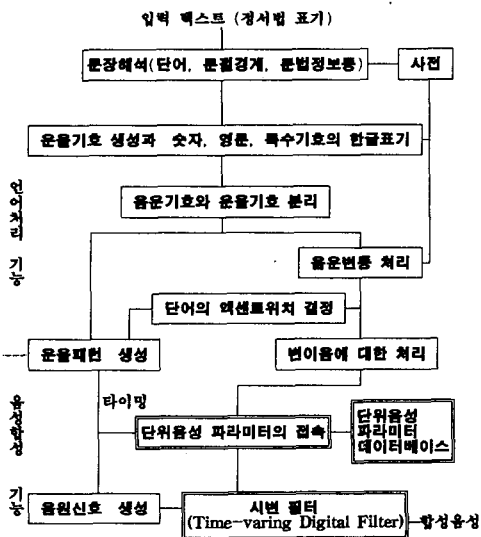


그림 2. 한국어 음성의 규칙합성 시스템의 블록도

성합성 기능으로 나누어 생각할 수 있다. 언어 처리 기능부에서는 정서법으로 쓰여진 입력 문장을 음운 기호와 운율 기호로 변환하여 출력하며, 음성 합성 기능부에서는 음운 기호와 운율 기호를 조음 파라미터 계열과 음원 파라미터 계열로 변환하고 또한 음성 생성 필터를 구동하기 위한 여기 신호를 생성한다.

이러한 텍스트 음성변환 시스템의 구현에는 음원과 조음의 파라미터 계열을 자연음성의 데이터를 사용하지 않고 완전히 규칙에 의해 음성을 합성하는 순수규칙합성 방식(포맷트 합성)과 조음 파라미터 계열의 일부를 직접 자연음성으로부터 추출하여 사용하는 보코더(Vocoder)를 이용한 분석합성 방식이 있다. [17] 순수합성 방식은 원리적으로는 이상적인 방식이지만, 조음파라미터계열을 생성하기 위한 규칙을 얻기가 매우 어렵다. 이에 반해, 분석합성계를 이용하는 규칙합성 방식은 적어도 단위음성의 데이터로서는 자연음성으로부터 직접 구한 음성특징파라미터를 사용할 수 있다. 따라서 파라미터 음성합성에서는 이미 실용화되고 있는 분석합성계를 이용할 수 있기 때문에 비교적 간단한 시스템으로 양질의 음성을 합성할 수 있다.

완전한 규칙 합성 시스템의 구축을 위해서는 먼저 음성 합성 기능부의 음성 생성을 위한 시변 필터(Time-varying digital filter)와 단위 음성 파라미터의 접속부, 단위 음성 데이터베이스 구축이 필요하며, 실험적인 방법에 의한 운율 규칙등의 부분 시스템이 확립되고 적용되어야 한다. 본 고에서는 주로 분석합성계를 이용한 규칙합성시스템에 대해서 논하고자 한다.

#### 1. 언어처리기능

언어처리 기능에서는 정서법으로 쓰여진 입력 문장을 형태소 및 구문해석, 의미해석을 거쳐 음운기호 및 운율 기호로 변환 할 해석기(Parser)가 필요하다. 입력된 문장은 이 해석기를 거쳐 운율 기호와 음운 기호로 변환되며 기호, 약어, 알파벳등도 여기서 발음 형태의 텍스트로 변환된다.

구문 분석에 의하여 문절 접속도에 따른 발화의 운율 세그먼트(운율 그룹, Pause 할당)가 생성되며 피치(Pitch)의 변화와 강세(Stress)기호가 생성되어, 이들 세그먼트에 적절한 운율 패턴, 즉 운율 기호열이 생성된다. 이러한 운율 기호열은 규칙에 의해 나타내어 진다. 그러나 아직 한국어에 있어서 음운 기

호, 운율 기호를 생성하기 위한 언어처리 기능이 미흡한 상태이므로 이 분야의 많은 연구가 요구된다.

1) 음운 기호 계열

입력된 문장은 사전 정보에 의한 형태소 및 구문, 의미해석을 기반으로 하여, 음운 변동 처리되어 소리나기 표기, 즉 음운 기호 계열로 변환된다. 이 음운 기호 계열은 변이음 처리와 장/단음 처리에 의해 음성 기호로 변환된다.

음운과 음운이 만날 때 양쪽 모두 또는 어느 한쪽의 소리가 변할 때 이를 음운 변동이라 하며 우리말의 음운변동은 규칙 음운변동과 불규칙 음운변동으로 나눌 수 있다. 음운변동의 규칙의 예외에 해당되는 불규칙 음운변동의 예로, 겹받침 /ㄹㅂ/은 음운변동의 규칙에 따르면 어말 또는 자음 앞에서 'ㄹ'로 발음하는 것이 원칙이지만(여덟 [여덜], 넓다. [널따]), '뱃다'의 경우 [뱃:따]로 발음된다.<sup>[20]</sup> 이러한 불규칙음운변동 단어는 예외발음사전에 등록시켜 음운변동 규칙들을 적용하기 전에 찾아 처리해 준다. 이와같이 불규칙 음운변동을 처리해 준 다음, 음운변동 규칙들을 적용하게 되는데, 여기에는 말음 법칙, 자음동화, 구개음화, 자음축약, 연음 법칙과 경음화등이 있다. 예를 들어 굳이->구지(구개음화), 앞->압(말음 법칙), 먹는->멍는(자음동화), 좋다->조타(자음 축약), 옷이->오시(연음 법칙), 국밥->국뺨(경음화) 등이 있으며, 이러한 규칙들이 연속하여 적용되어 지는 경우로서, 발머리->반머리(말음 법칙)->반머리(자음동화)와 같은 것이 있다.

텍스트로부터 음운기호열로 변환하기 위한 우리말의 음운변동처리 시스템은 많은 연구가 행하여져<sup>[11], [13]</sup>, 어느정도 구축되어 있다.

또한 우리말의 변이음 처리가 필요하다. 예를들어, /ㄹ/의 발음이 /천리/에서는 설측음화 되지만 /아리랑/에서는 탄설음화 되어 나타난다.<sup>[21]</sup> 또 다른 변이음의 예로, 자음 /ㄱ/은 무성음이나, 유성음화 되었을 때의 변이음이 그림 3에 각각 나타나있다. 무성의 /ㄱ/과 유성음의 /ㄱ/을 변별하지는 못하나 음성신호에 있어서는 크게 다르다. 이처럼, 파열음 /ㅈ//ㅊ//ㄱ/과 마찰음 /ㅎ/, 파찰음 /ㅅ/이 유성음 사이에서 유성음화<sup>[20]</sup> 하는 변이음의 처리는 합성음의 음질 특히, 자연성에 큰 영향을 준다.

그림 4는 문자열로부터 음소열로의 변환 알고리즘이다.<sup>[9]</sup> 먼저 마지막 음절이라면 폐쇄음이 있는 경우 말음 법칙을 적용하여 대표음 처리를 하며, 그렇지

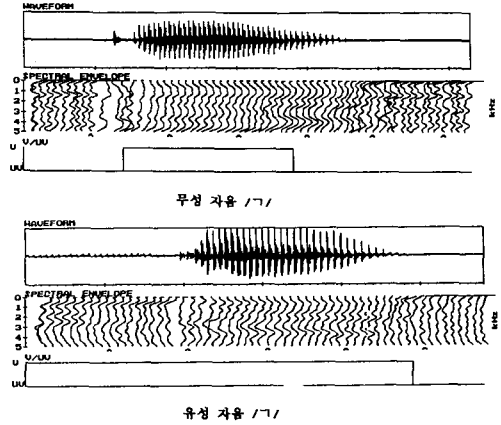


그림 3. 무성음/ㄱ/과 유성화 된 /ㄱ/의 파형과 스펙트럼 포락

않은 경우 끝난다. 마지막 음절이 아닌 경우, 예외 발음 사전을 참조하여 불규칙 음운변동이라면 뒤 음절의 초성 자음을 살펴 유성음화 또는 음운 첨가를 수행한다. 그렇지 않다면 단모음화와 구개음화등을 수행한다. 아래는 위의 알고리즘에 의해 문자열이 음소열로 즉, 소리나기 표현으로 변환된 예이다.

<예>

(변환 전)

할아버지와 할머니는 맛있는 것이 많이 생겼다.

(변환 후)

하라버지와 할머니는 마신는 거시 마니 생겡따.

2) 운율 기호 계열

음성언어의 운율은 단어의 액센트, 문장의 구조 또는 화자의 감정 상태등과 같은 언어학적 요소에 의해 변화하게 된다. 이러한 운율의 특징을 표현할 수 있는 음향학적인 요소로서는 기본 주파수(Fundamental frequency), 강세(Intensity) 또는 어절의 길이(Segmental duration)등이 있는데 이 가운데 기본 주파수는 음성 언어의 억양과 직접적인 관계가 있다.<sup>[17]</sup>

타 언어에서는 음운학적, 실험음성학적인 연구결과를 토대로 구문해석기법을 이용하여 입력 텍스트를 해석하므로서 음성합성을 위한 운율기호를 생성하는 처리 시스템이 구현되어 있으나<sup>[2], [3]</sup>, 우리말에는 이들에 대한 연구가 미흡한 실정이므로 자연스러운 한

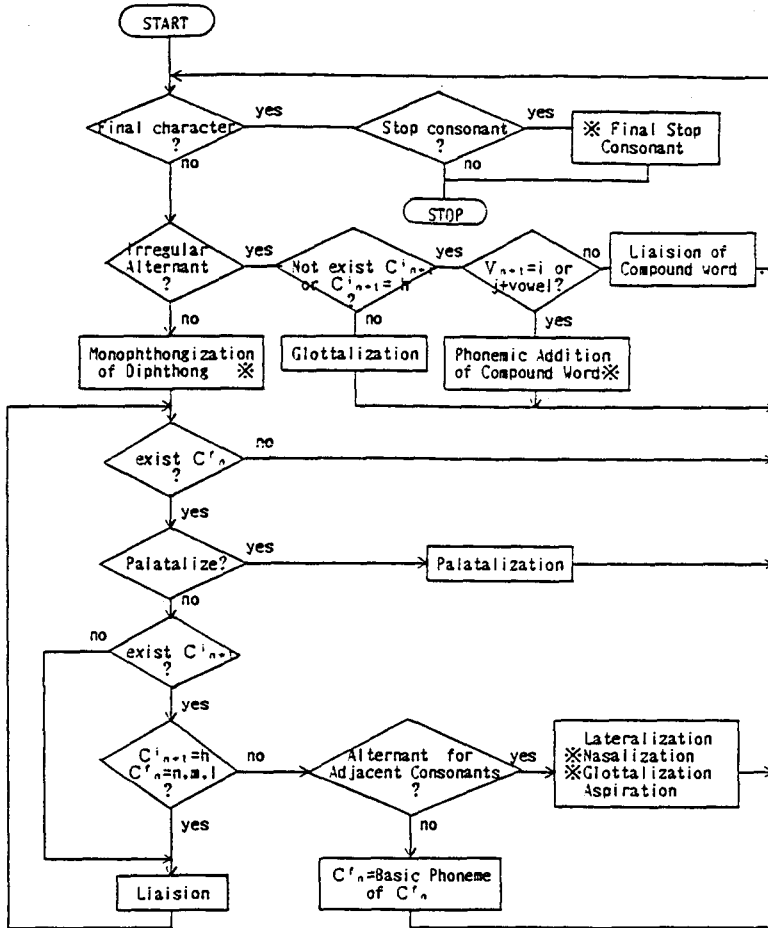


그림 4. 문자열을 음소열로 전환하는 알고리즘

국어 합성음 생성을 위해서는 한국어에 대한 운율 기호 생성 처리 시스템의 구현을 위한 기초 연구가 시급하다. 한국어의 경우 운율을 표현하기 위한 음향학적 요소로서 일반적으로 다음과 같은 것들이 있다.

① 기본 주파수

합성음에 자연성을 주는 운율요소중 기본주파수의 패턴은 가장 중요한 요소이다. 영어<sup>[2]</sup>와 일본어(후지사키 모델)<sup>[3]</sup>의 경우, 기본주파수 모델이 거의 확립되어 있는 상태이다. 문장에 대한 피치 패턴은 각각의 호흡 단락 그룹(Breath group) 단위로 만들어 지고, 미리 정의된 길이 만큼의 휴지가 호흡 단락들 사이에 삽입되어 피치 패턴이 만들어 진다.

일본어의 후지사키 모델은 피치 패턴의 생성 모델로서 시간 t의 함수인 다음의 식으로 정의하고 있다.<sup>[21]</sup>

$$\ln F_0 = \ln F_{\min} + A_{u1}G_u(t-T_0) + A_{u2}G_u(t-T_3) + A_aG_a(t-T_1) - G_a(t-T_2) \quad (1)$$

$G_u(t)$ 는 구 제어 지시(Phase control mechanism)의 임펄스 응답(Impulse response)함수이고는 액센트 제어 지시(Accent control mechanism)의 스텝 응답(Step response)함수이다.

여기서

- $F_{\min}$ : 액센트성분이 없을 때  $F_0$ 의 접근치
- $A_{u1}$ : positive utterance command의 크기
- $A_{u2}$ : negative utterance command의 크기
- $A_a$ : 액센트 명령의 진폭
- $T_0$ : positive utterance command의 timing
- $T_3$ : negative utterance command의 timing



$\alpha_{sc}(N)$ : 종성이 있는 N개의 단모음 음절 단어의 박동기

$\alpha_{lc}(N)$ : 종성이 있는 N개의 장모음 음절 단어의 박동기

N: 음절수, r: 상관계이다.

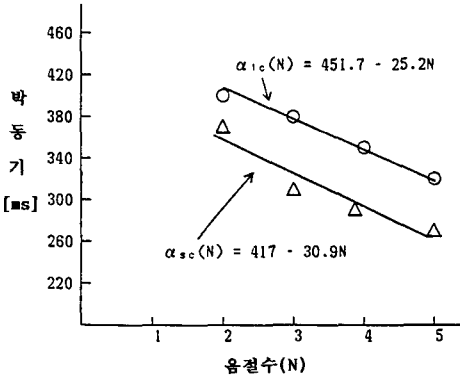


그림 7. VC, CVC음절(단모음 Δ 장모음 O)의 박동기( $\alpha_{sc}$ ,  $\alpha_{lc}$ )와 음절수(N)와의 관계

종성을 갖는 음절의 박동기에 있어서 종성이 차지하는 비율((후속음절모음개시점-종성개시점)/박동기)은 0.65(단모음), 0.47(장모음)이므로 종성의 길이를 다음과 같이 나타낼 수 있다.

$$\begin{aligned} \alpha_s^f(N) &= \alpha_{sc}(N) * 0.65 & (3) \\ \alpha_l^f(N) &= \alpha_{lc}(N) * 0.47 \end{aligned}$$

여기서,  $\alpha_s^f(N)$ 는 단모음 음절에 있어서 종성의 개시점부터 다음 음절의 모음 개시점까지의 길이이고  $\alpha_l^f(N)$ 는 장모음 음절에 있어서 종성의 개시점 부터 다음 음절의 모음 개시점 까지의 길이이다.

또, 최종 음절의 길이는 다음 (4)식으로 나타내었다.

$$\begin{aligned} \alpha_{s0}^1(N) &= 0.7 * \alpha_{s0}(N) \\ \alpha_{l0}^1(N) &= 0.6 * \alpha_{l0}(N) \\ \alpha_{sc}^1(N) &= 0.8 * \alpha_{sc}(N) \\ \alpha_{lc}^1(N) &= 0.7 * \alpha_{lc}(N) \end{aligned}$$

여기서  $\alpha_{s0}^1(N)$ 과  $\alpha_{l0}^1(N)$ 는 각각 최종 음절에 종성이 없는 단모음과 장모음 음절의 길이이고,  $\alpha_{sc}^1(N)$

과  $\alpha_{lc}^1(N)$ 는 각각 종성이 있는 단모음과 장모음 음절의 길이를 나타낸다.

(3) 휴지(Pause) 길이

단순한 묵음 구간 뿐만 아니라 말의 흐름이 끊긴다고 느껴지는 부분 또한 휴지라고 정의할 수 있다.

문절의 접속도에 따른 휴지 패턴은 자연음성의 분석을 통한 모델링에 의하여 합성음에 적용할 수 있다. 한국어 산문 낭독을 통한 휴지 분석의 연구 결과<sup>[26]</sup>, 묵음 구간 뿐만 아니라 길이 증가와 억양등도 휴지를 지각하게 하는 주요 요소이고, 하강조의 억양을 갖는 모든 문장과 단락 경계에서는 성질 변화가 관찰되며, 상위 경계로 올라갈수록 긴 묵음 구간, 하강 억양, 성질 변화, 길이 증가등으로 강하게 나타난다고 보고되었다.<sup>[26]</sup>

(4) 에너지 패턴

에너지는 음성의 인지도에 영향을 미치는 intensity를 나타내는 요소로서 스펙트럼 에너지나 파형의 크기를 합산하는 방법등으로 추정할 수 있다.<sup>[27]</sup> 문장내에서 에너지의 흐름을 에너지 윤곽선이라 하며 문장의 운율 성분과 관련이 있다. 음절의 에너지는 앞부분이 크며, 자음보다 모음이 크다는 경향이 있다.<sup>[27]</sup> 운율에 영향을 미치는 하나의 요소로서 에너지 패턴은 우리말의 액센트와 강세에 대한 분석을 행하여 모델화되어야 한다.

이상에 소개한 한국어의 운율패턴에 대한 연구에도 불구하고 아직은 미흡한 단계이므로, 더욱 섬세하게 제어할 수 있는 음운지속시간, 인토네이션, 휴지, 액센트등에 대한 물리적인 특성의 모델화에 대한 연구와, 구문과 운율기호와의 관계를 나타내기 위한 연구가 행해져야 한다.

## 2 음성 합성 기능

음성 합성 기능부에서는 변이음과 장단음과 같은 음운 변동의 처리와 길이, 에너지, 억양, 휴지등의 운율 패턴에 따라 단위 음성 파라미터를 접속하여 조음 파라미터와 음원 파라미터를 생성하고 이로서 시변 필터를 구동하여 합성음을 생성한다.

1) 음운 생성

성도의 변동을 나타내기 위한 규칙을 줄이기 위하여 음성의 단편을 미리 저장하며, 저장된 단편들을 접속하여 음성을 생성한다. 이를 위하여 최적의 단위 음성 선택 문제가 해결되어야 하고, 또 운율 파라미터의 수정이 음질에 영향을 미치지 않도록 하며, 접

속점에서 생기는 조음 파라미터의 불연속을 줄일 수 있고, 저장될 데이터 양을 효율적으로 감소 시킬 수 있는 기술이 요구된다.

단위 음성으로서 음소의 경우 데이터 양은 적으나 인접하는 음 사이의 매끄러운 천이의 실현이 어렵다. 음절의 경우 데이터 양이 많고 변이음 처리와 타이밍 제어가 어렵다. 다이폰(Diphone) 또는 반음절의 경우는 데이터 양이 음소보다 많으나 인접하는 음 사이의 포먼트 천이를 유지할 수 있으므로 접속점에서 불연속을 최소화할 수 있어 가장 많이 사용되고 있다. [8] [11] 한국어에 있어서 유성화하는 변이음의 경우, 전후음의 영향이 강하게 나타난다. 이러한 면에서 볼 때 변이음 처리를 위하여는 불연속성이 적고 포먼트의 천이를 유지하고 있는 다이폰이 좋은 합성 단위라 생각된다.

최근 연구에서는 연결문맥의 전후를 고려하여 접속 단위를 정의하는 COC(Context-Oriented Clustering) [4] [6] 와 CDU(Context-Dependent Unit) [18] 를 합성단위로 사용하여 합성음의 음질을 향상시키는데도 있다. 이 합성단위는 양질의 합성음을 생성하나 단위음성 데이터베이스의 크기가 커진다는 단점이 있다. 그러나 하드웨어의 발달로 이는 쉽게 극복할 수 있다.

한편, 선택한 단위음성의 데이터베이스 구축 시 단위 음성의 자동 작성은 매우 유용하다. 단위음성을 자동 작성할 경우 수작업에 비하여 균일한 데이터 작성이 가능하고 데이터베이스 구축 시간을 단축할 수 있으며, 다양한 화자의 데이터베이스 구축이 용이해진다는 장점을 갖는다. 독일어의 다이폰 자동 추출 [28], 일본어의 CV음절의 자동 작성법 [29] 과 한국어의 체계적인 반음절 작성법 [9] 과 다이폰 자동작성법 [30] 등이 제안되어 있으며 이에 의해 고품질의 음성 생성이 가능하다. 다이폰 자동 작성 연구 [30]에서는, 음성 시료로부터 켈스트럼 파라미터에 의한 특징 분석을 수행하여 이 파라미터의 시간적 변화량을 다이폰을 추출하기 위한 파라미터로 사용한다. 사용한 파라미터는 유성/무성 판별 파라미터, 에너지 레벨을 나타내는 0차 켈스트럼의 시간 변화량, 켈스트럼 고차항과 저차항에 의해 각각 얻어진 스펙트럼 포락의 시간 변화량, 영교차율과 켈스트럼의 유클리디안 거리이다. 위의 파라미터들을 사용하여 모음의 연쇄에서도 약 80%정도의 음소 경계 검출율을 얻었다. 다이폰 추출 파라미터와 다이폰 추출 결과는 그림 8과 같다.

실험음성시료는 남성화자가 발성한 /에베/이다.

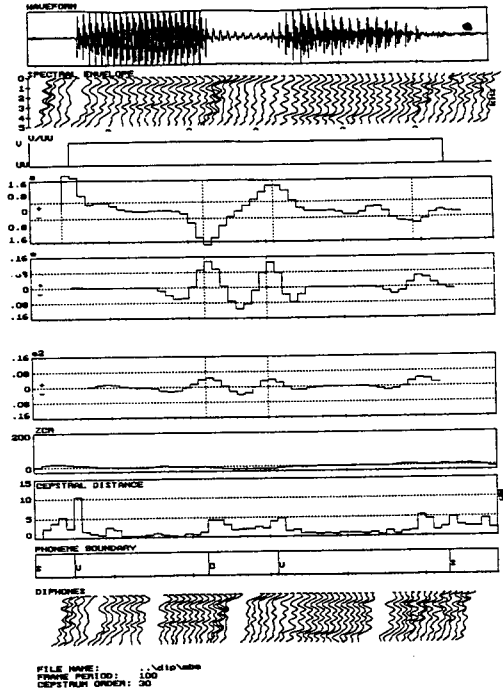


그림 8. 다이폰 추출 파라미터와 다이폰 추출 결과

단위 음성의 접속은 접속점에서 스펙트럼의 불연속성을 최소화할 수 있어야 한다. 독일어의 반음절 접속법으로서 파형의 정상 부분에서 2점 선형보간에 의한 접속방법과 한국어의 자음의 변이음에 있어서 전후의 음을 고려한 2점, 3점 선형보간 접속방법 [9] 등이 있다.

## 2) 운율 생성

자연 음성을 분석하여 모델링하고 이를 합성음에 운율로서 적용한다면 자연성을 살릴 수 있다. 합성음에 자연성을 부과하기 위한 운율 파라미터로서는 음원 패턴으로서 억양에 해당하는 피치 패턴, 리듬과 템포를 살리기 위한 타이밍 제어, 휴지(Pause) 길이 패턴과 에너지 패턴이 있다. 언어처리부에서 각 패턴에 맞게 생성된 운율기호열을 합성단위에 맞도록 조절하여 단위음성 접속 시 제어 파라미터로 사용한다.

## 3) 음성 신호 생성부

음성신호 생성부에서는 필터 계수가 조음 파라미터 계열인 시변 필터를 음원 파라미터계열로 구동하여

음성파형을 생성한다. 음성변환 방식의 선택은 시스템 전체의 각 부분과 밀접한 관계가 있으므로 합성음의 음질을 좌우한다고 할 수 있다. 일반적인 음성 변환방식은 크게 2가지로서 포만트(Formant)합성방식과 분석합성(보코더)방식이 있다.

포만트 합성방식은 순수 규칙합성 방식으로서 성도의 변화/필터특성은 각 음소 그리고 음소간의 포만트변화를 나타내는 규칙을 사용하여 기술한다.<sup>[2]</sup> 영어에 대해 개발된 Klatt 합성기는 상용화되어 쓰이고 있으나, 한국어 음성합성에 포만트 합성방식을 적용한 연구는 아직 우리말에 대해 충분히 분석하여 얻어진 데이터가 적기 때문에 한정된 단어를 합성하는 정도이다.<sup>[11], [13]</sup>

음성 분석합성계를 이용한 규칙 합성방식은 비교적 간단한 시스템으로 좋은 음성을 합성할 수 있으며, LPC계열의 PARCOR 분석합성, LSP 분석합성과 캡스트럼 분석합성 시스템 등이 있다.<sup>[17]</sup> 한국어 합성시스템으로 LPC계열을 이용한 시스템<sup>[10], [14], [28], [32]</sup>과 캡스트럼 방식을 이용한 시스템이 개발되어 있다.<sup>[9]</sup>

PARCOR 분석합성 시스템은 LPC 분석에 의해 얻어지는 스펙트럼 포락이 불안정 하기때문에 편자기상관함수(partial auto-correlation)를 이용하여 파라미터를 추정하는 방법이다.<sup>[17]</sup>

LSP 분석합성 시스템은 PARCOR 분석합성 시스템의 변형으로서 스펙트럼 포락의 파라미터는 주파수 영역의 파라미터이며, 시간적 보간 특성과 양자화 특성이 좋다.<sup>[1]</sup> LPC계열의 시스템들은 전극형(all-pole)모델인데 비해, LMA 필터를 이용하는 캡스트럼 보코더<sup>[17]</sup>는 극영모델(Pole-zero)로 zero성분도 포함하는 음성의 실현 가능성이 크다.

캡스트럼 보코더에서는 스펙트럼포락이 캡스트럼으로 표현되어 그 값이 그대로 필터계수로 이용된다. 또한 양자화 특성과 시간적인 보간특성이 매우 양호하므로 낮은 bit rate로 고품질의 음성을 합성할 수 있다. 음성의 대수진폭에 있어서 스펙트럼 포락  $H'(Ω)$ 는 최소위상 캡스트럼  $c(m)$ 에 의해 주어진다. 전달함수  $H(z)$ 가 식 (5)로 주어진 필터라면,

$$H(z) = \exp\left(\sum_{m=0}^M c(m)z^{-m}\right) \quad (5)$$

대수 진폭특성  $\ln H(ej)$  는 식 (6)과 같게 되어 대수 진폭에 있어서의 음성 스펙트럼과 같게 된다.

$$\begin{aligned} \ln|H(z)\exp(j\Omega) &= \operatorname{Re}\left(\sum_{m=0}^M c(m)\exp(jm\Omega)\right) \\ &= \sum_{m=0}^M c(m)\cos(m\Omega) \end{aligned} \quad (6)$$

전달함수인 지수함수  $\exp(W)$ 는 Pade' 근사식을 이용하여 전달함수를 근사시켜 필터를 설계하는데 이와같은 필터를 대수진폭근사(Log Magnitude Approximation, LMA)필터라 한다. 전달함수에 의한 LMA 필터의 근사 오차는 실제의 음성 캡스트럼  $c(m)$ 에 대해 0.2dB이하 이다. 캡스트럼은 양호한 양자화 특성, 보간특성을 갖는 대신 LPC파라미터 등에 비해 차수가 높는데, 멜캡스트럼을 이용하면 LPC정도로 낮아진다. 이때, data rate는 통상의 캡스트럼 보코더의 60-70% 정도로 된다. 그림 9는 캡스트럼 보코더에 의한 여성화자가 발성한 /동덕여자대학교/ 중 /대학/의 부분 파형으로, 상단의 파형은 분석합성음의 파형이고, 하단의 파형은 합성 단위로

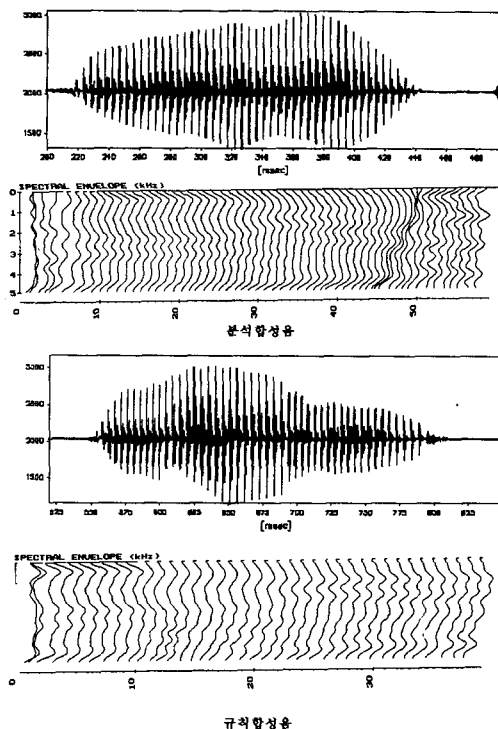


그림 9. 캡스트럼 보코더에 의한 /동덕여자대학교/의 분석합성음 및 규칙합성음 중 /대학/부분의 파형 및 스펙트럼 포락



다이폰을 사용하여 규칙합성한 같은 부분의 파형이다. 이를 보면, 다이폰 합성단위를 사용할 경우 모음의 정상부분에서 접속하므로 매우 자연스러운 천이구간을 갖음을 알 수 있다. 또한 캡스트럼 보코더의 사용으로 양질의 분석합성음을 얻음을 확인하였다.

자연스럽고 명료한 음성을 합성하기 위해, 규칙합성 시스템에 이용할 보코더의 선택은 먼저, 원 음성의 기본주파수와 다른 음원에 대해서도 합성음성의 왜곡이 작아야 하며, 스펙트럼 포락 파라미터의 시간적인 보간에 있어서 스펙트럼의 왜곡이 적어야 한다는 것과 음성의 기본 단위의 파라미터를 쉽게 추출할 수 있고, 규칙합성을 위한 단위음성 데이터 파일의 작성과 변경이 용이해야 한다는 점을 고려하여야 한다. 또한 분석계를 이용한 합성계의 경우 스펙트럼 제어는 고정된 단위 음성에서 벗어나 발성환경 즉, 음운 환경에 따른 단위음성을 사용하는 합성법이 고려된다면 좋은 음성을 생성할 수 있다.

#### IV. 결론

본 고에서는 한국어 규칙합성 시스템의 개요 및 기술과 규칙 합성 시스템의 부분 시스템인 언어 처리 기능과 음성 합성 기능을 소개하였다. 자연스러운 음성을 합성하기 위해서는 먼저 음성 신호 생성기술 뿐만 아니라 운율 처리에 대한 연구가 자연 음성의 분석을 바탕으로 하여 이루어져야 한다.

자연스러운 합성음을 생성하기 위하여, 운율기호로부터 운율패턴 즉 피치 패턴과 음운지속시간, 에너지 패턴등을 생성하는 모델에 관한 기초적인 연구가 많이 수행되어야 할 것이다. 또한 텍스트로부터 운율, 언어정보를 구하는 문제로서 기본주파수, 구경계, 액센트 위치등과 말의 부분, 어형변화의 타입과 형태, 구의존 구조등에 대한 연구 역시 많이 수행되어야 할 것이다. 이는 신호 처리, 실험 음성학과 언어 처리등 제 분야의 협조 체제를 절실하게 요구한다고 할 것이다.

#### 參 考 文 獻

[ 1 ] S.Saito, K.Nagata, Fundamentals of

speech signal processing. Academic press, 1985.

[ 2 ] D. H. Klatt, "Structure of phonological rule component for synthesis-by-rule program". IEEE vol. ASSP-24 no.5, pp. 391-398, 1976.

[ 3 ] K. Hirose, M. Yamaguchi, "Speech synthesis of sentence based on a model of fundamental Frequency contour generation", EIC vol. J72-A no.1, pp.32-40, 1989.

[ 4 ] K. Takeda 外, "選拓的に合成單位を用いる規則音聲合成", 電子情報通信學會論文誌(D-II), vol. J73 D-II no.12, pp. 1945-1951, 1990年 12月.

[ 5 ] Y. Sagisaka 外, "Concatenative Speech Synthesis By Minimum Distortion Criteria", IEEE. Trans. ASSP, pp.65-68, 1992.

[ 6 ] I. Satoshi, "Research Trends of Techniques of Speech Signal Processing in Japan", 제9회 음성 통신 및 신호처리 워크샵 논문집, pp.27-32, 1992.

[ 7 ] 이 양희, "한국어 음성의 규칙합성에 있어서 음절계속시간제어", 음향학회 음성통신 및 신호처리 workshop 논문집, pp118-123, 1989.

[ 8 ] Y. H. Lee and S. Imai, "Extraction and Concatenation of Demisyllables for Allphones in Korean Speech Synthesis-by-Rule", Trans. IEICE vol. J71-D no.6 pp.1020-1027, 1988.

[ 9 ] 李 亮熙, 半音節のメルケプストラムパラメータを用いる韓國語音聲の規則合成の研究, 東京工業大學博士論文, 1988.

[ 10 ] 이 재홍외, "반음절 단위 데이터베이스를 사용한 한국어 무제한 단위합성기의 제작", 제4회 신호 처리 합동 학술대회, pp.297-301, 1991.

[ 11 ] 이 정석외, "한국어 text-to-speech 합성기의 구현", 1992년도 한국정보과학회 봄 학술발표 논문집 vol.19 no.1 pp.107-110, 1992.

[ 12 ] N. K. Ha and S. R. Kim, "Development of Korean text-to-speech system", Korea-Japan joint Symposium on

- Acoustics, pp.261-267, 1991.
- [13] 조 철우외, "한국어 파열음의 규칙 합성을 위한 파라미터 추정에 관하여", 제1회 신호처리 합동 workshop 논문집, pp.51-54, 1988.
- [14] J. C. Lee et al, "A study of Text-to-speech synthesis based on LSP representations", Proceedings of Speech Communication and signal processing, pp.95-98, 1990.
- [15] 이 양희, "음성합성 기술 개발 현황 및 전망", 제9회 음성 통신 및 신호처리 워크샵 논문집, pp.88-93, 1992.
- [16] 이 용주, "음성합성 기술의 현황과 과제", Korea-Japan joint Symposium on Acoustics, pp.199-205, 1991.
- [17] Edited by Shuzo Saito, Speech Science and Technology, IOS Press, 1992.
- [18] 지민제의, "글소리 II 에서의 신호처리: PSOLA 합성방식", 제1회 ETRI 음성, 언어 및 음향정보처리 워크샵 논문집, pp.91-96, 1993.
- [19] D.O 'Shaughnessy, Speech Communication: Human and machine, Addison wesley press, 1990.
- [20] 이희승외, 한글맞춤법강의, pp.234-235 250, 신구문화사, 1991.
- [21] 이기문의, 국어음운론, pp.43-44, 학연사, 1991.
- [22] H. Fujisaki and K.Hirose, "Comparison of acoustic features of word accent in English and Japanese", J. Acoust.Soc. Jpn(E)7.1 pp.57-63, 1986.
- [23] 김 진영외, "한국어 억양에 관한 연구", Korea-Japan joint Symposium on Acoustics, pp.292-297, 1991.
- [24] 고 도홍, "음성합성 시 prosody의 처리: 한국어의 intonation을 중심으로", 음향학회 1990년도 음성통신 및 신호처리 논문집, pp.183-187, 1990.
- [25] 勾坂芳典, "音韻持續時間の制御と知覚", 日本音響學會, 1981年 8月.
- [26] 星野雅孝, "單語中における單語の持續時間の決定のための定量的モデル", 日本音響學會, 1982.
- [27] 지 민제의, "한국어 Pause Pattern의 음향음성학적 분석", 음향학회 1990년도 음성통신 및 신호 처리 논문집, pp.169-171, 1990
- [28] 이승훈, 한민수 "Audiotex 음질 개선에 관한 연구", 제 1회 ETRI 음성, 언어 및 음향정보처리 워크샵 논문집, pp.106-114, 1993.
- [29] H.Kaeslin, "A systematic approach to the extraction of diphone elements from natural speech", IEEE. Trans. ASSP, Vol. ASSP-34 NO.2 pp.264-271, 1986.
- [30] 古市, 今井, "音聲の規則合成のためのメルケプストラム CV音節データの自動作成", 電子情報通信 學會論文誌(D), j68-D.9, pp.1664-1672, 1985.
- [31] 정 인종의, "고품질의 한국어 합성 방식에 관한 연구", 제 1 회 ETRI 음성, 언어 및 음향정보처리 워크샵 논문집, pp.128-142, 1993.
- [32] 구준모외, "한국어 무제한 음성합성 시스템 : 가라사대", 제9회 음성통신 및 신호처리 워크샵 논문집, pp.201, 1992. ㉸

筆者紹介



李 亮 熙

1948年 10月 5日生

1976年 2月 동국대학교 전자공학과 졸업(학사)

1978年 2月 한양대학교 전자공학과 대학원(석사)

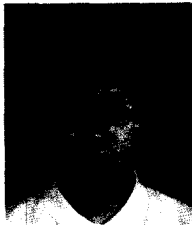
1984年 3月 동경공업대학 대학원 정보공학과 졸업(석사)

1988年 3月 동경공업대학 대학원 정보공학 전공(박사)

1988年 3月 ~ 1992年 2月 동덕여자대학교 전자계산학과 조교수

1992年 3月 ~ 현재 동덕여자대학교 전자계산학과 부교수

주관심분야: 음성정보처리, 자연언어처리



鄭 麟 淙

1968年 8月 2日生

1991年 2月 한양대학교 전자계산학과 졸업(학사)

1991年 8月 한양대학교 전자계산학과 대학원 입학

주관심분야: 음성합성, 음성인식



景 淵 靖

1970年 10月 1日生

1992年 2月 동덕여자대학교 전자계산학과 졸업(학사)

1992年 3月 동덕여자대학교 전자계산학과 대학원 입학

주관심분야: 음성합성, 자연언어처리, 음성인식