

# 포만트 합성방식을 이용한 문자-음성 변환에 관한 연구

## A Study on the Text-to-Speech Conversion Using the Formant Synthesis Method

최진산\*, 김민년\*\*, 서정욱\*\*\*, 배건성\*\*\*\*

(Jin San Choi, Yin Nyun Kim, Jeong Wook See, geun Sune Bae)

### ABSTRACT

Through iterative analysis and synthesis experiments on Korean monosyllables, the Korean text-to-speech system was implemented using the phoneme-based formant synthesis method. Since the formants of initial and final consonants in this system showed many variations depending on the medial vowels, the database for each phoneme was made up of formants depending on the medial vowels as well as duration information of transition region. These techniques were needed to improve the intelligibility of synthetic speech. This paper investigates also methods of concatenating the synthesis units to improve the quality of synthetic speech.

**Keywords:** formant synthesis, transition, conversion

### 1. 서론

멀티미디어가 급속도로 보급되고 있는 오늘날에는 인간 사이의 정보전달 뿐만 아니라 인간과 기계와의 정보전달도 중요한 상황이 되었다. 이에 따라 인간의 가장 자연스러운 정보전달 수단인 음성을 인간과 기계와의 정보전달 수단(man-machine Interface)에 이용하려는 연구가 계속되어 왔으며 이러한 연구는 컴퓨터의 발달과 더불어 음성합성, 음성 인식 등의 여러 분야에서 빠른 속도로 진행되고 있다[1,2]. 기계에서 처리된 내용을 인간의 목소리로 출력해 주는 음성합성은 합성할 수 있는 어휘의 범위에 따라 제한 어휘 합성과 무제한 어휘 합성으로 분류된다. 제한 어휘 합성은 합성하고자 하는 어휘들을 미리 분석하여 데이터베이스에 저장하였다가 이들의 조합에 의해 음성을 합성하는 방식으로, 현재 전화번호 안내, 은행잔고 확인, 지하철 안내방송 등에 쓰이고 있다 이 방식은 비교적 양호한 음질의 합성음을 얻을 수 있고 구현이 용이한

---

\* KBS안동방송국

\*\* LG전자기술원

\*\*\*경북대학교 전자·전기 공학부

장점이 있으나 합성 가능한 어휘가 한정적이라는 문제점이 있다. 무제한 어휘 합성은 음소, 반음소, 반음절 또는 음절 등을 기본적인 단위로 하는 합성단위의 데이터베이스를 구성하여 규칙에 의해 기본 단위들을 조합하여 음성을 만들어내는 방식이다. 무제한 어휘 합성기는 LSP(Line Spectrum Pair) 합성기, 포맷트 합성기, PSOLA(Pitch Synchronous Overlap and Add) 합성기 등이 있는데, 이들은 문자-음성 변환(TTS: Text-to-Speech) 시스템에 적용되며 합성음의 음질은 제한 어휘 합성 방식에 비해 떨어지나 임의의 문장을 합성할 수 있고 파라미터의 변화로 다양한 합성음을 만들 수 있는 장점이 있어서 국내외에서 많은 연구가 진행되어 왔다[3-6]. 포맷트 합성방식은 현재로는 LSP 합성방식에 비해 데이터베이스 구축에 많은 시간과 노력이 요구되며, PSOLA 합성방식에 비해 합성음의 음질이 다소 떨어진다는 단점이 있다 그러나, 합성 파라미터의 변환이 쉽고 파라미터 변환에 따른 다양한 형태의 합성음을 얻는게 용이하므로 멀티미디어 컴퓨터에서 적은 양의 데이터베이스로 다양한 음성 서비스가 가능한 포맷트 합성방식에 대한 꾸준한 연구가 요구된다.

본 논문에서는 포맷트 합성방식을 이용한 음소단위의 한국어 문자-음성 변환시스템 구현에 대하여 연구하였다. 합성음의 명료도를 높이기 위해 한국어 단음절의 초성과 중성에 대한 포맷트 궤적을 후속모음과 앞선모음에 따라 분석한 후, 분석된 결과를 포맷트 합성기에 적용시켜 반복적인 합성/청취 실험을 통해 천이구간에서의 길이정보와 함께 음소단위의 데이터베이스를 구축하였으며, 명료도를 확보하기 어려운 초성과 중성의 음소에 대해서는 후속모음이나 앞선모음에 따라 음소의 포맷트 값을 따로 구하였다. 또한, 합성음의 자연성을 개선하기 위해 초성과 중성, 중성과 중성사이의 합성단위의 접속 방법에 대해 실험하였는데, 포맷트 접속시에 선형 및 비선형의 연결방식에 따른 합성음의 음질을 비교하였다. 실험에 사용된 포맷트 합성기는 Klatt가 제안한 직/병렬 합성기를 이용하였으며[7], 운율정보는 후지사키 모델[8]을 수정하여 사용하였다. 본 논문의 구성은 다음과 같다. 먼저, 2장에서는 합성기를 구현하기 위한 음성 분석방법과 합성 파라미터의 데이터베이스의 구축방법에 대해 설명하고, 3장에서는 포맷트 방식을 이용한 한국어의 규칙 합성 방법에 대해 기술한다. 4장에서는 구현된 포맷트 합성기를 통한 합성 실험 및 결과에 대해 검토하고, 마지막으로 5장에서 결론을 맺는다.

## 2. 음성분석 및 합성 파라미터의 DB 구축

### 2.1 음성분석 및 단음절의 모델링

본 연구에서 분석의 대상으로 삼은 음절은 모음만으로 구성된 21개의 음절과, 초성과 중성의 결합으로 이를 수 있는 모든 음절, 그리고 중성과 중성이 결합되어 이를 수 있는 모든 음절로 하였다. 분석에 사용된 음성은 20대 남성 화자로부터 음절간에 충분한 간격을 두고 두 번씩 반복하여 녹음하였으며, 10 kHz로 샘플링하였다. 수집된 음성은 수작업을 통해 음절 단위로 자른 후 데이터베이스의 구성을 위해 포맷트를 추정하였다. Covariance 방법을 사용하여 얻어진 LPC 계수를 사용하여 root solving 방식으로 포맷트 주파수, 대역폭, 크기를 얻었으며, 포맷트

주파수가 잘못 구해진 부분에 대해서는 수작업으로 수정하였다. 모음의 포맷트 주파수와 대역폭은 root solving 방식으로 비교적 안정된 값을 구할 수 있었으며 합성을 위한 데이터베이스로 포맷트 궤적의 안정된 구간에서의 F1부터 F5까지 5개의 포맷트 값을 저장하였다. 이중모음의 경우는 음소의 시작 부분에서의 포맷트 값 5개, 음소의 안정된 구간에서의 포맷트 값 5개 그리고 이중모음 내에서의 포맷트 천이구간의 길이를 합성을 위한 데이터베이스에 저장하였다.

모음 합성시에는 직렬 합성기만을 사용할 것이므로 병렬 합성기의 파라미터로 쓰이는 각 포맷트의 크기는 따로 구하지 않았다. 자음의 포맷트 주파수는 후속 모음 또는 앞선 모음에 따라 따로 수집한 자음을 포함하는 단음절에 대해서 LPC 분석을 거친 후 root Solving 방식으로 F1부터 F5까지 5개의 포맷트값을 구해서 데이터베이스로 저장하였다. 종성의 포맷트 주파수는 root solving 방식으로 비교적 안정된 궤적을 얻을 수 있었으나, 초성의 경우 비음 /ㄱ/, /ㄴ/, /과 유음 /ㄹ/을 제외한 나머지 음소들은 root solving 방식으로 안정된 포맷트 궤적을 얻을 수 없어서 오차가 발생한 부분에 대해서는 수작업으로 수정하였다. 한편, 포맷트의 크기와 대역폭은 root solving 방식으로 안정된 값을 구하기가 어려우므로 합성음의 청취 결과와 합성음의 스펙트럼을 원음의 스펙트럼과 비교하는 등의 반복적인 실험을 통해 적절한 값을 구하였다. 데이터베이스에 저장할 포맷트의 대역폭은 후속모음 또는 앞선모음에 따라 따로 구분하지 않고 F1부터 F5의 각 포맷트에 대해서 음소당 하나만 구했고, 포맷트 진폭값은 반복적인 청취-분석 실험을 통해 후속모음 또는 앞선모음에 따라 A1에서 A5까지 각 포맷트 진폭값을 데이터베이스로 저장하였다.

그림 1은 '초성-중성'의 형태를 가지는 음성신호 /가/에 대해서 분석 후 수정을 거친 포맷트 궤적인데, 이 그림과 같이 각 음소들에는 포맷트 주파수의 안정된 구간과 천이구간이 각각 존재한다. 천이구간은 주로 음소의 경계부분에 존재하고, 안정된 구간은 주로 음소의 가운데에 위치하며 경우에 따라서는 안정된 구간이 존재하지 않을 수도 있다.

음소 내에서의 포맷트 궤적의 안정된 구간과 천이구간을 모델링하기 위해서 각 음소를 그림 2와 같이 포맷트의 앞부분 천이구간(T1), 정상상태 구간(T2), 뒷부분 천이구간(T3)의 세 구간으로 나누었고[9], 이에 따라 각 음소사이의 포맷트 궤적의 천이구간을 앞/뒤로 이웃하는 음소에 따라 따로 구했다. 또한, 각 음소의 지속시간은 주위 음소환경에 따라 변화하도록 했으나 천이구간의 길이는 일정하게 두었다.

## 2.2 음소간의 합성 파라미터 궤적

음성합성을 위해서는 데이터베이스에서 음소 단위로 얻은 포맷트 정보로부터 합성 프레임 단위의 포맷트 정보를 구해야 한다. 본 연구에서는 음소간의 포맷트 궤적을 선 형적으로 구성하였다. 그림 3에 '초성+단모음+중성'의 형태를 갖는 단음절의 포맷트 궤적을 구성하는 규칙을 보였다. 이때 초성의 앞부분과 종성의 뒷부분은 휴지(pause)로 가정하여 초성의 T1, 종성의 T3 값을 0으로 두었다. 한편 후속모음이 이중모음일 경우에 단모음에서 포맷트 궤적을 구성하는 규칙을 그대로 적용시키면 포맷트 궤적의 불연속을 가져와서 합성음 음질이 저하된다. 따라서 후

속 모음이 이중모음일 경우는 중성의 앞부분 천이구간을 0으로 두고 초성의 포먼트 타겟값을 이중모음 뒷부분의 정상상태 값으로 정함으로써 포먼트 궤적이 급격히 변하지 않도록 하였다.

그림 1. 음성신호 /가/의 파형 및 포먼트 궤적

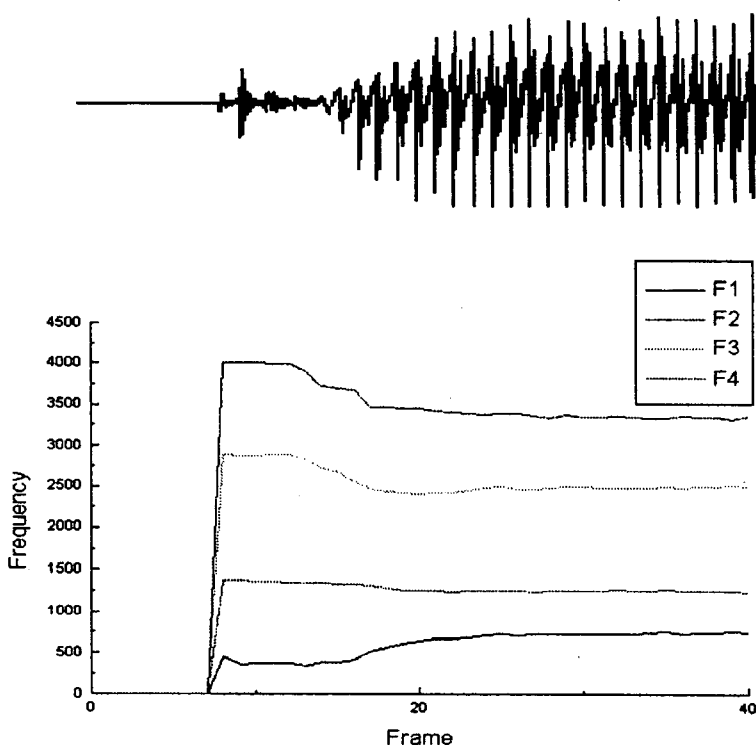


그림 2. 포먼트 궤적 변화에 따른 음소의 분할

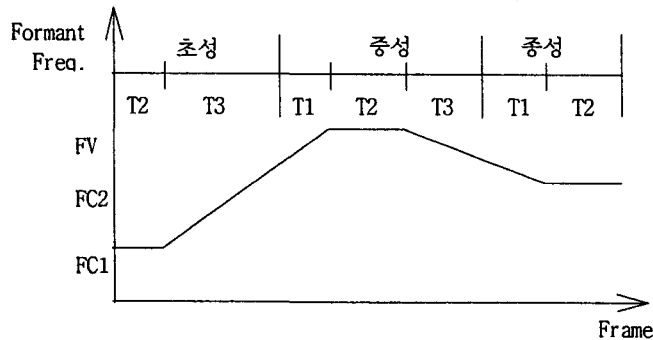
앞부분 천이구간 (T1)	정상상태 구간 (T2)	뒷부분 천이구간 (T3)
------------------	-----------------	------------------

$$T1 + T2 + T3 = T(\text{음소의 지속시간})$$

그림 4(a)는 '초성-중성'의 형태를 갖는 단음절에서 여기원 크기의 궤적을 구성하는 방법을 나타낸 것이다 이때 초성의 앞부분과 중성의 뒷부분은 휴지로 가정하였으며 초성은 무성음 여기원인 AF만을, 중성은 유성음 여기원인 AV만을 사용하는 경우에 해당된다. 초성과 중성이 같은 여기원을 사용하고 있을 때의 여기원 크기의 궤적을 구하는 규칙은 그림 4(b)에 나타내었다. 두 음소의 여기원이 겹치는 부분은 두 음소에 의한 여기원의 값 중 크기가 큰 값을 택했다. 한

편, 여기원의 값을 음소 내에서 직선으로 두었기 때문에 이득 궤적으로 음성의 파형을 성형하는 과정이 필요하다. 현재 이득 궤적은 비선형 평활화 방식을 일부 수정하여 사용하고 있다 [9,10]. 그러나, 실제 음성 합성 결과 파형의 모양이 합성음의 음질에 미치는 영향은 작음을 알 수 있었다.

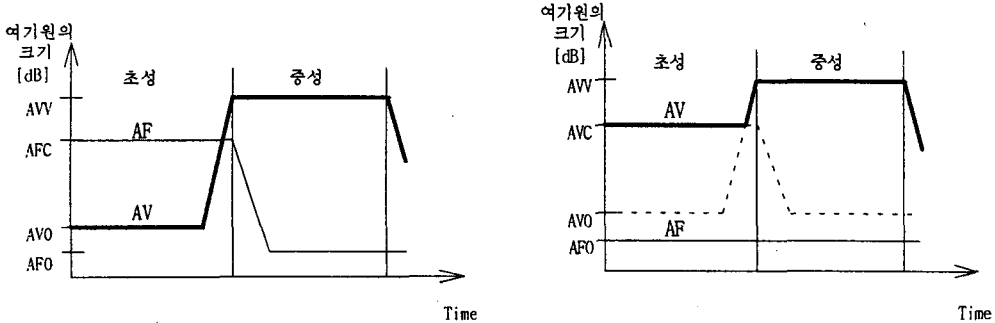
그림 3. 단음절의 포먼트 궤적 구성



FC1 : 초성포먼트값, FC2 : 중성포먼트값, FV : 중성포먼트값

병렬 합성기의 경우 포먼트 주파수와 대역폭 뿐만 아니라 포먼트의 진폭(A1~A5)궤적을 구하는 과정이 필요하다. 데이터베이스 내에 후속 모음이나 앞선 모음에 따라 구해진 자음의 각 포먼트의 진폭값(A1~A5)들로부터 음소 전체의 포먼트 진폭의 궤적을 구하는 과정은 여기원의 궤적을 구하는 과정과 거의 동일하며 해당 음소 내에서는 각 포먼트의 진폭값이 일정하도록 하였다. 단, 이 방법을 사용할 경우 거센소리나 마찰음 등이 인접하는 모음과 연결될 때 포먼트 진폭값의 불연속이 발생할 수 있다. 본 연구에서는 이와 같음소에 대해서는 예외 규정을 두기 위해 음소 단위의 데이터베이스 내에 음소의 포먼트 진폭값을 시간적으로 두 구간으로 분할하여 시작 값과 포먼트 천이구간을 정의함으로써 인접 음소와의 경계에서 포먼트의 진폭 궤적이 연속적으로 변할 수 있도록 하였다. 모음의 경우는 직렬 합성기를 사용하기 때문에 포먼트의 진폭값은 포먼트의 주파수와 대역폭에 의해 결정되므로 인접 음소 또는 인접 휴지와 경계에서의 포먼트 진폭의 불연속은 자음의 경우 보다는 작다. 그러나 해당 음소 내에서는 포먼트 진폭값들의 비가 이득 궤적과는 무관하게 전체적으로 일정하기 때문에 합성음의 부자연성을 가져온다. 이를 해결하기 위한 방안으로 모음의 앞부분에서는 고차의 포먼트 진폭 궤적이 서서히 증가되고 반대로 모음의 뒷부분에서는 고차의 포먼트 진폭값이 서서히 감소되도록 하였다.

그림 4. 단음절의 여기원 궤적 구성



(a) 인접 음소의 여기원이 서로 다른 경우

(b) 인접 음소의 여기원이 같은 경우

2.3 데이터베이스의 구성

분석된 결과를 구현된 직/병렬 포맷트 합성 시스템에 적용시켜 얻어진 합성음을 청취한 후 포맷트 주파수, 대역폭, 천이구간, 포맷트의 진폭값 등을 변화시켜 가면서 데이터베이스를 갱신하였다. 최종적으로 구해진 데이터베이스에서 자음의 경우 수작업을 통해 얻어진 파라미터의 값들이 분석 결과와는 다소 차이를 보였지만 합성시에는 더 명료한 음성을 얻을 수 있었다. 이는 포맷트 추정시 발생한 오류와 비선형적으로 변하는 포맷트 주파수, 대역폭 등의 여러 파라미터들을 선형적으로 모델링 하는 과정에서 발생된 것으로 여겨진다. 데이터베이스의 주요 구성은 아래와 같다

- 한국어 음소 48개에 대한 Klatt 직/병렬 합성기에 사용되는 40개의 파라미터
- 후속 모음에 따른 초성의 포맷트 주파수, 천이영역, 그리고 포맷트 진폭값
- 앞선 모음에 따른 중성의 포맷트 주파수, 천이영역, 그리고 포맷트 진폭값
- 이중모음 뒷부분의 데이터베이스
- 격음과 경음의 기본주파수 변화율

3. 포맷트 방식에 의한 규칙합성

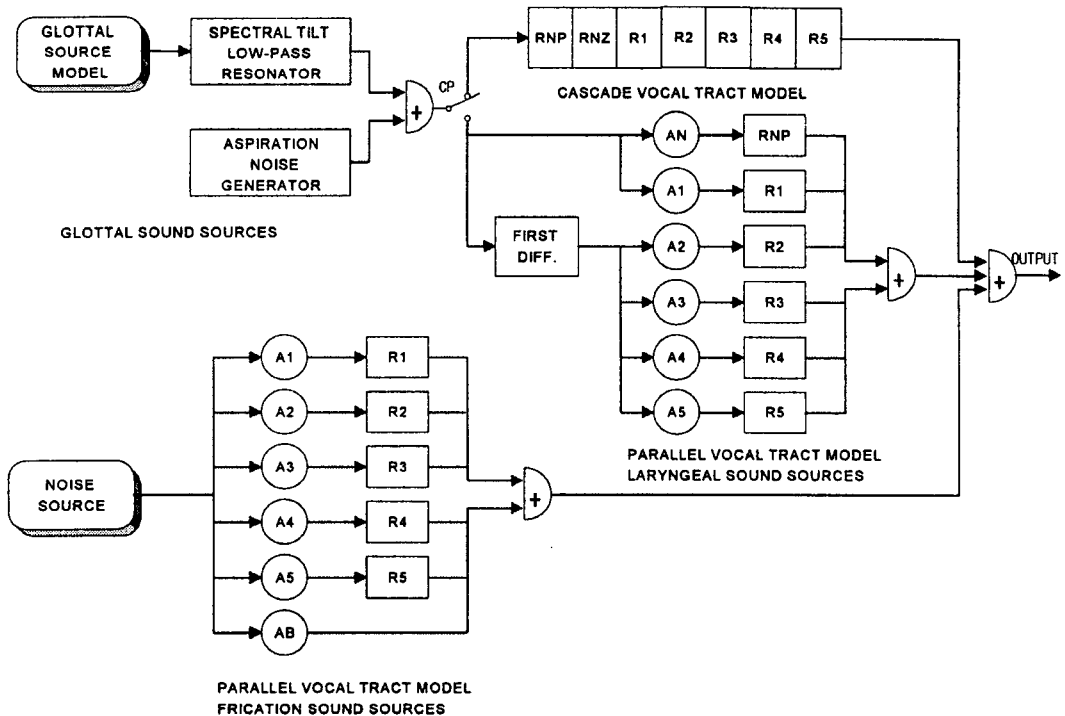
3.1 직/병렬 포맷트 합성 시스템

사람의 음성발생 과정을 모델링할 때 성도는 여러 가지 단면적을 갖는 관으로 모델링할 수 있는데, 각각의 관은 단면적에 따라 포맷트 주파수라 불리는 고유의 공진 주파수를 갖는다. 이 때, 개개의 공진 특성은 2차 디지털 공진회로로 모델링할 수 있으며 공진 회로를 직렬 또는 병렬로 연결하여 성도를 모델링 한다. 본 연구에서는 Klatt의 직/병렬 포맷트 합성기를 사용하여 음성합성을 하였는데[5], 합성기의 구조를 그림 5에 도시하였고, 표 1은 합성기에 사용되는 파라미터들을 나타내었다.

3.2 문장-음소 변환을 위한 전처리과정

자연스런 음성합성을 위해서는 운율처리 과정이 중요한 역할을 하며, 운율은 구문론적인 문장구조의 정보에 의해서 만들어진다. 본 연구에서는 음성학적 전처리과정을 문자열 정형화부, 음운변동 처리부의 두 부분으로 나누었다. 문자열 정형화부는 입력된 문자열의 숫자, 특수기호 및 영어 약자를 발음 가능한 한국어로 바꾸는 부분으로 입력문자가 한국어이면 아무런 처리를 하지 않고 그 밖의 경우에는 영어약자, 숫자열, 특수기호로 입력문자열을 분할하여 각각을 처리한 후 한글로 바꾸어 주어 다시 한국어 문자열로 합치게 된다. 음운변동 처리부에서는 한글 맞춤법 통일안에 맞게 작성된 것을 기준으로 한국어의 여러 음운변동을 고려하여 소리나는 데로 한국어 문자열을 바꾸어 주는 부분으로써 현재 고려하고 있는 음운변동으로는 받침의 발음법칙, 소리의 동화, 된소리 되기, 소리의 첨가 등이다[11]. 이때 규칙변동이 아닌 경우를 예외로 정하고 이를 예외사전으로 구성하였다. 입력된 문장을 단어별로 잘라낸 뒤 각 단어에서 “을, 를, 이, 가”라는 기본적인 조사를 떼어내고 단어의 첫 자음별로 기록된 예외사전을 검색하게 된다. 본 연구에서는 예외사전의 크기를 약 3,000단어 정도로 구성하였으며 예외사전으로 전환된 단어는 규칙변동을 하지 않도록 하였다[12].

그림 5. 포맷트 합성기의 구성



## 3.3 선형/비선형 연결에 의한 음소간의 포먼트 접속

문자-음성 변환의 규칙합성에서 합성단위를 음소로 할 경우 음소간의 천이영역에서 포먼트 주파수 및 대역폭의 궤적 형태는 합성음의 자연성 및 명료성에 영향을 미칠 수 있다. 음성신호의 분석/합성 과정에서 볼 수 있듯이 음소간의 포먼트 궤적은 갑작스럽게 변하지 않고 시간에

표 1. 포먼트 합성기의 제어 파라미터

Symbol	Name
F1	First formant frequency (Hz)
F2	Second formant frequency (Hz)
F3	Third formant frequency (Hz)
F4	Fourth formant frequency (Hz)
F5	Fifth formant frequency (Hz)
FNP	Nasal pole frequency (Hz)
FNZ	Nasal zero frequency (Hz)
B1	First formant bandwidth (Hz)
B2	Second formant bandwidth (Hz)
B3	Third formant bandwidth (Hz)
B4	Fourth formant bandwidth (Hz)
B5	Fifth formant bandwidth (Hz)
BNP	Nasal pole bandwidth (Hz)
BNZ	Nasal zero bandwidth (Hz)
A1	First formant amplitude (dB)
A2	Second formant amplitude (dB)
A3	Third formant amplitude (dB)
A4	Fourth formant amplitude (dB)
A5	Fifth formant amplitude (dB)
Blp	First formant parallel bandwidth (Hz)
B2p	Second formant parallel bandwidth (Hz)
B3p	Third formant parallel bandwidth (Hz)
B4p	Fourth formant parallel bandwidth (Hz)
B5p	Fifth formant parallel bandwidth (Hz)
F0	Fundamental freq. of voicing (Hz)
ASP	Amplitude of aspiration (dB)
TLT	Voicing spectral tilt(dB)
AV	Amplitude of voicing (dB)
ANP	Amplitude of nasal pole (dB)
AVP	Amplitude of parallel voicing (dB)

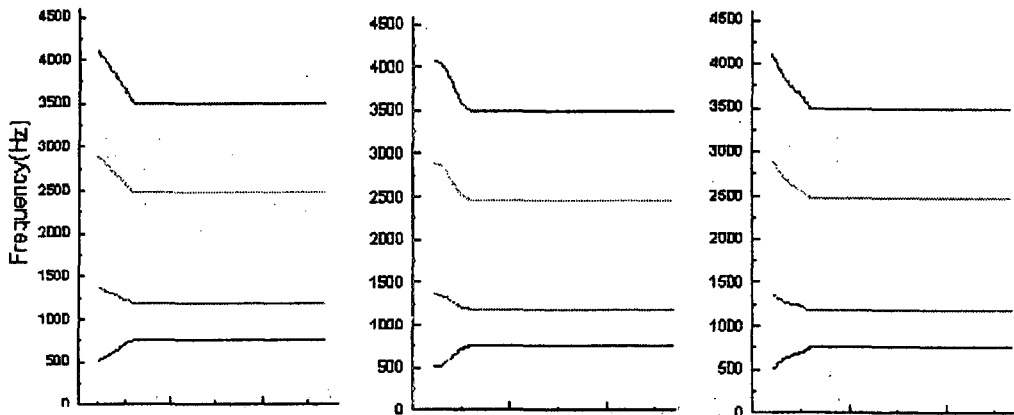


따라 완만하게 변화한다. 따라서 초성과 중성, 중성과 종성 사이의 궤적을 선형적인 방법과 비선형적인 방법으로 접속하는 방법을 제안하고 합성실험을 하였다. 한국어의 초성과 후속 모음으로 구성되는 단음절에서 선형적인 포맷트 접속은 초성의 안정된 포맷트 값에서 후속모음의 안정된 포맷트 값을 타겟값으로 하여 직선으로 연결하였다. 또한, 비선형적인 포맷트 접속은 포맷트의 시작점과 목표점 사이의 접속을 보다 자연스럽게 완만하게 만들어 주기 위하여  $\mu$ -law 특성과 지수함수 특성을 이용하여 천이 영역에서의 포맷트 값을 계산하였다 그림 6은 합성에 사용된 선형 및 비선형 방법을 이용한 포맷트 궤적을 그린 것이다.

### 3.4 기본주파수 궤적의 제어

유성음원의 기본주파수는 자연스런 합성음을 내기 위해 매우 중요한 요소이다 이는 지속시간이나 세기보다 기본주파수의 상승으로 액센트 성분을 가장 잘 나타낼 뿐만 아니라 문장의 형태, 구문 구조, 화제의 변경, 감정 등의 정보를 전달하는 역할을 하기 때문이다[13], 현재 기본주파수 궤적을 모델링하기 위하여 우선 문장의 형태(평서문, 의문문, 감탄문 등)에 따라서 그림 7과 같은 후지사키 모델[8]을 적용하여 기본주파수의 기본값을 나타내었다. 이렇게 주어진 기본주파수의 기본값에 어휘 및 구절에 따라 한국어의 특성을 고려하여 각각 악센트를 적용하였고, 평음에 비해 높은 시작 기본주파수를 가지는 경음과 격음에 대해서도 따로 처리하였으며, 마지막으로 각 문장의 종류(평서문, 의문문, 감탄문 등)에 따라 기본주파수를 처리하였다 [14,15].

그림 6. 합성에 사용된 선형 및 비선형 방법의 포맷트 궤적

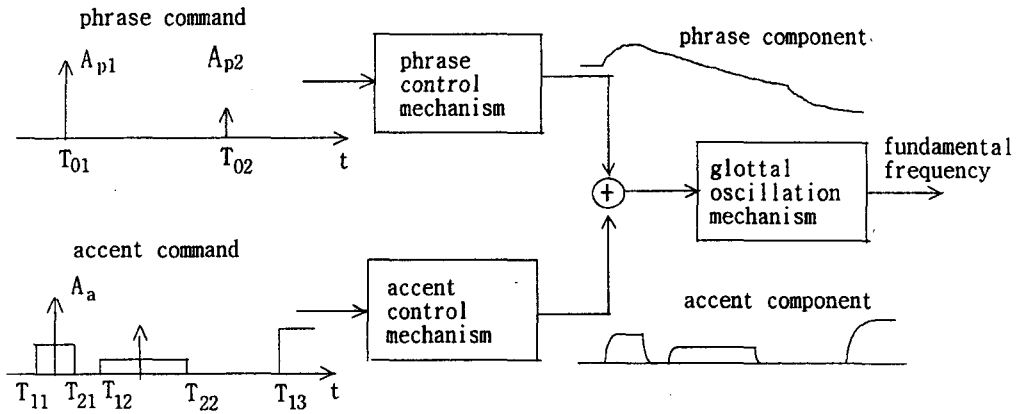


(a) 선형 특성의 포맷트 궤적

(b)  $\mu$ -law 특성의 포맷트 궤적

(c) 지수특성의 포맷트 궤적

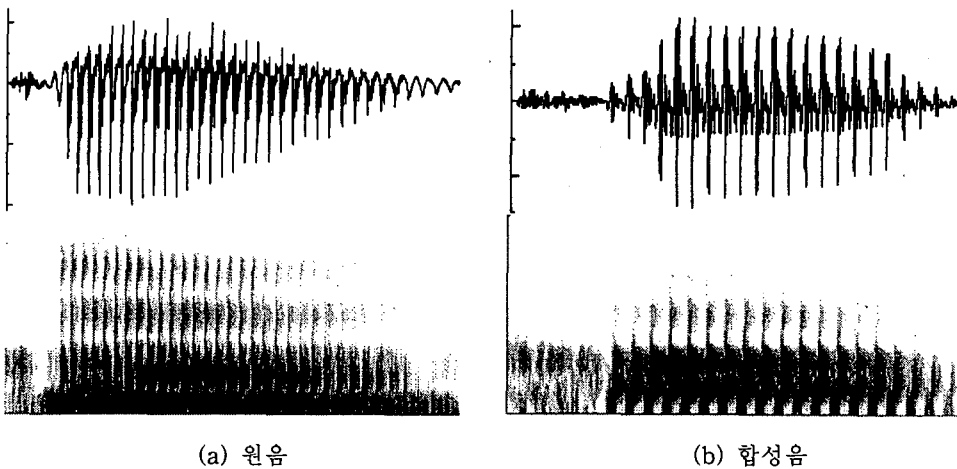
그림 7. 후지사키 모델의 구현



4. 실험 및 검토

2장에서 설명한 방법으로 얻어진 합성 파라미터의 DB를 이용하여 한국어 문자-음성 변환 실험을 하였다. 합성음성은 10 kHz의 샘플링 주파수와 샘플당 16 비트의 해상도를 가지도록 했으며 합성기의 파라미터는 10 msec마다 갱신하였다. 그림 8은 '초성파열음+단모음'의 형태를 가지는 단음절 /가/에 대한 원음과 합성음을 나타낸 것이고, 그림 9는 '거센소리+중성자음'의 형태를 가지는 단음절 /겨/에 대한 원음과 합성음을 나타낸 것이다

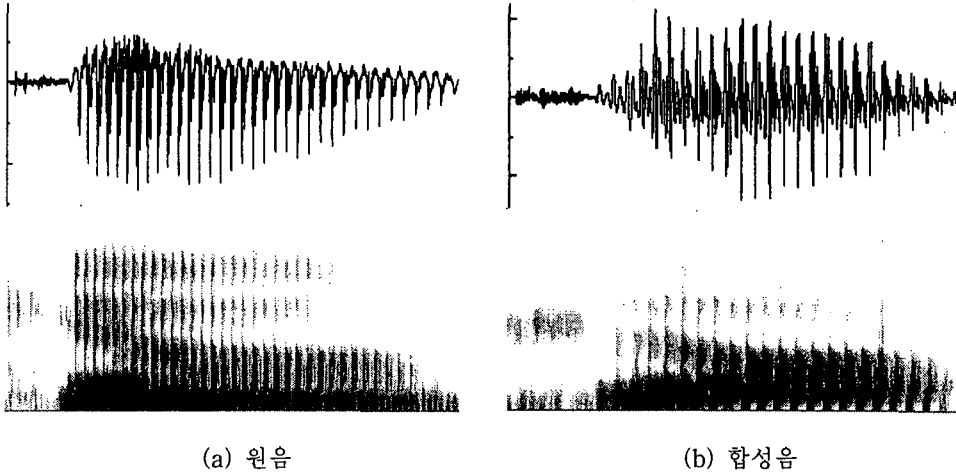
그림 8. 음성신호 /가/의 원음과 합성음



(a) 원음

(b) 합성음

그림 9. 음성신호 /겨/의 원음과 합성음

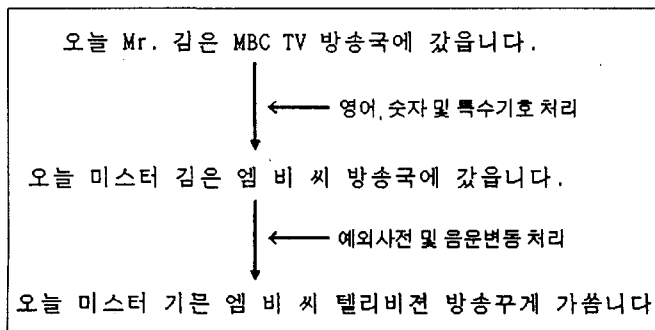


단음절 합성실험의 경우, 합성음에 익숙한 사람에게는 다소 부자연스러움이 느껴지지만 대체로 무난히 인식되었으나, 합성음에 익숙하지 않은 사람에게는 인식이 어려운 경우가 있었다.

분석 및 합성실험 결과, 자음의 포먼트 주파수 값은 특정한 값을 가진다기 보다는 후속모음 또는 앞선모음과 더 밀접한 관련이 있음을 알 수 있었으며, 또한 높은 차수의 포먼트가 자음의 명료도에 큰 영향을 미치는 일부 거센소리를 제외한 대부분의 자음은 두번째 및 세번째 포먼트가 다른 포먼트에 비해 자음의 명료도에 미치는 영향이 더 크다는 것을 알 수 있었다.

3장에서 설명한 문장-음소 변환을 위한 전처리과정과 선형/비선형 접속을 이용한 음소간 포먼트 접속, 그리고 기본주파수 궤적제어 등을 실제 합성실험에 적용하였다. 그림 10은 전처리과정에 의해 합성기에 입력되기 전의 문장-음소 변환과정의 예를 보인 것이다. 본 연구에서는 구문분석을 하지 않은 상태에서 전후 음소에 따른 음운변동만을 고려하였는데, 보다 개선된 음운변환을 위해서는 전처리과정 이전에 구문분석을 통해서 품사정보 및 장단의 정보를 따로 처리하는 부분이 고려되어야 한다

그림 10. 한국어 문장의 음운변동의 예

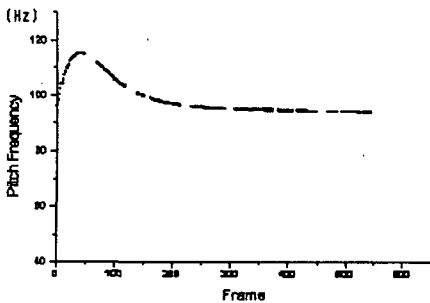


선형 및 비선형 함수를 사용한 포맷트 궤적의 접속에 따른 합성실험에서는 천이영역이 없이 급격한 포맷트 변화를 줄 경우에 비해 천이구간을 두어 접속한 경우에 합성음의 명료성이 많이 증가하였다. 그러나, 선형 및 비선형적인 접속방법에 따른 비교에서는 합성음의 음질 및 명료성에서 큰 차이를 느낄 수 없었다 이는 2장에서 합성 DB를 구할 때 포맷트 주파수 및 천이구간의 값이 선형적인 접속을 가정하여 얻어진 데에 기인하는 것으로 생각된다.

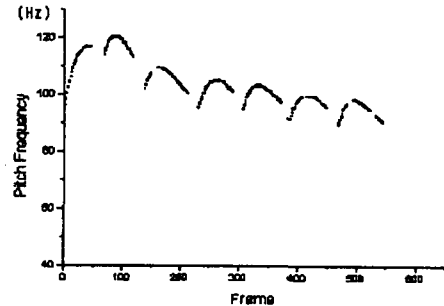
후지사키 모델을 사용하여 기본주파수를 제어하는 과정의 각 단계에서의 기본주파수 궤적을 그림 11에 나타내었다. 이렇게 기본주파수를 후지사키 모델을 이용하여 제어함으로써 보다 자연스러운 합성음을 만들 수 있었으나 후지사키 모델이 주로 액센트 언어인 영어나 일본어에 적합하므로 한국어의 기본주파수를 정확하게 제어할 수 없어서 합성음은 약간 어색한 부분이 드러나며 또한 기존의 기본주파수에 관한 연구가 보다 정형화 되고 수치화되어야 할 필요가 있다.

그림 12는 구현된 한국어 문자-음성 변환 시스템을 이용하여 출력된 한국어 문장 “여기는 경북대학교 신호처리 연구실입니다”에 대한 합성음의 파형 및 합성에 사용된 기본주파수와 포맷트의 궤적을 보인 것이다. Informal한 청취 실험에서 어느 정도의 명료성을 확보하였으나, 단 음절 및 단어 단위의 규칙합성에 비해 명료성이 감소하였다. 합성음의 명료성 및 자연성 향상을 위해 합성 DB 및 운율정보의 제어에 대한 개선의 여지가 많다고 하겠다.

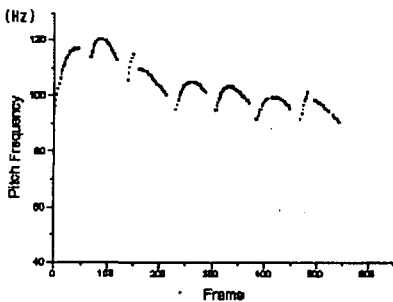
그림 11. 한국어 문장 “말은 사람의 특징이요 겨례의 보람이요 문화의 표상이다.”에 대한 기본주파수 궤적의 제어



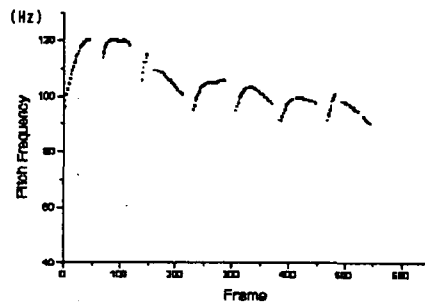
(a) 문장 전체에 대한 기본주파수 궤적



(b) 어절과 문장에 따른 기본주파수 처리



(c) 경음 및 격음의 기본주파수 처리



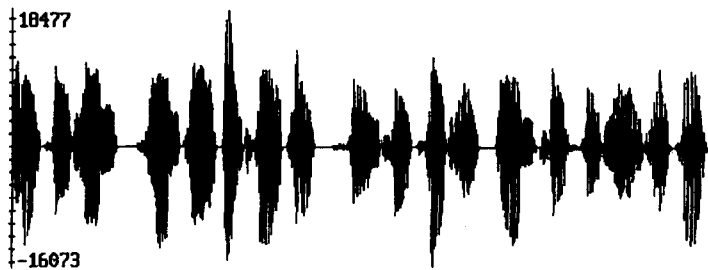
(d) 조사 및 음소환경에 따른 기본주파수 처리

## 5. 결 론

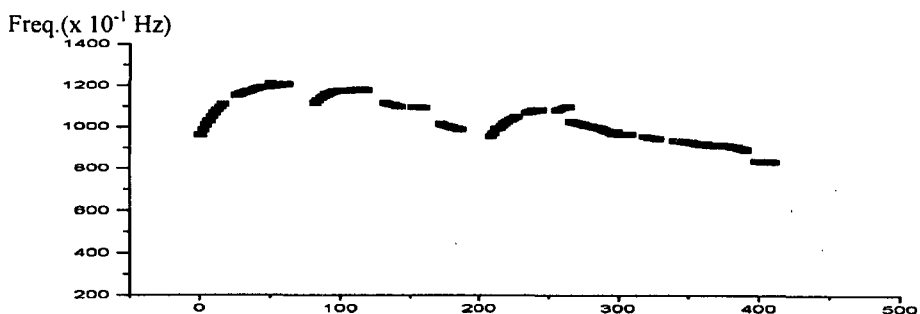
한국어 초성과 종성의 포만트 궤적을 후속모음과 앞선모음에 따라 분석하여 포만트 방식의 한국어 문자-음성 변환시스템에 필요한 DB를 구축하고 합성실험을 하였다. 분석과 합성 실험의 결과에서 초성과 종성의 포만트 값은 후속 모음과 앞선 모음에 따라 포만트 주파수의 변화가 현저하여 각 음소의 포만트 값을 하나로 둘 경우 명료성을 갖는 음성을 합성할 수 없었다. 따라서 반복적인 분석/합성 실험을 통해서 후속 모음과 앞선모음에 따른 초성과 종성의 포만트 천이구간의 길이 정보를 구하여 음성합성에 이용하고, 초성과 중성, 중성과 종성간의 연결규칙을 모델링 함으로써 합성음의 명료성을 개선하였다.

또한 한국어 문자-음성 변환시스템을 구현하기 위해 기본적으로 필요한 음운변동처리와 음소의 연결시 선형/비선형의 접속 실험 및 자연성 개선을 위해 기본주파수를 조정하였다, 즉, 음운변동처리에 있어서는 자음동화, 말음법칙, 구개음화, 경음화 등을 고려하여 규칙변화를 구현하였으며, 예외사전으로 약 3,000단어를 구성하고 많이 사용하는 숫자, 영어약자, 특수기호를 처리하도록 구성하였다.

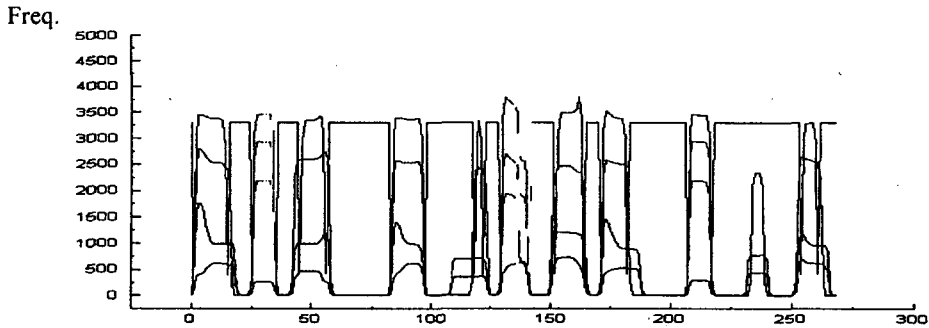
그림 I2. 한국어 문장 “여기는 경북대학교 신호처리 연구실입니다.”에 대한 합성 결과



(a) 합성음의 신호파형



(b) 기본 주파수 궤적



(c) 포만트주파수 궤적(F1~F4)

규칙합성 실험에서 초성 포만트와 중성의 포만트 사이의 접속을 선형적인 방법과 비선형적인 방법으로 모델링한 합성실험에서는 천이영역이 없이 급격한 포만트 변화를 줄 경우에 비해서는 명료성이 많이 증가하였으며, 선형적인 방법과 비선형적인 방법에서는 큰 차이를 느낄 수 없었다. 문장의 종류, 경음 및 격음의 특성, 조사의 유형 등을 고려하여 후지사키 모델을 이용한 기본주파수 궤적을 제어하여 보다 자연스러운 음성의 합성을 할 수 있도록 하였다.

앞으로 사용될 문자-음성 변환 시스템에 보다 나은 합성음을 얻기 위해서는 음절내에서의 음소간의 결합 규칙 뿐만 아니라 음절간의 결합 규칙에 대한 분석과 모델링이 필요하다. 또한 인간의 발성음과 같은 보다 자연스런 합성음을 얻기 위해서는 한국어의 운율에 대한 연구가 필연적이며, 또한, 문장의 전처리 분야에서는 보다 자세한 구문분석을 통한 운율처리가 필요하다.

본 논문은 한국과학재단의 핵심전문연구비(과제번호: 941-0900-015-2) 지원으로 수행되었으며 지원에 감사드립니다

### 참고 문헌

- [1] S. E. Levison, J. P. Olive, and J. S. Tschirgi. 1993. "Speech Synthesis in Telecommunications." *IEEE Commun.*, 31(11), 46-53.
- [2] D. B. Roe, J. G. Wilpon. 1993. "Whither Speech Recognition: The Next 25 Years." *IEEE Commun.*, 31(11), 54-62.
- [3] 정인중, 경연정, 이양희. 1993. "한국어 음성의 규칙합성." *전자공학회지*, 20(5), 587-596.
- [4] 이윤근, 안승권. 1993. "음성합성 기술분야." *전자공학회지*, 20(5), 16-25.
- [5] D. H. Klatt. 1987. "Review of text-to-speech conversion for English." *J. Acoust. Sec. Am.*, 82(3), 737-793.
- [6] 이용주. 1990. "통신처리를 위한 음성정보 변환기술 개발." *한국전자통신연구소 연구보고서*.
- [7] D. H. Klatt. 1980. "Software for a cascade/parallel formant synthesizer." *J. Acoust. Sec. Am.*, 67(3), 971-995.

- [8] H. Fujisaki and K. Hirose. 1984. "Analysis of voice fundamental frequency contours for declarative sentences of Japanese." *Acoust. Sec. Ja.*, 233-242, April.
- [9] 손일권. 1994. "포만트 합성방식에 의한 음소단위의 한국어 규칙합성." 경북대학교 석사학위 논문.
- [10] 김민년, 최진산, 손일권, 이준우, 배건성. 1995. "포만트 합성방식에 의한 한국어 문자/음성 변환에 관한 연구." *음성통신 및 신호처리 워크샵 논문집*. 12(1), 119-122.
- [11] 문교부 고시. 1994. *국어 어문 규정집*. 대한교과서 주식회사
- [12] 서울대학교 공과대학 전자공학과 응용전자연구실. 1989. "한국어 문자-음성변환을 위한 음성학적 전처리 프로그램 개발." *한국전기통신공사 최종보고서*.
- [13] 임운천. 1991. "한국어 규칙합성을 위한 운율법칙의 구현에 관한 연구." 고려대학교 박사학위 논문.
- [14] 안수길. 1988. "법칙합성을 위한 운율 및 음운 정보에 관한 연구." *서울대학교 생산기술연구소 연구보고서*.
- [15] 김형욱, 안승권. "한국어 문장-음성변환기의 운율 제어용 구문분석기." *음성통신 및 신호처리 워크샵 논문집*, 1992.
- [16] 김승곤. 1991. *음성학*. 과학사, 228-252.
- [17] J. N.Holmes. 1982. "Formant synthesizers: cascade or parallel?" *JSRU Research Report*, 1017.
- [18] 최진산, 배건성. 1995. "포만트 합성기를 이용한 한국어 초성과열음의 규칙합성에 관한 연구." *신호처리학회합동학술대회 논문집*, 8(1), 120-123,

접수일자 : '97. 8. 12.

- ▲ 최진산  
경상북도 안동시 한국방송공사  
Tel : (0562) 84-5100 (O) (053) 523-7825 (H)
- ▲ 김민년  
서울특별시 서초구 우면동 16  
LG 전자기술원 미디어통신연구소  
Tel. (02) 526-4408  
E-mail: kimmn@mmedia.sil.goldstar.co.kr
- ▲ 서정욱  
대구광역시 북구 산격동 1370  
경북대학교 전자전기공학부  
Tel: (053) 950-6500-627 Fax: (053) 950-5505  
E-mail: jwseo@palgong.kyungpook.ac.kr
- ▲ 배건성  
대구광역시 북구 산격동 1370  
경북대학교 전자전기공학부  
Tel: (053) 950-5527 Fax: (053) 950-5527  
E-mail: ksbae@ee.kyungpook.ac.kr