

영교차와 특징함수를 이용한 한국어 음성합성

홍 광 석*

Korean Speech Synthesis Using Zero Crossing and Feature Function

Hong Kwang-Seok*

Summary

Speech synthesis is a process which artificially produces speech for various applications, diminishing the dependence on using a person's recorded voice. The speech synthesis methods enable a machine to pass on instructions or information to the user through speaking.

This paper presents a study on Korean speech synthesis using zero crossing and feature function. Synthesized speeches were compared with the uttered speech sentence waveform by the speaker. Through computer simulation we confirmed that these methods improve the memory size and the response time.

서 론

음성에 의한 Man-Machine Interface는 멀티미디어 환경과 인간과 기계의 정보교환을 편리하게 하기 위한 중요한 기술이다. 음성처리에는 음성인식, 음성합성, 음성코딩 및 전송기법등이 있는데, 그중에서 음성합성 방법은 녹음편집, 분석합성, 규칙합성 등으로 나누어진다. 이들은 크게 제한어휘 합성과 무제한 어휘합성(안 등, 1993)으로 분리 되는데, 전자는 합성하고자 하는 어휘들을 미리 저장하였다가 이들의 조합에 의해 말을 합성하는 방법으로 합성대상 어휘가 제한되고, 후자는 언어의 기본단위인 음소, 반음절, 음절등의 조합에 의해 무

제한으로 말을 합성한다. 무제한 음성합성 분야는 Man-Machine Interface에 있어서 핵심적인 부분을 차지하고 있으며, 일부 선진국의 경우에는 실용화 단계에 와 있고, 국내에서도 최근에 상품화되기 시작 하였다. 그러나 합성음의 자연성이나 데이터의 크기, 계산량에 있어서 향상 되어야 할 부분이 문제로 남아있는 현실이다.

본 논문에서는 음소의 영교차 간격과 고유의 특징함수를 이용하여 무제한 한국어 음성합성을 행하는 방법을 제안한다. 이방법은 자연스럽게 발생된 음성의 각 음소를 관찰하여 영교차가 위치하는 지점과 그 파형을 나타낼 수 있는 특징함수를 정의하고, 이를 이용하여 규칙에 의해 입력된 문장을 합성음성으로 전환하는 방법이다. 이때 사용되는 데

* 공과대학 정보공학과(Dept. of Information Engineering, Cheju Univ., Cheju-do, 690-756, Korea)

이타는 각 음소의 영교차 간격과 특징함수 만을 사용하기 때문에 다른 파라미터를 이용한 합성방법보다 적은 데이터 만으로 합성음성을 만들어 낼 수 있다. 또한 각 음소별 하나의 특징함수를 이용하기 때문에 LPC, FORMANT, LSP(Furui 등, 1991)등을 이용한 합성방식 보다 계산량을 최소화 할 수 있다.

본 연구에서는 제안된 방법으로 음성합성 시스템을 구성하고, 음성합성 과정은 규칙에 의하여 합성하는 방법을 택하였다. 또한, 임의의 문장을 합성하여 화자가 발성한 같은 음에 대하여 시간영역에서 파형비교를 하였고, 청취의 방법으로 음질평가를 하였다.

음소의 영교차간격과 함수표현

음성의 합성단위는 구절, 단어, 음절, 반음절, 음소, diphone, triphone 등이 많이 사용되고 있는

Table 1. Korean phonemes.

(a) vowel

sound	front	middle	behind
	lingual	lingual	lingual
single vowel	ㅣ, ㄱ ㅣ, ㄴ ㅣ	ㅡ ㅣ ㅏ	ㅓ ㅗ

(b) consonant

sound		voiceless	voiced
plosive	unaspirated	ㅍ, ㅌ, ㅍ	ㅍ, ㅌ, ㅍ
	aspirate	ㅍ, ㅌ, ㅍ	
	hardening	ㅍ, ㅌ, ㅍ	ㅍ, ㅌ, ㅍ
affricate	unaspirated	ㅈ	ㅈ
	aspirate	ㅉ	
	hardening	ㅉ	ㅉ
frictional	simplicity	ㅅ	
	hardening	ㅆ	
	glottis	ㅎ	
nasal			ㅁ, ㄴ, ㅇ
lateral			ㄹ

데(이 등, 1993), 합성단위가 크면 자연성이 증가하지만 만들수 있는 합성음은 제한된다. 따라서 여기에서는 음소를 합성단위로 한다. 또한 모음인 경우 주기적인 신호가 반복하므로 한주기만을 설정 하여 반복하고, 자음인 경우 초성과 종성으로 나누어 표현한다. 한국어의 음소는 Table 1과 같이 분류 할 수 있으며, 음운변화는 자음과 자음이 연결되는 부분에서 새로운 천이음이 생성된다. 복모음은 단모음의 결합으로 생성하고, 음성합성시 초성 자음의 /O/음은 모음부에 흡수된 형태로 해석한다.

1. 영교차간격

음성합성을 위해 각 음소별 영교차 간격에 대한 정보를 데이터로 사용하였다. 영교차 간격에 대한 정보는 초성인 경우 /초성 + ㅏ/의 형태로 발성하여 시각적으로 음성패턴을 관찰한 후 적절한 위치를 세그먼트 하고, 세그먼트된 음성신호에서 영교차 간격을 설정하고, 종성인 경우 /ㅏ + 종성/의 형태로 발성하여 처리한다. 또한 단모음인 경우 모음을 단독발성하여 안정된 한주기에 대한 영교차 간격을 설정 하였다.

2. 특징함수

각 음소별 함수의 표현은 모음인 경우와 자음인 경우로 구분하여 처리한다. 단모음인 경우 한주기의 음성신호의 기본식은 다음과 같이 정의한다.

$$S(N) = \text{INT} [(-1)^I * A * \text{SIN} (\pi * J/Z)], \quad N=1, 2, \dots, N$$

여기에서 S(N)은 단모음 한주기 길이 N의 음성 신호열 이고, I는 영교차 순서, A는 진폭, J는 영교차 간격 사이의 샘플 순서, Z는 영교차 간격이다.

합성하고자 하는 음소가 /ㅏ/, /ㅣ/, /ㅓ/, /ㅣ/인 경우, 한주기의 음성신호가 비스듬히 경사를 갖고 있으므로 다음과 같이 경사를 주기 위한 식을 통한다.

$$S(N) = \text{INT} [S(N) * (M-N)/M], \quad M \geq N, \quad N=1, 2, \dots, N$$

또한, /ㅣ/, /ㅣ/, /ㅣ/는 음성신호 파형에 매끄

렵지 않은 고조파형의 변형이 있으므로 그것을 나타내기 위해 다음식과 같은 처리를 한다.

$$S(N) = S(N) + (-1)^N * G, \quad N=1, 2, \dots, N$$

여기에서 G는 변형의 피크 값이다. 이와같이 한 주기의 신호를 얻은후 필요한 시간길이 만큼 반복하면 합성된 단모음을 얻을 수 있다. 단모음 한 주기의 영교차 간격 사이의 샘플순서는 화자가 발성한 음성신호 파형의 관찰에 의해 Table 2와 같이 결정 하여 식에 적용하였다.

단모음의 /ㄱ/, /ㄴ/와 복모음의 경우 단모음의 결합으로 볼 수 있기 때문에 단모음 합성음을 결합하여 사용하는데, 복모음 발성시 뒤의 음이 더 길게 발성되기 때문에 앞의 음에 20-30%, 뒤의 음에 70-80%의 시간길이를 할당한다.

자음인 경우는 주기적이 아니기 때문에, 관찰된 음성파형으로 부터 자음부로 세그먼트된 전구간에 대하여 영교차 간격을 Table 3, 4와 같이 설정하여 처음의 S(N) 식을 사용하여 합성하였다. Table 3은 초성자음에 대한 것이고, Table 4는 종성자음에 대한 것이다. 여기에서 /ㄷ/, /ㅌ/를 제외한 나머지 자음은 피크-피크값을 ±5 양자화 레벨 이하로 제한한다.

Table 2. Zero crossing interval for 8 single vowels.

phoneme	zero crossing interval
ㅏ	5, 6, 6, 8, 6, 6, 6, 6, 6, 6, 7
ㅓ	15, 15, 15, 19
ㅗ	18, 17, 12, 16
ㅜ	9, 8, 9, 9, 7, 14
ㅡ	10, 11, 11, 10, 8, 11
ㅝ	6, 7, 7, 7, 6, 7, 6, 13
ㅟ	8, 6, 6, 9, 8, 6, 6, 14
ㅢ	12, 11, 13, 22

음성합성 알고리즘

1. 음성합성 시스템 구성

Fig. 1은 음성합성 시스템의 전체 구성도이다. 전

처리에서는 문자열을 음소열로 변환한다. 운율 제어부에서는 음성발성시 필요한 운율 파라미터를 추출하는데 억양, 강세, 리듬 등의 제어에 필요한 음소길이, 휴지기 길이 등의 파라미터를 추출한다. 합성부에서는 전단에서 추출한 운율 파라미터 및 음성 파라미터를 이용해 음성을 합성한다.

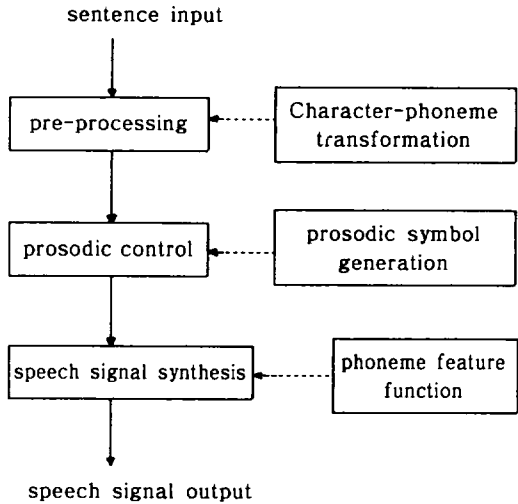


Fig. 1. Speech synthesis system.

2. 운율기호

문장의 의미를 올바르게 전달하기 위해서 문장을 적절히 끊어 읽는 쉼길기와 억양패턴을 생성하기 위한 운율패턴을 설정할 필요가 있다. 쉼 길이는 문장간 또는 홀문장간, 독립어 뒤, 숨을 쉬기위한 쉼, 의미 전달을 위한 쉼 등이 있는데, 각각의 경우에 맞는 쉼 길이를 정한다. 또한, 억양패턴을 발화 단위의 유성음 수와 연관관계에 따라 정의 하는데, 이러한 결정규칙을 바탕으로 운율기호 생성표를 어절단위로 작성 한다.

Table 5는 쉼길이를 표시 하였는데, 각 단계별로 20ms씩 차이가 나도록 하였다. Table 6은 운율패턴으로써 아는 음절을 나타낸다. 또한, ○표시는 억양이 올라가고, →는 평상, ↓는 내려가는 것을 의미한다. 운율의 조절은 모음부의 신호가 반복하는 피치주기를 10%이내로 증감함으로써 조정하였다.

Table 3. Initial consonant.

phoneme	zero crossing interval
ㄱ	14, 1, 19, 101, 4, 8, 3, 4, 2, 3, 3, 3, 3, 4, 1, 10, 2, 5, 2, 35, 4, 6, 3, 1, 4, 2, 4, 5, 4, 6, 3, 6, 4, 5, 4, 3, 7, 3 14, 5, 6, 4, 4, 7, 3, 4, 3, 2, 4, 1, 5, 1, 17, 4, 5, 5, 5, 4, 3, 4, 4, 3, 10
ㄷ	2, 8, 2, 5, 1, 1, 2, 8, 1, 63, 2, 8, 3, 3, 2, 2, 1, 17, 2, 7, 23, , 2, 2, 3, 4, 1, 2, 4, 7, 3, 1, 3, 2, 3, 1, 3, 3, 5, 1, 2, 2, 5, 6, 5, 6, 3, 6, 5, 3, 7, 3, 6, 3, 5, 2, 8, 4, 7, 3, 3, 3, 5, 4, 2, 1, 4, 3, 5, 4, 5, 3, 16, 3, 4
ㄴ	1, 4, 2, 4, 2, 4, 2, 22, 3, 35, 3, 3, 1, 4, 2, 4, 2, 9, 3, 3, 1, 2, 4, 2, 3, 3, 3, 2, 4, 2, 4, 2, 17, 1, 4, 3, 3, 2, 4, 2, 4, 2, 32, 1, 5, 1, 4, 1, 45, 4, 15
ㄹ	1, 3, 40, 101, 96, 2, 5, 3, 2, 2, 1, 1, 1, 3, 2, 4, 1, 6, 1, 5, 1, 56, 1, 17, 3, 3, 3, 3, 1, 4, 2, 5, 1, 13, 2, 8, 1, 8, 1, 5, 1, 5, 2, 4, 4, 9, 3, 3, 2, 4, 1, 4
ㅅ	2, 1, 1, 3, 2, 3, 2, 1, 8, 3, 1, 1, 1, 5, 1, 51, 1, 7, 1, 1, 15, 1, 4, 1, 3, 1, 1, 2, 3, 2, 2, 1, 5, 1, 6, 1, 1, 1, 3, 1, 4, 1, 4, 2, 4, 1, 2, 1, 1
ㅇ	1, 4, 2, 4, 2, 3, 3, 4, 1, 4, 2, 4, 1, 13, 1, 5, 1, 1, 4, 1, 4, 1, 4, 2, 4, 1, 4, 2, 3, 2, 13, 2, 5, 2, 1, 1, 1, 14, 1, 12, 1, 2, 3, 3, 2, 2, 3, 3, 2, 11, 1, 5, 1, 4
ㅈ	5, 1, 3, 55, 5, 9, 6, 6, 5, 9, 5, 9, 5, 4, 6, 6, 4, 6, 5, 8, 5, 5, 5, 7, 1, 6, 4, 5, 1, 2, 5, 4
ㅊ	2, 1, 1, 52, 9, 2, 7, 10, 7, 5, 1, 1, 4, 7, 4, 12, 3, 2, 2, 4, 7, 3, 2, 2, 3, 12, 3, 3, 1, 3, 8, 3, 7, 3, 2
ㅋ	1, 4, 2, 4, 2, 4, 2, 4, 3, 43, 1, 5, 1, 5, 1, 4, 1, 78, 9, 1, 5, 13, 8, 6, 7, 3, 5, 11, 5, 1, 2, 5, 8, 3, 4, 15, 5, 11, 5, 7, 4, 4, 11, 5, 7, 5, 1, 1, 5, 4
ㅌ	1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 3, 1, 1, 1, 1, 1, 4, 2, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 6, 1, 3, 1, 19, 1, 3, 1, 1, 1, 1, 1, 6, 1, 6, 1, 5, 1, 14, 1, 4, 1, 1, 1, 1, 6, 1, 7, 1, 10, 1, 16, 1, 2, 1, 10, 1, 1, 1, 1, 1, 6, 1, 1, 3, 8, 5, 1, 1, 5, 6, 9, 3, 7, 6, 6, 3, 6, 7, 7, 6, 28, 1, 1, 20, 26, 7, 2
ㅍ	2, 2, 1, 12, 1, 27, 1, 3, 1, 1, 1, 3, 1, 4, 1, 2, 1, 2, 1, 1, 1, 1, 1, 6, 1, 5, 1, 2, 1, 2, 1, 1, 1, 1, 2, 1, 10, 1, 1, 1, 1, 1, 10, 1, 2, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 3, 1, 4, 1, 1, 1, 6, 1, 1, 1, 7, 1, 7, 1, 12, 1
ㅑ	2, 2, 1, 3, 1, 2, 1, 2, 1, 20, 1, 4, 1, 2, 1, 4, 1, 5, 1, 8, 1, 3, 1, 3, 1, 15, 1, 3, 1, 1, 1, 5, 1, 1, 1, 4, 1, 1, 1, 20, 1, 6, 1, 12, 1, 1, 1, 3, 1, 13, 3, 1, 1, 23, 13, 17
ㅓ	13, 1, 8, 2, 3, 1, 3, 3, 3, 2, 3, 3, 3, 1, 3, 3, 4, 1, 4, 1, 19, 1, 24, 2, 7, 1, 3, 2, 7, 3, 14, 1, 1, 1, 7, 3, 5, 3, 5, 1, 2, 2, 5, 2, 4, 1, 3, 2, 7, 2, 4, 2, 3, 1, 23, 3, 12, 2, 25, 1, 7, 2, 7, 2, 6, 1, 2, 1, 5, 2, 17, 1, 1, 2, 4, 2, 18, 1, 4, 2, 3, 3, 3, 25, 3, 3, 12, 5, 3, 15, 4, 3, 14, 19
ㅕ	9, 19, 11, 8, 9, 8, 15, 8, 8
ㅗ	1, 5, 1, 6, 1, 1, 1, 6, 3, 7, 2, 22, 2, 46, 1, 10, 1, 26, 14, 18
ㅛ	2, 1, 1, 71, 2, 32, 18, 20, 17, 9, 1, 2, 5, 18, 15, 1, 2, 5, 8, 17, 2, 1, 14, 9, 2, 1, 6, 16, 1, 2, 15, 1, 1, 6, 9, 16, 1, 2
ㅜ	4, 32, 19, 27, 13, 10, 12, 20, 16, 8, 11, 19, 16, 9, 11, 19, 16, 10, 11, 18, 16, 10, 11, 19, 16, 9, 11, 18, 15, 11, 11, 18, 15, 11, 11, 18, 14, 12, 11, 12
ㅠ	14, 21, 21, 9, 2, 1, 6, 14, 4, 1, 17, 2, 2, 4, 2, 3, 4, 2, 2, 11, 3, 2, 4, 1, 11, 9, 2, 2, 4, 15, 3, 2, 17, 8, 3, 1, 5, 15, 2, 3, 14, 9, 8, 11, 2, 3, 1

Table 4. Final consonant.

phoneme	zero crossing interval
ㅂ	4, 12, 5, 3, 1, 2, 6, 4, 9, 2, 8, 19, 5, 9, 2, 5, 4, 6, 5, 32, 33, 5, 8, 3, 3
ㄷ	1, 4, 7, 2, 2, 7, 10, 24, 25, 4, 2, 8, 2, 5, 1, 22, 4, 1, 9, 3, 2
ㄱ	1, 5, 3, 23, 8, 2, 3, 13, 2, 21, 1, 7, 1, 72, 1, 17, 3, 29, 1, 6, 1, 11, 1, 57, 3, 8, 2, 28, 2, 16, 2
ㅋ	2, 8, 6, 20, 3, 3, 15, 6, 4, 4, 28, 4, 6, 6, 1, 9, 1, 13, 1, 26, 3, 6, 1, 9, 2, 5, 1, 6, 1
ㅌ	4, 7, 2, 2, 3, 19, 17, 13, 9, 22, 14, 17, 4, 27, 2, 3, 4, 8, 2, 8, 3, 4, 2, 3, 2
ㅋ	1, 9, 6, 11, 6, 3, 6, 15, 6, 2, 4, 5, 7, 4, 5, 6, 3, 7, 4, 3, 6, 4, 3, 6, 4, 5, 4, 6, 3, 6, 1, 7, 5, 5, 4, 6, 2, 7, 2, 6, 3, 6, 4, 6, 2, 5, 3, 7, 3, 6, 4, 24
ㅍ	1, 5, 2, 3, 4, 3, 2, 3, 3, 4, 2, 11, 1, 6, 2, 22, 2, 4, 2, 2, 2
ㅍ	8, 4, 8, 4, 1, 10, 9, 8, 10, 2, 4, 2, 2, 3, 3, 13, 10, 2, 3, 2, 4, 1, 4, 3, 8, 2, 4, 19, 15, 17, 3, 20, 2, 2, 11, 21, 5, 21, 3, 1, 9
ㅅ	17, 1, 3, 4, 2, 4, 2, 4, 2, 2, 1, 12, 22, 14, 1, 3, 33, 14
ㅆ	9, 4, 3, 3, 3, 2, 3, 2, 2, 4, 2, 1, 4, 7, 2, 3, 21, 3, 2, 12
ㅎ	13, 2, 1, 5, 4, 1, 2, 6, 1, 1, 4, 3, 2, 2, 2, 8, 32, 3, 2, 14, 16, 2, 2, 9, 3, 1, 35, 9, 1, 4, 19, 6, 3, 3
ㅇ	14, 17, 17, 12, 12, 16, 17, 11, 12, 15, 2, 1, 14, 10, 13, 15, 17, 7, 2, 1, 12, 15, 17, 10, 13, 15, 16, 17, 10, 15, 17, 17, 11, 14, 18, 16, 11, 16, 18, 16, 11, 16, 19, 15, 12, 17, 19, 16, 11, 17, 21, 16, 12, 1, 1, 15, 21, 15, 1, 1, 13, 1, 1, 15, 23, 15, 17, 15, 25, 14, 18, 17, 25, 15, 1, 1, 16, 18, 26, 14, 16, 1, 1, 18, 28, 17, 15, 18, 31, 14
ㄴ	16, 18, 16, 9, 13, 19, 15, 11, 14, 17, 15, 11, 14, 17, 15, 12, 14, 16, 15, 12, 14, 16, 16, 11, 14, 17, 16, 12, 13, 18, 15, 12, 14, 18, 16, 11, 15, 18, 16, 11, 16, 18, 16, 11, 15, 19, 16, 11, 15, 20, 16, 12, 15, 20, 17, 11, 16, 20, 16, 12, 16, 21, 16, 11, 16, 23, 16, 11, 18, 23, 14, 12, 18, 24, 14, 13, 17, 27, 4, 1, 1, 10, 17, 1, 1, 27, 14, 1, 1, 10, 2, 2, 16, 29, 14, 1, 1, 14, 16, 31, 13, 17, 16, 33, 16, 14, 16, 36, 16, 11, 16, 38, 16, 10
ㅇ	2, 1, 8, 20, 17, 10, 1, 1, 8, 21, 17, 10, 2, 1, 8, 5, 1, 14, 18, 10, 10, 21, 19, 9, 2, 1, 8, 5, 1, 15, 19, 9, 11, 6, 1, 14, 20, 10, 11, 5, 1, 14, 21, 10, 11, 5, 1, 2, 1, 12, 22, 9, 2, 1, 2, 1, 5, 5, 2, 1, 2, 11, 23, 9, 2, 1, 8, 6, 1, 2, 1, 14, 21, 9, 2, 1, 8, 5, 5, 12, 24, 10, 11, 5, 2, 1, 2, 13, 24, 10, 10
ㄹ	7, 1, 2, 3, 12, 1, 2, 7, 7, 3, 1, 4, 1, 2, 7, 10, 7, 6, 11, 1, 2, 2, 1, 4, 7, 11, 7, 10, 7, 7, 12, 2, 1, 8, 7, 11, 7, 10, 7, 8, 15, 2, 1, 5, 1, 2, 2, 1, 4, 2, 1, 7, 8, 10, 7, 8, 18, 11, 7, 11, 7, 11, 7, 7, 23, 9, 9, 10, 8, 10, 7, 9, 24, 10, 10, 8, 9, 10, 8, 7, 27, 11, 7, 11, 7, 7, 14, 1, 38, 2, 2, 4, 1, 8, 2, 2, 2, 3, 2, 16, 48, 12, 2, 14, 7, 7, 60, 11, 7, 12, 7, 3, 1, 2

Table 5. Pause length.

Pause length	case
4	sentence
3	isolated word
2	breath
1	meaning communication
0	no pause

Table 6. Prosodic pattern.

syllable case	1	2	3	4
1	o↑	o↑o	o↑o→o	o↑o→o→o
2	o→	o→o	o↑o↓o	o↑o→o↓o
3	o↓	o↓o	o→o↓o	o→o→o↓o
4			o↓o→o	o↓o→o→o

실험 및 고찰

음성합성 실험을 위한 장치 구성은 Fig. 2와 같다. 음소의 영교차 위치와 특징함수를 구하기 위한 음성 데이터의 취득은 발생된 음성을 마이크로 통하여 4kHz 차단 주파수를 갖는 저역통과 필터를 통한 후, ADC0800을 IBM-PC 386과 인터페이스 하여 10kHz로 샘플링된 입력 음성 신호를 받아 처리 하였다. 또한 키보드로 입력한 문장을 규칙에 의해 합성한 음성은 IBM-PC와 인터페이스 된 DAC0808을 통하여 저역통과 필터 처리 후 스피커를 통해 출력 하도록 구성 하였다.

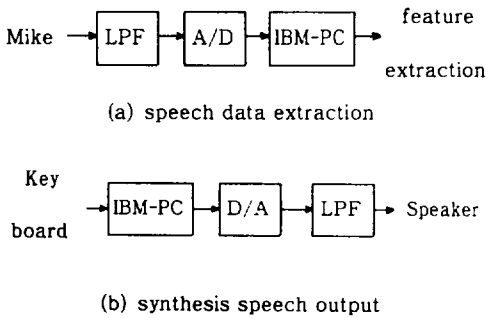


Fig. 2. Block diagram for speech synthesis system.

1. 규칙합성음 파형

Fig. 3은 사람이 발성한 단모음과 합성음을 비교 하기 위해 각각 안정된 구간을 20 ms씩 나타낸 그림이다. (a)는 /ㅏ/, (b)는 /ㅣ/, (c)는 /ㅓ/, (d)는 /ㅜ/에 대한 음성파형 인데, 위의 파형이 발성된 음이고 아래의 파형이 합성한 파형 이다. /ㅏ/와 /ㅣ/는 합성된 파형이 발성된 파형보다 부드러운 형태를 갖도록 함수를 설정 하였다.

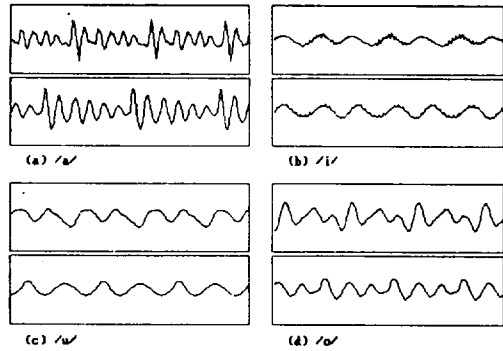


Fig. 3. Comparison between uttered and synthesized speech waveform.

Fig. 4는 사람이 발성한 음성파형과 합성에 의한 파형을 나타내었다. (a)는 /정보를 만드는 사람들/에 대한 음성 파형이고, (b)는 /새나라 새일꾼/에 대한 음성파형 인데, 위부분이 임의의 화자가 발성한 음성파형 이고, 아래가 합성에 의한 음성파형 이다.

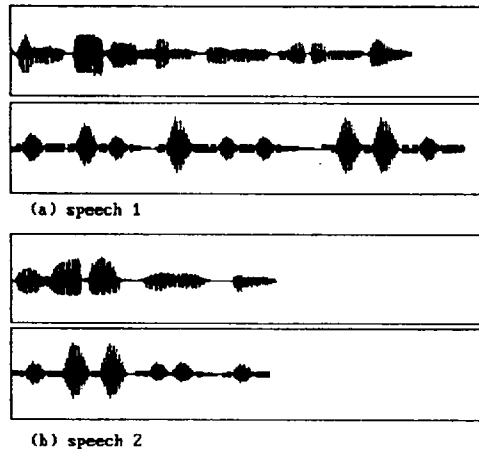


Fig. 4. Comparison between uttered and synthesized speech waveform.

발성된 음성과 합성된 음성을 비교하여 보면, 그림 (a)의 /정보를 만드는 사람들/의 경우 발성음의 길이가 1.62초, 합성음의 길이가 1.8293초의 길이를 갖는데, 같은 음성에 대해서 시간길이가 차이나는것은 화자 각각에 대해서 또는 같은화자가 같은 말을 반복하여 발성을 하여도 그때마다 발성음의 길이가 달라지므로 시간길이의 차이는 합성음질에 큰 영향을 주지 않는다.

합성된 음성이 발성된 음성보다 음절이 뚜렷하게 구분되어 나타나고 있는데, 이는 합성시 자연성을 추가하기 위한 기법을 좀더 개선해야 할 것이다.

또한, 이 음성합성 시스템에서 필요한 데이터 기억용량은 Table 2, 3, 4에 나타나 있듯이 임의의 음성을 합성하는데 최대 1744byte의 기억용량이 필요하다. 그러나 녹음편집 방식의 합성에서는 10kHz sampling 8bit A/D 변환인 경우 (a)의 /정보를 만드는 사람들/ 하나만 해도 16200byte의 기억공간이 필요하고, 기타의 합성방식에 있어서도 상당한 크기의 기억공간이 필요하다. 계산량에 있어서도 녹음편집 방식을 제외한 방법은 상관계수, 고차방정식, FFT 등의 많은 계산을 요하지만 제안된 방법은 하나의 함수에서 바로 음성샘플값을 얻기 때문에 실시간 처리가 가능한 방법이다.

2. 합성음질 평가

합성음의 청취에 의한 평가는 남,여 각 1인을 선정하여 MOS(mean opinion score)방법으로 하였다. 청취를 위한 단어는 1, 2, 3, 4음절 단어 각각 임의로 3단어씩 선택 하였고, 평가항목은 이해도, 명료성, 자연성의 3가지 항목으로 구분하고, 각 항목은 5단계의 평가등급을 부여하였다. 각 등급의 내용은 점수를 부여 하였는데, 아주 좋다(5), 좋은편이다(4), 보통이다(3), 나쁘다(2), 아주나쁘다(1)로 평가 하였다. Table 7에 나타난 바와 같이 합성음의 이해도는 음절이 많은 단어의 이해가 쉬웠고, 명료성은 단어의 음절수와는 무관하고, 합성된 음은 비교적 자연성이 떨어진다는 것을 알 수 있다.

Table 7. MOS test of synthesized word speech

candidate word synthesized word	1	2	3	word test		
				under-standing	clear	nature
1 syllable	일	삼	팔	3.83	4.0	3.83
2 syllable	조국	안부	사랑	4.17	3.17	3.50
3 syllable	어머니	아라동	막걸리	4.33	3.17	3.34
4 syllable	할아버지	정보공학	제주시청	4.67	4.0	3.83

적 요

본 논문은 무제한 음성합성시 필요한 새로운 음성합성 기법을 제안하여 실험 하였다. 제안된 방법은 각 음소가 갖는 영교차점과 고유의 특징함수를 정의하여 키보드 입력된 문장을 음성합성하는 방법이다.

기존의 합성방식은 많은 기억용량이 필요하면서도 제한된 어휘만을 합성해 내는 녹음편집형과 무제한 합성을 위한 파라미터로서 계산량이 많은 포먼트, LSP등, 선형예측계열의 분석 데이터를 보관하여 규칙에 의한 합성을 한다.

그러나 제안된 방법은 각 음소의 영교차 간격의 적은 양의 데이터를 기억하고, 이를 데이터로 하여 함수를 통한후 규칙에 의해 무제한 음성을 합성해 낼 수 있는 방법으로써, 실시간에 합성음을 만들어 낸다는 장점을 갖는다. 또한 합성음 파형 비교와 MOS 평가에서, 제안된 방법이 무리없이 합성음을 만들어 낸다는 것을 확인하였다.

제안된 방법에 의한 합성음의 음질을 더욱 개선하기 위해서는, 발성음에 따른 음운의 피치 변경과 발성 시간 길이 조절 알고리즘에 대한 연구와 천이음의 처리에 대한 연구가 필요하다.

참 고 문 헌

- Allen, J. M., S. Hunnicutt, D. Klatt, 1987, From text to speech : The Mitalk system, Cambridge University Press.
- 안승권, 이윤근, 1993, 음성합성 기술 분야, 대한 전자공학회지, 20권 5호, pp.16-25.
- Furui, S. 1992, Digital Speech Processing, Synthesis, and Recognition, Dekker.
- Furui, S. M., M. Sondt, 1991, Advances in Speech Signal Processing, Dekker.
- 허 용, 1985, 국어 음운학, 샘문화사.
- 진용욱, 강창희, 1994, 시간영역에서의 파형분석에 의한 무제한 어휘합성 및 음절유형별 규칙합성음 음질평가, 한국음향학회지, 13권 1호, pp.71-83.
- 김상용, 김정수, 1993, 형태소 해석을 이용한 합성 음성의 음운 및 운율처리, 대한전자공학회지, 20권 5호, pp.1-7.
- 이양희, 경연정, 정인종, 1993, 한국어 음성의 규칙 합성, 대한전자공학회지, 20권 5호, pp.80-90.
- Sagisaka, Y. 1990, Speech Synthesis from Text, IEEE Communication Magazine, pp.35-41.
- Yannakoudakis, E. J., P. J. Hutton, 1987, Speech Synthesis and Recognition Systems, Ellis Horwood.