

한국어 반음절단위 규칙합성의 개선을 위한 포만트천이의 변경규칙

An Alteration Rule of Formant Transition for Improvement of Korean Demisyllable Based Synthesis by Rule

이 기 영*, 최 창 석**
(Ki Young Lee*, Chang Seok Choi**)

“이 논문은 1995년도 한국학술진흥재단의 공모과제 연구비에 의하여 연구되었음”

요 약

본연구에서는 반음절단위 규칙합성에서 연속음성을 합성할 때 조음결합에 의한 천이구간이 없는 반음절의 연결로 겹속되어 부자연스러운 합성음이 되는 것을 개선하기 위하여 연쇄모음의 천이구간을 보상하는 방법으로 포만트천이의 변경규칙을 제안하였다. 반음절 단위만으로는 포만트천이가 발생하는 부분을 채울 수 없기 때문에 반음절단위의 음성데이터와 모음의 반음절 단위의 정상부위로부터 세그먼트한 정상모음 42개를 추가하여 데이터베이스를 구축하였으며 포만트를 변경하는 방법으로 포만트합성에서의 공진회로를 이용하였다. 제안한 방법의 타당성을 확인하기 위하여 음성합성시 연쇄모음 부분에 포만트천이의 변경규칙을 적용하여 원음성 및 변경규칙을 적용하지 않은 반음절단위 음성합성방식에 의한 합성음성의 스펙트로그램과 비교하고 MOS 테스트를 실시한 결과 보다 자연스러운 합성음성을 얻을 수 있음을 확인하였다.

ABSTRACT

This paper proposes the alteration rule to compensate a formant transition of several connected vowels for improving an unnatural synthesized continuous speech which is concatenated by each demisyllable without coarticulated formant transition for use in demisyllable based synthesis by rule. To fulfill each formant transition part, the database of 42 stationary vowels which are segmented from the stable part of each vowels is appended to the one of Korean demisyllables, and the resonance circuit used in formant synthesis is employed to change the formant frequency of speech signals. To evaluate the synthesized speech by this rule, we carried out the alteration rule for connected vowels of the synthesized speech based on demisyllable, and compare spectrogram and MOS tested scores with the original and the demisyllable based synthesized speech without this rule. The result shows that this proposed rule can synthesize the more natural speech.

I. 서 론

규칙합성이란 음운이나 음절기호 또는 문자의 열을 기반으로 무제한의 어휘를 음성으로 발성하도록 하는 방법이며 최근 음성신호처리기술의 발달과 함께 맨-머신 인터페이스(man-machine interface)에서 처리결과를 무제한 음성 메시지로 비교적 명확하고 자연스러운 목소리로 들

려주기 위한 합성방식으로 활발한 연구 및 개발이 진행되고 있다. 그러나 아직까지 규칙합성에 의한 합성음의 음질 수준이 자연음성에 비하여 매우 큰 열세를 보이고 있는 것은 자연스러운 음성의 합성을 위한 기본적 기술인 언어처리기술, 운율제어기술 및 음성생성기술 모두가 현재 충분한 수준에 도달하지 못하였기 때문이다[1, 2].

규칙합성을 구현하기 위한 음성생성기술에서 사용되는 합성단위[3]는 음소나 복합음소로 나눌 수 있는데 음소는 합성단위로서 가장 원리적이며 단위 수가 가장 적은 장점이 있지만 실제 음성파형은 여러가지 음소환경에 의한 천이부분이나, 발성방식 등에 의해 크게 달라지기

*관동대학교 전자통신공학과
Dept. Electronic Comm. Eng., Kwan Dong Univ.

**명지대학교 정보통신공학과
Dept. Information Comm. Eng., Myong Ji Univ.
접수일자: 1996년 6월 7일

때문에 실질적인 합성단위로서는 사용되지 않고 있으며 그 대신 복합음소, 즉, 음절, 반음절, 다이폰(diphone) 및 VCV 나 CVC 등의 단위를 채택하고 있다[4, 5]. 한국어의 경우 접속(concatination)하는 합성단위를 음절로 하면 약 2000 여개 정도의 단위로 무제한 합성이 가능하나 기본단위인 음절 사이의 천이구간이 없기 때문에 연속음성의 합성시 부자연스러운 음성으로 들린다[6, 7]. 반음절단위 음성합성은 약 600여개의 비교적 적은 반음절 단위 수로 무제한 합성을 실현할 수 있는 방식이기는 하나 합성된 음성의 각 음절은 후반부 반음절의 접속으로 이루어지기 때문에 최종에는 단독음절들이 연결된 것 같은 합성음성을 얻게 되어 자연음성처럼 연속적으로 발생된 음성으로 청취되지 않을 뿐만 아니라 각 음절 사이에 천이구간이 없는 음성파형이 되므로 기계적인 음성으로 들리게 된다. 이러한 단점을 보완하기 위하여 음운의 천이부분을 포함한 다이폰 단위를 이용하여 연속발성된 음절들의 경계를 자연스럽게 연결시킬 수 있다. 그러나 다이폰단위의 합성음성의 경우 명료성이 어느 정도 확보되나 전문적인 수준의 다이폰 경계를 정밀하게 확보하지 못하면 조음 결합에 의한 음운 변화를 완전히 포함하지 못하며 합성단위의 수도 약 1200개 이상이 필요하다. 또한 앞과 뒤에서 연쇄하여 발생되는 모음이나 자음의 영향에 의한 조음결합상태를 한 단위로 하는 VCV나 CVC 를 합성단위로 하는 경우 필요한 단위의 총수는 20000에서 40000여개로 천이구간을 내포한 장점은 있으나 합성단위의 총수가 너무 방대해 지는 단점이 있다. 여기서 합성단위의 종류에 관계없이 공통적으로 대두되는 문제점은 연속음성의 합성시 조음결합에 의한 천이구간을 어떻게 처리하여 각 단위를 접속하느냐로 귀결되며 조음결합은 하나의 음운에서 다른음운으로 연쇄발성할 때 인간의 발성부인 조음기관의 한 음운에 대한 고유위치가 다른 음운의 고유위치로 이동하면서 나타나기 때문에 특히 음성의 대부분을 차지하고 있는 모음에서 어떠한 음운연쇄로부터 추출된 모음인가에 따라 그 모음의 포먼트구조의 천이특성은 판이하게 달라진다.

본연구에서는 원음성의 모음과 모음의 연쇄부분에서 각 모음의 파형은 서로 독립되어 있지 않으며 서로 다른 모음이 연결될 때에도 포먼트구조의 천이를 내포한 파형이 모음과 모음을 연결한다는 점과 인위적으로 포먼트구조를 변경한 파형을 재구성할 수 있다는 점에 착안하여 포먼트천이의 변경을 이용하여 합성단위의 수가 비교적 적으면서 무제한 어휘의 합성이 가능한 반음절단위 규칙 합성방식에 연쇄모음의 포먼트천이구간을 보상하는 변경규칙을 적용함으로써 기존의 방법보다 자연성을 개선하는 방법을 제안하였다. 이방법에서 포먼트구조를 변경하는 방법으로는 포먼트합성[2]에서 이용하는 공진회로를 사용하였으며 연쇄모음의 모음과 모음을 접속할 단위에 변경된 포먼트구조의 음성파형을 삽입하기 위해 반음절단위의 데이터베이스의 모음에서 안정된 부분의 정

상모음들을 추가하였다. 이추가된 정상모음은 42개 정도이며 전문적인 지식없이 데이터베이스를 구축할 수 있는 장점이 있다. 제2장에서는 자연음성에서 나타나는 포먼트천이의 근사식에 대하여 서술하였고 제3장에서는 이를 이용하여 포먼트변경을 분석합성하는 방법을 서술하였으며 제4장에서는 포먼트천이의 변경규칙을 이용한 개선된 반음절단위 음성합성방식에 대하여 서술하였다. 제5장의 실험 및 고찰에서는 변경규칙을 이용하여 포먼트천이를 수행한 합성음성의 파형과 스펙트로그램들을 원음성 및 변경규칙을 적용하지 않은 합성음성과 비교 검토하였으며, 제6장에서 결론을 맺고 있다.

II. 포먼트천이의 근사식

음성의 포먼트는 성도를 통해 발생되는 음성신호의 공진주파수들이다. 이공진주파수는 음성의 조음현상과 조음결합과 밀접한 관계에 있을 뿐만아니라 성대의 떨림으로 생성되는 여기신호의 피치주기가 빠르고 느림에 따라서도 영향을 받는다. 즉, 피치주기가 빠르면 포먼트도 비교적 높아지며 피치주기가 느리면 포먼트도 비교적 낮아진다.

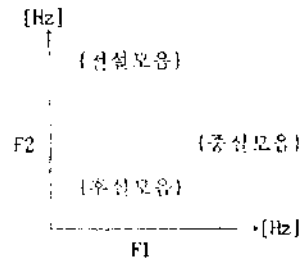


그림 1. 제1, 제2포먼트주파수와 모음의 관계

그림 1에는 제1, 제2포먼트주파수와 모음사이의 관계를 나타내고 있다. 이같은 현상은 모음과 모음이 연속으로 발생되는 연쇄모음의 조음결합에 따른 포먼트의 천이와도 밀접한 관계가 있다. 예를 들어 전설모음 '이'에서 중설모음 '아'로 연속 발생할 경우 '이'에서 낮았던 제1포먼트는 '아'로 변하면서 중설모음 '아'의 안정상태에 이르기 까지 높아지며 아에 대해 제2포먼트는 전설모음 '이'에서 높았다가 중설모음 '아'의 안정상태에 이르기 까지 낮아지는 천이를 보이게 된다.

즉, 연쇄모음 "이아"의 천이구간에서 '이'의 원래 포먼트가 연쇄될 모음 '아'의 포먼트까지 포먼트천이를 일으킨다.

그림 2에 원음성 '이', '아' 및 '우'을 연속발성한 파형과 그의 스펙트로그램을 보이고 있다. 여기서 각 모음이 연쇄된 부분의 스펙트로그램의 각 포먼트주파수가 천이를 일으키며 연결되고 있는 상태를 볼 수 있는데 그 모양은

각 포먼트주파수가 각속도 0에서 π 까지 이동하는 cosine 파형으로 근사시킬 수 있다. 이과정에서 제n포먼트의 천이값 $F_n(t)$ 를 근사식으로 표현하면 다음과 같다.

$$F_n(t) = F_n(0) + \{F_n(0) - F_n(T)\} \cdot [1 - \{1 - \cos(\pi \cdot t/T) + 1\} / 2] \quad (1)$$

$$0 \leq t \leq T$$

여기서 $F_n(t)$ 는 천이과정 중의 제n포먼트이며, $F_n(0)$ 는 연쇄모음에서 앞선 모음의 안정된 상태에서의 제n포먼트, $F_n(T)$ 는 앞선 모음에 연쇄될 모음의 안정된 상태에서의 제n포먼트, $0 \leq t \leq T$ 는 앞선 모음에서 연쇄될 모음 사이의 포먼트천이가 발생하는 시간범위이다.

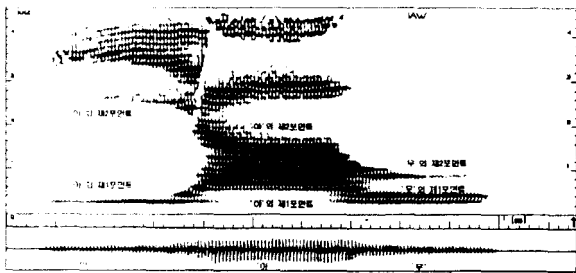


그림 2. 음성파형의 포먼트천이 "아아우"

Ⅲ. 포먼트변경을 위한 분석합성

모음은 폐로부터 나오는 공기가 성도를 통과하며 각 모음에 따라 서로 다른 공진주파수인 포먼트를 발생시킨다. 역으로 포먼트를 달리하면 서로 다른 모음으로 들리게 된다. 본장에서는 발생된 모음음성을 공진회로와 분석합성을 이용하여 포먼트천이를 포함하면서 다른 모음으로 들리도록 하기 위한 포먼트의 변경방법에 대하여 서술한다. 다음 그림 3은 포먼트의 변경을 위한 분석합성시스템이다.

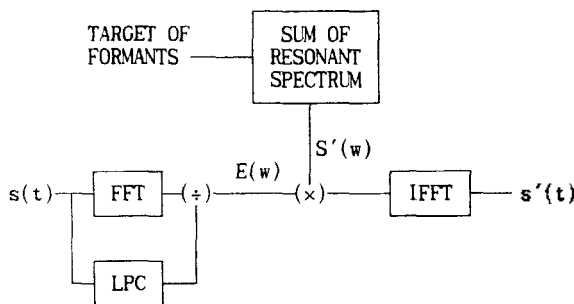


그림 3. 포먼트변경 분석합성시스템

그림 3의 분석과정에서는 음성신호 $s(t)$ 를 선형예측분석(LPC)한 스펙트럼포락으로 푸리에변환(FFT)한 복소스펙트럼을 나누어 여기신호의 스펙트럼 $E(w)$ 를 만든 후, 타겟이 되는 포먼트값으로부터 생성한 공진스펙트럼의 합 $S'(w)$ 를 얻을 수 있는데 이공진스펙트럼은 포먼트주파수와 대역폭으로부터 공진특성곡선을 제한할 수 있는 공진회로를 사용하였으며 다음 식과 같다.

$$RS(w) = \frac{a}{1 - bz^{-1} - cz^{-2}} \quad (2)$$

여기서 z 는 e^{jw} 이고 계수 a, b, c 는 포먼트특성 즉, 공진주파수, 대역폭에 의하여 구해지며, 제m포먼트까지의 합을 구하려면 식(2)를 m번 사용하고 이를 합해준다. 이를 식으로 나타내면 다음과 같다.

$$S'(w) = \sum_{i=1}^m RS_i(w) \quad (3)$$

식(2), (3)을 이용하여 포먼트가 변경된 스펙트럼 $S'(w)$ 이 재현되면 분석과정에서 구한 여기신호의 스펙트럼 $E(w)$ 를 곱하여 다시 역푸리에변환하므로써 포먼트의 천이과정이 포함된 음성파형 $s'(t)$ 를 얻는다. 이를 식으로 나타내면 다음과 같다.

$$s'(t) = \text{IFFT} [S'(w) \times E(w)] \quad (4)$$

여기서 IFFT는 역푸리에변환이다.

Ⅳ. 개선된 반음절단위 합성방식

본장에서는 다이폰단위의 데이터베이스를 이용하지 않고 반음절단위를 기본으로 하며 포먼트천이의 변경을 이용하여 반음절단위 합성음성의 연속발성음성의 경계를 자연스럽게 연결시키는 개선방법에 대하여 기술하였다. 이방법에서는 특히 연속음성의 연쇄모음인 경우 포먼트구조의 천이를 포함한 파형이 모음과 모음을 연결한다는 점과 인위적으로 시간축과 포먼트구조를 변경할 수 있다는 점에 착안하여 2장과 3장에서 포먼트천이의 변경을 규칙화하여 반음절단위 음성합성방식을 개선한다. 여기서 변경된 포먼트천이의 파형을 접속합성하기 위해 모음에 대한 반음절단위의 데이터베이스에서 안정된 부분을 천이구간의 길이 만큼 세그먼트하고 그부분을 정상모음이라 하여 데이터베이스에 추가하여 이용한다. 그림 5는 포먼트천이의 변경규칙을 이용한 개선된 반음절 규칙합성의 구성도이다.

포먼트천이는 연쇄모음 부분에서 발생하며 반음절단위의 합성음성에서는 앞선 전모음의 후반부 반음절과 뒤에 연결될 모음의 전반부 반음절이 접속될 부분에 해당된다. 이부분에서 포먼트천이를 포함한 모음의 음성파형이 접속되도록 하기 위하여 식(1)부터 식(4)까지를 모두

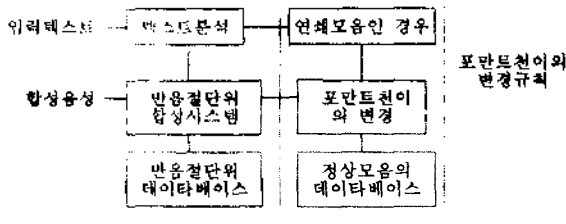


그림 4. 포맷트천이의 변경규칙을 적용한 반응절 규칙합성 구성도

이용하는 포맷트천이의 변경규칙을 적용한다. 즉, 모음 x와 모음 y가 연쇄될 경우 포맷트천이의 변경규칙은 다음과 같다.

STEP 1. 앞선 모음의 후반부 반응절 대신 해당된 모음의 전반부 반응절에서 안정상태 부위를 시간축으로 확장하여 생성하거나 안정상태 부위만을 절취한 정상모음으로 대체한다.

STEP 2. 연쇄될 모음의 전반부 반응절 대신 해당된 모음의 후반부 반응절에서 안정상태 부위를 시간축으로 확장하여 생성하거나 안정상태 부위만을 절취한 정상모음으로 대체한다.

STEP 3. STEP 1에서 대체된 정상모음의 천이될 포맷트 값을 다음 식으로 구한다.

$$Fn(t) = Fn(0) + \{Fn(0) - Fn(T)\} \cdot [t - \{ \cos(\pi \cdot t/T) + 1 \} / 2] \quad (5)$$

$$0 \leq t \leq T/2$$

STEP 4. STEP 2에서 대체된 정상모음의 천이될 포맷트 값을 다음 식으로 구한다.

$$Fn(t) = Fn(0) + \{Fn(0) - Fn(T)\} \cdot [t - \{ \cos(\pi \cdot t/T) + 1 \} / 2] \quad (6)$$

$$T/2 \leq t \leq T$$

STEP 5. STEP 3과 4에서 구한 천이될 포맷트 값을 식(2), (3), (4)에 대입하여 각각 포맷트 변경된 음성파형을 구하고 이들을 각각 앞선 모음의 후반부 반응절 대신, 연쇄될 모음의 전반부 반응절 대신 접속한다.

이와 같은 포맷트천이규칙에서 STEP 3에서는 앞선 모음에서 연쇄될 모음 까지의 전체 천이구간의 처음부터 중간까지 포맷트천이의 변경될 포맷트 값들을 얻은 것이며 STEP 4에서는 전체 천이구간의 중간부터 끝까지 포맷트천이의 변경될 포맷트 값들을 얻은 것이다. 또한 STEP 5에서는 STEP 3과 4에서 구한 포맷트 값들을 이용하여 포맷트천이를 내포하도록 포맷트변경된 천이구간의 파형을 얻게 되고 이들을 반응절단위와 접속하게 된다. 여기서 앞선 모음에서 연쇄될 모음까지 전체 천이구

간의 절반씩의 포맷트변경을 수행하는 것은 포맷트변경을 많이하면 할 수록 왜곡이 많이 생기기 때문이기도 하지만 반응절단위의 접속규칙을 그대로 이용할 수 있기 때문이다.

V. 실험 및 결과

본연구에서는 포맷트천이 부분을 포함하지 않은 기존의 반응절단위 규칙합성방식을 개선하기 위하여 포맷트천이의 변경규칙을 적용하였다. 본장에서는 이를 확인하기 위하여 포맷트변경을 수행한 결과를 원음성과 비교하였으며 데이터베이스의 구축내용과 포맷트천이의 변경규칙을 적용한 경우의 합성된 음성파형과 스펙트로그램을 원음성, 기존의 반응절단위 합성음성과 비교 검토한다.

(1) 데이터베이스의 구축 및 분석합성 조건

데이터베이스는 반응절단위의 음성과 정상모음의 음성과 각각의 LPC계수 및 잔차에너지로 구성하였다. 반응절단위는 초성자음 19개와 중성모음 21개의 조합에 의한 전반부 반응절 399개와 중성모음 21개와 중성 8개의 조합에 의한 후반부 반응절 168개의 합인 567개이며 추가된 정상모음 42개는 전반절 모음 21개와 후반절 모음 21개로 구성되며 이들을 합하여 총 609개 단위로 데이터베이스를 구축하였으며 접속될 합성단위의 연속성을 위하여 OLA[8]를 이용하였으며 포맷트천이구간의 길이와 대응하기 위한 시간축 신축 방법으로는 SOLA[9]를 이용하였다. 여기서 분석과 합성을 위해 사용한 실험조건은 표 1과 같다.

표 1. 실험조건

A/D data	10kHz sampling, 16-bit
window function	Hamming
window(frame) length	256 samples
window shift	64 samples
LPC analysis order	15

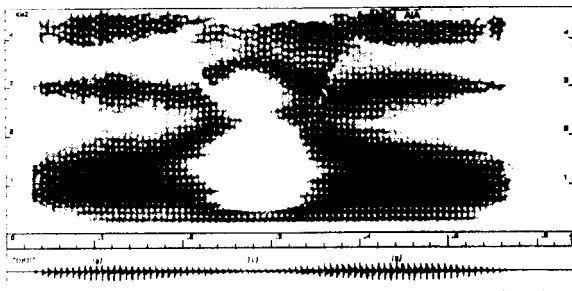
(2) 포맷트를 변경한 결과

연쇄모음의 경우 앞선 모음의 정상부위에서 연결될 모음의 정상부위 사이에 포맷트천이가 존재하며 포맷트천이를 포함한 음성파형이 앞선 모음음절과 연쇄될 모음음절을 연결한다. 이부분을 인위적으로 만들어 주기 위해 본연구에서는 2장과 3장의 포맷트천이 근사식과 포맷트 변경 분석합성시스템을 이용하여 포맷트천이를 내포한 음성파형을 작성하였으며 이를 원음성의 파형 및 스펙트로그램과 비교하고자 한다.

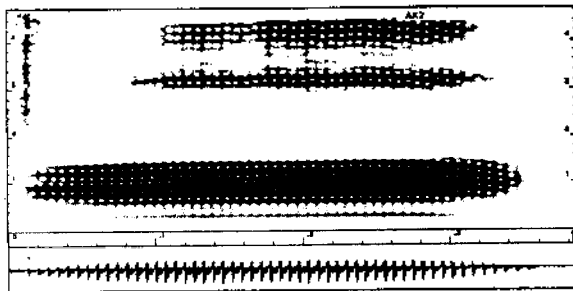
그림 5는 포맷트천이의 근사식 식(1)과 식(2) 부터 (4)의 포맷트변경을 위한 분석합성시스템을 이용하여 원음

성 모음 '아'의 제1, 제2 포먼트를 변경하였을 때 원음성 "아야"와 같은 포먼트천이를 갖도록 음성파형을 변경하였다. 이를 위해 포먼트천이가 시작되는 모음 '아' 부분(식(1)의 $F_n(0)$)에서 연쇄될 모음 '이'가 안정되는 부분(식(1)의 $F_n(T)$)까지와 다시 모음 '이' 부분(식(1)의 $F_n(0)$)에서 연쇄될 모음 '아'가 안정되는 부분(식(1)의 $F_n(T)$)의 포먼트천이가 끝나는 부분까지 식(1)을 이용하여 근사적인 포먼트천이 값을 구하였으며 식(2),(3),(4)를 이용하여 포먼트천이를 내포하는 음성파형으로 변경하였다.

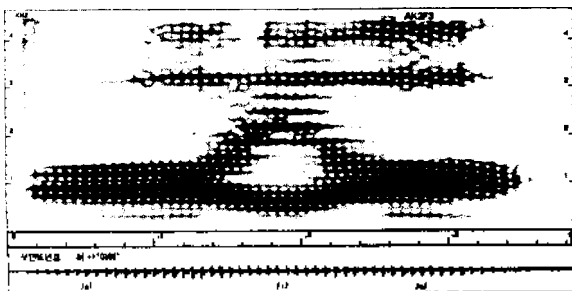
그림 5의 (c)는 원음성 (a)의 "아야"와 같도록 하기 위하여 원음성 (b)의 '아'의 파형을 포먼트변경한 결과이다. 여기서는 제1포먼트와 제2포먼트만을 변경하였지만 원음성 "아야"와 같은 음성으로 청취되었으며 음성파형에 내포된 스펙트로그램도 원음성과 같은 모양으로 변경되었음을 확인할 수 있다.



(a) 원음성 '아'



(b) 원음성 '아'



(c) 포먼트변경한 음성 '아'-'아야'

그림 5. 포먼트변경을 수행한 음성파형과 스펙트로그램의 비교

(3) 합성결과와 비교검토

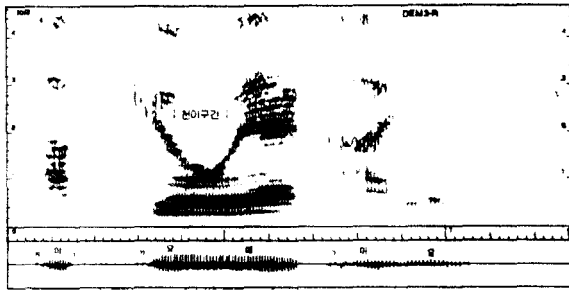
본절에서는 4장에서 서술한 포먼트천이의 변경규칙을 기존의 반응절단위 규칙합성에 적용하여 합성한 결과를 비교검토한다. 그림 6에서는 "학교에 가요"라는 원음성과 기존의 반응절단위 규칙합성에서 접속될 반응절의 파형들을 그대로 나열한 합성음성 및 본연구의 포먼트천이 규칙에 의해 변경된 파형을 그대로 나열한 합성파형을 비교하였으며, 그림 7에서는 접속될 각 합성단위를 LPC 계수의 열로 하여 LPC 디지털 필터로 합성할 경우 기존의 반응절단위 규칙합성에 의한 합성음성과 포먼트천이 규칙을 적용하여 합성한 경우의 합성음성의 파형과 스펙트로그램을 비교하였다.

그림 6의 (a)는 "학교에 가요"의 원음성으로서 특히 "-교에-"부분에 "요에"라는 연쇄모음이 있으며 그 부분에 두 모음사이의 포먼트천이구간이 그림(a)의 '천이구간'으로 표시한 부분과 같이 나타나고 있다. 그림 6의 (b)는 기존의 반응절단위의 규칙합성에 의해 합성할 음성파형을 원형 그대로 나열한 것이며 그림(a)의 원음성에서 관찰되었던 "-요에-"부분과 같은 포먼트의 천이구간이 포함되어 있지 못하다. 이는 반응절단위의 규칙합성에서는 이중모음 '요'를 처음부터 자체의 포먼트천이가 끝난 이후의 부분까지를 전반부 반응절이라 하고 그 이후부터 '요'의 모음파형의 끝부분까지를 후반부 반응절로 하여 접속규칙에 의해 규칙합성하게 된다. 따라서 기존의 반응절단위의 규칙합성에서 연쇄모음이 연결되는 경우 두 모음사이의 천이는 포함되지 않기 때문에 그림(a)의 원음성의 연쇄모음 "-요에-"인 '천이구간'과 같은 포먼트천이를 포함시키지 못하게 되므로 합성음성의 "-교에-"는 '교'음절과 '에'음절을 따로따로 발성하여 붙여 놓은 것처럼 합성되어 부자연스럽게 들린다.

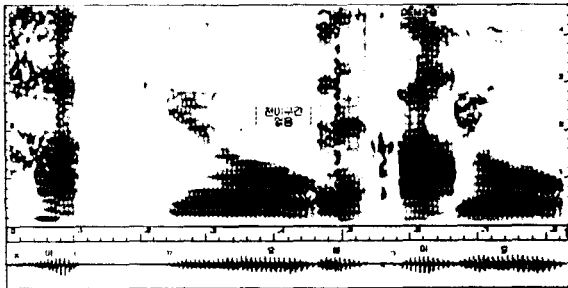
그림 6의 (c)는 본연구의 4장에서 제안한 포먼트천이의 변경규칙을 적용하여 기존의 반응절단위 규칙합성에서 연쇄모음 부분에 포먼트천이규칙을 적용하여 천이구간을 포함하도록 한 합성단위 파형을 접속한 것이다. 그림 (c)의 '요'의 후반부 반응절 부위는 본래 그림(b)의 '요'의 후반부 반응절을 STEP 3에 따라 '요'가 안정된 부분의 정상모음을 포먼트천이변경하여 대체한 것이며 '에'의 전반부 반응절 부위는 본래 그림(b)의 '에'의 전반부 반응절을 STEP 4에 따라 '에'가 안정된 부분의 정상모음을 포먼트천이변경하여 대체한 것이다. 이합성음성에서는 기존의 반응절단위 규칙합성음성과 같이 포먼트천이가 없어 기계적으로 발성된 것처럼 들리던 것을 변경규칙에 의해 포먼트천이를 내포하게 한 것이며 청취시 기존의 반응절단위 규칙합성에 의한 합성음성보다 자연스러운 면이 향상되었음을 확인할 수 있었다.

(4) 청취 MOS 테스트 평가

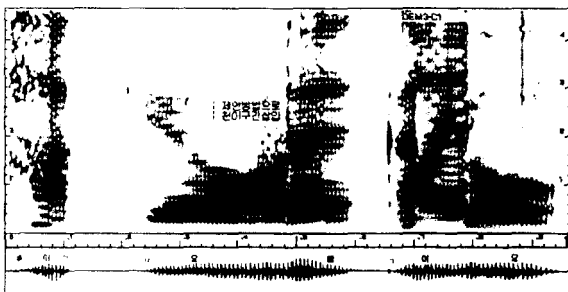
실험결과로부터 본연구에서 제안한 포먼트천이의 변경규칙을 적용하여 반응절단위 규칙합성에 의한 문장음



(a) 원음성



(b) 반음절접속합성



(c) 포만트변경접속합성

그림 6. 원음성과 파형접속에 의한 합성음성의 비교

성을 합성하였을 때 반음절단위의 접속만으로 이루어진 합성음성보다 자연성이 얼마나 개선되었는지를 평가하기 위하여 MOS 테스트로 비교하는 실험을 하였으며 여기서는 합성음성의 연쇄모음 부분에서 포만트천이에 의해 합성된 문장음성의 자연성이 어느 정도 향상되었는지를 평가하기 위한 것이며 운율조절은 하지 않았다. 다음 표 2는 MOS 테스트에서 사용한 문장이며 주로 연쇄모음이 포함된 문장으로 하였다.

MOS 테스트의 실험대상은 대학교의 연구자 10명으로 하였으며 반음절단위의 동일문장에 대해 접속만으로 이루어진 반음절단위 규칙합성에 의한 합성문장과 포만트천이 변경규칙을 적용했을 때의 합성문장을 무작위 순서로 청취시킨 후, 포만트천이의 변경규칙을 적용한 합성음성의 음질이 좋으면 1, 두 합성음성이 동일하면 0, 포만트천이의 변경규칙을 적용한 합성음성이 나쁘면 -1로 하는 상대적인 점수로 누적시킨 후 평균을 내어 명료성과 자연성을 비교하였다. 표 3은 MOS 테스트의 결과를 보이고 있다.

표 3의 MOS 테스트의 결과로 나타나는 수치는 연쇄모음을 포함하고 있는 비교적 짧은 문장을 연쇄모음에서 포만트천이가 포함된 경우와 그렇지 않은 경우 각각 청취시킨 결과를 상대적으로 나타내고 있으며 값의 크기가 0 보다 작으면 포만트천이의 변경규칙을 포함한 합성음성이 포만트천이를 포함하지 않은 합성음성보다 음질이 떨어진다는 것이고 0 보다 크면 클수록 포만트천이의 변경규칙을 적용한 합성음성이 좋다는 것을 나타낸다. 전 문장에 대한 MOS 테스트의 평균이 명료성과 자연성에서 모두 0 보다 큰 수치인 0.60과 0.95를 보인 것은 포만트천이의 변경규칙을 적용한 합성음성이 포만트천이를 포함하지 않은 합성음성보다 명료성과 자연성에서 비교적 양호해졌음을 나타내고 있음을 확인할 수 있었다. 또한 변경규칙에 의해 포만트천이가 접속됨에 따라 포만트천이가 없는 합성음성보다 자연성이 향상된 합성음이 명료성에서도 비교적 향상된 결과를 보이고 있는 것은 자

표 2. 실험문장과 그문장에 포함된 연쇄모음

실험 문장	포함된 연쇄모음
a. 온주는 <u>아름다운</u> 여인이다.	{“-아우-”, “-어이-”}
b. 그 아이는 착하다.	{“-아이-”}
c. <u>하이델이나 천리안</u> 을 쓰는 이유	{“-아이-”, “-어이-”, “-이유-”}
d. 형보다 <u>아우</u> 가 낫다.	{“-아우-”}
e. 대대로 <u>이</u> 어온 가문이다.	{“-어어-”, “-어오-”}
f. 여우는 <u>바</u> 이올린	{“-어우-”, “-아이-”, “-이오-”}
g. 원도우즈는 <u>무엇</u> 인가	{“-오우-”, “-우어-”}
h. <u>오</u> 이가 맛있다.	{“-오이-”}
i. 감사할 <u>어</u> 라	{“-어어-”}
j. <u>이</u> 에 <u>나</u> 야갈 <u>바</u> 를 <u>뵈</u> 려야 한다.	{“-이에-”, “-아이-”, “-다야-”}
k. 컴퓨터 인터 <u>페이스</u>	{“-에이-”}
l. 소프트웨어와 하드 <u>웨어</u>	{“-으워-”, “-웨어-”, “-어와-”}

표 3. MOS 테스트 결과

구분	a문장	b문장	c문장	d문장	e문장	f문장	g문장	h문장	i문장	j문장	k문장	l문장	평균
명료성	0.7	0.3	0.8	0.4	0.7	0.8	0.7	0.3	0.5	0.8	0.5	0.7	0.60
자연성	0.9	0.9	1.0	0.8	1.0	1.0	1.0	0.8	1.0	1.0	1.0	1.0	0.95

연스러운 합성음성에 의해 발음된 합성음의 명료성도 향상될 수 있는 가능성을 보인 것으로 사료된다.

Ⅶ. 결 론

본 연구에서는 반음절단위 음성합성방식에서 연속음성을 합성할 때 연속발성될 부분이 단독 발성된 음절의 연결로 구성되어 부자연스러운 합성음이 되는 것을 개선하기 위하여 포만트천이의 변경규칙을 적용하여 포만트천이를 생성하여 보상해 줌에 의해 반음절단위의 규칙합성방식을 개선하는 방법을 제안하였다. 반음절단위만으로는 포만트천이가 발생하는 부분을 채울 수 없기 때문에 반음절단위의 음성데이터 외에 정상모음 42개를 추가하여 데이터베이스를 구축하였으며 포만트를 변경하는 방법으로 포만트합성에서의 2차 공진회로를 사용하였다. 제안한 방법의 타당성을 확인하기 위하여 음성합성시 연쇄모음을 포함한 문장에 대하여 포만트천이의 변경규칙을 적용하여 비교검토한 결과 기존의 반음절단위 음성합성방식에서 합성음성이 단독모음으로 연결되어 부자연스럽게 청취되던 것을 연속발성된 음성처럼 포만트천이를 포함시켜 보다 비교적 자연스러운 합성음성을 얻을 수 있었다. 따라서 포만트천이의 변경규칙을 추가하여 반음절단위 음성합성방식을 개선할 수 있음을 확인하였다.

References

1. J. L. Flanagan, Speech analysis, Synthesis and Perception, 2nd. Ed., Springer-Verlag, Berlin, 1972
2. S. Furui, M. M. Sondhi, Advances in Speech Signal Processing, Marcel Dekker, Inc., New York, Basel, Hong Kong, pp. 741-853, 1992
3. Eric Keller, Fundamentals of Speech Synthesis and Speech Recognition, John Wiley & Sons, pp. 69-127, 1995
4. 배명진, "TTS 음성합성기술," 한국통신학회지, 제11권, 9호, pp. 761-772, 1994
5. 이용주, "반음절단위, LSP방식에 의한 한국어 음성의 규칙 합성에 관한 연구," 박사학위논문, 고려대학교, 1992. 7.
6. 공병구, 김상룡, 김정수, "이음절 접속에 의한 음절저하 및 극복대책 연구," 음성통신 및 신호처리 워크샵, pp. 279-284, 1993
7. 이기영, 최창석, 최갑석, 이현수, "타화자의 여기신호를 이용한 억양변환," 한국음향학회지, 제14권 제4호, pp. 21-28, 1995
8. F. J. Chapentier, M. G. Stella, "Diphone Synthesis Using Overlap-add Technique for Speech Waveforms Concatination," ICASSP 86, pp. 2015-2018, 1986
9. S. Roucos, A. M. Wilgud, "High Quality Time-scale Modification for Speech," ICASSP 85, pp. 493-496, 1985

▲이 기 영(Ki Young Lec) 1961년 5월 7일생



1984년 2월: 명지대학교 전자공학과 졸업(공학석사)

1986년 2월: 명지대학교 전자공학과 석사과정 졸업(공학석사)

1992년 2월: 명지대학교 전자공학과 박사과정 졸업(공학박사)

1993년 3월~현재: 관동대학교 전자통신공학과 조교수

▲최 창 석(Chang Seok Choi) 1954년 7월 15일생

1978년 2월: 홍익대학교 전자공학과 졸업

1988년 2월: 일본 金澤(가나자와)대학교 전기정보공학과 석사 과정 졸업(공학석사)

1991년 2월: 일본 金澤(가나자와)대학원 전기정보공학과 박사 과정 졸업(공학박사)

1984년 1월~1992년 2월: 산업기술정보원 책임연구원

1993년 3월~현재: 명지대학교 정보통신공학과 부교수