합성음 구현을 위한 음의 억양과 장단변환 연구

° 하정호 *, * 정재호
* 인하대학교 전자공학과 디지털 신호처리 연구실

A Study on Pitch-Rate and Time-Rate Modifications for Speech Synthesis

° Jung-Ho Ha *, * Jae-Ho Chung
* Inha University Dept. of Electronic Engr. DSP Lab.

요 약

합성의 궁극적 목표는 어휘의 제한 없이 어떠한 말이라도 자연스럽게 다양한 음색과 속도로 합성해 내는 것이다. 따라서 음성합성 (text-to-speech) 시스템의 성능은 전하고자 하는 정보를 얼마나 정확한 발음으로, 자연스럽게합성음을 만들 수 있는가에 달려 있다. 우수한 성능을 갖는 음성합성 시스템을 구현하기 위해서는 문율법에서 산출된 음의 억양과 장단변환을 효과적으로 적용시킬 수 있는 음향신호처리 알고리즘이 필요하다. 본 논문은 문율법에 따라 합성음을 적은 계산량을 유지하면서 시간영역에서 음색은 그대로 유지하면서 억양변환하고, 알맞는 속도로 장단변환하는 알고리즘을 개발하였다. 이를 이용하여 음편인 기본음(도미솔)만을 가지고 원하는 음 높이와 길이의 합성음을 산출하였다.

본 논문에서는 음의 억양과 장단변환을 위한 알고리즘을 제안하였으며, 이를 아카펠라음의 합성에 응용하였으며, 이러한 알고리즘은 자동음성서비스(ARS)나 예약 시스템 등을 적은 데이타 베이스로 다양하게 합성할 수 있음을 보였다.

1. 서 론

음성합성 시스템의 성능은 전하고자 하는 정보를 얼마나 정확한 발음으로, 자연스럽게 합성할 수 있는가에 달려 있다 [1]. 정확한 발음은 이해성(intelligence), 선명성(clarity)과 직결되며, 합성한 음성이 명확하지 않을 경우 의사전달에 문제점으로 제기 될 것이다.

자연스러운 합성음을 만들기 위해서는, 첫째, 원하는 출력(output)의 언어학적 분석, 둘째, 알맞는 억양 (intonation)과 장단(duration)을 산출 해내는 운율법, 셋째, 산출된 억양과 장단을 적용하여 합성음을 만들어 내는 디지탈 신호처리의 알고리즘 등이 성공적으로 적용되어져야 한다. 이를 위하여 S. Roucos와 A. Wilgus가 주파수영역에서 음의 억양변환 알고리즘을 제안한 바 있으며, 프랑스 CNET사에서는 음의 장단변환방법으로서 PSOLA 방법을 제안하였다[2,6,7].

본 논문에서는, 원하는 합성음을 산출해내기 위하여, 음의 억양변환은 S. Roucos와 A. Wilgus가 제안한 알고리즘 중, 필터링과정을 시간영역 FIR 필터로 구현하여 계산량을 줄였고, 장단변환은 프로토타입 내삽법(prototype waveform interpolation, PWI)을 적용하여 음성정보의 손실을 줄일 수 있도록 처리하였다. 2절에는 음의 억양변환 알고리즘 및 적용결과가 수록되어 있고, 3절에는 음의 장단변환을 위하여 적용된 프로토타입 내삽법이 설명되어 있으며, 4절에서는 음의 억양과 장단변환을 동시에 적용, 합성음을 구현하였다. 마지막으로 본 논문을 정리하는 결론과, 문제점 그리고 앞으로의 연구 방향에 대하여 제안하였다.

2. 음의 억양변환

2.1 음의 억양변환 알고리즘

음의 억양변환은 화자의 음색(tone-color)을 변환시키지 않고, 원음의 억양을 자유로이 높이거나 낮출 수 있는 알고리즘의 개발을 목표로 한다. 억양변환은 음성처리시 샘플링비(sampling rate)만 바꾸는 것을 의미하지 않는다. 즉, 샘플링비만 바꾸면 화자의 음색이 크게 달라지기 때문에, 다른 화자가 발음한 것으로 인식 되어지는 현상이 발생한다. 따라서 음색을 그대로 유지하면서 억양을 변환하려면, LPC분석에 의하여 음성신호를 구강에 대한 정보와 음원신호(residual signal)로 나누어 억양변환을 해야한다.

음의 억양변환 알고리즘은 각각 주파수 영역과 시간 영역으로 나누어서 분석을 할 수 있는 데, 이미 S. Roucos와 A. Wilgus는 저역통과 필터를 주파수 영역에서 처리하여 합성음을 산출해 냈다[2]. 그러나, 이러한 저역통과 필터는 FFT(fast Fourier transform)과정이 적어도 2회를 거쳐야 하므로 계산량이 많아지게 된다[3,4,5].

본 논문에서는 필터의 위상을 선형으로 만들 수 있고, 주파수 영역 IIR(infinite impulse response) 필터의 사용에 비하여 같은 성능을 유지하면서 계산량을 크게 줄일 수 있는 FIR(finite impulse response) 필터를 시간영역에서 설계하였다.

(그림1)에 음의 억양 변환을 위하여 본 논문에서 사용한 시스템이 설명되어져 있다. 음의 억양을 p/q배 만큼 바꾸고자 할때, 우선 LPC(linear predictive coding) 분석방법에 의하여 주어진 음성신호 $s[n]$을 음원신호($e[n]$, residual signal)와 구강에 대한 정보($h[n]$, vocal tract information)로 분리한다. 음색의 변질을 방지하기 위하여 구강에 대한 정보, 즉 LPC 계수들은 그대로 유지한다. 반면에, 분리한 음원신호에 내삽법 (interpolation), 필터링, 그리고 외삽법(decimation)을 적용하여 억양이 변환된 음원신호 $e'[n]$을 구한다. 마지막으로 변환된 음원신호 $e'[n]$는 구강에 대한 정보 $h[n]$와 함께, 원하는 만큼의 억양이 변환된 음성신호 $s'[n]$을 생성한다.

본 논문은 (그림1)에서의 필터링과정을 시간 영역에서 작동하는 FIR 필터로 설계하였다. 이러한 FIR 필터는 크게 다음과 같은 두가지의 장점을 기대할 수 있다. 첫째, FIR 필터를 사용하므로 필터의 위상을 정확하게 선형으로 유지 할 수 있다. 즉, 위상의 선형성은 비선형에 의한 위상의 왜곡(distortion) 을 막을 수 있으며, 이는 합성음의 자연성 유지를 위하여 매우 바람직한 사항이다. 둘째, 알고리즘 구현시 필터링 과정에서 요구되어지는 계산량을 줄일 수 있다. 세밀히 말하면 주파수 영역에서 필터링을 할 경우, 음원신호에 대한 FFT와 IFFT의 적용이 요구되는데, 이런 과정에서 필요한 계산량은 일반적으로 음성신호의 프레임 길이와 FFT의 크기에 비례함을 알 수 있다[3,4,5]. 따라서 주파수 영역에서의 필터링은 FFT 크기에 의존을 하고, 그 크기는 계산량과 아주 밀접함을 알 수 있다. 반면에 FIR 필터를 시간영역에서 적용할 경우에는 음원신호와 필터사이의 중첩(convolution)의 연산이 요구되므로, 필요로 하는 계산량은 FIR 필터의 길이에 비례한다. 결국 FIR 필터의 사용은, 위상을

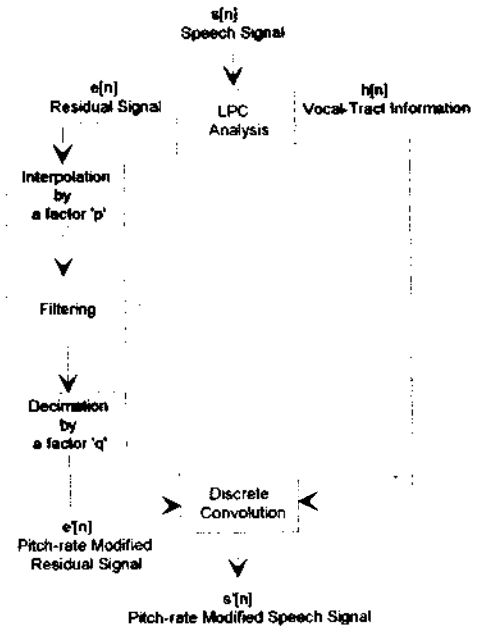

그림 1 음의 억양변환을 위한 시스템

선형으로 유지하며 계산량을 적게하면서 필터링을 이행할 수 있다. 마지막으로 필터의 임펄스 응답이 비순환(non-recursive) 하므로 필터계수의 수는 유한하기 때문에 시스템 구현이 능률적이고, 또한 효과적으로 실현할 수 있다는 장점이 있다.

2.2 음의억양 변환 알고리즘의 적용

이미 FIR필터의 차단 주파수가 π/4인 경우에 필터의 계수의 수는 전체길이의 1/10만으로도 IIR필터와 유사한 결과를 나타냄을 보였고 억양의 변환율은 0.7에서 부터 1.7까지에서 자연성을 유지함을 보였다[8]. (그림2)와 (그림3)은 이러한 변환율내에서 FIR필터계수의 개수 10개를 차단주파수 π/4에서 사용하여 구현한 합성음이다. 화자의 발음은 " 두 나라 사이에 불가침 조약이 조인 되었습니다 " 이다. (그림2)에서 원래의 음성파형이 아래이고, 위의 파형은 억양변환율을 5/4배로 변환한 합성음이다. (그림3)은 원래의 음성파형이 아래이고, 위의 파형은 억양변환율을 4/5배로 변환한 합성음이다. 합성음은 SNR이나 SEGSNR(Segmental SNR)로 비교할 수 없으므로, 일반인에 대하여 청음테스트로 하였다. 그 결과 자연성을 유지하면서 음색이 바뀌지 않음을 보였다. (그림2)와 (그림3)에서 보듯이 전체의 음의 길이는 억양변환율에 따라 늘어나거나 줄어듦을 알 수 있다. 억양변환율이 1보다 적으면 음색이 올라간 높은 소리를 내고, 반면에 1보다 크면 음색이 이와 반대현상이 나타난다.

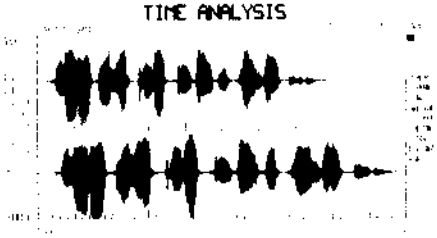

그림 2 음성신호의 억양변환의 예
(위) 변환율 4/5
(아래) 원래음성

그림 3 음성신호의 억양변환의 예
(위) 변환율 5/4
(아래) 원래음성

3. 음의 장단 변환

3.1 음의 장단 변환 알고리즘

음의 장단변환 알고리즘의 제안은, 프로토타입 파형 내삽법(prototype waveform interpolation, PWI)의 방법을 적용하였다. 프로토타입 파형 내삽법이란, 이웃 구간들 사이의 음의 장단을 바꾸고자 할 때, 먼저 각 프레임에서 그 프레임을 대표할 수 있는 한 주기의 음성 파형을 검출한다. 검출된 음성 파형들은 합성하고자 하는 음의 길이시간(duration)을 고려하여 짜집기 형식으로 이어진다.

프로토타입 파형 내삽법은 원래 저 전송율이 요구되는 환경에서의 음성 코딩을 위하여 Kleijn이 제안하였다. 본 논문은 이러한 알고리즘을 음의 장단변환에 적용하고자 시도하였다.

음성합성을 위한 프로토타입 파형 내삽법의 적용을 위하여, 먼저 20~30ms의 길이를 갖는 ,각각의 음성 프레임들로부터, 각 프레임을 대표할 수 있는 한 피치의 파형만을 뽑아 낸다. 이를 각 프레임의 프로토타입 파형(prototype waveform)이라 하고, 이러한 파형들을 운율법에서 계산된 합성음의 길이를 고려하여 선형적으로 짜집기(interpolation) 시킴으로 장단이 변환된 음을 발생시킨다. 프로토타입 내삽법은 대표적 파형들간의 짜집기이므로, 다른 음성 합성 알고리즘들에 비하여 음성 데이타의 저장량을 상당히 줄일 수 있고, 자연성을 크게 떨어뜨리지 않으면서 장단변환을 효과적으로 이행 할 수 있다는 장점이 있다.

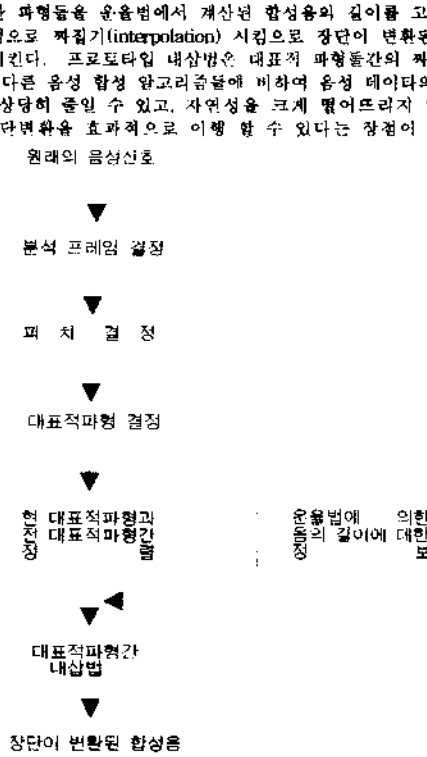

그림 4 음의 장단변환을 위한 시스템

이러한 음의 장단 변환을 위한 PWI의 알고리즘이 (그림4)에 나타나 있다. (그림4)에서는 주어진 음성에 대하여 프로토타입 파형을 결정하기 위하여 프레임별 피치를 산출하고, 결정된 피치값의 길이로 프로토타입 파형을 프레임별로 산출하여 저장한다. 결정된 프로토타입 파형들 사이에 내삽법을 적용하여 음의 장단이 변환된 합성음을 산출한다.

피치를 결정하는 방법에는 지금까지 다양한 방법이 대두되어 왔다. 그러나 본 논문에서는 AUTOC(Modified Autocorrelation Method)를 사용하였다[9]. 피치를 결정하는 블럭 다이어그램이 (그림5)에 나타나 있다. 여기에서 사용한 초기피치값을 결정하기 위한 윈도우 $W_1[n]$는 301개이고, 피치 보상을 위한 $W_2[n]$는 221개이다. 본 논문에서는 음의 장단 변환 알고리즘 개발시 한 프레임의 길이를, 8KHz의 샘플링 주파수를 갖는 음성신호의 경우, 20ms(160 샘플/프레임)씩 분석을 하였으며, 피치 결정시 사용된 윈도우들은 Kaiser윈도우를 사용하였다.

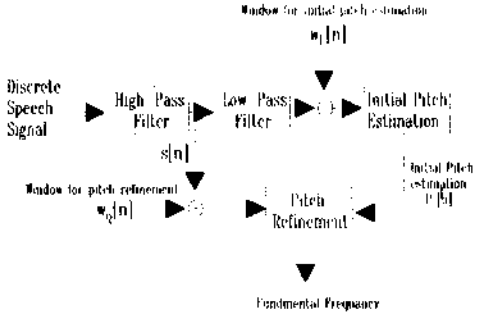

그림 5 피치 결정 알고리즘

3.2 PWI에서의 피치 사이클과 옵셋의 결정

임의의 피치 사이클을 k 라 하고, 피치 피리어드 (pitch period)를 $p(k)$라 하면, 현재의 interpolation의 시작점은 피치 사이클 k=0의 중심이 된다. 즉, 시작점에서의 피치 피리어드는 $p(0)$가 되고, interpolation의 마지막 점인 k=K에서의 피치 피리어드는 $p(K)$가 된다. 피치 피리어드들 사이의 interpolation은 식 (1)을 사용하여 k의 순서에 따라 선형적으로 수행한다.

$$p(k) = \frac{K-k}{K} p(0) + \frac{k}{K} p(K), \quad k = 0, 1, \ldots, K. \qquad (1)$$

또한, k번째 피치 사이클의 중앙을 시간 옵셋 t_k라 하면, 시간 옵셋 t_k는 첫번째 부터 k번째 까지의 피치 피리어드들을 사용하여 다음과 같이 나타낼 수 있다.

$$t_k = t_0 + \frac{1}{2} \sum_{l=1}^{k} [p(l) + p(l-1)] \qquad (2)$$

식 (2)를 사용하여 프레임당 피치사이클의 옵셋을 정한다.

전체의 interpolation영역은 처음과 마지막의 피치 피리어드 값 $p(0)$와 $p(K)$를 사용하여 아래 식과 같이 결정할 수 있다.

$$t_K - t_0 = \frac{K}{2}(p(0) + p(K)) \qquad (3)$$

따라서, PWI란 프로토타입 파형들의 선형적 삽입법임을 알 수 있다. 식 (3)에서 구한 t_k와 $p(K)$의 옵셋이 정수배로 떨어지지 않을 경우, $p(K)$ 의 옵셋과 식 (3)의 t_k의 차를 반올림하여 새로운 옵셋으로 보정한다.

3.3 PWI에서의 프로토타입 파형의 결정

현재 프레임의 프로토타입 파형을 $pw(0, t)$라 하고, 다음 프레임의 프로토타입 파형을 $pw(K, t)$라 하면, 현 프레임의 프로토타입 파형의 구간 $pw(0, t)$는 [-1/2 p(0), 1/2 p(0))이고, 다음 프레임의 프로토타입 파형의 구간 $pw(K, t)$는 [-1/2 p(K), 1/2 p(K))이다. 그러나 이 두 파형들 간의 interpolation시, 구간이 서로 다르기 때문에 둘중의 하나는 한쪽으로 피치 피리어드를 맞추어야 한다. 이런 과정을 zero-padding이라 한다. 어떠한 부분을 영으로 채울 것인가는 현 프레임의 대표적 파형과 다음 프레임의 대표적 파형을 비교하여 다음과 같은 관계식을 사용하여 결정 한다.

$$pw(m,t) = \begin{cases} pw(m,t), & -1/2p(m) \le t < 1/2p(m), \\ 0, & elsewhere. \end{cases} \qquad (4)$$

식 (4)에서 m은 0 또는 K이며, $pw(m,t)$는 새로 만들 파형을 나타낸다.

3.4 PWI를 사용한 음의 장단 변환

(그림6)은 '*canoe*'라는 음성신호로서 여성 화자에 의하여 발음되었다. 여성음의 경우에서는 20ms동안에 4개 내지 5개의 피치 주기가 있으므로, 프레임 간 분석시 10ms로 분석을 하였다. (그림6)의 아래 그림은 원래의 음성이고 윗 음성은 PWI를 사용하여 원음의 길이(duration)을 그대로 유지하면서, 프로토타입 파형들이 interpolate된 결과이다. (그림6)에서 원래의 음과 합성음을 비교하여 보면, 피치가 약간 증가한 모습을 보인다. 이러한 이유는, 마지막의 피치피리어드의 옵셋과 식(2)를 이용하여 구한 시간 옵셋을 비교하면 정수로 정확히 떨어지지 않으므로 보정이 필요하다. 따라서 이런 경우에는 옵셋을 비교하여 반올림을 하였다. 결국 (그림6)의 합성음은 이러한 영향으로 인하여, 원래의 음과 비교시 음의 길이가 약간 늘어났음을 볼 수 있다. 청음테스트에서는 원음과 프로토타입의 파형들을 interpolate하여 합성한 음사이에는 전혀 차이점을 발견할 수 없었다.

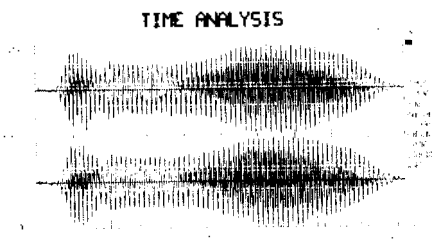

그림 6 음성신호의 장단변환의 예 (1) : " caneo "
(위) PWI를 적용하여 재구성한 음
(아래) 원래음성

한편, (그림7)와 (그림8)은 PWI 방법을 사용하여, 원래의 음의 길이를 각각 0.7 배와 1.7 배로 변환한 결과가 각각 상단에 나타나 있다. 원음의 발음시간은 샘플링 주파수가 8kHz시 374.5 ms이었다. 음의 길이를 0.7배로 바꾸었을때, (그림7)에서의 음의길이는 222.0ms이었다. 이는 0.7배로 정확히 떨어지지 않음을 알 수 있으며, (그림8)에서도 음의 길이가 584.1ms이었다. 이는 앞에서 제시한 피치 피리어드들이 정수배로 떨어지지 않음에 기인한 결과이다.

청음 테스트를 통하여 음의 길이가 줄었을때에는 잡음을 감지 할 수 없었다. 반면에 음의 길이가 늘어난 경우에는 약간의 잡음을 감지 할 수 있었다.

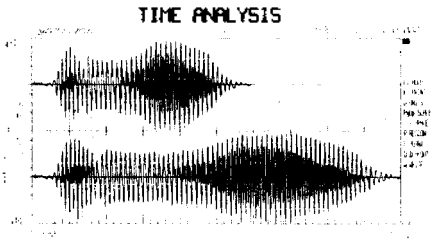

그림 7 음성신호의 장단변환의 예 (1):" caneo "
(위) 변환율 0.7
(아래) 원래음성

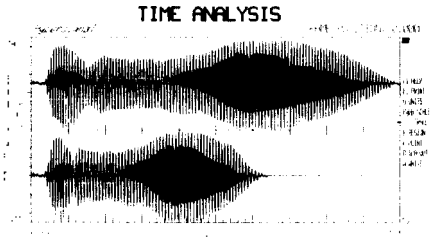

그림 8 음성신호의 장단변환의 예 (1):" caneo "
(위) 변환율 1.7
(아래) 원래음성

(그림9)와 (그림10)은 남성 화자에 의하여 발음된 음성신호로서, 발음의 내용은 '두 나라 사이에 불가침 조약이 조인되었습니다.'이다. (그림9)와 (그림10)에서는 각 프레임들의 프로토타입 파형과 피치 피리어드를 사용하여 각각 음의 길이를 0.7배와 1.7배로 하여 합성한 결과가 상단에 나타나 있다. 합성과정에서, 무성음 부분에서는 피치의 의미가 없으므로, 무성음 부분에 한하여 원음을 사용하였다. 음의 장단이 변한 합성음에서는, 자연성은 만족스럽게 유지가 되었으며, 약간의 잡음을 감지하였다.

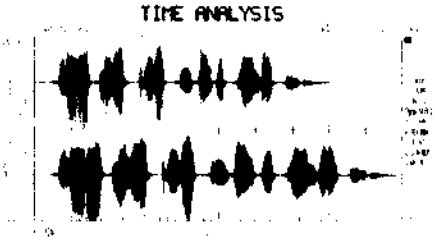

그림 9 음성신호의 장단변환의 예 (1):
"두 나라 사이에 불가침 조약이 조인 되었습니다."
(위) 변환율 0.7
(아래) 원래음성

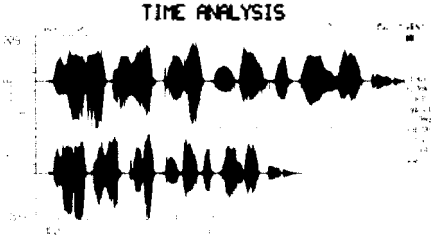

그림 10 음성신호의 장단변환의 예 (1):
" 두 나라 사이에 불가침 조약이 조인 되었습니다. "
(위) 변환율 1.7
(아래) 원래음성

3.5 음의 장단변환 알고리즘의 성능

PWI 방법에 의하여 장단이 변환된 합성음은, 원음과 비교하여 자연성이 그대로 유지됨을 알 수 있었다. 음의 길이가 짧아졌을 때는 잡음이 감지되지 않았으며, 음의 길이가 원음에 비하여 길어졌을때는 약간의 잡음이 감지되었다. 변환된 음의 길이와 잡음의 정도는 서로 비례하였다.

본 논문에서 시도한 PWI에 의한 장단변환 알고리즘을 실제의 음성합성기에 성공적으로 적용하기 위해서는, 다음의 두가지 사항들에 대한 연구가 집중적으로 필요하다. 첫째, 본 논문에서 시도한 PWI 방식은 프로토타입 파형과 함께, 피치 피리어드에 대한 정보가 요구된다. 따라서, 유성음일 경우 본 논문에서 제안한 알고리즘에 의하여 음의 장단을 변환시킬 수 있다. 반면에 무성음의 경우에 있어서는, PWI 방식에 의한 음의 장단변환은 (현재로서는) 불가능하다. 따라서, 무성음의 장단 변환시 적용할수 있는 알고리즘개발이 요구된다. 둘째는 청음테스트에서 감지된 약간의 잡음이다. 이에대한 해결을 위하여서는, 프로토타입 파형들간의 좀더 정교한 interpolation이 요구된다. 따라서, 프로토타입 파형들간의 interpolation을 시간 영역이 아닌 주파수 영역에서 행하는 방식도 고려되어져야 하리라 사료된다.

4. 아카펠라에의 응용

본 논문에서 구현한 알고리즘은 음성 합성분야에서 다양하게 이용될 수 있다. 이번 절에서는 아카펠라(acappella)곡에 앞에서 제안한 알고리즘을 적용 하고자 한다.

2절에서는 음색을 변환하지 않고, 억양을 바꾸는 알고리즘을 문장에 대하여 적용 하여 억양 변환율이 0.7 에서 부터 1.7 사이까지는 주파수 영역이나 시간영역에서의 변환이 자연성을 그대로 유지 할수 있음을 알았고, 3절에서는 원하는 장단으로 음을 생성할 수 있음을 보였다. 이번 절에서는 이를 이용하여 하나의 발성에 대하여 적용을 하고자 한다. 발성음으로는 비숙련자에 의한 "아"의 발성음으로 기본음(도, 미, 솔, 높은도)만 발성하였다. 기본음만으로 자연음계를 구성할 경우, 필요한 "레, 파, 라, 시, 높은 레, 높은 미" 음들을 합성하고, 이러한 자연음계로 다양한 음성을 구현하고자 한다.

4.1 기본음에 대한 억양변환

표1은 비숙련된 음성 데이타로 음성의 억양을 각각 도, 미, 솔, 높은 도로 부터 "레, 파, 라, 시, 높은 레, 높은 미"를 합성하여 비교 분석한 비교표이다. 레의 경우 발성음 '도'에서 합성한 음으로 프레임의 길이는 원음인 '도'보다는 작아짐을 알 수 있다. 이는 2절에서 보였다. 이러한 발원음과 생성음에 대한 분석이 표2에 설명되어있다.

'솔'음으로 생성한 음의 반음을 내린 '파'를 보면, 억양변환율이 10/9배이고, 온음을 올린 '라'의 억양변환율은 9/10임을 알 수 있다. 또한 2계음을 올리려면 억양변환율을 4/5로 해야 됨을 알 수 있다. (그림11)과 (그림12)는 표2에서 설명한 예로, 솔음에서 생성한 '파'음과 '라'음을 피치를 보여준다.

(그림11)에서 하단 그림'파'음이 원음인 상단 그림 '솔 음보다 피치가 늘어남을 알수 있다. (그림12)도 하단 그림 '라'음이 상단 그림 '솔'음보다 피치가 감소 됨을 볼 수 있다. 합성음에서는 길이가 기본음과 비교시 조금 짧아지거나 길어짐을 알 수 있다. 이는 음의 억양 변환시 피치의 길이가 늘어나고 줄어 드는 현상에 따라 나타나는 현상이다. 결국 이를 보상하여 주는 방법은 장단의 길이를 알맞게 바꾸어 주어야 함을 알 수 있다.

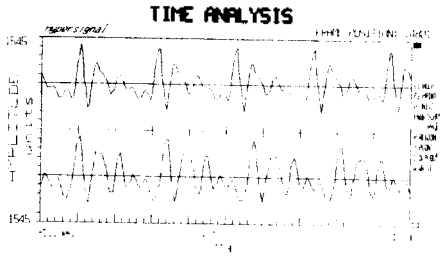

그림 11 피치 비교 : "아"
(위) 발성음 "솔"
(아래) 생성음 "파"

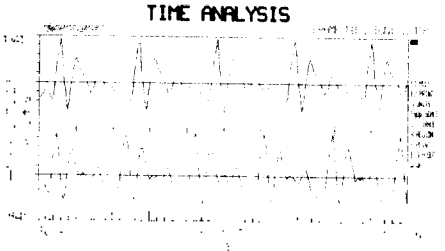

그림 12 피치 비교 : "아"
(위) 발성음 "솔"
(아래) 생성음 "라"

4.2 기본음에 대한 장단의 변환

3절에서 적용한 알고리즘을 이용하여 발성음 '아'에 대하여 장단을 변환하였다. 비숙련자에게서 받은 데이타는 순수한 유성음이므로, 4분음표의 길이라고 가정을 하면, 장단 변환율이 0.5배시 8분음표로 변환되고, 장단 변환율이 2.0배시 2분음표가 됨을 알 수 있다. 따라서 이를 이용하여 지속시간을 조절할 수 있다.

4.3 아카펠라음의 구현

4.1과 4.2에서 합성한 생성음과 발성음을 사용하여 원하는 악곡을 화음을 넣어 노래를 산출하였다. (그림13)은 자연계음을 보인 것이고 ,(그림14)는 고향의 봄을 2중주로 엘토와 소프라노로 연주한 것이다. 그 결과 피치가 연속적이 성질을 가지면 장단변환시 발생한 잡음도 줄어듬을 볼 수 있었다.

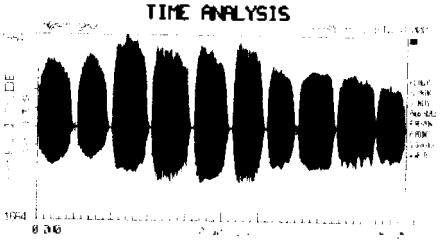

그림 13 음악에의 적용 (1):자연계음

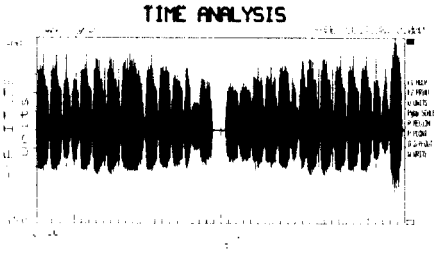

그림 14 음악에의 적용 (2):고향의 봄

5. 결 론

본 논문에서는 억양 변화과 장단 변환의 알고리즘을 먼저 문장에 대하여 각각 적용하였으며, 또한, 기본음 "아"음을 음원으로하여 원하는 음을 생성하였다. 음의 장단의 변환시 일정한 주기를 갖는 음성에 대하여서는 잡음이 적은데, 일정한 주기를 갖지 않는 음성, 예를 들면 문장이나 멜럼이 있는 음성이나 잡음이 심한 음성등에 대하여서는 아직도 약간의 잡음이 산출 되었다. 이는 프로토 타입 내삽법에 의하여 짜집기시 발생하는 잡음에 기인한다.

본 논문에서는 음의 억양과 장단변환을 위한 알고리즘을 제안하였으며, 이를 아카펠라음의 합성에 응용하였다. 본 논문에서 개발한 알고리즘을 적용하여 자동음성서비스(ARS)나 예약 시스템 등을 적은 데이타 베이스로 다양하게 합성할 수 있을 것으로 사료된다. 앞으로 개선되어야 할 과제는 무성음에 대한 장단의 변환처리와 시스템의 실시간 구현이다.

참고문헌

[1] J. L. Flanagan, *"Voice of Men and Machines"*, Journal of the Acoustical Society of America, 1972.

[2] S. Roucos and A. Wilgus, "Waveform Segment Vocoder : A New Approach for Very-Low Rate Speech Coding," *Inter. Conf. on Acoust., Speech, and Signal Proc.*, pp.236-239, 1985.

[3] A. V. Oppenheim and R. W. Schafer, *"Discret-Time Signal Processing"*, Prentice Hall, NJ, 1989.

[4] L. R. Rabiner and R. W. Schafer, *"Digital Processing of Speech Signal"*, Prentice-Hall, Inc., Englewood Cliffs, NJ 1978.

[5] Emmanuel C. Ifeachor and Barrie W. Jervis, *"Digital Signal Processing A Practical Approach"*, Addison-Wesley

[6] F. J. Charpentier and M. G. Stella, "Diphone Synthesis Using An Overlap-Add Technique for Speech Waveforme Concatenation," *IEEE Proc. Inter. Conf. on Acoustics, Speech, and Signal Proc.*, pp.2015-2018, 1986.

[7] F. Charpentier and E. Moulines, "Text-to-Speech Algorithms Based on FFT Synthesis," *IEEE Proc. Inter. Conf. on Acoustics, Speech, and Signal Proc.*, pp.667-670, 1988.

[8] J.H. Ha, J. Chung, " A Study on Low-Pass Filtering Applied to Pitch Rate Modification " *The 4th Joint Conference on Com. & Inf.* pp.460-466, 1994.

[9] Michael S. Brandstein, Peter A. Monta, John C. Hardwick, and Jae S. Lim, "A Real-Time Implementation Of The Improved MBE Speech Coder," *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Proc.*, 1990

표 1. 기본음에 대한 피치 분석

측정값	도	레	미	파	솔	라	시	높은도
1	0	0	0	0	0	0	0	0
2	67	58	55	49	43	40	35	31
3	64	57	54	48	44	39	35	32
4	64	58	53	48	44	39	35	33
5	64	57	53	49	44	39	35	33
6	64	57	54	49	44	40	35	33
7	64	57	54	48	44	40	35	33
8	64	57	54	49	44	40	35	33
9	64	57	53	49	44	39	35	33
10	64	57	52	49	44	39	35	33
11	63	57	51	49	43	39	35	32
12	64	58	51	48	43	39	35	32
13	64	58	51	48	43	40	35	32
14	64	58	52	48	44	40	35	32
15	64	58	52	49	44	40	34	32
16	64	58	52	49	45	39	34	33
17	64	58	52	50	44	39	34	32
18	66	58	51	49	44	38	35	32
19	64	59	50	49	43	38	35	32
20	64	58	50	48	42	39	34	32
21	65	58	51	47	42	39	34	32
22	65	58	51	47	43	39	33	32
23	65	58	52	47	44	38	36	32
24	65	57	52	48	44	37	0	32
25	66	58	52	48	43	37		32
26	65	63	51	49	43	42		32
27	64	0	50	48	42	0		32
28	64		50	48	41			32
29	66		51	47	44			32
30	69		55	46	50			32
31	0		61	47	0			34
32			0	49				0
33				58				
34				0				
평균	65	58	52	48	43	39	35	32

표 2. 발원음과 생성음의 억양변환 표

합성음	발원음	억양변환율	FIR필터개수
레	도	9/10	20
파	솔	10/9	20
라	솔	9/10	20
시	솔	4/5	15
높은 레	높은도	9/10	20
높은 미	높은도	4/5	15