혼합 위상 정보를 이용한 TTS 합성음 생성 알고리즘*

Speech Synthesis Algorithm Using Mixed Phase Information for TTS Systems

권 철 홍** · 이 민 규***
Chul Hong Kwon · Minkyu Lee

ABSTRACT

New speech synthesis algorithms capable of flexible prosody (especially F0) modification are desired for a high quality TTS system. TD-PSOLA is the most popular synthesis algorithm. The algorithm shows very high quality when F0 modification is limited. However, the quality degradation due to pitch epoch detection error becomes severe as the F0 modification factor becomes large. On the other hand, the vocoder framework is very flexible in F0 manipulation. The synthesized speech quality from the vocoder is far from natural human speech and suffers from buzziness. To remedy the buzzy quality from the vocoder and make more natural synthetic speech, we propose a mixed phase vocoder.

Keywords: TTS, F0 modification, Homomorphic Vocoder, Mixed Phase

1. 서 론

현재 대용량의 음성 DB를 요구하는 코퍼스 기반 TTS에 대해 활발한 연구가 진행되고 있다. 코퍼스 기반 TTS 방식은 DB가 보유하고 있는 운율을 그대로 이용하므로 운율의 제어가 필요하지 않다. 그러나 임베디드 TTS는 DB가 소용량이므로 운율의 제어가 필수적이다. 따라서 임베디드용 고품질 TTS 시스템을 위해 F0(또는 피치) 제어 범위가 넓은 합성음 생성 알고리즘이 요구된다. 현재 TTS 합성음 생성을 위한 음성신호처리 알고리즘 중에서 TD-PSOLA가 가장 널리 채택되고 있다[1]. 이 알고리즘은 구현이 간단하고 실시간 합성이 가능하며 제한된 F0 제어 범위에서 고품질의 TTS 합성음을 만들 수 있다는 장점을 갖고 있다. TD-PSOLA는 피치 동기에 맞춰 음성신호에 창 함수(window function)를 적용한다. 이 창 함수는 성문 폐쇄 시점에 중앙을 두고 있고, 그 길이는 피치 주기의 두 배이다. 이 알고리즘은 두 가지 주요 단점이 있다. 첫째, 피치 제어에 제한이 있다. TD- PSOLA는 피치 제어를 제한시킬 때는 고품질 합성음을 만든다. 이를 위해서 매우 정확하게 성문 폐쇄 시점을 찾아

* 본 연구는 과학재단 박사후 해외연수지원 결과 중의 일부입니다.
** 대전대학교 컴퓨터정보통신공학부
*** 미국 루슨트, 벨 연구소

서, 정확하면서 일관성이 있는 피치 위치의 마킹이 요구된다. 그러나 성문 폐쇄 시점의 에러는 피치 제어 범위가 증가하는 경우 합성음을 심각하게 저하시킨다[2][3]. 둘째, DB 구축에 시간이 많이 필요하다. 이 알고리즘에 필수적인 피치 위치는 피치 추출 알고리즘을 이용하여 유성음 구간에 피치 주기적으로 마킹된다. 현재의 기술 중에서 TD-PSOLA 방식에 적합한, 피치 위치를 정확히 검출할 만한 자동적인 방법은 없는 것으로 알려져 있다. 따라서 TD-PSOLA 방식을 채용한 대부분의 TTS 시스템은 성문 폐쇄 시점을 수작업으로 수정하는데, 이것은 상당한 시간이 소요되는 어려운 작업이다.

보코더(vocoder)는 음성신호에서 스펙트럼(spectrum)과 음원(excitation source) 정보를 분리한다. 보코더 방식은 TD-PSOLA 방식에 비해 두 가지 주요 이점을 갖고 있다. 첫째, 보코더는 피치 제어 범위가 넓다는 능력을 보유하고 있다. 스펙트럼과 음원 사이에 상관 관계가 적다는 것은 잘 알려진 사실이다. 보코더에서는, 스펙트럼과 음원을 분리하고, F0 제어를 음성신호가 아닌 음원에 적용하므로, F0 제어가 대부분의 음성 지각 정보를 포함하고 있는 스펙트럼 정보에 영향을 미치지 않는다. 따라서 피치 제어 범위가 좁든 넓든 상관없이 항상 유사한 품질의 고품질 합성음을 만들 수 있다. 둘째, 음성신호를 피치 비동기적으로 분석하기 때문에 피치 마킹이 필요 없어 DB 구축에 시간이 적게 든다는 이점도 갖고 있다. 그리고 보코더 방식은 음성의 특징을 대표하는 파라미터를 추출하여 음성 DB를 구축하므로 음성 DB의 크기가 작다는 이점도 갖고 있다. 그렇지만 보코더 합성음은 buzzy하다는 단점이 있어 기계음이 강한 면이 있다.

본 논문에서는 기본적으로 homomorphic 보코더 방식을 따르는 합성음 생성 알고리즘을 제안한다. 보코더의 buzzy한 음성 품질을 개선하기 위해 위상(phase) 정보를 이용한다. 합성 단계에서는 TD-PSOLA에서와 같이 overlap-add 방식을 이용하여 합성음을 생성한다.

본 논문의 순서는 다음과 같다. 서론에 이어 2장에서 켑스트럼(cepstrum) 분석과 homomorphic 보코더에 대해 기술하고, 3장에서는 본 논문에서 제안한 혼합 위상 보코더(mixed phase vocoder)의 분석 및 합성 단계를 설명하고, 4장에서 제안한 방식의 성능을 평가하고 그 결과에 대해 토론한 뒤, 5장에서 결론을 맺는다.

2. 켑스트럼 분석 및 homomorphic 보코더

2.1 켑스트럼 분석

푸리에 변환(Fourier transform) $X(e^{jw})$를 갖는 음성 신호 $x(n)$을 생각하자. $X(e^{jw})$는 다음과 같이 극 형태(polar form)로 나타낼 수 있다.

$$X(e^{jw}) = |X(e^{jw})|e^{j\arg[X(e^{jw})]} \tag{1}$$

여기에서 $|(\cdot)|$과 $\arg[(\cdot)]$는 각각 $(\cdot)$의 진폭(magnitude)과 위상(phase)을 나타낸다. $X(e^{jw})$의 복소 대수(complex logarithm) $\hat{X}(e^{jw})$는 다음과 같이 정의할 수 있다.

$$\hat{X}(e^{jw}) = \log|X(e^{jw})| + j\arg[X(e^{jw})] \tag{2}$$

위 식에서 우측의 실수부분은 우함수(even function)이고 허수부분은 기함수(odd function)이다. 참고로 $\log|X(e^{jw})|$와 $\arg[X(e^{jw})]$ 사이에는 Hilbert 변환 관계가 성립한다[4]. 음성 신호 $x(n)$의 복소 켑스트럼(complex cepstrum) $\hat{c}(n)$은 다음과 같이 복소 대수의 역 푸리에 변환(inverse Fourier transform)으로 구할 수 있다.

$$\hat{c}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi}[\log|X(e^{jw})| + j\arg[X(e^{jw})]]e^{j\omega n}d\omega \tag{3}$$

$e^{jwn} = \cos(wn) + j\sin(wn)$이므로, 식 (3)은 다음과 같이 쓸 수 있다.

$$\hat{c}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi}[\log|X(e^{jw})|\cos(wn) - \arg[X(e^{jw})]\sin(wn)]dw \tag{4}$$

식 (4)에서 실함수(real function)의 복소 켑스트럼은 실함수라는 것을 알 수 있다. 실수 켑스트럼 $c(n)$ (이후에는 켑스트럼 이라고 부르기로 하자.)은 푸리에 변환 진폭의 대수의 역 푸리에 변환으로 얻어지고, 다음과 같이 수식으로 나타낼 수 있다.

$$c(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\log|X(e^{jw})|e^{jwn}dw \tag{5}$$

식 (3)과 (5)를 비교하고, $\arg[X(e^{jw})]$가 기함수라는 사실로부터, 복소 켑스트럼 $\hat{c}(n)$과 켑스트럼 $c(n)$ 사이에는 다음 관계가 성립한다.

$$c(n) = \frac{\hat{c}(n) + \hat{c}(-n)}{2} \tag{6}$$

즉, 식 (6)을 이용하면 복소 켑스트럼 $\hat{c}(n)$에서 켑스트럼 $c(n)$을 간단히 구할 수 있다.

$X(e^{jw})$의 위상 $\arg[X(e^{jw})]$에 대해 살펴보자. FFT 처리에 의해 얻어진 원 위상(principle phase)은 modular 2π 연산 때문에 불연속적이고, 따라서 유일성(uniqueness) 문제가 발생한다. 이 문제 해결을 위해 원 위상이 연속성을 갖도록 해야 하고, 이런 위상을 unwrapped 위상이라고 부른다. 본 논문에서는 Tribolet가 제안한 위상 unwrapping 알고리즘을 이용하였다[5].

2.2 Homomorphic 보코더

음성신호는 스펙트럼과 음원으로 구성되어 있고, 다음과 같이 두 성분의 컨볼루션(convolution)으로 표현할 수 있다.

$$s(n) = h(n) * e(n) \tag{7}$$

여기에서 $s(n)$은 음성 신호, $h(n)$은 성도 필터의 임펄스 응답, $e(n)$은 음원을 나타내고, *는 컨볼루션을 의미한다. 우측의 두 성분은 그림 1에서 보인 바와 같이 homomorphic deconvolution 과정에 의해 분리될 수 있다. $s(n)$, $h(n)$, $e(n)$의 켑스트럼을 각각 $c(n)$, $\hat{h}(n)$, $\hat{e}(n)$이라고 하자. 세 성분 사이에는 다음과 같은 관계가 성립한다.

$$c(n) = \hat{h}(n) + \hat{e}(n) \tag{8}$$

여기에서 $\hat{h}(n)$은 켑스트럼 $c(n)$의 저 차수(low-time) 성분이고, $\hat{e}(n)$은 고차수(high-time) 성분이다. 그리고 $\hat{h}(n)$은 적당한 창 함수를 이용하여 $c(n)$에서 구할 수 있다. 이 함수를 리프터링 필터(liftering filter)라 부르고, 다음 식과 같다.

$$\begin{aligned} l(n) &= 1, \quad n \le P \\ &= 0, \quad otherwise \end{aligned} \tag{9}$$

P는 음성 신호의 피치 주기이다. 또한, $\hat{e}(n)$은 다음 창 함수를 이용하여 $c(n)$에서 구할 수 있다.

$$\begin{aligned} l(n) &= 1, \quad n > P \\ &= 0, \quad otherwise \end{aligned} \tag{10}$$

위에서 구한 $\hat{h}(n)$을 이용하여 음성 신호를 합성하는 과정이 그림 2에 보이고, 이를 homomorphic 보코더라고 부른다. $\hat{h}(n)$의 푸리에 변환은 음성 스펙트럼의 대수 진폭, 즉 스펙트럼 포락(envelope)이다. 임펄스 응답 $h(n)$은 스펙트럼 포락의 지수(exponential)의 역 푸리에 변환으로 구해진다. 최종적으로 임펄스 응답 $h(n)$이 음원 $e(n)$과 컨볼루션되어 합성 음성신호가 얻어진다.

위 경우에 $\hat{h}(n)$의 푸리에 변환의 위상은 모두 영이고, 따라서 합성 음성신호는 대칭적인 성질을 갖는다. 이런 임펄스 응답을 영 위상 임펄스 응답(zero phase impulse response)이라고 부른다. 영 위상 임펄스 응답과 동일한 스펙트럼 포락을 갖는 최소 및 최대 위상 임펄스 응답(minimum and maximum phase impulse response)을 구할 수 있다. 최소 위상 임펄스 응답은 식 (9)의 창 함수 대신에 다음 함수를 이용하여 구할 수 있다.

$$\begin{aligned} l(n) &= 1, \quad n = 0 \\ &= 2, \quad 0 < n \le P \\ &= 0, \quad otherwise \end{aligned} \tag{11}$$

최대 위상 임펄스 응답은 다음 창 함수를 이용하여 구할 수 있다.

$$l(n) = 1, \quad n = 0 \qquad (12)$$
$$= 2, \quad -P \le n < 0$$
$$= 0, \quad otherwise$$

최소 위상 임펄스 응답이 다른 두 경우에 비해 지각적으로 더 좋은 합성음을 만든다는 사실이 알려져 있다[6].

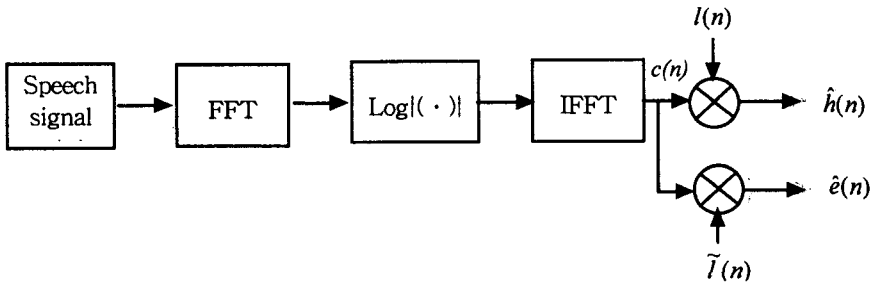


그림 1. Homomorphic deconvolution

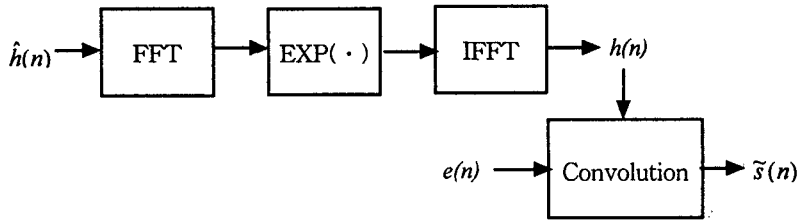


그림 2. Homomorphic vocoder

3. 혼합 위상 보코더

본 논문에서, 고품질 TTS 시스템을 위해 새로운 분석 및 합성 알고리즘인, homomorphic 보코더에 기반을 둔 혼합 위상 보코더를 제안한다. Homomorphic 보코더는 입력 음성신호를 피치 비동기적으로 분석하며, 2.2절에서 기술한 바와 같이 입력 음성에서 스펙트럼과 음원 정보를 분리한다. 입력 음성을 두 성분으로 분리함으로써 음원과 스펙트럼 정보가 독립적으로 자유롭게 제어될 수 있다.

F0 제어를 음성신호가 아닌 음원에 적용하므로, F0 제어가 대부분의 음성 지각 정보를 포함하고 있는 스펙트럼 정보에 영향을 미치지 않는다. 즉, 원 음성신호에서 음원의 F0가 변경되더라도 수정된 음성신호는 원 음성신호와 같은 스펙트럼 포락을 유지한다. 따라서 피치 제어 범위가 좁든 넓든 상관없이 항상 유사한 품질의 고품질 합성음을 만들 수 있다. 한 가지 단점은 homomorphic 보코더는 buzzy한 합성음을 만든다는 것이다. 이것은 주로 고 주파수 영역에

서 불규칙성(randomness)의 결여 때문이다. 고 주파수 영역에서 원 위상(original phase) 정보를 보유하는 것이 지각적으로 중요하다는 연구결과가 보고되어 있다[7][8]. 따라서 제안 알고리즘은 homomorphic 보코더에서 발생하는 buzzy한 특징을 제거하기 위해 위상 정보를 이용한다. 즉, 저 주파수 영역에 최소 위상을, 고 주파수 영역에 원 위상을 혼합한다. 합성 단계에서는 TD-PSOLA에서와 같이 overlap-add 방식을 이용하여 합성음을 생성한다.

3.1 분석

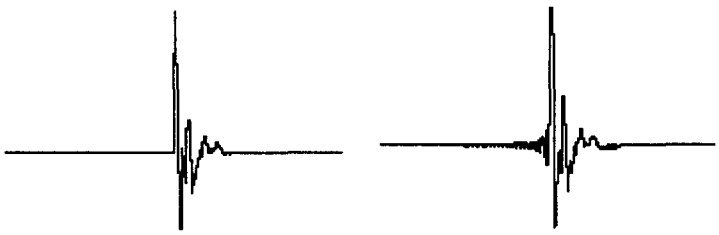
먼저 음성 프레임을 유성음과 무성음으로 나눈다. 무성음 구간은 처리하지 않고 입력 음성신호를 그대로 저장하고, 유성음 구간은 제안 알고리즘을 이용하여 처리한 뒤 혼합 위상 임펄스 응답을 저장한다. 음성신호를 단구간으로 나누어 분석하기 위해 음성신호에 창 함수를 적용한다. 이 함수는 시간과 주파수 도메인에서 등가적인 분해능을 갖는 것이 바람직하다. 본 논문에서는 이런 특성을 갖는 가우시안 창 함수(Gaussian window)를 사용했다.

$$w(n) = \exp[-\pi(t - t_c)/t_w)^2] \tag{13}$$

여기에서 t_c는 창 함수의 중앙을 나타내고, t_w는 그 함수의 길이의 반을 가리킨다. 창 함수의 길이는 평균 피치 주기의 2배가 바람직하다.

혼합 위상 임펄스 응답을 구하는 과정은, 먼저 최소 위상 임펄스 응답을 구하고 이 임펄스 응답에 원 위상 정보를 이용하여 불규칙성을 더한다. 여기에서 복소 켑스트럼을 먼저 구하고, 켑스트럼은 식 (6)을 이용하여 복소 켑스트럼에서 구할 수 있다. 최소 위상 임펄스 응답은 식 (11)의 창 함수를 사용하여 켑스트럼을 최소 위상 리프터링(minimum phase liftering) 함으로써 구할 수 있다. 이렇게 구한 최소 위상 임펄스 응답은 중앙에 최대 펄스를 갖는 특징을 갖고 있다. 이 특징은, 최대 펄스가 어디에 위치하는지 항상 알 수 있기 때문에 F0 제어에 매우 중요하다. 대조적으로, TD-PSOLA 방식은 성문 폐쇄 시점을 정확히 구하기가 매우 어려운 난점을 갖고 있다. 그런데, 최소 위상 임펄스 응답으로 합성한 음성신호는 buzzy한 음색을 갖는다. 이것은 주로 최소 위상 특성으로 생긴 인접 주기간의 불규칙성의 결여 때문이다.

본 논문에서는 최소 위상 임펄스 응답에 불규칙성을 더하기 위해 원 위상 정보를 이용한다. 복소 켑스트럼이 원 위상을 보유하고 있으므로, 복소 켑스트럼을 FFT하여 원 위상을 구한다. F0 제어를 용이하게 하기 위해, 저 주파수 영역은 원 위상 대신에 최소 위상을 사용한다. 즉, 저 주파수 영역은 최소 위상을, 고 주파수 영역은 원 위상을 혼합한다. 이것이 혼합 위상 보코더라는 이름이 붙은 이유이다. 이렇게 구한 임펄스 응답은, 최소 위상 임펄스 응답에서처럼 중앙에 최대 펄스를 갖고 있으며, 동시에 고 주파수 영역의 원 위상으로 불규칙성을 보유하게 된다. 그림 3에서 보듯이, 최소 위상과 혼합 위상 임펄스 응답 모두 중앙에 최대 펄스가 보인다. 그런데, 최소 위상 임펄스 응답은 최대 펄스의 좌측이 모두 0인 반면에, 혼합 위상 임펄스 응답은 약간의 에너지를 갖고 있다. 이것은 고 주파수 영역의 원 위상 때문에 생긴 것이다. 이것이 합성음에 인접 피치간의 불규칙성을 가져와 지각적으로 더 좋은 합성음을 만든다.



(a) 최소 위상 임펄스 응답, (b) 혼합 위상 임펄스 응답

그림 3. 임펄스 응답

3.2 합성

합성하려는 음소 열, 목표 피치와 지속시간 등이 주어지면, 단구간 음성 단위를 연결하여 목표 음소 열을 합성한다. 원 음성 단위는 상이한 피치, 지속시간을 갖는 다양한 문맥으로 녹음된다. TTS 시스템에서는 원래의 억양을 목표 억양으로 수정하는 것이 필요하다. 지속시간 제어에 관해, 원 음소 지속시간은 목표 지속시간으로 선형적으로 보정된다. 지속시간이 수정된 후, 피치 제어는 overlap-add 방식을 기초로 수행된다. 분석 단계에서 구한 혼합 위상 임펄스 응답은 피치 정보를 갖고 있지 않다. 따라서, 합성 단계에서 피치는 자유롭게 제어될 수 있다. 또한, 그림 3에서 보인 바와 같이 임펄스 응답의 감쇠(decaying) 특성 때문에 overlap-add 과정에서 스펙트럼 정보의 손실은 매우 적다.

상이한 단어에서 추출한 두 음성 단위를 연결할 때 경계에서 스펙트럼 포락의 불일치가 발생한다. 이것은 연결형 합성 방식(concatenative synthesis)의 가장 중요한 단점의 하나로 잘 알려진 사실이다. 본 논문에서 제안한 방식에서, 시간 도메인에서 인접한 두 임펄스 응답의 단순한 선형 보간(linear interpolation)은 스펙트럼의 심각한 불연속을 발생시키지 않는다. 인접한 두 임펄스 응답은 유사한 스펙트럼 포락을 갖고 있고, 그렇다면 그 둘은 또한 유사한 위상 정보를 갖게 된다. 왜냐 하면 2.1절에서 언급했듯이 위상은 Hilbert 변환을 통해 스펙트럼 포락에서 구할 수 있기 때문이다. 따라서 유사한 스펙트럼 포락과 위상을 갖는 인접한 두 임펄스 응답의 시간 도메인에서의 단순한 선형 보간은 스펙트럼의 매끄러운 연결(smooth concatenation)로 결과한다. 이것은 Dutoit의 연구 결과와 일치한다[2].

계산의 복잡도는 문제가 되지 않는다. 합성 단계에서 사용되는 혼합 위상 임펄스 응답은 분석 단계에서 off-line으로 모든 음성 단위에 대해 한 번 수행된다. DB 저장 메모리의 감축은 DPCM 방식으로 얻어질 수 있다. 일반적으로 DPCM 방식은 매 샘플에 적용되나, 본 논문에서 제안한 방식은 매 피치 주기마다 적용된다. 즉, 현 피치 주기의 샘플과 이전 피치 주기의 대응하는 샘플간의 차이가 저장된다.

4. 실험 및 평가

혼합 위상 보코더 방식과 TD-PSOLA 방식의 성능을 비교하기 위하여 청취 테스트(informal listening tests)를 수행했다. 합성 방식을 제외한 나머지 TTS 과정은 두 시스템이 같다. TTS 분야에서 십 수년간 연구경력이 있는 전문가들이 두 방식의 합성음을 평가했다. 청취 테스트 결과 몇 가지 중요한 사항이 공통적으로 지적됐다. 첫째, 혼합 위상 보코더 방식이 TD-PSOLA 방식보다 좋은 합성음을 만든다는 사실에 전문가들이 동의했다. 혼합 위상 보코더 방식의 합성음이 원 음성에 지각적으로 더 가까웠다. 또한, 혼합 위상 보코더 방식이 F0 제어의 넓은 범위에서 훨씬 더 지속적인 좋은 합성음을 생성했다. 둘째, TD-PSOLA 시스템에서는 연결 부분에서 약간의 스펙트럼 불일치가 청취자의 귀에 매우 거슬리는 합성음을 생성했다. 반면에, 혼합 위상 보코더 방식에서는 스펙트럼 포락의 매끄러운 연결이 시간 도메인에서의 단순한 선형 보간으로 얻어졌다.

주관적인 청취 테스트도 수행했다. 이 테스트는 A-B 선호도 테스트이다. 두 방식에 의해 합성된 10 문장이 10 인의 청취자에게 제시되었다. TD-PSOLA 방식에 대한 혼합 위상 보코더 방식의 상대적인 선호도를 다음과 같이 5 등급으로 채점하였다. 1) 매우 나쁘다, 2) 나쁘다, 3) 비슷하다, 4) 좋다, 5) 매우 좋다. 표 1에서 보듯이, 선호도의 전체 평균은 4.1이다. 따라서 혼합 위상 보코더 방식의 우월성이 주관적인 청취 테스트로 확인됐음을 알 수 있다.

F0 제어 범위에 따른 MOS 테스트도 수행했다. 자연 음성의 원 F0 곡선을 1/2 배에서 2 배까지 변경했다. 합성음 10 문장이 10 인의 청취자에게 제시됐다. MOS 테스트 결과가 표 2에 보인다. 제어 범위 1/1.5 배에서 1.5 배에 대한 MOS 값은 4.2이고 이는 매우 좋은 결과이다. 1/2 배와 2 배에 대한 MOS 값은 이보다 약간 낮으나 여전히 좋은 성능을 보여준다.

표 1. TD-PSOLA 방식과 비교한 혼합 위상 보코더 방식의 선호도 테스트(MOS)

Listener	1	2	3	4	5	6	7	8	9	10	Average Score
Preference Score	4.3	4.1	4.0	4.2	3.8	4.3	4.4	3.7	4.2	4.4	4.1

표 2. F0 제어 범위에 따른 MOS 테스트 결과

Modification Factor	1/2.0	1/1.5	1/1.2	1.2	1.5	2.0
MOS Value	4.0	4.2	4.2	4.2	4.2	d4.1

5. 결 론

본 논문에서는 TTS 합성 알고리즘으로 혼합 위상 보코더 방식을 제안하고, 이 방식의 분석 및 합성 과정에 대해 설명하였다. 이 방식은 homomorphic 보코더 방식에 기초를 둔다. 켑

스트럼에서 최소 위상 임펄스 응답이 구해진다. 이 임펄스 응답은 중앙에 최대 펄스를 갖는다. 따라서 F0를 제어하기가 수월하다. 그런데 최소 위상 임펄스 응답으로 합성한 음성은 buzzy한 특징을 갖는다. 이런 buzzy한 특성을 제거하기 위해, 복소 켑스트럼에서 얻을 수 있는 원 위상 정보를 이용한다. 혼합 위상 임펄스 응답은 저 주파수 영역에 최소 위상을, 고 주파수 영역에 원 위상을 혼합하여 만든다. 비공식인 청취 테스트 및 주관적인 테스트로부터 제안한 방식이 TTS 분야에서 매우 좋은 품질의 합성음을 생성한다는 것을 알 수 있었다.

참 고 문 헌

[1] Hamon, C., E. Moulines. & F. Charpentier. 1989. "A diphone system based on time-domain prosodic modifications of speech." *Proc. of ICASSP 89*, 238-241, Glasgow.

[2] Dutoit, T. 1997. *An introduction to text-to-speech synthesis.* Kluwer academic publishers, Dordrecht, The Netherlands, Ch. 10.

[3] Takano, S. & M. Abe. 1999. "A new F0 modification algorithm by manipulating harmonics of magnitude spectrum." *Proc. of Eurospeech 99*, 1875-1878, Budapest.

[4] Oppneheim, A. V. & R. W. Schafer. 1989. *Discrete-time signal processing.* Prentice Hall, NJ, Ch. 12.

[5] Tribolet, J. M. 1977. "A new phase unwrapping algorithm." *IEEE Trans. on acoustics, speech, and signal processing*, vol. 25, no. 2, 170-177.

[6] Rabiner, L. R. & R. W. Schafer. 1978. *Digital processing of speech signals.* Prentice Hall, NJ, Ch. 7.

[7] Banno, H., J. Lu., S. Nakamura., K. Shikano & H. Kawahara, 1998. "Efficient representation of short-term phase based on group delay." *Proc. of ICASSP 98*, 861-864. Seattle.

[8] Pobloth, H. & W. B. Kleijn. 1999. "On phase perception in speech." *Proc. of ICASSP 99*, Phoenix.

접수일자: 2001. 10. 24.
게재결정: 2001. 12. 2.

▲ 권철홍
대전 동구 용운동 96-3 (우: 300-716)
대전대학교 컴퓨터정보통신공학부
Tel: +82-42-280-2555 Fax: +82-42-284-0109
E-mail: chkwon@dju.ac.kr

▲ 이민규
600 Mountain Ave., Murray Hill, NJ, 07974, USA
Bell Labs, Lucent Technologies
Tel: +1-908-582-7085 Fax: +1-908-582-3306
E-mail: minkyul@research.bell-labs.com