켑스트럼 분석합성형 음성합성 시스템에서의 피치변경에 따른 스펙트럼 포락 왜곡 현상에 관한 연구

○ 김 연준, 오 영환
한국과학기술원 전산학과

Distortion of Spectrum Envelope with Change of Pitch Period in the Cepstrum Analysis-synthesis System

○ Kim, YunJoon, Oh, YungHwan
Dept. of Computer Science, Korea Advanced Institute of Science and Technology

요 약

음성합성에 있어서 음의 자연성을 향상시키는 문제는 크게 두가지로 나누어진다. 첫째는 합성음을 원음에 가깝게 구현하려는 합성방법 자체의 문제로, 언어 합성이 가지고 있는 일반적인 문제이다. 또 다른 문제는 운율에 관한 것으로 날말 또는 문장 내에서의 운율에 따라 합성음의 자연성이 좌우된다. 이러한 운율의 조절에는 지속시간, 피치, 그리고 음의 세기 등이 이용된다.

켑스트럼을 이용하여 분석합성을 하는 경우, pole-zero 모델로 스펙트럼 포락을 근사하므로 원음에 충실하고, 필터계수와 구동정보를 분리하여 분석, 합성하므로 인위적인 운율의 조절이 용이하여 음성합성이 가지는 위의 두가지 문제를 해결하는데 적합하다고 판단된다.

본 연구에서는 켑스트럼을 이용하여 분석합성 시스템을 구성하였다. 음성 합성 과정에서, 운율 조절 파라미터중의 하나인 피치 주기의 변경에 따른 스펙트럼 포락의 왜곡에 대해 살펴보고, 왜곡을 최소화하는 방안을 제안한다.

1. 서 론

음성합성은 인간의 정보 전달 수단인 음성을 기계에 의하여 합성하는 기술로 오랜동안의 연구로 실용화 단계에 있다. 근래에는 음성합성에 대한 수요가 점차 늘어나고 그 적용범위도 넓어짐에 따라 인간의 음성과 같이 자연스러운 합성음에 대한 사용자의 요구가 늘어나고 있다. 그러나 현재까지 개발된 대부분의 음성합성 시스템은 합성음의 명료성(intelligibility)의 측면에서는 어느 정도의 수준에 이르렀으나 자연성(naturalness)면에서는 만족스러운 결과를 얻지 못하고 있다.

자연스러운 합성음을 얻기 위해서는 운율처리 단계가 필요하다. 운율제어요소에는 억양, 음의 길이, 강세, 휴지기 등이 있으며, 이들을 제어하기 위하여 각각 피치 주기, 지속시간, 세기 등이 이용된다[2]. 피치 주기는 음성의 억양을 제어하는데 사용되며 문장의 의미론적인 요인과 구문론적인 요인에 의하여 영향을 받게 된다. 따라서 자연스러운 합성음을 얻기 위해서 적절한 피치제어가 필요하다.

파형코딩에 있어서의 피치 제어는 피치반분법을 사용하여 피치를 늘이고 줄이는 복잡한 과정을 반복한다[4]. 반면, 분석합성 방식의 경우에는, 분석과정에서 음원정보와 피치정보를 분리 추출, 저장하므로, 합성음의 피치변경이 용이하여 반음절 또는 음절단위의 음성합성을 가능하게 한다. 그러나 분석과정에서 추출한 원래의 피치가 아닌 피치 생성 규칙에 의하여 생성된 피치를 이용하여 합성하는 경우, 스펙트럼의 왜곡이 발생하여 합성음의 음질을 저하시키는 요인이 된다.

본 연구에서는, 피치 제어가 쉬우며 원음의 음운성을 잘 나타내는 켑스트럼 분석합성 시스템에서

피치변경에 의한 스펙트럼 왜곡에 대해 조사하고 이를 최소화하는 방안을 제안한다.

2. 합성시스템의 구성

분석합성법은 합성단위의 임펄스정보와 필터정보를 분리하여 독립적으로 저장하는 방법으로 흔히 전문적인 남성화자 1인이 독립발성한 음성을 분석하여 필터 파라미터로 저장하였다가 합성시에 필요한 파라미터를 선택하여 접속하고 이에 운율정보를 더하여 합성하게 된다[2].

합성시 사용하는 피치정보는, 분석시에 구한 피치가 아닌 합성할 문장의 의미분석 또는 구문분석을 통하여 얻은 정보를 이용하여 자동 생성되는 것이므로 원래의 피치와는 많은 차이가 있다. 따라서 문서음성변환 (Text-to-Speech) 시스템에서의 피치변경은 필수적이다.

기존의 LPC, PARCOR, LSP 등의 LPC계열의 분석합성방식은 저장해야 할 데이타의 양이 적어 메모리의 효율이 높고, 또한 분석시에 추출된 임펄스정보나 필터 정보를 합성시에 인위적으로 변경시킬 수 있기 때문에 음절이나 음소 단위의 접속에 의한 음성합성이 가능하다.

그러나, 이들 LPC 계열의 분석합성 방식은 포만트의 밴드폭이 좁기 때문에, 원음의 피치주기를 그대로 사용하지 않고 다른 피치를 사용하여 합성할 경우, 합성음의 왜곡이 큰 단점이 있다 [2].

본 연구에서 이용한 합성시스템은, 캡스트럼을 필터의 입력으로 사용하는 분석합성법으로 pole-zero 모델로써 비교적 원음과 가까운 음성의 합성이 가능하며 스펙트럼 포락과 캡스트럼 계수와의 관계가 명확하므로 주파수 변경에 의한 스펙트럼 포락의 왜곡을 관찰하기 쉽고 이를 최소화할 수있을 것으로 판단된다[3].

캡스트럼 합성시스템의 구성은, [그림 1]에서 나타난 것과 같이 기존의 LPC 계열의 분석합성방식의 형태를 따르고 있다.

기존의 LPC 계열의 분석합성방식과 다른 점은, 구동신호로부터 음성을 합성해내는 시변필터의 입력 계수가 LPC 계수대신 캡스트럼 계수라는 점이다[1].

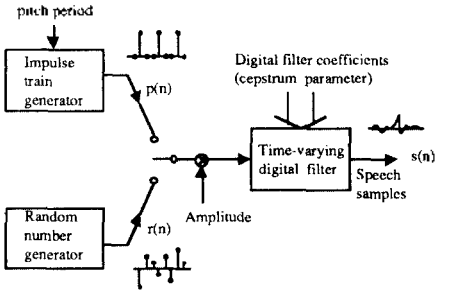


그림 1. 음성합성시스템의 구성

3. 캡스트럼 계수의 추출

본 연구에서 이용한 분석합성 시스템의 시변필터의 입력으로 사용되는 캡스트럼 파라미터를 추출하는 과정은 [그림 2]과 같다.

분석할 음성신호를 일정한 시간 간격으로 표본화하고, 표본화한 신호에 시간창을 씌워 퓨리에 변환하여 power spectrum을 구한다. power spectrum에 log를 취하여 역 퓨리에 변환한 후, 이에 lifter를 씌워 캡스트럼 계수를 구한다[1].

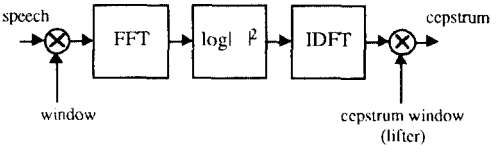


그림 2. 캡스트럼 추출과정

캡스트럼을 lifter에 통과시키면 스펙트럼 포락으로 나타내는 음운 정보와 피치 주기로 표현되는 음원 정보를 분리할 수 있다. 기존의 lifter는 [그림 2]에서 나타난 것과 같이 스펙트럼을 역 퓨리에 변환하여 구한 캡스트럼으로부터 음운 정보를 가지는 30차 계수까지만 취하여 이를 합성 필터의 계수로 사용한다.

분석합성 시스템에서 피치 주기가 변경될 경우,

스펙트럼 포락의 왜곡에 의한 합성음의 이그러짐이 나타난다. 이때 발생하는 스펙트럼 포락의 왜곡은, 스펙트럼 포락의 개괄적인 형태의 변화보다는 스펙트럼 포락의 미세구조의 변화에 의한 요인이 크므로, 미세구조의 형태를 반영하는 켑스트럼 계수의 고차부분에 가중치를 적게 준다면 왜곡을 줄일 수 있을 것으로 생각된다.

본 연구에서는 위에서의 관측 결과에 따라 다음과 같은 lifter를 제안한다.

식 (1)의 W(n)은 Blackman창 함수를 나타낸다. 제안한 켑스트럼 lifter는 식 (2)과 같이 Blackman창의 값이 0.5~1.0 사이에 위치하도록 조절하였다. 기존의 lifter와 제안한 lifter를 비교하면 [그림 3]과 같다.

$$W(n) = 0.42 - 0.5\cos\left(2\pi\frac{n}{N-1}\right) + 0.08\cos\left(2\pi\frac{2n}{N-1}\right) \quad (1)$$

$$W'(n) = \frac{1}{2} + \frac{W(n)}{2} \quad (2)$$

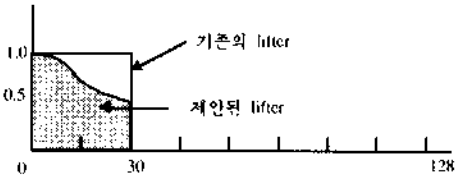


그림 3. 켑스트럼 lifter

4. 실험 및 결과 고찰

합성음의 음질을 평가하는 방법에는 주관적인 평가 방법과 객관적인 평가 방법으로 크게 분류할 수 있다. 주관적인 평가방법은 품질을 평가하기 위해 소비되는 시간이 길며, 품질 평가 결과가 평가하는 사람의 심리적 환경에 매우 민감하게 작용하는 문제점이 있다[5].

본 연구에서는 이러한 문제점을 고려하여 합성음의 품질을 객관적으로 평가하는 방법을 적용하였다. 객관적으로 음질을 평가하는 방법은 기준음성(원음성)과 비교음성(합성음)과의 신호특성의 거리차로 왜곡을 정의할 수 있으며, 그 방법에는 파형 왜곡 비교, 스펙트럼 왜곡 비교, 그리고 스펙트럼 포락 왜곡의 비교 등이 있다.

피치변경에 의한 합성음과 원음과의 비교는, 스펙트럼 포락의 왜곡에 대하여 비교하였다. 그 비교 척도로는 켑스트럼 거리 척도(cepstrum distance measure)를 사용하였다.

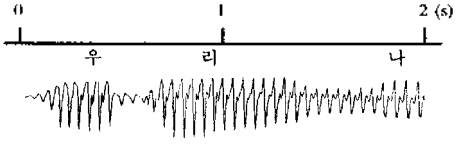
식 (3)은 켑스트럼 거리 척도를 관계식으로 나타낸 것으로, 합성음과 원음의 단구간 스펙트럼 포락의 비교를 위하여 각각의 스펙트럼 포락을 근사하는 켑스트럼 계수 간의 유클리디안 거리를 측정하였다. 평가는 임의의 구간에서의 표본을 30개 추출하여 켑스트럼 계수간의 유클리디안 거리를 계산하고, 이의 평균거리 $\bar{d}$와 분산을 구하였다.

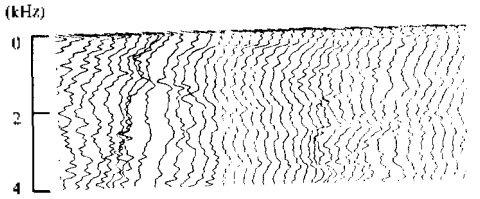
$$\bar{d} = \frac{1}{N}\sum_{i=1}^{10}(c^i - c_i^0)^2 \quad (3)$$

실험결과는 [표 1]에서 나타난 것과 같이 제안된 lifter에 의하여 분석합성한 경우가 기존의 방식으로 분석합성한 경우보다 원음과의 평균거리면에서 더 가깝게 분포하는 것으로 나타났다. [그림 4]는 원음과 합성음들의 스펙트럼 포락을 나타낸 것이다.

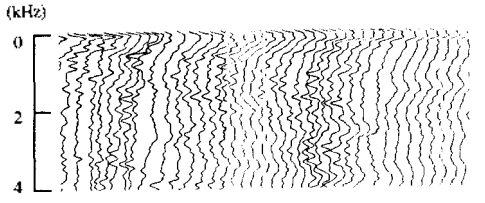
pitch	기존 lifter		제안된 lifter	
	평균거리	편차	평균거리	편차
org.	0.088	0.011	0.031	0.01
x 0.5	0.263	0.021	0.132	0.018
x 0.7	0.158	0.02	0.113	0.02
x 1.2	0.101	0.012	0.100	0.013
x 1.5	0.99	0.011	0.089	0.01

표 1. 기존의 lifter와 제안된 lifter와의 비교

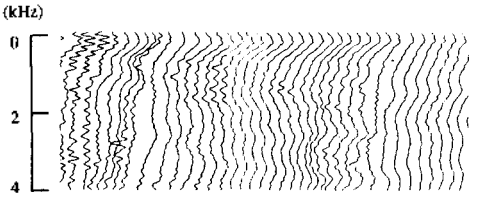




(a) 원 음



(b) 기존 lifter에 의한 합성음



(c) 제안된 lifter에 의한 합성음

그림 4. 스펙트럼 포락에 의한 비교

5. 결 론

본 연구에서는 피치변화에 따른 스펙트럼의 왜곡에 대해 조사하였고 이를 최소화하는 캡스트럼 lifter를 제안하였다.

기존의 lifter와 제안된 lifter의 비교 평가를 스펙트럼 포락의 근사치인 캡스트럼 계수 간의 유클리디안 거리를 이용하여 수행한 결과, 제안된 lifter에 의해 분석합성하였을 경우에 원음에 더 가까운 합성음을 생성한다는 것을 알 수 있었다.

앞으로의 연구에서는 좀더 많은 어휘를 대상으로 하여 실험하여 제안한 방식에 대하여 성능 평가를 수행하여, 피치 변화율과 캡스트럼의 가중치간의 명확한 관계를 규명하여야 할 것이다.

6. 참고 문헌

[1] A. V. Oppenheime, R. W. Schafer ; Discrete-Time Signal Processing, Prentice-Hall, 1989

[2] S. Furui ; "Digital Speech Processing, Synthesis, and Recognition", Dekker, 1992

[3] S. Imai ; "Log Magnitude Approximation Filter", IEICE pp. 886~893, December 1980

[4] 민경중, 배명진, 윤회상, 안수길 ; "음성 파형코딩의 음원피치 변경에 관한 연구", 한국음향학회 학술발표회 논문집 pp. 45~49, 1991

[5] 홍진우, 유경완, 김순협 ; "규칙합성음의 객관적 품질평가에 관한 연구", 한국음향학회 학술발표회 논문집 pp. 67~72, 1991