Duration Control에 의한 G.723.1 보코더 전송률 개선에 관한 연구

장 경 아, 유 영 민, 배 명 진
숭실대학교 정보통신공학과
156-743 서울시 동작구 상도5동 1-1

A Study on Improvement of Bit Rate using Duration Control of Speech in G.723.1 Vocoder

KyungA Jang, YoungMin You, MyungJin Bae
Dept. of Telecomm. Engr., Soongsil Univ. Seoul 156-743, Korea

요 약

CELP계열의 부호화기인 G.723.1 5.3kbps ACELP를 기반으로 하여 음질을 유지하면서 전송률을 낮출 수 있는 새로운 부호화 방법을 제안한다. 본 논문에서 적용한 부호화 방법은 음성 합성시 파라미터로 사용되는 지속시간 변경에 의해 CELP형 보코더의 전송률을 감소하고자 한다. 먼저 음성을 보코더 입력단에 입력하기 전 지속시간을 FFT 변환 특성을 이용해 음색의 변경 없이 지속시간을 줄임으써 계산시간을 줄이고 진폭과 위상 각각 1/2"배의 interpolation과 Decimation을 수행하여 부호화한다. 이렇게 부호화된 데이터는 G.723.1 복호화를 거치고, 다시 FFT point의 1/2"배 point로 IFFT과정을 수행함으로써 스팩트럼의 변경 없이 지속시간을 변경하여 원 음성을 합성하게 된다. G.723.1 보코더를 통과한 후 파형을 복원 실험한 결과 기존의 5.3kbps ACELP보다 46%정도 감소하였다.

I. 서 론

현재까지 발표된 음성부호화기 중 가장 많은 연구가 이루어지고 있는 방식은 CELP(Code Excited Linear Prediction)구조이다. 이러한 구조의 보코더들은 낮은 전송율에서 양호한 음질을 얻을 수 있으며 ITU-T 국제표준화 기구를 통해 다양한 응용분야에서 표준화가 이루어지고 있다. 특히 PCS 및 전화기 라인상에서의 인터넷을 통한 화상회의를 위하여 낮은 전송율에서 고음질을 가지는 코덱이 많은 주목을 받고 있다.

이러한 CELP 계열 보코더들 중에서 인터넷폰 및 화상통신용 음성 부호화기로 ACELP/MP-MLQ(algebraic CELP/Multipulse Maximum Likelihood Quantization)의 5.3/6.3kbps dual rate를 G.723.1 권고안으로 선정하였다.

CELP계열의 부호화기인 G.723.1 5.3kbps ACELP를 기반으로 하여 음질을 유지하면서 전송률을 낮출 수 있는 새로운 부호화 방법을 제안한다.

본 논문에서 적용한 부호화 방법은 음성 합성시 파라미터로 사용되는 지속시간 변경에 의해 CELP형 보코더의 전송률을 감소하고자 한다. 논문의 2장에서는 FFT의 변환 특성을 이용하여 음색의 변경 없이 지속시간을 변경해 주는 방법을 설명하고, 3장에서는 제안한 알고리즘에 대해 설명하고, 4장에서는 제안된 알고리즘의 실험 및 결과를 설명하고 마지막으로 결론을 내리겠다.

II. 지속시간 변경법

일반적으로 지속시간은 음성의 속도나 말의 리듬을 결정하며, 강세나 의미의 강조 등을 나타내는 중요한 정보를 포함하고 있다.

지속 시간 변경법은 음성을 합성시 자연성을 부가시키기 위한 파라미터로 사용되는 중요한 요소이며 이에 대한 방법에는 다음과 같다.

1. Klatt가 제안한 지속시간에 관한 규칙

지속시간을 변화시키는 요인을 언어학 외적인 요인, 구문론 적인 요인, 음성학적인 용인, 생리적인 요인 등

으로 구분하고 각각의 요인들에서 화자의 심리와 발성 속도, 강조, 구/절의 경계, 간어내의 음절의 위치, 장모음, 단모음, 파열음, 마찰음, 휴지기 등에서의 지속시간 규칙을 정하여 합성 시에 적용하려 하였다 이 모델은 아래의 식과 같다.

$$DUR=((INH역-MIN역)*PRCNT)/100+MINDUR \quad (2.1)$$

여기서 INDUR은 한 세그먼트의 고유 지속시간을 msec 단위로, MINDUR은 한 세그먼트의 최소 지속시간을 msec로 표현한 것이고 PRCNT는 1부터 10까지에 의해 정해지는 백분율의 약어이다.

2. Lindblom이 제안한 swlthrtlrks에 관한 규칙

주어진 음절의 앞에 있는 음절의 개수와 뒤에 따라오는 음절의 개수에 관한 식으로 단지 음절의 영향만을 고려하였다.

$$s=\frac{D}{(a+1)^{\alpha}+(a+b)^{\beta}} \quad (2.2)$$

여기서 S : 음절의 지속시간, a: 뒤따라 오는 음절의 개수이고, b : 선행하는 음절의 개수, α, β : 조절 파라미터이다.

3. PWI(Prototype Waveform Interpolation) 방법

시간영역 변경법 중에서 비교적 간단한 방법으로 프로토 타입 파형 내 삽입 및 삭제에 의한 지속시간 변경법이 있다.

이 방법은 저 전송률이 요구되는 환경에서의 음성 코딩을 위하여 klejin에 의해 제안되었다. 파형 내 삽입 및 삭제란, 이웃 구간들 사이의 음의 장단을 바꾸고자 할 때, 먼저 각 프레임에서 그 프레임을 대표할 수 있는 한 주기 음성파형을 검출하고 검출된 음성 파형들은 합성하고자 하는 지속시간을 고려하여 선형적으로 파형 내 삽입 및 삭제를 통해 지속시간을 변경시키는 방법을 말한다.

이 방법은 다른 음성합성 알고리즘에 비해 음성데이터의 저장량을 줄일 수 있고, 자연성을 크게 떨어뜨리지 않으면서 장단의 변화를 효과적으로 수행 수 있다는 장점을 가지고 있다. 하지만 정확한 피치 검출에 의한 프로토 타입 파형이 결정되지 않으면 파형의 삽입 및 삭제 시 음질이 열하가 심하게 나타난다. 또한 단순히 삽입 및 삭제를 통해 지속

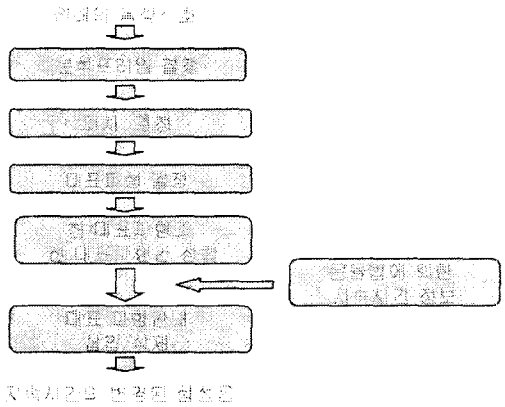

그림 2-1. PWI 지속시간 변경시스템 블록도

시간을 변경하기 때문에 변경율이 정수배가 되지 않을 경우, 정확한 지속시간 변경이 어렵다는 단점을 가지고 있다. 그림 2-1은 PWI 방법에 의한 지시간 변경 시스템에 관한 블록도를 나타낸다.

III. 제안한 알고리즘

본 논문에서는 보코더 단에 입력 전에 지속시간을 변경하는 방법으로 FFT변환 특성을 이용해 음색의 변경 없이 지속시간을 변경하는 방법을 사용하였다. 본 방법은 주파수 영역에서의 지속시간 변경 법으로 FFT를 이용하여 계산시간을 줄이고 진폭과 위상에 각각 1/2"배의 Decimation을 수행한 다음 G.723.1 보코더 입력하여 부호화시킨 후 G.723.1 복호화 단을 통과 후 FFT point의 1/2" point로 IFFT과정을 수행함으로써 스펙트럼의 변경 없이 지속시간을 변경, 음성을 합성하였다. 우선 Frame 단위로 Segment된 음성신호를 FFT 통해 진폭성분과 위상성분으로 나눈다. 프레임 단위의 음성신호에 대하여 DFT를 하면 아래의 식과 같다.

$$S_{fr}(k)=\sum_{n=0}^{N-1}s(n)e^{-j2\frac{\pi}{N}kn},\ 0\le n\le N-1 \quad (3.1)$$

이 신호는 실수부와 허수부로 나눌 수 있는데 아래의 식과 같다.

$$s_{fr}(K)=Re[S_{fr}(K)]+jIm[S_{fr}(K)] \quad (3.2)$$

진폭과 위상은 다음 식과 같이 정의된다.

$$(K)=10\log S_{fr}^{2}(K) \quad (3.3)$$

$$\varphi(K) = \tan^{-1}(Im[S_{fr}(K)]/Re[S_{fr}(K)]) \quad (3.4)$$

이렇게 얻어진 각 진폭과 위상선분에 1/2"배로 주파수 축에서 Decimation과정을 수행한다.

그런 다음 1/2" 배 point로 FFT를 수행하여 지속시간이 1/2정도 줄어든 음성을 얻어낼 수 있었다.

다음그림은 FFT된 음성의 진폭과 위상스펙트럼에 원신호의 1/2배로 Decimation하기 전과 Decimation한 후의 진폭과 위상 스펙트럼을 나타낸 것이다. 각각 진폭과 위상스펙트럼은 음질에 크게 영향을 주지 않는 범위 내에서 거의 변하지 않았으며 비율만 1/2배로 된 것을 알 수 있다.

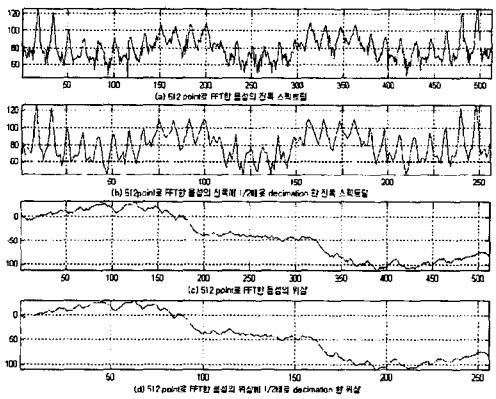

그림 3-1. 지속시간이 변경된 음성의 진폭과 위상 스펙트럼

(a) 512point로 FFT한 음성의 진폭 스펙트럼

(b) 512point로 FFT한 음성의 진폭에 1/2배로 Decimation 한 진폭 스펙트럼

(c) 512point로 FFT한 음성의 위상

(d) 512point로 FFT한 음성의 위상에 1/2배로 Decimation 한 진폭 스펙트럼

/일기예보 아나운서 음성시료(여자 아나운서)/

Ⅳ. 실험 및 결과

본 논문에서 제안한 방법을 시뮬레이션하기 위해 IBM-PC/PentiumIV(1GHz)에 마이크 입력이 가능한 16비트 A/D변환기를 인터페이스하여 8kHz의 표본화율로 16비트 양자화하여 저장하였다. 시뮬레이션 시 피치분석 프레임단위를 240표본으로 사용하였으며, 부프레임 길이는 60표본으로 하였다. 피치주기 단위로 부호화하였다. 처리결과의 성능을 측정하기 위해 다음의 대표적인 문장을 연령층이 다양한 남녀 5명의 화자가 각 5번씩 발성하여 시료로 사용하였다. 시료는 두드러진 피크를 가지지 않고 잡음이 30dB를 가진 방에서 녹음하였다.

발성1: "인수네 꼬마는 천재소년을 좋아한다."
발성2: "창공을 날으는 인간의 도전은 끝이 없다."
발성3: "예수님께서 천지창조의 교훈을 말씀하셨다. "
발성4: 일기예보 아나운서 음성시료

원 음성은 일반인을 이용하여 채취하였는데 그 이유는 훈련된 발화자보다 일반 사용자의 음성을 정확히 반영한다고 볼 수 있기 때문이다. 그리고 주관적 음질 평가를 하기 위해 MOS(Mean Opinion Score)를 사용하였으며 P.81을 준수한 MNRU Ver.2.0을 사용하였다. 제안한 방법을 C-언어와 MATLAB으로 구현하여 5.3kbps ACELP (ITU-T 표준안 G.723.1) 보코더에 적용하였다. CELP형 보코더인 G.723.1 보코더에 음성을 입력하기 전 전처리 과정으로 지속시간을 음질과 위상의 왜곡없이 1/2"배로 줄이기 위해 다음과 같은 과정을 수행해 준다. 부호화단에서는 입력 음성이 G.723.1 보코더 encoder에 입력되기 전 진폭과 위상을 1/2"배로 Decimation을 수행하여 처리하고, G.723.1 복호화단 통과한 후 부호화단에서 지속시간에 대해 변경한 만큼의 2"배의 interpolation을 처리하여 음성을 복원시킨다. 본 논문에서 제안한 알고리즘을 적용한 CELP형 보코더는 G.723.1 dual rate 5.3-6.3kbps 보코더이다. 제안한 알고리즘을 실행한 결과는 다음과 같다. 표 4-1에서는 G.723.1 보코더를 통과한 후 측정된 전송률에 대한 표이다. 전송률은 제안한 방법이 기존의 5.3kbps ACELP보다 46% 감소하였다. 그림 4-3에서는 지속시간 변경 전의 녹음된 원래의 음성 파형과 G.723.1 보코더 입력하기 위해 1/2배의 지속시간을 변경한 음성 파형을 나타내고 있다. 마지막 그림은 G.723.1 보코더를 통과한 후 음성 파형을 나타낸다.

V. 결 론

CELP 부호화기는 선형 예측 합성에 의한 분석 부호화의 원칙에 기본을 두고 있다. 이 중 G.723.1은 5.3/6.3kbps의 이중 전송률을 갖는 구조로 되어있다. 그러나 G.723.1 역시 음성신호를 성분 분리하여합성하는 방식인 CELP 보코더 계열의 합성에 의한 분석방법을 사용하기 때문에 많은 계산량으로 인한 처리 시간의 소모를 피할 수 없다는 문제점을 갖고 있다. G.723.1은 두개의 서로 다른 보코더를 포함하고 있어 DSP칩으로 구현 시 많은 내부 메모리와 계산량을 필요로 한다.

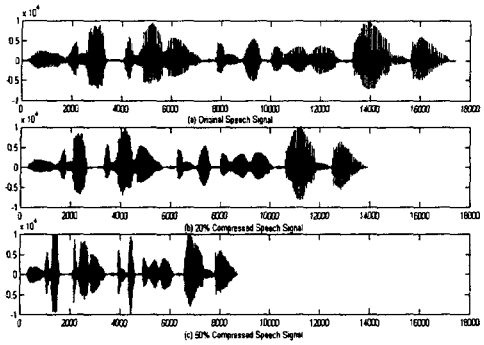

그림 4-1. 원음성 시료를 일정한 비율로 Decimated 한 음성파형
/인수네 꼬마는 천재소년을 좋아한다/(발성1)
(a) 원음성파형
(b) 20% decimated한 음성파형
(c) 50% decimated한 음성파형 (인코더 들어가기 전의 음성파형)

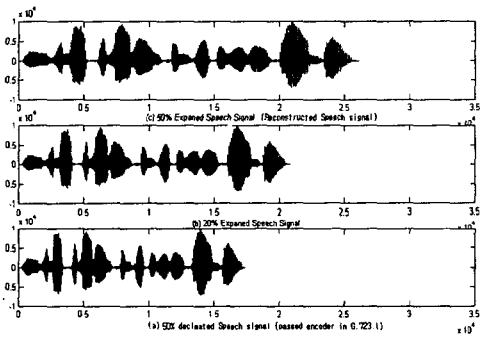

그림 4-2. 원음성 시료를 일정한 비율로 Interpolat -ed한 음성파형
/인수네 꼬마는 천재소년을 좋아한다/(발성1)
(a) 원음성파형
(b) 20% Interpolated한 음성파형
(c) 50% Interpolated한 음성파형 (디코더 나온 후음성파형)

논문에서는 G.723.1 5.3kbps ACELP를 기반으로 하여 음질을 유지하면서 전송률을 5kbps정도로 낮출 수 있는 새로운 부호화 방법을 제안한다. 본 논문에서는 음성 데이터를 G.723.1 보코더 입력하기 전에 전처리 단을 이용하여 전송율을 감소하고자 한다. 여기서 적용한 음성부호화법은 기존의 파형 압축방법과는 전혀 다른 방법으로 입력 음성의 지속시간을 FFT변환 특성에 의해 변경하는 방법으로 본 방법은 주파수 영역에서 FFT를 이용하여 계산시간을 줄이고 진폭과 위상에 각각 1/2"배의 Decimation을 수행한 다음 G.723.1 보코더 입력하여 부호화 시킨 후 G.723.1 복호화단을 통과 후 FFT point의 1/2" point로 IFFT과정을 수행함으로써 스펙트럼의 변경없이 지속시간을 변경, 음성을 합성한 실험 결과 기존의 5.3kbps ACELP보다 46%정도 감소하였다.

표 4-1. 전송률 비교

	G.723.1 (5.3kbps)	Proposed Method	Degradation bs
발성 1	5.251	2.6529	2.0478
발성 2	4.656	2.0486	2.6074
발성 3	5.044	3.7241	2.6734
발성 4	4.999	2.9954	2.0036

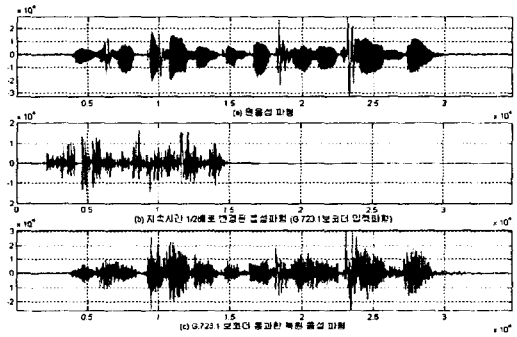

그림 4-3. 원 음성파형과 G.723.1 보코더를 통과한 파형과의 비교(남성 화자)
/인수네 꼬마는 천재소년을 좋아한다/(발성1)
(a) 원 음성시료
(b) G.723.1 보코더 입력 전 음성파형
(c) G.723.1 보코더 통과 후 복원된 음성파형

참고 문헌

[1] 배명진, "디지탈 음성분석", pp.95-120, 동영출판사, 1998. 4.
[2] L R. Rabiner, R.W Schafer, " Digital Processing of Speech Signal", pp.38-115, Prentice Hall, 1978.
[3] A. M. Kondoz, "Digital Speech", pp. 84-92, John Wiley & Sons Ltd, 1994.
[4] ITU-T Recommendation G.723.1, March, 1996.
[5] Panos E. Papamichalis, Practical Approaches to Speech Coding, Prentice-Hall, 1987.