가변전송률 CELP 부호화기 설계를 위한 발성률 비교 분석에 관한 연구

장경아, 민소연, 배명진

A Study on a comparison and analysis of Speaking rate estimation for adaptive bit rate on CELP vocoder

JangKyungA, MinSoYeon, BaeMyungJin

Abstract

음성 부호화 기술은 전송률과 복잡도를 줄이고 음질을 향상시키는 방향으로 진행되고 있다. 현재 상용화되고 있는 CELP형 보코더는 낮은 전송률에 비해 우수한 음질을 제공한다. 본 논문에서는 기존의 방식과 다르게 보코더 단에 입력 음성이 들어가기 앞서 전처리 기법을 수행하는 전처리단을 부가하여 전송률을 낮추는 방법을 소개하고, 소개된 방법들을 각기 비교하고 분석하고자 한다. 전처리기법들은 음성 인식이나 합성에서 사용되는 파라미터들을 적용시켰으며, 처리시간이나 계산시간에 있어 기존의 방식에서 많은 영향을 미치지 않은 간단한 알고리즘으로 구현하였다. 소개하는 전처리단에서는 기존의 코딩방식에서 사용하지 않은 파라미터들, 발성율, 지속시간, PSOLA 방식들을 이용하였다.

1. 서론

G.723.1 보코더는 인터넷 폰이나 화상회의, voice mail system, voice-pager 등에 응용이 가능하며 현재 상용버전으로 나와 사용되고 있다[2]. 이 중 G.7723.1은 5.3/6.3kbps의 이중 전송률을 갖는 구조로 되어있다[1]. 최적의 전송 환경을 위하여 두 개의 전송률을 사용하기 때문에 다른 보코더 표준안들에 비해서 더욱 융통성이 높다. 그러나 G.723.1 역시 음성신호를 성분 분리하여 합성하는 방식인 CELP 보코더 계열의 합성에 의한 분석방법을 사용하기 때문에 많은 계산량으로 인한 처리시간의 소모를 피할 수 없다는 문제점을 갖고 있다[3]-[5]. G.723.1은 두개의 서로 다른 보코더를 포함하고 있어 DSP칩으로 구현시 많은 내부 메모리와 계산량을 필요로 한다. 논문에서는 G.723.1 5.3kbps ACELP를 기반으로 하여 음질을 유지하면서 전송률을 낮출 수 있는 새로운 부호화 방법을 소개하고, 이 방법들의 분석하고자 한다. 본 논문에서는 음성 데이터를 G.723.1 보코더 입력하기 전에 전처리단을 이용한다. 본 논문에서는 발성속도에 따라 전송률을 달리 적용하여 효율적으로 부호화 방법과 LSP의 특성을 이용하여, 음성파형 내 음소가 변화하는 구간을 측정함으로써 발성율을 이용하여 압축한 방법 두 가지를 비교 분석해 본다. 스펙트럼 기울기를 이용한 방법은 발성속도가 빠를 때는 프레임 크기를 줄여 시간적으로 빠르게 변화하는 성질에 초점을 맞추고, 반대로 발성속도가 느릴 때는 프레임 크기는 키우고 각 프레임의 파라미터 표현에 비트를 더 할당한다. LSP를 이용하여 발성속도를 측정한 방법은 발성율 측정에 따라 음성구간 지속시간을 변경하여 압축한다.

2. 스펙트럼 기울기에 의한 발성측정법

발성속도는 1초의 음성신호 중 포함된 음소의 수로 정의할 수 있다. 본 논문에서는 음소를 직접측정하지 않고 단지 스펙트럼의 변화를 추정하여 음소의 변화를 측정하였다. 그리고 스펙트럼 변화가 3프레임 60msec동안 발생하지 않으면 발성속도가 느린 부분으로 간주하고, 3프레임동안 스펙트럼의 변화가 발생하면 빠른 발성 부분으로 간주한다. 식 1은 스펙트럼 변화를 추정하기 위해 사용한 파라미터를 나타낸 것이다

$$D_n = \frac{R_n(0)}{R_n(1)}, \quad R_n(j) = \sum_{m=0}^{N-1-j} s_n(m)s_n(m+j) \quad (1)$$

D_n은 스펙트럼 기울기를 나타내는 파라미터이고, $R_n(j)$는 자기상관계수이다. 그림 1에서 4까지는 60msec동안의 스펙트럼 기울기의 변화도를 나타낸 것이다. 프레임 크기를 30msec로 하여 15msec 겹쳐 분석하였다. 그림 1, 2, 4의 경우에는 변화가 일어나는 부분이고, 3은 변화가 일어나지 않는 구간이다. 그림 (b)는 각 구간의 스펙트럼 기울기를 나타내고 있다. 세 개의 그래프가 차이가 없으면 발성속도가 느린 부분으로 간주하고, 차이가 크면 발성속도가 빠른 구간으로 간주한다.

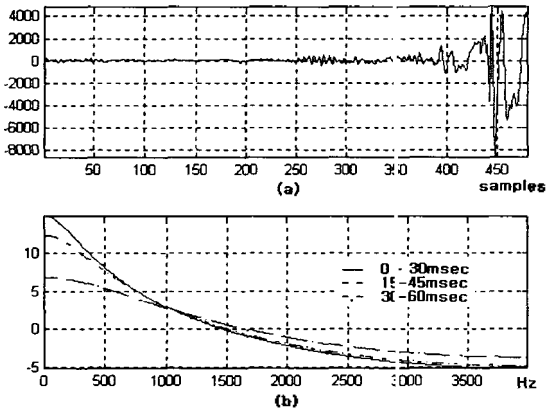


그림 1. 스펙트럼 변화의 추정
(a) 음성파형 (b) 스펙트럼 기울기

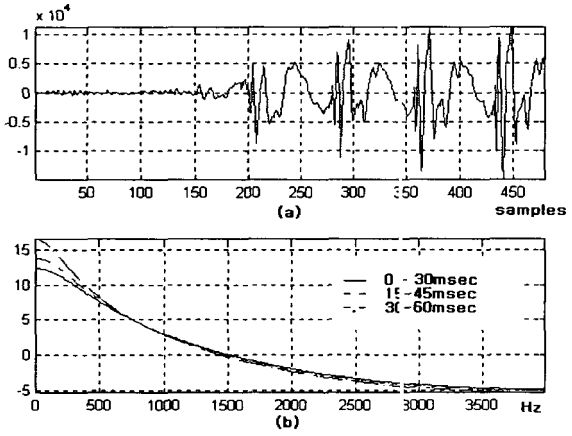


그림 2. 스펙트럼 변화의 추정
(a) 음성파형 (b) 스펙트럼 기울기

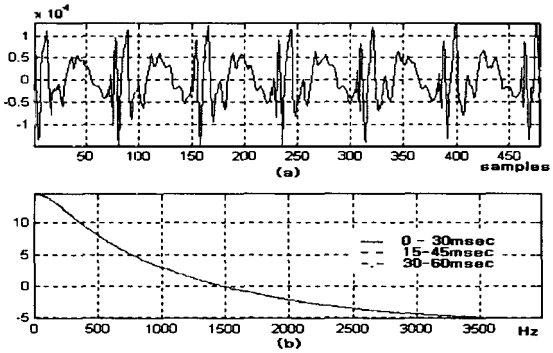


그림 3. 스펙트럼 변화의 추정
(a) 음성파형 (b) 스펙트럼 기울기

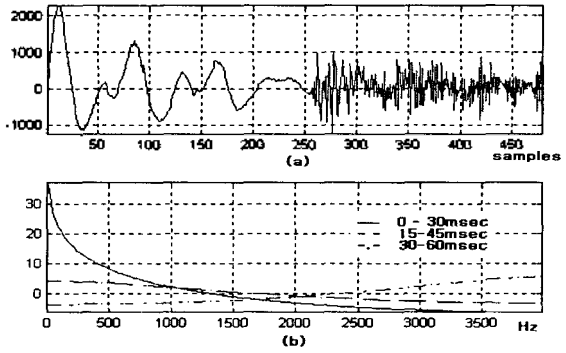


그림 4. 스펙트럼 변화의 추정
(a) 음성파형 (b) 스펙트럼 기울기

3. LSP 위치에 따른 발성률 측정법

본 논문에서 사용되는 파라미터는 LSP 변화도를 이용한 발성율 파라미터이다. 여기서 발성율 측정하고자 하는 방법은 기존의 연구와 달리 처리시간과 알고리즘 자체가 간단하게 측정할 수 있도록 사용하였다. 보코더 자체 내부의 처리이나 계산량이 많으므로 인해 처리시간이 많이 소모되어, 적용한 알고리즘을 적용시킴으로써 보코더 내에서 계산량이나 처리시간을 더 부가시키지 않을 수 있는 알고리즘을 사용하였다. 본 논문에서 고려하는 발성속도는 묵음 부분이 제거된 음성신호에서의 발성속도이다. 본 논문에서는 먼저 묵음구간의 에너지와 LSP 파라미터를 정보를 이용하여 음성 검출을 수행하고, 파라미터를 추출하는 묵음구간은 발성시료의 처음부분을 이용하였다. 60msec동안의 평균 LSP 값을 사용하여 거리를 측정하기 위해 유클리디안 거리측정법을 사용하였다.

$$D(n) = \frac{1}{P} \sum_{i=0}^{P} |LSP_n(i) - LSP_{n+1}(i)|^2 \qquad (2)$$

D(n)는 n번째 분석구간과 n+1번째 분석구간과의 LSP 거리를 나타내고, P는 LSP 분석 차수이다. 입력음성의 발성속도를 계산하기 위해서는 먼저 현재 처리되는 분석구간이 묵음인지 판정해야 된다. 묵음의 판정은 미리 구한 에너지 문턱값과 LSP 파라미터를 이용한다. 묵음 판정이 끝난 후 인접 분석구간과의 LSP 거리를 측정한다. 측정된 거리 값이 문턱값을 넘는 경우는 음소의 변화가 일어난 것으로 판정하고 이전에 음소가 변화된 구간에서 진행된 시간을 계산한다.

$$SPR = \frac{Fs}{VST(n) - VST(n-1)} \qquad (3)$$

3.1 지속시간 변경을 이용한 압축 합성 방법

앞서 발성속도 측정에 따라 느린 발성을 한 구간에서는 지속시간을 변경을 하여 보코더 입력한다. 지속시간 변경은 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^n배의 Decimation을 수행한 다음 G.723.1 보코더 입력하여 부호화시킨 후 G.723.1 복호화 단을 통과 후 FFT point의 1/2^n point로 IFFT과정을 수행함으로써 스펙트럼의 변경 없이 지속시간을 변경, 음성을 합성하였다.

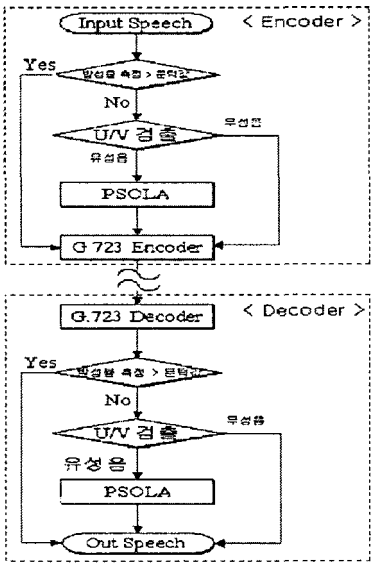


그림 5. 제안한 알고리즘

4. 실험 및 결과

컴퓨터 시뮬레이션에 이용한 장비는 IBM-PC 586(300MHz)에 상용화된 AD/DA 컨버터를 인터페이스한 시스템이다. 처리결과의 성능을 측정하기 위해 다음의 대표적인 문장을 연령층이 다양한 남녀 5명의 화자가 각 5번씩 발성하여 시료로 사용하였다. 음성 시료는 SNR이 30dB인 환경 하에서 녹음하였다. 음성 시료는 다음과 같다.

발성1: /인수네 꼬마는 천재소년을 좋아한다./
발성2: /예수님께서 천지창조의 교훈을 말씀하셨다./
발성3: /창공을 헤치 나가는 인간의 도전은 끝이 없다./
발성4: /숭실대학교 정보통신과 음성통신 연구팀이다./

제안한 알고리즘의 시뮬레이션은 G.723.1 ACELP 부호화기에 C-언어로 구현하여 수행하였다. 제안한 알고리즘의 성능 비교는 G.723.1 Annex A를 통과한 음성과 제안한 알고리즘을 통과한 음성의 MOS 측정하였다. 부호화기는 240샘플 프레임마다 처리한다(8kHz 샘플링에서 30ms). 각 프레임은 DC 성분을 제거하기 위해 하이패스 필터를 통과하고, 60샘플의 4개의 부프레임으로 나누어지고 모든 두 부프레임에서 발성속도를 측정한다. 윈도우 크기는 240, 480의 2종류를 사용하였다. 그림 7은 스펙트럼 기울기를 적용하여 음성파형 내에서 발성속도가 느린 부분으로 간주된 부분을 나타낸 그림이다. 발성속도가 느린 부분은 480샘플(60msec)크기의 윈도우를 적

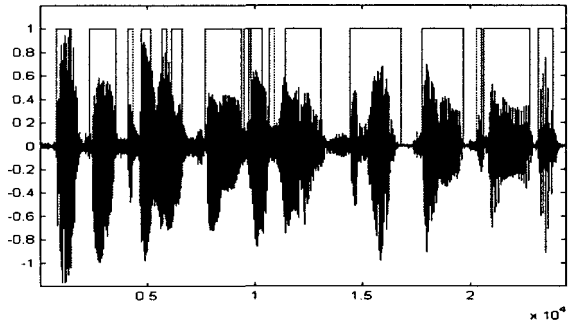
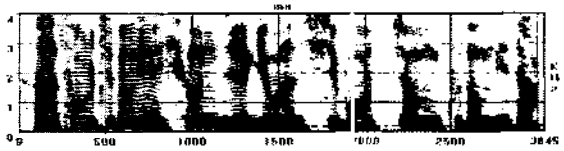
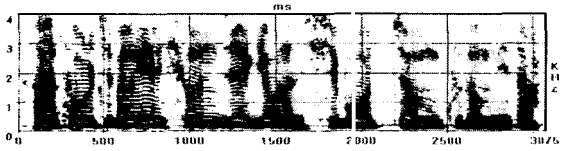


그림 7. 느린 발성속도를 가진 위치 측정 결과

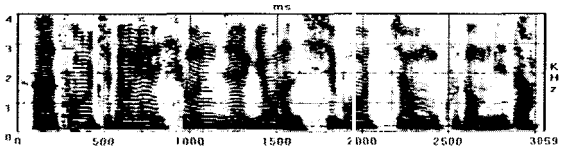
용하여 분석하고 부호화하였다. 그리고 발성속도가 빠르게 측정된 부분은 5.3kbps로 부호화 하였고, 발성속도가 느린 부분은 6.3kbps로 부호화 하였다.



(a) 음성파형의 스펙트로그램



(b) G.723.1 보코더를 통과한 신호의 스펙트로그램



(c) 제안한 보코더를 통과한 신호의 스펙트로그램

그림 8. LSP를 이용한 발성율 측정에 의한 보코더를 통과한 음성 스펙트로그램의 비교

5. 결 론

CELP 부호화기는 선형 예측을 통한 합성에 의한 분석을 기본원리로 한 부호화기이다. 그러나 고정 윈도우를 사용하여 분석함으로써 시간적으로 발성속도가 변하는 음성신호를 적응적으로 분석하지 못하는 단점이 있다. 따라서 본 논문에서는 스펙트럼 변화를 측정하여 발성속도를 빠르고, 느린 부분으로 분류한 후 각각의 경우에 따라 전송률을 달리하여 부호화 하였다. 결과 전송률은 감소하고, 음질에서는 큰 차이를 보이지 않았다. 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 분리하여 합성하는 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. 논문에서는 G.723.1 5.3kbps ACELP를 기반으로 하여 음질을 유지하면서 전송률을 5kbps정도로 낮출 수 있도록 몇 가지 파라미터들을 적용한 알고리즘을 소개하였다. 입력 음성이 보코더단에 들어가기 전 전처리과정을 수행함으로써 전송률을 개선하고자 한 결과, 발성율에 따른 지속시간을 변경을 적용한 알고리즘과 스펙트럼 기울기를 적용하여 발성을 측정하고 보코더에 적용한 알고리즘은 각각 16.2%과 5.6%씩 개선되었다.

표 1. 스펙트럼 기울기를 적용한 전송률 비교

	G.723.1 (5.3kbps)	Proposed Method	Degradation bps
발성 1	5.306	4.495	0.811
발성 2	5.324	4.332	0.992
발성 3	5.322	4.585	0.737
발성 4	5.326	4.391	0.935

표 2. 발성율을 적용한 알고리즘의 전송률 비교

	G.723.1 (5.3kbps)	Proposed Method	Degradation bps
발성 1	5.299	5.019	0.28
발성 2	4.759	4.457	0.302
발성 3	5.245	4.944	0.301
발성 4	5.019	4.748	0.271

참 고 문 헌

[1] 이미숙, "IMT-2000을 위한 음성부호화연구", 가입자망연구소 정보통신연구 제13권 제1 호, 1999.3

[2] A.M. Kondoz, "Digital Speech",, John Wiley & Sons, 1994.

[3] ITU-T Recommendation G.723.1, March, 1996.

[4] 정찬중, 나덕수, 신동성, 배명진, "스펙트럼 누설에너지를 이용한 음성신호의 창함수 적용에 관한 연구", 한국통신학회, 하계종합학술대발표회 논문집 (상), PP. 487-490, 1999.7

[5] W. B. Klejin et. al, "Speech Coding and Synthesis", Elsevier Science B.V., 1995.

[6] "음성 신호 처리 기술", 한국과학기술원, 삼성첨단기술센터, 제 3권 음성부호화, PP.171

[7] L.R. Rabiner, and R.W. Schafer *"Digital Processing of Speech Signals"*, Prentice-Hall, Englewood Cliffs, New Jersey, 1978.