정현파 모델을 이용한 2.4kbps 음성부호화 알고리즘

정회원 백 성 기*, 배 건 성**

2.4kbps Speech Coding Algorithm Using the Sinusoidal Model

Sung-gi Baek*, Keun-sung Bae** *Regular Members*

요 약

STC(Sinusoidal Transform Coding) 방식은 주파수 영역에서 음성신호의 스펙트럼 피크치들을 정현파로 모델링하여 합성하는 음성부호화 방식을 말한다. 저전송률 STC 방식에서는 스펙트럼의 모든 피크를 이용하는 대신, 기본주파수와 고조파에 해당하는 스펙트럼 포락선에서의 크기와 그때의 위상을 이용하여 음성을 합성한다. 본 논문에서는 정현파 모델에 기반한 2.4kbps 음성부호화 알고리즘을 제안한다. 피치정보는 모든 스펙트럼 피크를 사용한 합성음과 선택된 주파수와 고조파를 이용한 합성음과의 평균자승에러를 이용하여 추정하고, 위상정보는 여기신호 펄스의 시작시기를 나타내는 onset time과 성도 모델 전달함수의 위상을 이용하여 얻는다. 크기정보는 SEEVOC 알고리즘과 선형예측계수를 이용하여 추정한다. 실험결과, 합성음의 스펙트럼 특성은 원음성의 포만트 정보를 대부분 가지고 있으며, 위상정보도 원음성의 위상을 잘 따라감을 확인하였다. 합성음의 음질평가를 위해서 informal한 MOS(Mean Opinion Score) 테스트를 시행하였으며, 2.0kbps의 HVXC와 비교하여 대체적으로 MOS 3.1 이상의 음질을 얻을 수 있었다.

ABSTRACT

The Sinusoidal Transform Coding(STC) is a vocoding scheme based on a sinusoidal model of a speech signal. The low bit-rate speech coding based on sinusoidal model is a method that models and synthesizes speech with fundamental frequency and its harmonic elements, spectral envelope and phase in the frequency region. In this paper, we propose the 2.4kbps low-rate speech coding algorithm using the sinusoidal model of a speech signal. In the proposed coder, the pitch frequency is estimated by choosing the frequency that makes least mean squared error between synthetic speech with all spectrum peaks and speech synthesized with chosen frequency and its harmonics. The spectral envelope is estimated using SEEVOC(Spectral Envelope Estimation VOCoder) algorithm and the discrete all-pole model. The phase information is obtained using the time of pitch pulse occurrence, i.e., the onset time, as well as the phase of the vocal tract system. Experimental results show that the synthetic speech preserves both the formant and phase information of the original speech very well. The performance of the coder has been evaluated in terms of the MOS test based on informal listening tests, and it achieved over the MOS score of 3.1.

I. 서 론

초고속통신망, ATM망, 인터넷, 이동통신망 등과 같은 다양한 형태의 디지털 통신망에서 여러 종류의 음성통신 서비스를 제공하기 위해서는 디지털화된 음성을 더욱더 효율적으로 전송하고 저장할 수 있는 음성압축 알고리즘, 즉 음성부호화기가 요구된다. 특히, 이동통신과 함께 인터넷폰, 음성메일과 같

*삼성전자 정보통신, **경북대학교 전자·전기공학부(ksbae@ee.knu.ac.kr)
논문번호 : 010211-0804, 접수일자 : 2001년 8월 21일

은 컴퓨터 통신망에서의 음성통신 서비스를 위해서는 양호한 음질을 가지면서 보다 높은 압축률을 얻을 수 있는 저전송률 음성부호화 알고리즘에 대한 연구가 필요하다[1]. 1960년대부터 전화통신망에 적용되고 있는 전송속도 64kbps PCM(Pulse Code Modulation) 방식은 일반적으로 '압축되지 않은 디지털 음성'으로 고려되어 다른 음성부호화 방식과의 성능 비교를 위한 기준으로 사용되었다. 이러한 PCM 방식보다 전송되는 정보량을 줄이기 위해 8~16kbps 정도의 중간 전송률 음성부호화 알고리즘과 2.4~8kbps 정도의 낮은 전송률, 그리고 2.4kbps 이하의 아주 낮은 전송률 부호화 알고리즘에 대해 많은 연구가 진행되고 있다.

음성신호의 압축을 위해 주로 사용되는 음성부호화 알고리즘으로 유럽에서는 RPE-LTP(Regular-Pulse Excitation Long-Term Prediction) 방식을 사용하며, 북미와 일본, 한국에서는 CELP (Code-Excited Linear Prediction) 방식 계열의 혼합부호화(Hybrid Coding) 방식을 사용한다. CELP 방식[1,2]은 1985년 Schroeder와 Atal에 의해 제안된 부호화 방식으로 음성합성을 위한 여기신호를 벡터 양자화된 코드북에서 합성에 의한 분석(Analysis-by-Synthesis) 방식으로 선택하여 양호한 음질을 유지하면서도 64kbps PCM방식에 비해서 약 8배 정도의 압축률을 얻을 수 있다. 따라서, 4~8kbps 정도의 전송률을 가지면서 우수한 음질을 얻고자하는 음성부호화기에 대한 연구는 주로 CELP 방식 계열의 압축알고리즘이 주 연구대상이 되어 왔다. 그러나, CELP 방식은 4kbps 이하의 낮은 전송률에서는 유성음에 존재하는 주기성분을 정확하게 모델링하지 못하므로 음질이 급속도로 나빠지는 단점이 있다[1-3]. 따라서, 4kbps 이하의 전송속도에서 양질의 합성음을 얻기 위해 새로운 음성부호화 방식이 요구되며, 이러한 전송률에서 확정된 표준안이나 제안되고 있는 방법들은 LPC(Linear Predictive Coding) 음성부호화기 계열이나 하모닉 계열이 주류를 이루고 있다. LPC 계열에는 미국방성의 2.4kbps 음성부호화기 MELP(Mixed Excitation Linear Predictive vocoder)[4]가 1996년 표준안으로 채택되었고, 하모닉 계열에는 CELP와 하모닉 모델을 결합한 HVXC(Harmonic Vector eXcitation Coding)[5]가 MPEG-4 표준안으로 채택되었으며, 정현파 모델을 기반으로 하는 STC(Sinusoidal Transform Coding) 방식[6-10]이나 MBE(Multi-Band Excitation) 방식[11], WI(Waveform Interpolation)[12] 방식들이 제안되고 있다.

MBE 방식은 음성신호의 스펙트럼을 유/무성음의 이원적인 모델로 구분하여 분석/합성하는 방식이다. 그러나 이 방식은 피치 추정 과정이 복잡하고 계산량이 많으며, 피치 추정 과정에서 많은 지연시간을 필요로 한다는 단점이 있다. HVXC 방식은 무성음일 경우에는 기존의 CELP 방식을 사용하며, 유성음일 경우에는 여기신호를 정현파를 이용하여 모델링하는 방식이다. 그러나, HVXC 또한 무성음 구간에서는 기존의 CELP 방식을 그대로 사용하고, 유성음구간에서는 MBE 방식에서 사용되는 피치 추정 방식을 이용함으로써 알고리즘이 복잡하고 계산량이 많다는 단점이 있다[3]. 이에 비해 STC 방식은 간단한 알고리즘으로 양질의 합성음을 얻을 수 있다.

STC 방식은 정현파 모델을 이용한 음성부호화 방식으로, 주파수 영역에서 스펙트럼의 모든 피크치들을 정현파로 모델링하여 합성하는 방식을 말한다[1]. 그러나, 정현파 모델에 기반한 저전송률 음성부호화기에서는 전송되는 정보량을 줄이기 위해 스펙트럼의 모든 피크를 사용하는 대신 스펙트럼 포락선 상에서 기본주파수와 고조파에 해당하는 크기 정보를 이용한다[10]. 즉, 음성신호가 유성음인 경우에는 기본주파수와 고조파에 해당하는 스펙트럼 포락선 성분들을 정현파로 표현하고, 무성음인 경우에는 주기성분이 일정하지 않으므로 스펙트럼 포락선에서 일정 간격으로 크기를 찾아서 그에 해당하는 정현파를 생성하게 된다[10].

본 논문에서는 정현파 모델을 이용한 2.4kbps 음성부호화 알고리즘을 제안하고, 합성음의 파형과 스펙트럼 특성, 그리고 MOS(Mean Opinion Score) 테스트를 이용한 합성음의 음질을 비교/분석하였다. 제안한 알고리즘은 25ms 분석프레임마다 음성신호를 분석하고 정현파 모델 파라미터를 추출하여 부호화하며, 피치정보, 스펙트럼 크기정보, 위상정보를 전송파라미터로 한다. 피치정보는 먼저 자기상관함수를 이용한 개루프 피치 검색을 통해 대략적인 피치를 검출한 후[13], 대략적인 피치 주위에서 정확한 피치를 찾는다. 정확한 피치는 모든 스펙트럼의 피크에 해당하는 주파수를 이용한 합성음과 피치에 의한 기본주파수와 고조파를 이용한 합성음과의 에러가 최소가 되도록 주파수 영역에서 찾는다[14]. 스펙트럼 크기정보는 스펙트럼 포락선을 이용하며, SEEVOC(Spectral Envelope Estimation VOCoder)[15] 알고리즘과 14차 LPC 계수[16]를 사용하여 추정한다[17]. 위상정보는 여기신호 펄스의 시작 시기

를 나타내는 onset time과 성도모델 전달함수의 위상을 이용하여 추정하였다[18]. 실험결과, 스펙트럼 특성에서 합성음이 원음성의 포만트 성분을 대부분 가지고 있으며, 합성음의 위상 또한 원음성의 위상을 잘 따라감을 확인하였다. 그리고, 2.0kbps HVXC와의 informal한 MOS 테스트결과 대체적으로 MOS 3.1 이상의 음질을 얻을 수 있었다.

본 논문의 구성은 다음과 같다. 2장에서는 제안한 정현파 모델을 이용한 2.4kbps 음성부호화 알고리즘에 대해서 설명한다. 3장에서는 음성부호화 알고리즘에 의해 얻어진 합성음을 비교/분석하며, 마지막으로 4장에서 결론을 맺는다.

Ⅱ. 정현파 모델을 이용한 2.4kbps 음성부호화 알고리즘

1. 정현파 모델을 이용한 음성부호화 알고리즘

8kHz, 16bits로 샘플링되고 양자화된 25ms의 분석프레임을 갖는 입력음성은 전처리과정을 거친 후, 유/무성음의 이원적인 모델로 나뉘어져 처리되며, 12.5ms의 overlap을 갖는다. 부호화기에서는 피치, 스펙트럼 포락선을 위한 선형예측 계수와 이득, 추정한 위상정보가 양자화 된 후 전송되며, 부호화기에서는 전송된 파라미터를 이용하여 식 (1)과 같이 음성을 합성한다.

$$s(n)=\sum_{l=1}^{L} A_l \cos(lw_o n+\phi_l) \qquad (1)$$

여기서 A_l과 ϕ_l은 기본주파수 w_o와 고조파 lw_o에 해당하는 스펙트럼의 크기 및 위상을, L은 합성에 사용되는 정현파의 수를 나타낸다. 그림 1과 2는 각각 본 논문에서 제안한 부호화기와 복호화기의 블럭도를 나타낸다.

입력음성은 DC 성분과 불필요한 저주파 성분을 제거하기 위해서 140Hz의 cutoff 주파수를 갖는 고역 여파기를 거치며, overflow를 방지하기 위해 2배의 down-scaling 과정을 거치게 된다[13]. 식 (2)는 고역 여파기와 2배의 down-scaling 과정을 나타낸 것이다.

$$H_{hl}=\frac{0.46363718-0.92724705z^{-1}+0.46363718z^{-2}}{1-1.9059465z^{-1}+0.9114024z^{-2}} \qquad (2)$$

전처리과정을 거친 입력음성은 대략적인 피치추정 후, 유/무성음으로 나뉘어 처리된다. 대략적인 피치추정은 개루프 피치 검색 알고리즘[13]을 사용하며, 여기서 구해진 피치에서의 자기상관함수값과 분석프레임에서의 영교차율, 에너지를 이용하여 유/무성음을 결정한다. 유성음의 경우에는 피치정보와 위상정보, 스펙트럼정보를 찾으며, 무성음은 스펙트럼정보만을 찾고 위상정보로는 80Hz 간격으로 $-\pi \sim \pi$ 범위에서 uniformly distributed random number를 사용한다[10]. 유성음의 경우 대략적인 피치와 이전 프레임의 피치를 이용하여 평균피치를 구하며, 평균피치의 2.5배에 해당하는 길이를 적응 윈도우의 길이로 정한다[10]. 전처리과정을 거친 음성에 구해진 적응 윈도우를 적용하여 얻어지는 음성신호로부터 정확한 피치와 스펙트럼 포락선, 위상정보를 구하게 된다. 그림 3은 분석프레임에 적응윈도우를 적용하

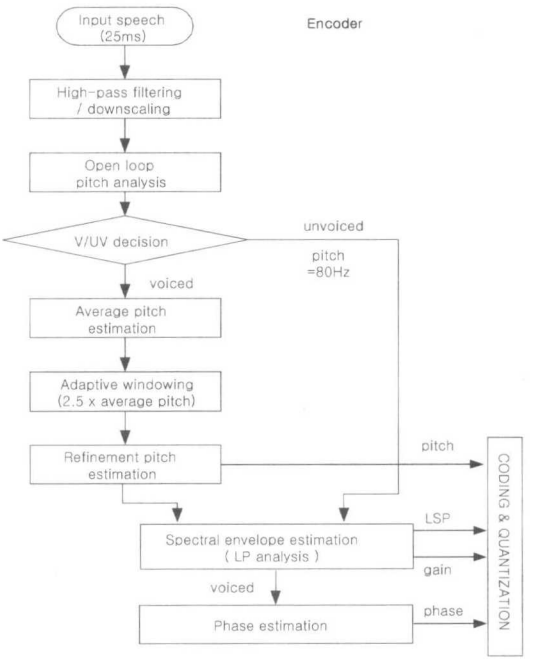


그림 1. 제안한 STC의 부호화기 블록도

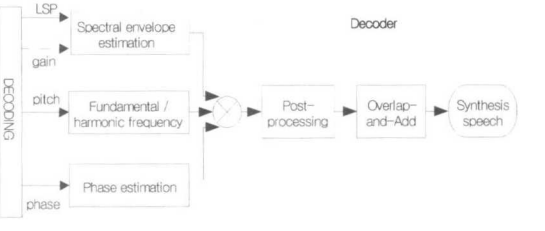


그림 2. 제안한 STC의 복호화기 블록도

는 과정을 보인 것이다.

정확한 피치정보는 모든 스펙트럼의 피크에 해당하는 주파수를 이용한 합성음과 피치에 의한 기본주파수와 고조파를 이용한 합성음과의 에러가 최소가 되도록 주파수 영역에서 피치를 찾는다[14]. 정현파 모델에 기반한 저전송률 음성부호화기는 기본주파수와 스펙트럼 포락선을 이용하여 각 고조파에 해당하는 스펙트럼 크기를 얻는다. 따라서, 스펙트럼 포락선은 기본주파수와 고조파에 해당하는 스펙트럼 피크를 잘 따라가야 한다. 본 논문에서는 SEEVOC 알고리즘을 사용하여 스펙트럼의 모든 피크 중에서 기본주파수와 고조파에 해당하는 특정 피크를 검출한 후[15], 이 피크를 사용하여 14차 선형예측계수[16]를 구하는 방법을 이용하였다. 위상정보는 여기신호 펄스의 시작 시기를 나타내는 onset time과 최소위상시스템(minimum phase system)을 가정한 성도모델 전달함수의 위상을 이용하여 추정한다[17]. 이 때, 성도모델 전달함수의 위상은 전달함수가 최소위상시스템이라는 가정 하에 구해진 것이므로, onset time과 성도모델 전달함수의 위상만으로는 입력 음성 파형의 부호를 결정하지 못한다. 이러한 문제는 전달함수의 위상에 $\beta\pi$를 더하고, 원음성의 위상을 이용한 합성음과 추정한 위상을 이용한 합성음과의 평균자승에러가 최소가 되는 n_o와 β를 선택함으로써 해결할 수 있다. 여기서 β는 0 또는 1의 값을 나타내며, 추정된 위상정보는 식 (3)과 같다.

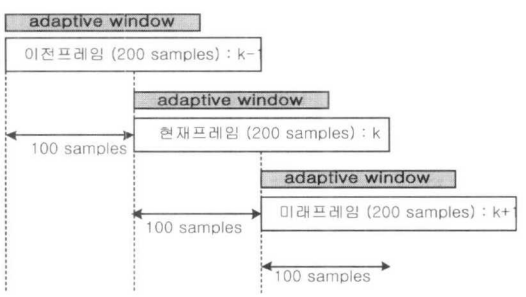


그림 3. 적응 윈도우의 적용 과정

$$\hat{\phi}_l(w_l) = -\hat{n}_o w_l + \Phi_s(w_l) + \beta\pi \qquad (3)$$

여기서, $\hat{\phi}(w_l)$은 추정한 위상정보이고, $\hat{n}_o$는 onset time, $\Phi_s(w_l)$는 성도모델 전달함수의 위상을 나타낸다.

부호화기에서 구해진 유/무성음 정보, 피치, 선형예측계수, 이득, 위상정보는 25ms마다 60bits로 양자화되어 전송되며, 복호화기에서는 전송된 파라미터를 이용하여 합성음을 구한 후, 후처리 과정을 거친다. 후처리는 100Hz의 cutoff 주파수를 갖는 고역여파기와 전처리의 down-scaling을 보상하기 위한 up-scaling으로 이루어져 있다[13]. 식 (4)는 사용된 고역 여파기이다. 또한, 프레임간 연속성을 위하여 overlap-and-add 합성방식을 수행하여 최종적인 합성신호를 생성하게 된다.

$$H_{hi2} = \frac{0.93980581 - 1.8795834z^{-1} + 0.93980581z^{-2}}{1 - 1.9330735z^{-1} + 0.93589199z^{-2}} \qquad (4)$$

2. 파라미터의 전송과정

부호화기에서 구해진 유/무성음 정보, 피치정보, 스펙트럼 포락선 정보, 위상정보는 25ms 마다 60bits로 양자화되어 전송된다. 전송되는 파라미터는 유성음일 경우에는 유성음 정보, 피치정보, 위상정보($\hat{n}_o$, β), 스펙트럼 포락선 정보(선형예측계수, 이득)이며, 무성음일 경우에는 무성음 정보, 스펙트럼 포락선 정보, 이득정보이고, 피치와 위상정보는 아무런 의미 없는 0을 전송한다. 파라미터의 전송은 그림 4와 같이 25ms 마다 이루어지며, 2개의 분석프레임을 갖는다. 본 논문에서는 첫 번째 분석프레임을 frame A, 두 번째 분석프레임을 frame B라 명한다. Frame B가 미래의 12.5ms(100samples)를 필요로 하기 때문에 지연시간은 분석프레임 25ms를 포함하여 전체 37.5ms가 된다.

파라미터의 전송은 다음과 같이 이루어진다. Frame B의 기본주파수는 양자화되어 전송되며, frame A의 기본주파수는 인덱스가 선택되어 전송된다. 즉, frame A의 기본주파수는 현재 분석프레임

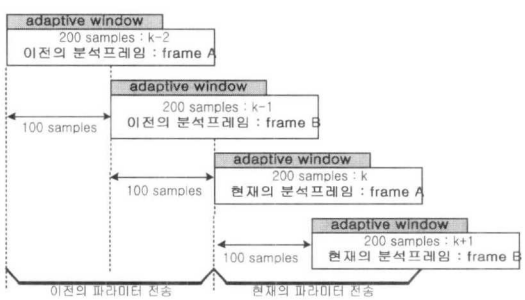


그림 4. 분석프레임과 파라미터 전송 과정

frame B와 이전 분석프레임 frame B의 기본주파수를 식 (5)와 같이 보간 하여 현재 분석프레임 frame A의 기본주파수와 비교하여, 식 (6)의 오차가 작은 인덱스를 전송한다.

$$\begin{aligned} \widehat{w_o}^{\,k}_{\,0} &= {w_o}^{k-1} \\ \widehat{w_o}^{\,k}_{\,1} &= {w_o}^{k+1} \\ \widehat{w_o}^{\,k}_{\,2} &= 0.5\,{w_o}^{k-1} + 0.5\,{w_o}^{k+1} \\ \widehat{w_o}^{\,k}_{\,3} &= 0.25\,{w_o}^{k-1} + 0.75\,{w_o}^{k+1} \end{aligned} \tag{5}$$

여기서 ${w_o}^{k-1}$은 이전 분석프레임의 frame B에 대한 기본주파수이며, ${w_o}^{k+1}$은 현재 분석프레임의 frame B에 대한 기본주파수 이다.

$$err = |\,{w_o}^{k} - \widehat{w_o}^{\,k}\,| \tag{6}$$

선형예측계수는 frame B의 선형예측계수만이 LSF(Line Spectral Frequency)[19]로 바뀐 후 양자화되어 전송된다. Frame A의 선형예측계수는 전송되지 않고, 복호화기에서 현재 분석프레임 frame B의 LSF와 이전 분석프레임 frame B의 LSF를 식 (7)과 같이 보간 하여 사용한다. 그 이외에 이득과 유/무성음 정보, 위상정보($\widehat{n_o}$, β)는 frame A, frame B에서 구해진 파라미터가 각각 양자화되어 전송된다.

$$\begin{aligned} &frame\ A\text{: } LSF^{\,k}_{\,i} = 0.5\,LSF^{\,k-1}_{\,i} + 0.5\,LSF^{\,k+1}_{\,i} \\ &frame\ B\text{: } LSF^{\,k+1}_{\,i} = LSF^{\,k+1}_{\,i} \end{aligned} \tag{7}$$

3. 전송 파라미터의 2.4kbps 양자화 기법

부호화기에서 구해진 파라미터는 25ms마다 60bits로 양자화되어 전송되므로 본 논문에서 제안한 음성부호화기는 2.4kbps의 전송률을 가진다. 표 1은 제안한 STC 음성부호화기의 2.4kbps 비트 할당을 나타낸 것이다.

표 1. 2.4kbps STC의 비트 할당

	frame A	frame B	Total
유/무성음	1	1	2
L S F		20	20
피 치	2	8	10
onset time	8	8	16
β	1	1	2
이 득	5	5	10
Total			60

유/무성음 정보는 frame A, B 각각 1bit를 할당하여, 0 또는 1을 전송한다. Frame B의 14차 선형예측계수는 LSF로 바꾸어[19], 그림 5와 같이 유/무성음에 따라 multistage split VQ를 사용하여 20bits로 양자화 한다[20]. LSF 특성이 유성음일 경우와 무성음일 경우 차이가 있으므로 유/무성음을 구별하여 양자화 한다. 먼저 14차의 LSF에 8bits를 할당하여 양자화한 후, 양자화되지 않은 LSF와 8bits로 양자화된 LSF의 에러를 (6, 8)로 분리하여 각각 6bits로 양자화 한다.

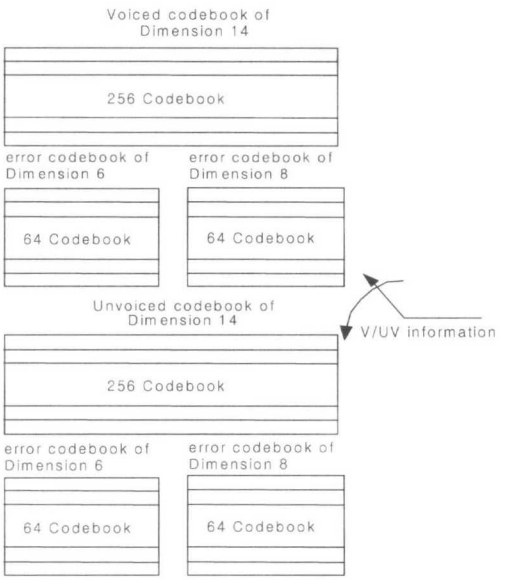


그림 5. LSF의 코드북 구조

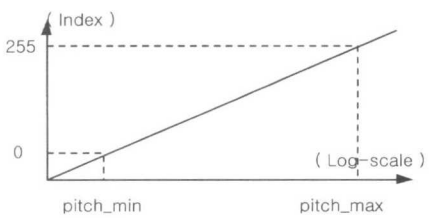


그림 6. 피치(기본주파수)의 양자화

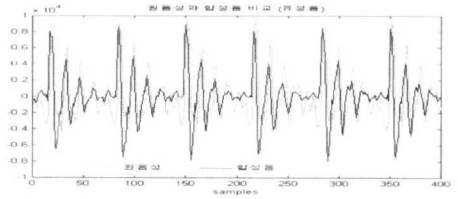
피치의 양자화는 피치의 역수인 기본주파수를 사용한다. Frame B의 피치는 그림 6과 같이 8bits를 할당하여 대수적으로 양자화 한다. Frame A의 피치는 식 (5)와 식 (6)에서 선택된 인덱스가 2bits로 양자화되어 전송된다. 위상정보 ($\widehat{n_o}$, β)는 frame A,

B에 대해서 같은 양자화 비트가 할당되며, 스칼라 양자화 된다. 이득은 frame A, B에 대해서 각각 6bits를 할당하며, Lloyd-Max 알고리즘[21]을 이용하여 비선형적으로 양자화 한다.

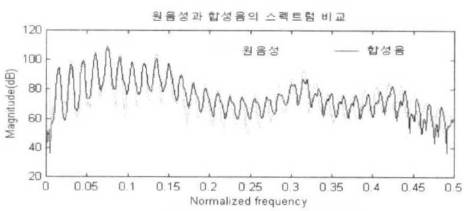
Ⅲ. 실험 및 고찰

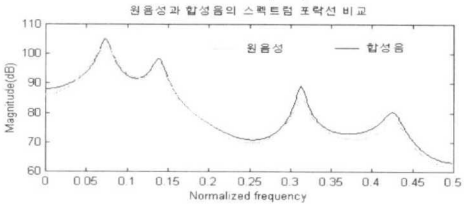
본 연구에서는 8kHz로 샘플링되고, 16bits로 양자화된 임의의 음성데이터를 이용하여 원음성과 합성음의 파형, 스펙트럼, 위상을 비교/분석하였으며, 합성음의 음질평가를 위해서 주관적인 음질평가 방법인 MOS 테스트를 이용하였다. 그림 7, 8, 9는 원음성과 정현파모델에 기반한 2.4kbps 음성부호화기를 이용한 합성음의 파형, 스펙트럼, 그리고 위상을 나타낸 것이다. STC 방식이 파형부호화 방식이 아니므로 합성음의 파형은 원음성에 비해 약간의 왜곡이 발생한다. 그러나, 스펙트럼 특성에서는 원음성의 포만트 정보를 대부분 가지고 있으며, 위상에서는 원음성의 위상과 합성음의 위상이 거의 일치함을 알 수 있다.

표 2는 2.0kbps HVXC(Harmonic Vector eXcitation Coding)와 제안한 정현파 모델을 이용한 2.4kbps 음성부호화 알고리즘에 대해 남·여 10명의 청취자를 대상으로 합성음의 음질을 주관적인 음질평가 방법인 MOS 테스트를 이용하여 informal한 청취실험을 수행한 결과를 보인 것이다. 실험에 사용된 음성은 남성화자가 발성한 2개의 문장과 여성화자가 발성한 2개의 문장을 사용하였으며 각 음

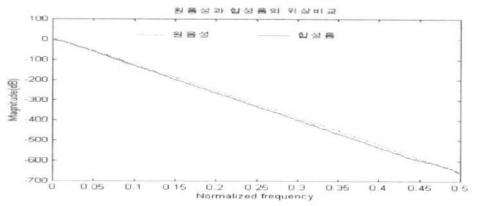


(a) 원음성과 합성음의 파형 비교



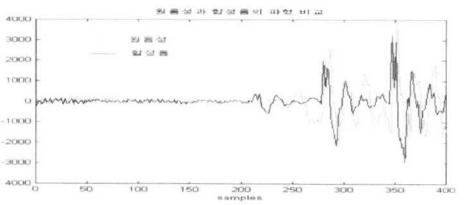


(b) 원음성과 합성음의 스펙트럼 비교

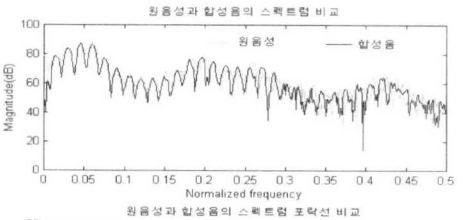


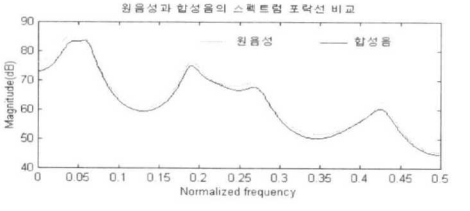
(c) 원음성과 합성음의 위상 비교

그림 7. 원음성과 합성음 비교 (유성음 구간)

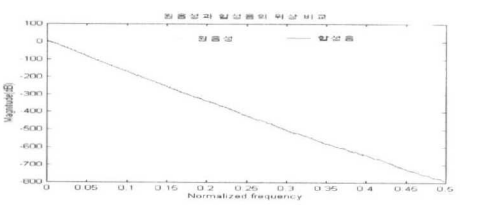


(a) 원음성과 합성음의 파형 비교



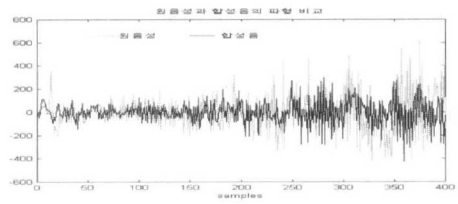


(b) 원음성과 합성음의 스펙트럼 비교

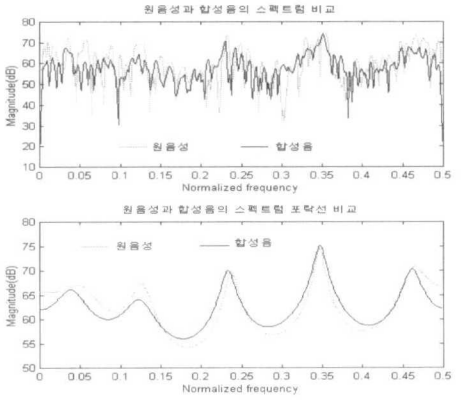


(c) 원음성과 합성음의 위상 비교

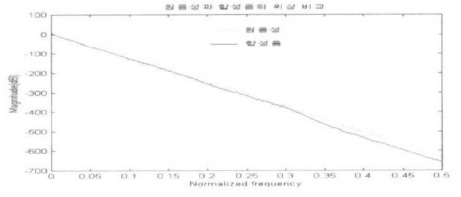
그림 8. 원음성과 합성음 비교 (천이구간)



(a) 원음성과 합성음의 파형 비교



(b) 원음성과 합성음의 스펙트럼 비교



(c) 원음성과 합성음의 위상 비교

그림 9. 원음성과 합성음 비교 (무성음 구간)

성데이터는 8kHz, 16bits로 양자화 되었다. 아래의 문장 1, 2는 남성화자가 발음한 음성이고, 문장 3, 4는 여성화자의 것이다.

문장 1 : 하늘을 날고자 하는 인간의 욕망은 끝이 없습니다. (남성화자)

문장 2 : 한국의 가을하늘은 참으로 맑고 푸르릅니다. (남성화자)

문장 3 : 사람은 어릴 때 좋은 습관을 들이는 것이 중요하다. (여성화자)

문장 4 : 김대중 대통령은 오늘 대북포용정책은 ... (여성화자)

실험결과, 제안한 STC 방식은 비록 2.0kbps HVXC 방식에 비해 음질은 조금 떨어지지만, 대체적으로 MOS 3.1 이상의 합성음을 얻을 수 있었다. 그러나, HVXC가 CELP와 하모닉 모델의 장점만을 복합하여 만듦으로써 음질은 뛰어나지만 알고리즘이 복잡하다는 단점[3]이 있는데 비해, 제안한 STC 방식은 전형적인 하모닉 모델에 기반하고 있으므로 알고리즘이 간단하고 계산량이 적어 실시간 구현이 용이하다는 장점이 있다.

표 2. 합성음의 MOS 테스트

문장	1	2	3	4	Total
원음성	5.0	5.0	4.4	4.6	4.75
2.4kbps STC	3.2	3.0	3.1	3.0	3.08
2.0kbps HVXC	3.4	3.9	3.2	3.3	3.45

Ⅳ. 결 론

STC 방식은 정현파 모델을 이용한 음성부호화 방식으로, 주파수영역에서 음성신호의 스펙트럼 피크 성분들을 정현파로 모델링하여 합성하는 방식을 말하며, 일반적으로 저전송률에서는 음성신호의 고조파 성분들을 정현파로 모델링하여 합성한다. 본 논문에서는 정현파 모델을 이용한 2.4kbps 음성부호화 알고리즘을 제안하고, 합성음의 파형과 스펙트럼 특성, 그리고 MOS 테스트를 이용한 합성음의 음질을 비교/분석하였다. 제안된 알고리즘은 25ms 분석 프레임마다 음성신호를 분석하여 정현파 모델 파라미터를 추출하고 부호화 한다. 피치정보는 개루프 피치검색과 평균자승에러를 이용하여 추정하고, 위상정보는 여기신호 펄스의 시작 시기를 나타내는 onset time과 성도 모델 전달함수의 위상을 이용하여 예측한다. 크기정보는 스펙트럼 포락선에 의해 나타내며, SEEVOC 알고리즘에 의해 구해진 스펙트럼 피크를 이용하여 선형예측계수를 구하고, 이를 이용하여 스펙트럼 포락선을 추정한다.

실험결과, 합성음의 파형에서는 STC 방식이 파형 부호화 방식이 아니므로 원음성에 비해 약간의 왜곡이 발생하지만, 스펙트럼 특성에서는 합성음의 스펙트럼 특성이 원음성의 포만트 정보를 대부분 가지고 있으며, 위상정보도 원음성의 위상을 잘 따라가는 것을 볼 수 있다. 또한, 합성음은 2.0kbps의 HVXC와 비교하여 대체적으로 MOS 3.1 이상의 음질을 가짐을 확인하였다. 제안한 STC 방식은 2.0kbps의 HVXC 방식에 비해 음질은 조금 떨어지

지만 알고리즘이 간단하고 계산량이 적어서, 고압축율이 요구되는 멀티미디어 시스템에서 음성압축 알고리즘으로 유용하게 사용될 수 있을 것이다.

참 고 문 헌

[1] A. S. Spanias, "Speech Coding : A Tutorial Review", Proc. of IEEE, Vol.82, No.10, pp. 1541-1582, Oct. 1994.

[2] M. Schroeder, and B. Atal, "Code-Excited Linear Prediction (CELP) : High Quality Speech at Very Low", Proc. 1985 IEEE Int. Conf. on Acoustics, Speech and Signal Proc., pp. 937-940, 1985.

[3] 최용수, 강홍구, 윤대희, "저전송률 음성부호화 연구 동향", 제 15회 음성통신 및 신호 처리워크샵, pp. 113-120, 1998.

[4] A. V. McCree, T. P. Barnwell III, "A Mixed Excitation Vocoder LPC Model for Low Bit Rate Speech Coding", IEEE Trans. on Acoustics, Speech and Signal Proc., Vol.3, No.4, pp. 242-250, 1995.

[5] Technical Description of Sony IPC's Proposals for MPEG-4 Audio and Speech Coding, Nov. 1995.

[6] R. J. McAulay and T. F. Quatieri, "Speech Analysis/Synthesis Based on a Sinusoidal Representation", IEEE Trans. on Acoustics, Speech and Signal Proc., Vol.34, No.4, pp. 744-754, Aug. 1986.

[7] T. F. Quatieri and R. J. McAulay, "Speech Transformation Based on a Sinusoidal Representation", IEEE Trans. on Acoustics, Speech and Signal Proc., Vol.34, pp. 1449-1464, Aug. 1986.

[8] E. B. George and M. J. T. Smith, "Speech Analysis using an Analysis/Overlap-Add Sinusoidal Model", IEEE Trans. on Acoustics, Speech and Signal Proc., Vol.5, No.5, pp. 389-406, Sep. 1997.

[9] R. J. McAulay and T. F. Quatieri, "Sinusoidal coding", Speech Coding and Synthesis, Chapter 4, W. B. Kleijn and K. K. Paliwal Eds., Elsevier, 1995.

[10] R. J. McAulay and T. F. Quatieri, "Low-Rate Speech Coding Based on Sinusoidal Model", Advances in Speech Signal Processing, Chapter 6, S. Furui and M. M. Sondhi Eds., Dekker, Inc., New York, 1992.

[11] J. C. Hardwick, "A 4800bps Multi-band Excitation Speech Coder", S. M. Thesis, E.E.C.S. Department, M.I.T., May, 1988.

[12] W. B. Kleijn, K. K. Paliwal, "Waveform Interpolation for Coding and Synthesis", Speech Coding and Synthesis, Elsevier, 1995.

[13] R. Salami, C. Laflamme, J.P.Adoul, and D. Massaloux, "A toll quality 8kbps speech codec for the personal communications system(PCS)", IEEE Trans. on Veh. Technol., Vol.43, pp. 808-816, 1994.

[14] R. J. McAulay and T. F. Quatieri, "Pitch estimation and voicing detection based on a sinusoidal model", 1990 IEEE Int. Conf. Rec. on ASSP, pp. 249-252, 1990.

[15] D. B. Paul, "The spectral envelope estimation vocoder", IEEE Trans. on Acoustic, Speech and Signal Proc., Vol.29, pp. 786-794, 1981.

[16] A. El-Jaroudi and J. Makhoul, "Discrete all pole modeling", IEEE Trans. on Acoustics, Speech and Signal Proc., Vol.39(2), pp. 411-423, Feb. 1991.

[17] 백성기, 배건성, "음성신호의 정현파 모델에 기반한 저전송률 부호화를 위한 스펙트럼 포락선 추정 기법에 관한 연구", 제17회 음성통신 및 신호처리 학술대회, pp. 197-200, 2000.

[18] R. J. McAulay and T. F. Quatieri, "Sine-wave phase coding at low data rates", Proc. 1991 IEEE Int. Conf. on Acoustics, Speech and Signal Proc., Vol.1, pp. 577-580, May 1991.

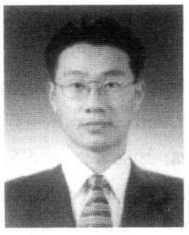
[19] P. Kabal and R. P. Ramachandran, "The computation of line-spectral frequencies using Chebyshev polynomials", IEEE Trans. on Acoustics, Speech and Signal Proc., Vol.34, pp. 1419-1426, 1986.

[20] S. Ahmadi and A. S. Spanias, "New techniques for sinusoidal coding of speech at 2400bps", Signals, Systems and Computers, Conference Record of the Thirtieth Asilomar Conference on, Vol.1, pp. 770-774, 1997.

[21] S. P. Lloyd, "Least Squares Quantization in

PCM", IEEE Trans. on Information Theory, pp. 129-137, Mar. 1982.

백 성 기(Sung-gi Baek) 정회원



1999년 2월 : 경북대학교 전자공학과 졸업
2001년 2월 : 경북대학교 전자공학과 석사
2001년 3월~현재 : 삼성전자 정보통신 연구원

<주관심 분야> 디지털 신호처리, 통신공학 등

배 건 성(Keun-sung Bae) 정회원

1977년 2월 : 서울대학교 전자공학과 졸업
1979년 2월 : 한국과학기술원 전기 및 전자공학과 석사
1989년 5월 : University of Florida 공학박사
1979 3월~현재 : 경북대학교 전자·전기공학부 교수

<주관심 분야> 음성분석 및 인식, 디지털 신호처리, 디지털 통신, 음성 부호화, 웨이브렛 이론 등