

G.723.1, G.729 부호화기와 MLT 방법을 이용한 광대역 음성 부호화기 설계

김정중, 김종학, 이인성
충북대학교 전자공학과

Design of Wideband Speech Coder Using the G.723.1, G.729 Combined with MLT

Jeongjoong Kim, Jonghark Kim, Insung Lee
Dept. of Radio Engineering, Chungbuk National University
email: kjjfa@hanmail.net

요 약

본 논문에서는 ITU-T G.723.1, G.729 부호화기와 MLT(Modulated Lapped Transform) 방법을 이용한 광대역 음성 부호화방법을 제안한다. 제안된 광대역 음성 부호화 방법은 16 kHz로 샘플링된 입력신호를 QMF(Quadrature Mirror Filter) 사용하여 저대역과 고대역으로 나누며, 각 대역은 8 kHz의 샘플링을 갖는 협대역 음성 신호로 변환된다. 고대역은 MLT변환 후 벡터 양자화하며 또한 MLT를 사용한 ATC(Adaptive Transform Coding)방법을 적용하여 표현하며 저대역은 G.723.1과 G.729 부호화기를 사용한다.

설계된 광대역 음성부호화기의 성능을 평가하기 위하여 MOS (Mean Opinion score) 실험을 수행하였다. MOS 실험을 통해 16 kbps G.729-MLT VQ방식이 G.722 56kbps 와 비슷한 음질을 나타내었다.

1. 서 론

IMT-2000 시스템에서는 기존의 음성위주의 이동 통신 서비스에서 데이터, 영상 서비스 등을 포함한 멀티미디어 통신이 가능하게 된다. 또한 앞에서 제시된 기본

적인 음성통신 서비스를 바탕으로 하는 무선 화상회의 서비스나 오디오 전송 서비스도 가능하게 된다. 일반적으로 화상회의에서 사용되는 음성 대역폭은 공중전화망(PSTN)에서의 200Hz-3400Hz 대역폭보다 넓은 50Hz-7000Hz의 대역폭을 갖고, 샘플링 주파수로 16kHz를 사용한다. 위와 같은 시스템에 사용되는 광대역 음성부호기로는 ITU-T 권고안으로 결정된 64Kbps 음성부호화 방식인 G.722가 있다. 광대역음성을 전송하기 위해서 음질의 저하가 없는 저 전송률의 부호화기가 요구된다. 이런 광대역 음성코딩에 대한 표준안은 16 kbps 근처의 전송률에서 G.722의 48 kbps의 성능과 비교하여 같거나 더 좋은 성능을 요구한다. 현재 16 kbps 정도의 전송률에서 개발된 광대역음성 부호화기의 구조를 살펴보면, 주로 대역 분할을 이용한 코딩 방법이다.

본 논문에서는 저대역에 CELP(Code Excited Linear Predictive)구조를 기반으로 한 G.723.1, G.729 부호화기를 사용하였고 고대역에서는 DCT를 기반으로 한 MLT방법을 선택하여 음성부호화기를 설계하였다. 8 kbps 전송률의 G.723.1, G.729부호화기의 엔코더에서는 8000 sample/sec에서 80샘플을 가지는 10ms의 프레임마다 음성신호를 분석하여 CELP모델 파라미터는 잔여신호와 합성필터 파라미터를 복원하는데 사용하고, 음성은 LP 합성 필터를 통과한 잔여신호의 필터링에 의해 복원된다. 고대역 신호는 엔코더에서 80샘플의 신호마다 MLT 변환 후 양자화 시켜 전송한다. 수신단에서는 ATC방법으로 역부호화하여 출력신호를 만든다. 본 논문은 기존

**본 과제는 과학재단(과제번호 2000-1-30200-003-3)에 기초연구과제비로 지원됨

의 저대역 음성 부호화기인 G.723.1 과 G.729를 저대역에서 사용함으로써 기존의 음성통신 시스템과의 호환성을 유지함을 보였다. 제안된 광대역 음성부호화기의 성능은 ITU-T G.722를 포함하여 MOS 시험을 수행한다.

2. 광대역 음성 부호화기의 인코더 알고리즘

광대역 음성 부호화기는 저 전송률에서 음질 개선을 위하여 고대역/저대역부분으로 나누어 각 부분에 따라 코딩 방법을 다르게 적용하여 각각 8kbps의 전송률을 가져 16kbps의 전송률을 가지는 음성부호화기이다. 이 음성부호화기는 음성신호를 QMF 분석 Bank에 의하여 두 개의 대역폭으로 나누어지고, 각각 decimation 되어 두 개의 8kHz 샘플링신호로 변환된다. 저대역에서는 ITU-T 협대역표준인인 G.723.1, G.729를 각각 사용하여 음성부호화되고 고대역의 신호는 ATC(Adaptive Transform coding)방법으로 8kbps의 전송률로 부호화되고, 변환은 MLT(Modulated Lapped Transform) 알고리즘을 이용한다. 저대역과 고대역의 부호화된 정보는 전체적으로 13.3kbps,13.3kbps,16kbps의 전송률을 갖는다.

2.1 저대역 인코더 알고리즘

QMF에 의해서 분리된 0Hz-3400Hz 대역의 음성 신호로 G.723.1 과 G.729의 알고리즘의 인코더로 부호화한다. G.729 알고리즘의 인코더는 10ms 음성 프레임마다 음성신호를 분석하여 CELP 모델 파라미터를 추출한다. 추출된 음성파라미터들은 프레임 당 80 비트가 할당되어 부호화된다. 입력되는 음성신호는 고역필터를 통과한 후, 자기상관관계에 의해 선형예측계수(LPC)가 추출되며, LPC계수의 계산 및 전송은 10ms 프레임마다 한번 수행한다. LPC 계수의 전송을 위하여, LSP(Line Spectrum Pair)파라미터 변환되며 파라미터 값들은 2단계 예측 벡터로 양자화된다. CS-ACELP에서 고정코드북과 적응 코드북 파라미터 값들은 부프레임 마다 검색되어 전송된다. 피치지연에 대하여 페구간 검색을 하지 않고, 개구간 피치검색 방법을 통해 대략의 피치지연을 예측한 후, 페구간 검색을 수행하며 페구간 검색에서는 개구간 방법에 의하여 예측된 피치 지연의 주변 값들만을 검색하여 정확한 피치 지연값을 찾아낸다. 최적의 피치지연 값을 찾아낸 후, 음성의 여기신호 추출을 위한 고정코드북 검색이 수행되는데 이때 코드북의 구조는 각 프레임마다 4개의 펄스만이 지정된 위치에 +1,-1의 값을 갖게되므로 고정 코드북 검색시 계산량을 줄일수 있다. 적응코드북과 고정코드북의 이득 양자화에는 Conjuate 구조를 갖는 2개의 코드북을 사용한다. 인코딩의 최종단계에서는 다음 프레임의 목적신호를 구하기 위해 합성

및 가중 필터의 메모리 갱신을 수행한다.

반면에 G.723.1은 기본적으로 합성에 의한 분석법을 이용한 선형예측법을 사용하여 240샘플 단위로 프레임을 처리하며 60샘플씩을 가지는 부프레임으로 나누어져 각각의 부프레임에 대해 선형예측 분석을 수행한다. 마지막 부프레임은 LSP계수로 바뀌어지고 이미 구해져 있는 음성의 LSP계수 DC값과의 차이를 PSVQ(predictive split vector quantization)로 양자화 한다. 두 개의 부프레임마다, 지각가중 필터를 통과한 음성을 이용하여 개구간 피치를 찾고 개구간 피치검색 방법을 통해 대략의 피치지연을 예측한 후, 페구간 검색을 수행한다. 페구간 검색에서는 개구간 방법에 의하여 예측된 피치 지연의 주변 값들만을 검색하여 정확한 피치 지연값을 찾아낸다. 페구간 피치예측 후 잔류 신호는 5.3kbps 모드에서는 ACELP, 6.3kbps 모드에서는 MP-MLQ방식을 사용한다. 이와같은 구조로 저대역 신호는 인코딩 되어 8kbps의 전송률을 갖는다.

다음 표는 G.729 와 G.723.1의 비트할당이다.

표1 G.729의 비트할당(8kbps)

Parameter	부프레임0	부프레임1	부프레임2	비트수
LSP	LO,L1,L2,L3			18
적용코드북지연	P1,P2	8	5	13
피치지연parity	P0	1		1
고정코드북인덱스	C1,C2	13	13	26
고정코드북부호	S1,S2	4	4	8
코드북이득1	GA1,GA2	3	3	6
코드북이득2	GB1,GB2	4	4	8
Total				80

표 2 G.723.1 대한 비트할당 (6.3 kbit/s, ()은 5.3kbit/s)

Parameter	부프레임 0	부프레임 1	부프레임 2	부프레임 3	total
LPC					24(24)
적용코드북지연	7(7)	2(2)	7(7)	2(2)	18(18)
모든이득의 결합	12(12)	12(12)	12(12)	12(12)	48(48)
펄스위치	20(12)	18(12)	20(12)	18(12)	73(48)
펄스사인	6(4)	5(4)	6(4)	5(4)	22(16)
Grid index	1(1)	1(1)	1(1)	1(1)	4(4)
Total					189(158)

2.2 고대역 인코더 알고리즘

QMF에 의해 분리된 3400Hz-7000Hz 대역의 음성신호는 DCT를 기반으로 하는 MLT변환되어 벡터양자화된다. MLT는 인접한 MLT프레임에 대한 기본함수 사이에 50% 오버랩을 가지는 선형변환의 전형적인 샘플이다. 그림1 는 이를 보여주고 있다. 이러한 성질을 이용하여 고대역 인코더에서 사용한 MLT변환은 다음과 같다.

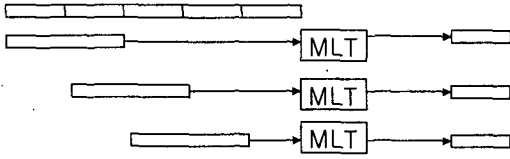


그림 1. 50%의 오버랩을 갖는 MLT

$$mlt(m) = \sqrt{\frac{2}{M}} \sum_{n=0}^{M-1} \cos\left[\frac{\pi}{M}(n + \frac{1}{2})(m + \frac{1}{2})\right] v(n) \quad \dots(1)$$

$$v(n) = w\left(\frac{M}{2} - 1 - n\right)x\left(\frac{M}{2} - 1 - n\right) + w\left(\frac{M}{2} + n\right)x\left(\frac{M}{2} + n\right)$$

$$\text{for } 0 \leq n \leq \frac{M}{2} - 1 \quad \dots(2)$$

$$v\left(n + \frac{M}{2}\right) = w(M - 1 - n)x(M + n) + w(n)x(2M - n)$$

$$\text{for } 0 \leq n \leq \frac{M}{2} - 1 \quad \dots(3)$$

where

$$w(n) = \sin\left(\frac{\pi}{2M}(n + 0.5)\right) \quad \text{for } 0 \leq n < M \quad \dots(4)$$

여기서 M은 데이터의 블록의 길이를 나타낸다. (1)식에서 MLT변환으로 얻은 계수를 나타내고 있다. (2)에서 $v(n)$ 은 현재 프레임과 과거 프레임의 윈도우값의 합을 나타내고 있으며 여기서 $w(n)$ 은 저 전송반대역필터로 MLT변환을 위한 윈도우이다. 위의 식들은 DCT4를 기반으로 한다는 사실을 알 수 있다. 고대역의 입력신호는 한 프레임 단위로 MLT 변환되어 주파수영역의 값으로 변환된다. 한 프레임 내에서 음성의 스펙트럼은 모든 대역에 걸쳐 균등하게 분포되어 있지 않으므로, 모든 MLT 계수에 같은 크기의 비트를 할당하여 양자화 하는 것은 비효율적이다. 즉, 한 부분의 스펙트럼이 더 큰 영역에 더 많은 비트를 할당하는 것이 효율적이다. MLT 계수와 연관하여 표현된다. 한 주파수영역에 해당하는 MLT 계수의 에너지는 다른영역의 에너지보다 큰 값을 갖는다. 본 연구에서는 이러한 음성의 특성과 MLT 계수의 관계를 이용하여 고대역 신호를 부호화한다. 다음은 블록에너지와 블록마다 할당되는 비트의 관계이다.

각 부프레임의 MLT계수들은 에너지 벡터 코드북의 인덱스에 해당되는 에너지 값에 의해 정규화가 되고, 비트할당 정보를 이용하여 벡터 양자화된다. 에너지 벡터 코드북의 인덱스는 MLT 계수의 벡터 양자화에 할당된 비트정보와 이득을 제공한다.

그림 2은 에너지 벡터 코드북과 비트 벡터 코드북 그리고 벡터 코드북의 관계 양자화 구조를 나타낸다. 이와

같은 구조로 고대역의 신호는 부호화되어 8Kbps의 전송률을 갖는다.

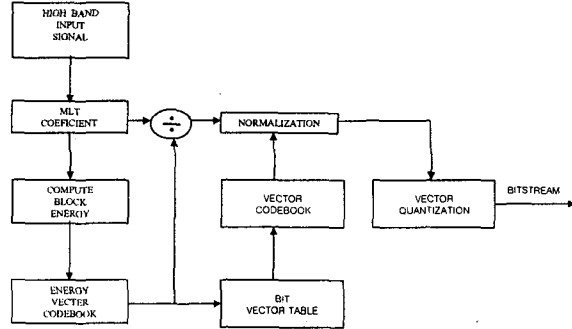


그림 2. 벡터 양자화의 구조

고대역 신호의 인코딩 과정을 요약해 보면 다음과 같다.

- 1) 10ms의 음성신호를 매 프레임마다 MLT변환 한다.
- 2) 변환된 MLT 계수는 몇 개의 부프레임으로 나누어지고, 각 부프레임마다 에너지가 계산된다.
- 3) 계산된 각 서브 블록은 에너지 값들은 에너지 벡터 코드북을 이용하여 양자화되고, 코드북의 인덱스값을 얻는다. 에너지 벡터 코드북의 인덱스값은 부가정보로 전송된다. 에너지 벡터 코드북은 각 서브블록에 할당되는 비트 정보를 포함하는 비트 벡터 코드북과 연결되어 설계된다. 그러므로 에너지 벡터 벡터코드북의 인덱스는 MLT계수를 벡터 양자화할 때, 각 서브 블록에 대하여 할당되는 비트 정보를 제공한다.
- 4) 각 서브 블록내의 MLT계수들은 에너지 벡터 코드북의 에너지값으로 나누어짐으로써 정규화된다.
- 5) 비트 벡터 코드북의 비트 할당 정보를 이용하여, 각 블록의 정규화된 MLT계수는 벡터 양자화 된다.

3. 광대역음성 부호화기의 디코더 알고리즘

디멀티플렉서에서 출력되는 저대역은 G.729, G.723.1으로 역부호화 하고 고대역 정보는 인덱스를 이용하여 신호를 구성하고 IMLT과정을 통해 출력신호를 만든다. 역부호화된 두 신호는 QMF 합성 bank에 의하여 합성되어 최종 출력신호를 만들어 낸다.

3.1 저대역 디코더 알고리즘

G.729의 경우 LSP계수들은 보간된 후 각각의 부프레임에 대한 선형 합성 필터계수로 변환된다. 적응 코드북과 고정 코드북의 여기신호에 이득이 곱해진 후, 단구간 합성필터 및 후단여파기를 통과하여 음성신호는 합성된다. G.723.1의 경우 LSP 계수들은 보간되고 합성필터

계수로 변환하고 포먼트 필터를 통과한다. 피치 검색과 여기신호는 피치여기필터를 거쳐 합성필터계수로 변환한다. 각각의 계수는 이득 크기 단위로 계산되어 출력된다.

3.2 고대역 디코더 알고리즘

고대역의 정보는 Inverse VQ 와 IMLT 과정에 의해 역부호화된다. 고대역 역부호화기에서 사용한 IMLT 의 식은 다음과 같다.

$$u(n) = \sum_{m=0}^{M-1} \sqrt{\frac{2}{M}} \cos\left[\frac{\pi}{M}\left(m+\frac{1}{2}\right)\left(n+\frac{1}{2}\right)\right] ml(m) \dots(5)$$

$$y(n) = w(n)x\left(\frac{M}{2}-1-n\right) + w\left(\frac{M}{2}-1-n\right)u_old(n) \dots(6)$$

$$\text{for } 0 \leq n \leq \frac{M}{2}-1$$

$$y\left(n+\frac{M}{2}\right) = w\left(\frac{M}{2}+n\right)u(n) - w\left(\frac{M}{2}-1-n\right)u_old\left(\frac{M}{2}-1-n\right)$$

$$\text{for } 0 \leq n \leq \frac{M}{2}-1 \dots(7)$$

$$\text{where } w(n) = \sin\left(\frac{\pi}{2M}(n+0.5)\right) \text{ for } 0 \leq n < M-1$$

$$u_old(n) = u(n+160) \dots(8)$$

위의 (5)에서 u(n)은 IMLT의 계수를 구한다. 식(6)(7)은 u(n)에서 구한 계수를 넣어서 각각을 윈도우하여 계수를 구한다. 식(8)은 u_old(n)에 대해 사용하는 안은 받은 다음 프레임에서 사용하기 위해 u(n+160)과 같이 저장한다.

4. 실험 및 결과

설계된 고대역 부호화기의 성능 평가는 G.722 48kbps G.723.1 와 G.729를 사용하여 합성된 광대역 신호를 비교하였다. 실험에 사용된 문장은 남자 16문장과 여자 16문장을 사용했으며 9명의 사람이 참가 했다. 본 논문에서 제안한 16kbps 광대역 음성 부호화기의 성능 평가는 ITU-T의 표준안인 G.722를 포함한 MOS(Mean Option Score) 시험을 통하여 수행되었다. MOS 시험에는 ITU-T 의 표준안인 G.722의 48kbps, 56kbps, 64kbps 전송률을 가지는 음성이 포함되었으며, 음질 시험의 개관적인 지표를 위하여 MNRU(Modulated Noise Reference Unit) 의 잡음 크기가 28dB, 31dB, 36dB인 음정도 포함되었다.

실험결과 G.723.1-MLT VQ인 경우는 G.722 48kbps와 비교했을 때 5.3kbps는 MOS 0.26 , 6.3kbps는 MOS 0.17의 음질 저하를 보였다. 하지만 G.729-MLT VQ인 경우엔 약 MOS 0.09 음질이 나음을 확인하였다. 음질

시험의 결과는 표 3에 나타나 있다.

표3. MOS 시험결과

	MNRU					G.722			설계된 코더		
	원음	28dB	31dB	33dB	36dB	48	56	64	G.723.1		G.729
						Kbps	Kbps	Kbps	5.3	6.3	
남자	4.31	3.30	3.34	3.92	4.19	4.15	4.24	4.28	3.94	3.99	4.20
여자	4.31	3.36	3.68	3.85	4.04	4.01	4.22	4.19	3.72	3.84	4.14
전체	4.31	3.32	3.70	3.89	4.11	4.08	4.23	4.24	3.82	3.91	4.17

5. 결론

본 연구에서는 13.3kbps, 14.3kbps, 16kbps의 전송률을 갖는 광대역 음성부호화기를 설계하였다. 기존의 저대역 음성 부호화기인 G.723.1 과 G.729를 저대역에서 사용함으로써 기존의 음성통신 시스템과의 호환성을 유지할 수 있다. 고대역 음성을 전송하기 위해 MLT 변환을 이용한 벡터 양자화 방법을 제안하였다. 제안된 광대역 음성부호기의 성능평가는 MOS 시험을 통하여 수행되었고 14.3kbps G.723.1-MLT VQ인 경우는 48kbps G.722 보다 약간의 음질 저하를 보였고, G.729-MLT VQ인 경우엔 56kbps G.722와 비슷한 음질을 보였다.

참고문헌

- [1] ITU-T Recommendation G.729. "Coding of speech at 8kb/s using Conjugate Structure Algebraic Code Excited Linear Prediction", Mar,1996
- [2] ITU-T Recommendation G.723.1. "Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3kbps", Mar,1996
- [3] S. Shlien, "The Modulated lapped transform, its time - varying forms, and its applications to audio coding standards" IEEE Trans. on speech and audio processing, vol. 5, No.4, July 1997.
- [4] H. S. Malvar, "Lapped transforms for efficient transform/subband coding." IEEE Trans. on Acoustics, Speech and signal processing, vol. 38, No.6, pp.969-978, June 1990.
- [5] A. M. Kondoz, Digital speech, John Wiley & Sons, 1996.