

푸리에 후처리 보상 기법을 이용한 향상된 MELP 음성부호화기

고 봉 옥 , 김 종 교
전북대학교 전자정보공학부

전화 : (063) 272-1177 / 팩스 : (063) 270-2400

Improved MELP Coder Using Fourier Post Processing Compensation Method

Bong-Ok Ko , Chong-Kyo Kim
Div. of Electronics and Information Engineering, Chonbuk National University

E-mail : moakdosa@hanmail.net

Abstract

This paper presents an improved MELP Coder using Fourier magnitude compensation method chosen the new 2.4 kbit/s U.S. Federal Standard.

Although the MELP is quite good, it has some distortion for low-pitch male speakers.

An improved MELP coder includes a post processing for the Fourier magnitude model that allows the MELP to reconstruct the lower frequency spectrum more accurately and improve the speech quality.

In this new compensation algorithm, the harmonic magnitudes in the low frequencies are adaptively modified by removing the effect of the two filters. Also, the bit rate of the improved MELP coder is the same as that of the Federal Standard MELP coder.

Formal quality tests show that the improved MELP coder was preferred over the Federal

Standard MELP coder by 80.8%.

1. 서 론

통신기술의 중심이 아날로그에서 디지털로 이동함에 따라서 무선통신, 음성메일, 멀티미디어 메시징, 네트워크 통신, 합성음 기술 등 많은 응용분야에서 음성부호화에 대한 연구가 행해지고 있다.

음성부호화는 효율적인 전송 및 저장을 위하여 음성 신호를 압축된 디지털 데이터로 표현하는 것으로써, 인지적 음질을 저하시키지 않고 최소 비트수로 나타내는 것이다.

1976년에 미 국방성은 2.4 kbit/s에서의 보안통신을 위한 알고리즘으로 LPC(linear prediction coefficient)를 추천했다. 1993년 DoD Digital Voice Processing Consortium(DDVPC)은 2.4 kbit/s에서의 새로운 미 연방 표준을 선택하기 위한 최소 요구조건과 실험계획을 만들었고, 1996년 3월에 채택되었다[1][2][3].

MELP는 2.4 kbit/s에서 4.8 kbit/s FS-1016보다 음질이 좋으며, 기존의 LPC에 기반하여 5가지의 특징을 추가하여 성능을 향상시켰다[4][5][6][7].

본 논문에서는 미 연방 표준인 MELP의 음질개선을 위해 퓨리에 크기를 보상하는 기법을 제안했다.

본 논문의 구성은 다음과 같다. 2장에서는 MELP 음성코덱을 살펴보고, 3장에서는 음질개선 기법에 대해서 설명하며, 4장에서는 모의실험 및 성능평가에 대해서 언급한다. 마지막으로 5장에서 결론을 맺는다.

2. MELP 음성코덱

2.1 부호화 과정

부호화기에서는 파라미터 모델을 얻기 위해서 본래의 음성이 분석된다. 세부적인 분석과정은 그림 1과 같다[8].

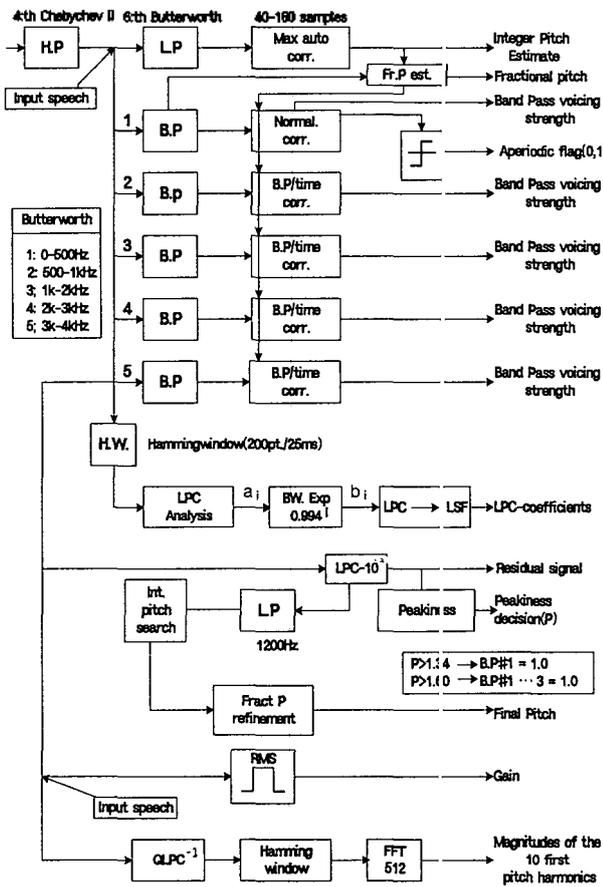


그림 1. MELP 부호화 과정

먼저 4차 Chebychev 2형 고역통과 필터로 직류성분을 제거하고, 6차 Butterworth 저역통과 필터로 전역피치(pitch)를 검색한다. 입력음성을 5개의 주파수 대역으로 나누어서 대역통과 보이싱(voicing)을 계산하고, 10차 LPC 분석을 한다. LPC 잔차(residual)신호로부터 피크를 계산하고, 최종 피치 분석을 하게 된다. 이득

분석/양자화 단계를 거친 후 처음 10개의 피치 고조파들의 퓨리에 크기를 계산하고 양자화한다.

2.2 복호화 과정

복호화기는 비트 스트림(bit-stream)을 언팩(unpack)하여 고품질의 합성음을 만든다. 세부과정은 그림 2와 같다.

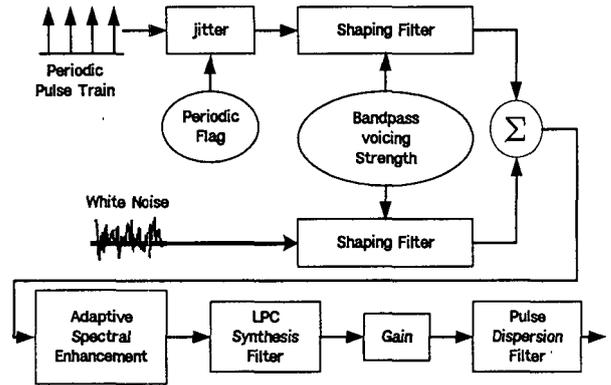


그림 2. MELP 복호화 과정

모든 MELP 파라미터들은 각각의 합성된 피치 주기가마다 보간(interpolate)된다. 필터된 펄스와 잡음 여기들을 더함으로써 혼합여기를 생성하고, 10차 pole/zero 필터인 적응 스펙트럼 향상과정을 거친 후에 10차 all-pole 필터로 합성하며, 이득 조절을 하게된다. 최종적으로 스펙트럼을 평활화 한 삼각 펄스들로부터 유도된 65차 FIR 필터를 통과한다.

2.3 전송 포맷

MELP 부호화기는 매 22.5ms의 프레임마다 54 bit를 전송하여 2.4 kbit/s의 전송률을 갖는다. 유성음과 무성음 프레임에 대한 비트 할당은 표 1과 같다[9].

표 1. MELP 부호화기의 비트할당

파라미터	유성음	무성음
LSF	25	25
퓨리에 크기	8	-
이득(프레임마다 2개)	8	8
피치	7	7
대역통과 보이싱	4	-
비주기 플래그	1	-
에러 정정	-	13
동기 비트	1	1
합계 / 22.5ms 프레임	54	54

3. 음질개선 기법

3.1 MELP의 문제점

푸리에 크기를 모델링하는 기법은 저주파 스펙트럼을 좀 더 정확하게 재생하고 음질을 향상시키는 장점이 있는 반면에, 피치가 낮은 남성화자들의 합성음의 경우 여전히 고역통과 필터된 특징이 나온다.

이것은 MELP에서 사용되는 두 개의 서로 다른 필터에 의한 것으로서, 적응 스펙트럼 향상필터와 60Hz의 차단주파수를 가진 고역통과 필터 때문이다.

3.2 푸리에 크기 보상기법

피치가 낮은 남성화자의 합성음을 향상시키기 위해 저주파에서의 푸리에 크기들은 적응 스펙트럼 향상 필터와 고역통과 필터를 제거함으로써 적극적으로 수정한다.

수정된 고조파 크기는 아래와 같다.

$$|\tilde{M}(e^{jw_i})| = |M(e^{jw_i})| \frac{\sqrt{E}}{|H(e^{jw_i})|} \quad (1)$$

여기서, w_i 는 i 번째 고조파 주파수, E 는 평균 푸리에 스펙트럼 에너지, $|M(e^{jw_i})|$ 는 수정하지 않은 푸리에 크기를 뜻한다.

향상된 MELP 부호화기는 부호화된 신호 $M(e^{jw_i})$ 의 푸리에 크기를 사용하기 보다 수정된 고조파의 푸리에 크기를 사용한다. 두 개의 필터 크기 응답인 $|H(e^{jw_i})|$ 은 아래와 같다.

$$|H(e^{jw_i})| = |H_1(e^{jw_i})| |H_2(e^{jw_i})| \quad (2)$$

여기서, $|H_1(e^{jw_i})|$ 는 적응 스펙트럼 향상필터의 크기응답이고, $|H_2(e^{jw_i})|$ 는 고역통과 필터의 크기응답이다.

또한 적응 스펙트럼 향상필터의 이점을 살리기 위해서 제 1고조파보다 200Hz 작은 고조파들에 대해서만 적용을 하였다. 제 1고조파 F_1 은 양자화한 LSF(line spectrum frequencies)를 이용하여 구하게 된다.

$$F_1 = \begin{cases} \frac{\hat{f}_1 + \hat{f}_2}{2} & \text{if } \hat{f}_2 - \hat{f}_1 < \hat{f}_3 - \hat{f}_2 \\ \frac{\hat{f}_2 + \hat{f}_3}{2} & \text{otherwise} \end{cases} \quad (3)$$

여기서 \hat{f}_i 는 i 번째 양자화한 LSF이다.

알고리즘의 전체적인 흐름도는 그림 3과 같다.

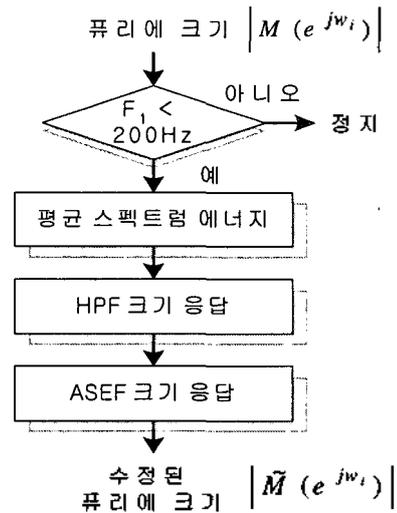


그림 3. 알고리즘 흐름도

4. 모의실험 및 성능평가

아래 3개의 문장에 대해서 각각 5명의 남성화자에 대해서 푸리에 크기들을 비교한 뒤, 주관적 평가 방법인 A/B 비교방법을 통해 성능향상을 입증하였다[10].

- ① /여기는 전북대 음성처리연구실입니다/
- ② /음성기술은 컴퓨터 그 자체의 미래입니다/
- ③ /하나님은 당신을 사랑하십니다/

남성화자에 대하여 피치를 변화시키면서 푸리에 크기를 보상한 결과는 표 2와 같다.

표 2. 실험의 결과

	문장①		문장②		문장③	
피치[Hz]	125.7	95.4	118.9	88.7	122.4	93.3
차이(보상전)	-2.2	-4.8	-7.2	-12.2	-3.6	-5.6
차이(보상후)	-0.6	0.8	1.8	3.6	-0.2	1
보상폭[dB]	1.6	4	5.4	8.6	3.4	4.6

피치가 낮아질수록 원음과 합성음간의 퓨리에 크기는 커짐을 알 수 있었고, 복원력 또한 높음을 확인할 수 있었다.

음질평가는 총 15개의 문장 쌍에 대해서 미 연방 표준 MELP 코더를 참조모델로 하고, 20명의 비전문 청취자가 각각의 문장에 대해서 합성음을 들은 후 두 개의 코더 중에 하나를 강제로 선택하는 방식을 사용하였다.

주관적 평가 결과는 표 3과 같으며, 향상된 MELP 코더가 참조모델인 표준 MELP보다 80.3%의 선호도를 가진다.

표 3. 주관적 평가 결과

	문장①		문장②		문장③	
피치[Hz]	125.7	95.4	118.9	88.7	122.4	93.3
표준 MELP	30%	15%	20%	10%	25%	15%
향상된 MELP	70%	85%	80%	90%	75%	85%

향상된 MELP의 선호도를 그래프로 표시하면 그림 4와 같다.

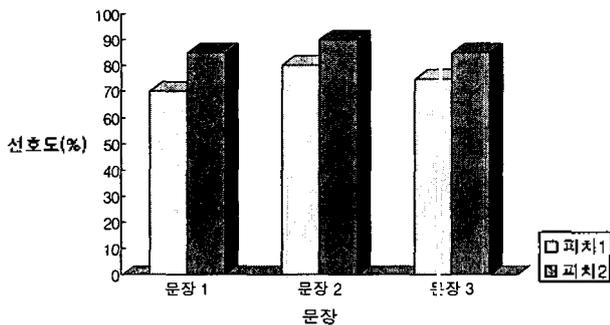


그림 4. 향상된 MELP의 선호도

5. 결론

본 논문에서는 2.4 kbit/s 미 연방 표준 음성부호화기인 MELP의 음질개선을 위해 퓨리에 크기를 후처리적으로 보상하는 기법을 제안했다.

본 알고리즘은 퓨리에 크기 값을 수정하기 때문에 원 MELP의 전송률에는 영향을 끼치지 않는 특징이 있다.

제안한 보상 알고리즘의 성능을 측정하기 위해 5명의 남성화자에 대해 피치를 변화시키면서 퓨리에 크기를 비교 분석하였다. 음질평가는 주관적 음질평가 방법인 A/B 비교방법을 통해 약 80.8%의 선호도를 가짐을 알 수 있었다.

앞으로 더욱 향상된 음질을 달성하기 위하여 더욱 효과적인 후처리 기법에 대한 연구와 MELP의 문제점을 좀더 보완하는 피치 탐색 알고리즘 및 다양한 잡음 환경에서의 연구가 필요하다.

참고문헌

- [1] M. Bielefeld, "1995 Terms of Reference (TOR) for the Selection of a New Government Standard Voice coder at 2400 bps," MP -8932D-95-05, The MITRE Corporation, 1995.
- [2] M. Bielefeld, L. M. Supplee, "Developing a Test Program for the DoD 2400 bps Vocoder Selection Process," Proceedings of ICASSP, IEEE, pp. 1141-1144, 1996.
- [3] M. Bielefeld, "1995 Test and Evaluation Plan (TEP) for the Selection of a New Government Standard Voice coder at 2400 bps," MTR MTR95W0000136, The MITRE Corporation, 1995.
- [4] A. V. Macree, K. Truong, E. B. George, T. P. Barnwell III, and V. Viswanathan, "A 2.4 kbits/s MELP Coder Candidate for the New U.S. Federal Standard," in Proc. ICASSP, vol. 1 pp. 200-203, May 1996.
- [5] L. M. Supplee, R. P. Cohn, J. S. Collura, and A. V. Macree, "MELP: The New Federal Standard at 2400 bps," in Proc. ICASSP, vol. 2 pp. 1591-1594, April 1997.
- [6] 문인섭, 김종교, "VSELP 음성 부호화기의 성능 향상에 관한 연구," 전북대학교, pp. 14-15, 1995.
- [7] J. P. Campbell, Jr., T. E. Tremain, and V. C. Welch, "The Federal standard 1016 4800 bps CELP voice coder," Digital Signal Processing, 1(3):145-155, 1991.
- [8] Lynn M. Supplee, Ronald P. Cohn, John S. Collura, Alan V. McCree MELP: The New Federal Standard At 2400 BPS.
- [9] "Specification for the Analog to Digital Conversion of Voice by 2400 bps Mixed Excitation Linear Prediction" Draft May 28, 1998. www.plh.af.mil/ddvpc
- [10] F. Chiou, "User-Interactive Speech Enhancement Using Fuzzy Logic," PhD thesis, Georgia Institute of Technology, 1998.