

8kbps에 있어서 ACFBD-MPC에 관한 연구

이시우

상명대학교 정보통신공학과

A Study on ACFBD-MPC in 8kbps

See-Woo Lee

Information and Telecommunication, SANGMYUNG University

요약 최근 무선네트워크의 효율을 높이기 위하여 신호압축 방식의 사용이 증가되고 있다. 특히, MPC 시스템은 비트율을 줄이기 위하여 피치추출 방법과 유성음과 무성음의 음원을 사용하였다. 일반적으로, 유성음원과 무성음원을 사용하는 MPC 시스템에 있어서, 같은 프레임 안에 모음과 무성자음이 있는 경우에 재생 음성파형에 일그러짐이 나타난다. 이것은 대표구간의 멀티펄스를 피치구간마다 복원하는 과정에서 재생 음성파형이 정규화 되는 것이 원인으로 작용한다.

본 논문에서는 재생 음성파형의 일그러짐을 제어하기 위하여 피치구간 마다 멀티펄스의 진폭을 보정하고, 특정 주파수를 이용하는 ACFBD-MPC(Amplitude Compensation Frequency Band Division-Multi Pulse Coding)를 제안하였다. 실험은 남자와 여자 음성에서 각각 16개의 문장을 사용하였으며, 음성신호는 10kHz 12bit로 A/D 변환하였다. 또한 8kbps의 부호화 조건에서 ACFBD-MPC 시스템을 구현하고, ACFBD-MPC의 SNR를 평가하였다. 그 결과 ACFBD-MPC의 남자 음성에서 14.2dB, 여자 음성에서 13.6dB 임을 확인할 수 있었으며, ACFBD-MPC가 기존의 MPC에 비하여 남자음성에서 1dB, 여자음성에서 0.9dB 개선되는 것을 알 수 있었다. 이 방법은 셀룰러폰이나 스마트폰과 같이 낮은 비트율의 음원을 사용하여 음성신호를 부호화하는 방식에 활용할 수 있을 것으로 기대된다.

Abstract Recently, the use of signal compression methods to improve the efficiency of wireless networks have increased. In particular, the MPC system was used in the pitch extraction method and the excitation source of voiced and unvoiced to reduce the bit rate. In general, the MPC system using an excitation source of voiced and unvoiced would result in a distortion of the synthesis speech waveform in the case of voiced and unvoiced consonants in a frame. This is caused by normalization of the synthesis speech waveform in the process of restoring the multi-pulses of the representation segment.

This paper presents an ACFBD-MPC (Amplitude Compensation Frequency Band Division-Multi Pulse Coding) using amplitude compensation in a multi-pulses each pitch interval and specific frequency to reduce the distortion of the synthesis speech waveform. The experiments were performed with 16 sentences of male and female voices. The voice signal was A/D converted to 10kHz 12bit. In addition, the ACFBD-MPC system was realized and the SNR of the ACFBD-MPC estimated in the coding condition of 8kbps. As a result, the SNR of ACFBD-MPC was 13.6dB for the female voice and 14.2dB for the male voice. The ACFBD-MPC improved the male and female voice by 1 dB and 0.9 dB, respectively, compared to the traditional MPC. This method is expected to be used for cellular telephones and smartphones using the excitation source with a low bit rate.

Keywords : Speech Signal Processing, Speech Coding Condition, Speech Coding System, Coding Parameter

본 논문은 상명대학교 연구과제로 수행되었음.

*Corresponding Author : See-Woo Lee(SANGMYUNG Univ.)

Tel: +82-41-550-5352 email: swlee@smu.ac.kr

Received April 26, 2016

Revised (1st June 13, 2016, 2nd June 27, 2016, 3rd July 1, 2016, 4th July 6, 2016)

Accepted July 7, 2016

Published July 31, 2016

1. 서론

여 스마트폰 보급확산과 더불어 통신기술뿐만 아니라 멀티미디어 관련 기술도 다양한 형태로 진보하고 있다. 그러나 스마트폰 사용자 증가에 따른 통신용량의 과부하에 대한 대책은 미흡한 편이다.

통신용량의 과부하에 대책으로 멀티미디어 신호를 효율적으로 압축하는 기술이 대안이 될 수 있다. 신호를 압축하는 것은 신호를 전송하는 비트율을 낮춤으로서 통신용량의 과부하를 줄이는 것이다. 음성신호는 유무선 통신에서 반드시 다루어야만 하는 신호로서 다양한 방법으로 신호를 압축하며, 일반적으로 음성신호의 주기성과 비주기성의 특성을 이용하여 음성신호를 압축한다 [1,2,3]. 음성신호의 주기성과 비주기성은 성대 진동이 주기적으로 이루어지는 유성음과 혀와 치아에 흐르는 비주기적인 공기흐름에 의하여 발생하는 무성음의 특성에 따라 주어지는 것이다. 이러한 특성을 이용하여 주기적인 유성음 재생에는 펄스성 음원을, 비주기적인 무성음 재생에는 백색잡음의 음원을 사용하는데, 특히 펄스성 음원은 피치정보를 이용하여 대표구간의 음원을 피치구간 마다 재구성함으로써 비트율을 실현할 수 있다[4,5].

그러나 프레임마다 대표구간의 피치정보[6,7,8]를 이용하여 펄스성 음원을 재생하는 경우, 펄스성 음원이 대표구간의 진폭정보에 의해 정규화되는 현상과 유성음과 무성음의 중간특성을 갖는 음성신호, 즉 무성음에서 유성음으로 천이되는 과정에서 발생하는 음성신호는 유성음원과 무성음원 어느 한쪽의 음원으로도 완전한 재생이 어렵다.

본 연구에서는 정규화된 펄스성 음원의 진폭을 보상하고 특정주파수 정보를 전송하여 무성음과 유성음의 중간특성을 갖는 프레임의 음성신호를 재생하는 방법으로 ACFBD-MPC(Amplitude Compensation and Frequency Band Division-Multi Pulse Coding)을 제안하고 개선된 신호대 잡음비(SNR)를 알아보려고 한다.

2. 진폭과 주파수 보정

ACFBD-MPC는 Fig. 1에 나타낸바와 같이 원 음성신호 $x(n)$ 와 멀티펄스 $v(n)$ 에 의하여 재생된 음성신호 $\bar{x}(n)$ 로부터 식(1)이 최소가 되도록 펄스의 진폭(g_i)과

위치(m_i)를 결정한다.

$$E = \sum_{n=0}^{N-1} [x(n) - \bar{x}(n) * w(n)]^2 \quad (1)$$

$$\bar{x}(n) = \sum_{i=1}^K g_i \cdot h(n - m_i)$$

여기에서 K , g_i , m_i , $h(n)$ 는 멀티펄스의 수, i 번째의 펄스진폭과 위치, 음성재생 필터(STP)의 임펄스 응답을 나타낸다. 식(1)의 $w(n)$ 은 Weighting 필터로서 다음 식과 같은 특성을 갖는다.

$$W(z) = (1 - \sum_{i=1}^p a_i z^{-i}) / (1 - \sum_{i=1}^p a_i r^i z^{-i}), (0 < r < 1) \quad (2)$$

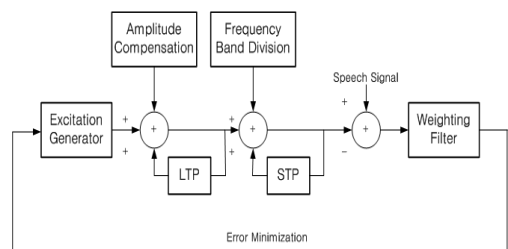


Fig. 1. ACFBD-MPC

아울러, ACFBD-MPC의 전달함수 $H(z)$ 는 다음 식으로 나타낼 수 있다.

$$H(z) = 1 / (1 - \sum_{i=1}^Q b_i z^{-(T+i-1)}) \cdot (1 - \sum_{i=1}^P k_i z^{-i}) \quad (3)$$

여기에서 b_i , T , a_i 는 자기상관함수(R)와 상호상관함(Φ)수로부터 구하여진 피치계수, 피치주기, 예측계수를 나타내며, 식(1)를 최소화하는 멀티펄스의 진폭과 위치는 다음 식을 만족한다.

$$\sum_{j=1}^K g_j \cdot R(|m_i - m_j|) = \Phi(m_i), (1 \leq i \leq K) \quad (4)$$

윗 식으로 표현되는 멀티펄스는 음성신호의 에너지 증감에 따라서 멀티펄스의 진폭도 증감하는 펄스로 구성된다. 그러나 정보 전송률을 낮추기 위하여 피치정보를 이용한 대표구간의 멀티펄스를 사용하여 구동음원을 재

구성함에 따라 멀티펄스의 진폭이 정규화 된다. 이러한 정규화된 진폭을 셀룰러폰과 스마트폰에 적용된 CELP 방식에서는 코드북의 구동음원을 재구성할 때 발생하는 에러신호를 이용하여 음원이득을 제어하여 음질을 개선하고 있다[9]. 또한 유성음과 무성음의 중간특성을 갖는 음성신호를 유성음원과 무성음원 어느 한쪽의 음원으로 재생하기 어렵다. 따라서 본 연구에서는 음질을 개선하기 위하여 정규화 된 멀티펄스의 진폭의 이득을 제어하고, Fig. 2에 나타낸바와 같이 피치정보와 영교차율(Zero-crossing Rate)을 비교하여 선정된 프레임에 대해 특정 주파수 정보를 이용하여 음질을 개선한다.

우선, 유성음을 다음식과 같은 n 개의 파형으로 나타낸다.

$$s(t) = \sum_{k=1}^M (A_V(k) e^{-\pi B_V(k)t} \cdot \cos(2\pi F_V(k)t + \theta_V(k))) \quad (5)$$

여기에서 $F_V(k)$ 는 공진주파수, $B_V(k)$ 는 대역폭, $A_V(k)$ 와 $\theta_V(k)$ 는 파형의 진폭과 위상을 나타낸다.

일반적으로 음성파형은 진폭레벨이 증가하거나 감소하는 패턴을 갖는데, 이러한 진폭레벨 증감 패턴을 고려한 멀티펄스의 진폭과 위치를 다음 식으로 나타낼 수 있다.

$$A_V(k) = \gamma \cdot \sum_{k=1}^M \alpha_k \quad (6)$$

$$\theta_V(k) = \sum_{k=1}^M g_k \cdot \delta(n - m_k) \quad (7)$$

여기에서 γ 는 각 프레임의 진폭레벨의 증감패턴, α_k 는 피치위치에 있는 멀티펄스의 진폭 값을 나타낸다. 아울러 α_k 는 피치위치의 멀티펄스 진폭 값 g_k 와 g_{k+1} 의 차이를 다음 식과 같이 계산하여 수신측에 전송한다.

$$\alpha_k = |g_k - g_{k+1}| \quad (k = 1, 2, 3, \dots, M) \quad (8)$$

수신측에 전송된 대표구간의 멀티펄스는 피치구간마다 멀티펄스의 진폭과 위치를 보정하여 멀티펄스를 재구성한다. 여기에서 α_k 를 전송하는데 사용하는 비트를 줄이기 위해 g_k 와 피치위치에 있는 멀티펄스의 최대 진폭 값(g_{\max})으로부터 다음 식과 같은 진폭보정계수(β_k)를

구하여 사용한다. 즉, 전송된 β_k 와 g_{\max} 의 정보는 수신측에서 α_k 를 구하여 적용할 수 있고, α_k 에 비해 β_k 가 작은 값이기 때문에 보다 작은 비트를 할당할 수 있다.

$$\beta_k = |\alpha_k / g_{\max}| \quad (9)$$

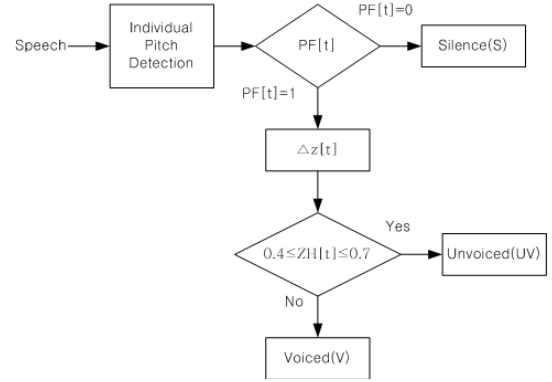


Fig. 2. V/UV/S Extraction

수신측에 전송한 g_{\max} 와 β_k 로 피치위치에 있는 멀티펄스의 진폭을 보정하여 생성된 멀티펄스 음원을 다음 식으로 나타내었다.

$$v(n) = \sum_{k=1}^M (g_k + \gamma |\beta_k \cdot g_{\max}|) \cdot \delta(n - m_k) \quad (10)$$

윗 식으로 표현된 멀티펄스 음원을 피치간격마다 재구성하여 음성재생필터를 구동하여 음성신호를 재생한다.

3. 부호화 조건

ACFBD-MPC에 있어서, 프레임 길이는 FFT를 고려하여 25.6ms로 하였으며, 3.4kHz LPF로 주파수 대역을 제한하고, 10kHz, 12bit로 표본화 및 양자화 하였다.

ACFBD-MPC를 이동통신 음성부호화 방식에 적용하기 위해서는 일반적으로 음성신호의 전송률이 8kbps 이하의 되어야 하기 때문에 이번 연구에서 ACFBD-MPC의 전송률이 8kbps가 되도록 Table 1과 같이 부호화 하였다. 부호화 조건을 살펴보면, V/UV/S 선택정보에 2bit를 할당하였고, 음성재생필터의 차수는 10차를 사용하였으며, 음성재생필터 계수의 변화가 스펙트럼의 변화에

미치는 영향은 낮은 차수의 계수일수록 영향이 크기 때문에 낮은 차수일수록 많은 bit를 할당하였다.

Table 1. Coding Condition

Parameter[bit]	MPC	ACFBD-MPC
V/UV/S	2	2
Voiced Segment		
Coefficient of Synthesis Filter	7,6,5,5,4 3,3,3,3,3	7,6,5,5,4 3,3,3,3,3(42)
g_{max}	10	10
g_k	8	6
m_k	8	6
β_k		3
γ		2
Number of Multi-Pulse	10	10
Pitch	8	8
Unvoiced Segment		
Max Amplitude		7
Frequency Signal of Low Frequency Band		3
Frequency Signal of High Frequency Band		3
	222	214
kbps	8	8

4. 실험결과

음성통신에 사용할 수 있는 약 8kbps의 부호화 시스템으로서 ACFBD-MPC를 Fig. 3에 나타내었다. 음성 10kHz 12bit A/D, D/A 변환기와 3.4kHz LPF를 입출력단에 구비하였다.

FIR-STREAK 필터로부터 피치정보를 얻고, 자기상관함수(ACF)와 상호상관함수(CCF)로부터 멀티펄스를 추출하고 피치정보를 이용하여 대표구간의 멀티펄스 정보를 수신측에 전송한다. 또한 영교차율(ZCR)과 피치정보를 이용하여 FFT를 실행한 후 음성재생에 필요한 주파수정보를 전송한다. 아울러 영교차율과 피치정보로부터 V/UV/S의 선택정보를 추출한다. V인 경우에 멀티펄스 음원을 사용하여 음성신호를 재생하고, UV인 경우에는 선택된 주파수 정보만을 이용하여 음성신호를 재생한다. 그리고 S인 경우에는 음성신호가 없는 프레임이므로 특별한 신호처리를 실시하지 않고 25.6ms의 프레임처리를 시행한다.

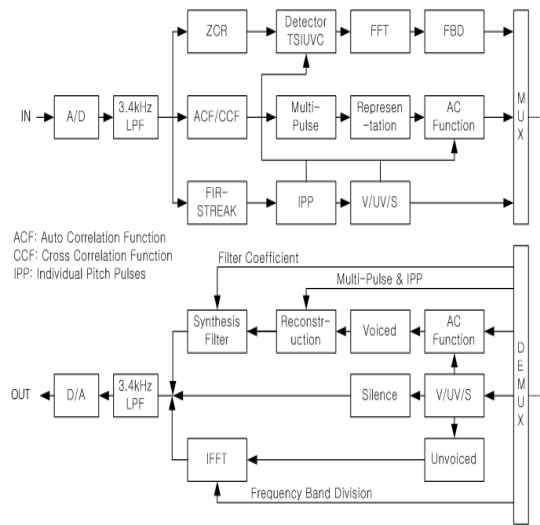


Fig. 3. System of ACFBD-MPC

이번에 새롭게 제안한 ACFBD-MPC와 기존 MPC의 비교평가는 청각적인 비교평가인 주관평가 보다는 신호 왜곡의 개선정도를 확인할 수 있는 객관평가인 SNR만을 사용하여 비교하고자 하였다. 이번 실험에서 사용한 음성샘플은 일반인 남녀 4명이 조용한 환경에서 10kHz 12bit로 녹음한 것으로서 Table 2에 제시된 음성샘플을 사용하여 SNR를 측정하였다.

Table 2. Speech Sample

	male	female
Participants	4	4
Time	54.4 sec	54.4 sec
Sentences	16	16
Vowel	145	145
Unvoiced Cosegment	34	34

Table 3. SNR_{seg} of MPC and APFBD-MPC

Method [dB]	kbps	male	female
MPC	8	13.2	12.7
ACFBD-MPC	8	14.2	13.6

SNR를 측정한 결과, Table 3에 나타낸바와 같이 약 8kbps에 있어서 ACFBD-MPC는 남자음성에서 14.2dB, 여자음성에서 13.6dB를 얻었다. 기존의 MPC에 비하여 남자음성에서 1dB, 여자음성에서 0.9dB 정도 개선되었

으며 ACFBD-MPC와 MPC의 *SNR* 결과를 Fig. 4에 나타내었다. *SNR*의 결과로부터 알 수 있듯이 MPC에 비하여 ACFBD-MPC의 음질이 개선된 것을 알 수 있었다.

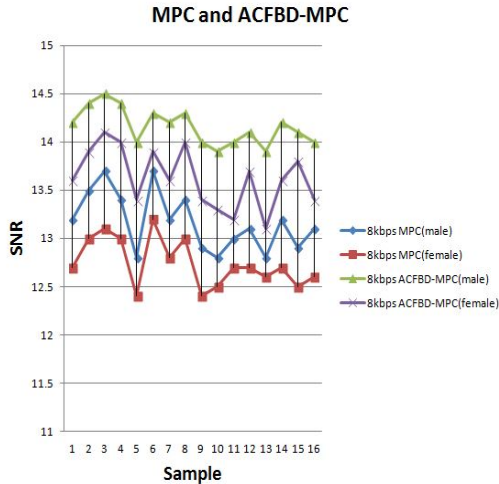


Fig. 4. SNR of MPC and ACFBD-MPC

5. 결론

본 연구에서는 멀티펄스 음성부호화 방식의 구동음원인 멀티펄스 음원의 진폭을 보정하는 동시에 특정 주파수 정보를 이용하여 재생 음성신호의 음질을 개선하는 ACFBD-MPC를 새롭게 제안하였다. 실험결과, 기존의 MPC에 비하여 남자음성에서 1dB, 여자음성에서 0.9dB 정도 개선된 것을 알 수 있었다. 다만, *SNR*을 좀더 개선하기 위해서는 보다 다양한 음성부호화 파라미터를 적용할 필요가 있다. 따라서 본 연구의 향후 과제로서 8kbps에서 다양한 음성부호화 파라미터 적용한 새로운 음성부호화 시스템으로부터 보다 개선된 *SNR*을 얻고자 한다.

REFERENCES

- [1] Selma Ozaydm, Buyurman Baykal: "Matrix quantization and mixed excitation based linear predictive speech coding at very low bit rates", *Speech Communication* 41, pp.381-392, 2003.
DOI: [http://dx.doi.org/10.1016/S0167-6393\(03\)00009-8](http://dx.doi.org/10.1016/S0167-6393(03)00009-8)
- [2] K.Krishna, V.L.N.Murty, K.R.Ramakrishnan: "Vector quantization of excitation gains in speech coding",

Signal Processing 81, pp. 203-209, 2001.

DOI: [http://dx.doi.org/10.1016/S0165-1684\(00\)00200-0](http://dx.doi.org/10.1016/S0165-1684(00)00200-0)

- [3] Ghaemmaghami, S., Sridharan, S.: "Very low rate speech coding using temporal decomposition," *IEEE Electron. Lett.* 35(6), pp. 456-457, 1999.
DOI: <http://dx.doi.org/10.1049/el:19990316>
- [4] LeBlanc, W.P., Bhattacharya, B., Mahmoud, S.A.: "Efficient search and design procedures for robust multi stage vector quantization of LPC parameters for 4kbps speech coding," *IEEE Trans. Speech Audio Process*, pp. 373-385, 1993.
DOI: <http://dx.doi.org/10.1109/89.242483>
- [5] McCree, A.V., Barnwell, T.P.: "A mixed excitation LPC vocoder model for low bit rate speech coding", *IEEE Trans. Speech Audio Process*, pp. 242-250, 1995.
DOI: <http://dx.doi.org/10.1109/89.397089>
- [6] L. Hodgson, M. E. Jernigan, B. L. Wills: "Nonlinear Multiplicative Cepstral Analysis for Pitch Extraction in Speech," *IEEE, S4b.11*, 1990.
DOI: <http://dx.doi.org/10.1109/ICASSP.1990.115616>
- [7] Lawrence R. Rabiner, Michael J. Cheng, Aaron. Rosenberg, Carol A. McGonegal: "A Comparative Performance Study of Several Pitch Detection Algorithms," *IEEE, Vol. ASSP-24*, 1976.
DOI: <http://dx.doi.org/10.1109/TASSP.1976.1162846>
- [8] Chong Kwan Un, Shin-Chien Yang: "A Pitch Extraction Algorithm Based on LPC Inverse Filtering and AMDF," *IEEE, Vol. ASSP-39*, 1991.
DOI: <http://dx.doi.org/10.1109/TASSP.1977.1163005>
- [9] K.Krishna, V.L.N.Murty, K.R.Ramakrishnan: "Vector quantization of excitation gains in speech coding", *Signal Processing* 81, pp. 203-209, 2001.
DOI: [http://dx.doi.org/10.1016/S0165-1684\(00\)00200-0](http://dx.doi.org/10.1016/S0165-1684(00)00200-0)

이 시 우(See-Woo Lee)

[정회원]



- 1990년 3월 : 日本大學(Nihon Univ) 전자공학과 (공학석사)
- 1994년 3월 : 日本大學(Nihon Univ) 전자공학과 (공학박사)
- 1994년 3월 ~ 1998년 2월 : (주) 삼성전자 통신연구소/멀티미디어 연구소
- 1998년 3월 ~ 현재 : 상명대학교 정보통신공학과 교수

<관심분야>

유무선통신, 음성신호처리