

8kbps 비트율을 갖는 ACFBD-MPC와 LMS-MPC를 통합한 ACLMS-MPC 부호화 방식[☆]

An ACLMS-MPC Coding Method Integrated with ACFBD-MPC and LMS-MPC at 8kbps bit rate.

이 시 우^{1*}

See-woo Lee

요 약

본 논문에서는 합성 음성파형의 일그러짐을 제어하기 위하여 V/UV/S(Voiced / Unvoiced / Silence)의 스위칭을 사용하고, 피치구간마다 멀티펄스를 보정하며, 무성자음(Unvoiced)의 근사합성에 특정주파수를 이용하는 ACFBD-MPC(Amplitude Compensation Frequency Band Division - Multi Pulse Coding)와 LMS-MPC(Least Mean Square - Multi Pulse Coding)를 통합한 8kbps ACLMS-MPC(Amplitude Compensation and Least Mean Square - Multi Pulse Coding) 부호화 방식을 제안하고자 한다.

여러 방식을 통합하는데 있어서, 음성파형의 일그러짐을 줄이면서 유성음과 무성음의 비트율을 8kbps로 조정하는 것이 중요하다. 유성음과 무성음의 비트율을 8kbps로 조정함에 있어서, 개별피치를 이용하여 대표구간의 멀티펄스를 피치구간마다 복원함으로써 음성파형을 효율적으로 합성할 수 있다. 8kbps의 부호화 조건에서 ACLMS-MPC 방식을 구현하고 SNR를 평가한 결과, ACLMS-MPC의 SNR는 남자음성에서 15.0dB, 여자음성에서 14.3dB 임을 확인할 수 있었다. 따라서 ACLMS-MPC가 기존의 MPC, ACFBD-MPC, LMS-MPC에 비하여 남자음성에서 0.3dB~1.8dB, 여자음성에서 0.3dB~1.6dB 정도 개선된 것을 알 수 있었다.

이러한 방법들은 셀룰러폰이나 인터넷폰과 같이 낮은 비트율의 음원을 사용하여 음성신호를 부호화하는 방식에 활용할 수 있을 것으로 기대된다. 향후 멀티펄스 음원의 진폭과 위치를 동시에 보정하는 6.9kbps 음성부호화 방식의 음질평가를 수행하고자 한다.

☞ 주제어 : 셀룰러폰, 스마트폰, 인터넷폰

ABSTRACT

This paper present an 8kbps ACLMS-MPC(Amplitude Compensation and Least Mean Square - Multi Pulse Coding) coding method integrated with ACFBD-MPC(Amplitude Compensation Frequency Band Division - Multi Pulse Coding) and LMS-MPC(Least Mean Square - Multi Pulse Coding) used V/UV/S(Voiced / Unvoiced / Silence) switching, compensation in a multi-pulses each pitch interval and Unvoiced approximate-synthesis by using specific frequency in order to reduce distortion of synthesis waveform.

In integrating several methods, it is important to adjust the bit rate of voiced and unvoiced sound source to 8kbps while reducing the distortion of the speech waveform. In adjusting the bit rate of voiced and unvoiced sound source to 8 kbps, the speech waveform can be synthesized efficiently by restoring the individual pitch intervals using multi pulse in the representative interval. I was implemented that the ACLMS-MPC method and evaluate the SNR of APC-LMS in coding condition in 8kbps. As a result, SNR of ACLMS-MPC was 15.0dB for female voice and 14.3dB for male voice respectively. Therefore, I found that ACLMS-MPC was improved by 0.3dB~1.8dB for male voice and 0.3dB~1.6dB for female voice compared to existing MPC, ACFBD-MPC and LMS-MPC.

These methods are expected to be applied to a method of speech coding using sound source in a low bit rate such as a cellular phone or internet phone. In the future, I will study the evaluation of the sound quality of 6.9kbps speech coding method that simultaneously compensation the amplitude and position of multi-pulse source.

☞ keyword : Cellular Phone, Smart Phone, Internet Phone

1. 서 론

¹ Smart Information and Telecommunication Engineering, SANGMYUNG UNIV, Cheonan, 330, Korea.

* Corresponding author (swlee@snu.ac.kr)

[Received 2 May 2018, Reviewed 21 May 2018(R2 5 September 2018), Accepted 18 September 2018]

☆ 본 논문은 상명대학교 연구과제로 수행되었음.

최근 음성인식과 음성통신을 결합한 형태의 지능형 IoT 기기 시장이 블루오션으로 각광받고 있다. 음성인식을 위해서는 음성통신 방식을 이용하여 음성신호를 전달하는 과정이 필요하고, 전달된 음성신호에 오류가 없어야

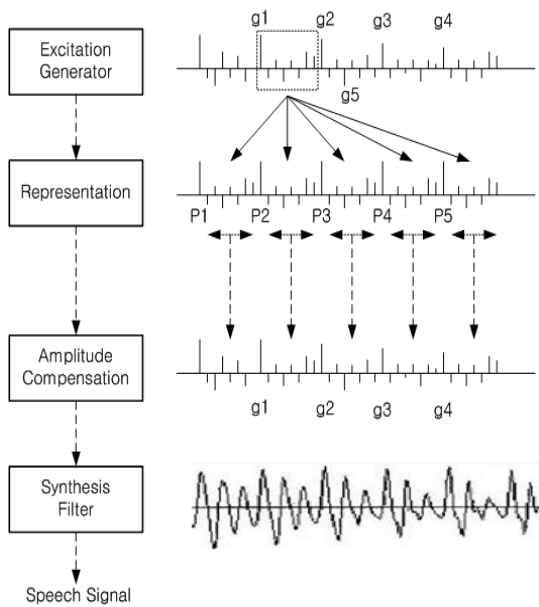
식과 같은 특성을 갖는다.

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

아울러, 멀티펄스의 진폭과 위치는 다음 식을 만족한다.

$$\sum_{j=1}^K g_j \cdot R(|m_i - m_j|) = \Phi(m_i), (1 \leq i \leq K) \quad (3)$$

여기에서 R , Φ 는 자기상관함수와 상호상관함수를 나타낸다.



(그림 2) 멀티펄스의 음원
(Figure 2) Exciting Source of Multi-Pulse

이렇게 해서 만들어진 멀티펄스 유성음원을 그대로 수신측에 전송하면 전송량이 크게 늘어나게 된다. 그러면 8kbps의 전송량을 맞추기 어렵게 되므로, 유성음의 주기성을 이용하여 추출한 개별피치의 피치구간마다 대표구간의 멀티펄스를 재구성하여 사용하면 유성음원의 정보 전송량을 줄일 수 있다[9]. 그러나 그림 2와 같이 대표구간의 멀티펄스를 이용하여 피치구간마다 멀티펄스 음원을 생성하는 경우, 정보 전송량을 줄일 수 있는 반면 재구성한 멀티펄스 음원의 정규화에 따른 재생 음질이

저하되는 원인으로 작용한다. 이러한 문제를 해결하기 위하여 ACFBD-MPC에서는 그림 2와 같이 재생된 음성 신호 $\bar{x}(n)$ 가 최소가 되도록 진폭(g_i)을 원래의 멀티펄스 음원에 가까운 근사 값을 갖도록 보장한다.

3. 천이구간의 재생

음성신호는 일반적으로 음성신호가 존재하지 않는 무음 구간, 모음과 같은 주기성을 갖는 유성음 구간, 자음과 같이 비주기성을 갖는 무성음 구간 그리고 유성음과 무성음의 중간특성을 갖는 천이구간(TSIUVC)으로 구성되어 있다. 이러한 TSIUVC를 연속적인 음성신호에서 자동으로 추출하는 동시에 효과적으로 재생하는 방식으로 TSIUVC 재생에 유용한 주파수정보를 이용하는 방식[7]과 TSIUVC 재생에 유용한 주파수 대역의 근사값을 구하는 LMS 알고리즘을 적용한 방식이 있다[8]. 전자는 개별피치정보와 ZCR(Zero Crossing Rate)를 이용하여 TSIUVC 구간을 추출한 후 FFT하여 얻은 주파수 대역을 여러 주파수 영역으로 분할하고, 분할한 각 영역의 주파수 정보를 이용하여 TSIUVC를 재생한다. 재생한 TSIUVC의 SNR을 측정하여 TSIUVC 재생에 유효한 주파수 정보가 0.547kHz 이하의 낮은 주파수 영역(LF: Low Frequency)과 2.813kHz 이상의 높은 주파수 영역(HF: High Frequency)에 존재함을 입증한 바 있다[7]. 한편 LMS 알고리즘은 TSIUVC의 주파수를 LF와 HF의 여러 블록으로 나누어 근사값을 구한다음 TSIUVC를 재생하는 방법이다. 이러한 방법은 유성음과 무성음의 중간 특성을 갖고 있는 음성신호를 유성음원 혹은 무성음원을 사용하지 않고 재생할 수 있는 방법이라 할 수 있다.

4. 부호화 조건과 시스템 구성

ACFBD-MPC와 LMS-MPC를 통합한 새로운 방식인 ACLMS-MPC는 음성신호를 8kbps로 부호화하였다. 신호 처리 코딩은 C/C++ 언어를 사용하였으며, 음성신호는 10kHz, 12bit로 표본화 및 양자화 하였다. 그리고 실제 음성통신 시스템에서 적용되고 있는 3.4kHz LPF를 사용하였으며, 프레임의 길이는 FFT 처리를 고려하여 25.6ms로 하였다.

기존의 MPC, ACFBD-MPC, LMS-MPC와 ACLMS-MPC의 음질평가를 위해서 전송률이 8kbps가 되도록 표 1과 같이 부호화 조건을 설정하였다. 표 1의 부호화 조건을 살

퍼보면, 유성음/무음/TSIUVC (V/S/TSIUVC)의 선택정보에 2bit를 할당하였고, 유성음(V)은 피치정보와 멀티펄스 음원을 사용하고, 음성합성 필터의 차수는 10차를 사용하였다. 따라서 음성합성 필터의 계수로 10개가 생성되며, 이 10개의 계수는 음성신호의 스펙트럼에 미치는 영향은 낮은 차수일수록 음성신호의 스펙트럼에 미치는 영향이 크기 때문에 낮은 차수의 계수일수록 많은 bit를 할당하였다.

유성음(V)의 경우, 원 음성신호에서 산출한 멀티펄스의 진폭(g_k)과 위치(m_k), 멀티펄스의 최대 진폭 값(g_{max})에 할당하는 bit는 원 음성파형에 가깝게 재생하는데 필요한 정보이다. 본 연구에서 음원으로 사용한 멀티펄스의 수는 10개이며, 멀티펄스의 최대 진폭 값(g_{max})에 할당된 bit는 MPC와 ACFBD-MPC 경우 10bit, LMS-MPC와 ACLMS-MPC 경우 6bit를 할당하였다[7][8]. 멀티펄스의 진폭(g_k)과 위치(m_k)는 표 1과 같이 할당하였으며, 재생 파형의 진폭을 보정하기 위하여 프레임의 진폭레벨의 증

감패턴(γ)을 고려하여 진폭보정 계수(β_k)를 산출하고, 이를 재생 파형의 진폭보정에 적용하여 기존의 MPC에 비하여 SNR이 떨어지지 않는 범위에서 보다 적은 bit를 할당하였다[7].

TSIUVC의 경우, 천이구간을 추출하기 위한 최초 개별 피치의 위치(P_0), 개별피치 간격의 평균(I_{AV}), 개별피치 간격이 편차(DP_i)는 25.6ms 내에 존재할 수 있는 피치수를 고려하여 bit를 할당하였다. 여기에서 MPC는 자기상관법에 의해 구해진 평균피치 정보(Pitch)를 적용하기 때문에 TSIUVC 구간 추출은 불가능하므로 25.6ms의 프레임에 존재할 수 있는 피치를 고려하여 8bit를 할당하였다. ACFBD-MPC, LMS-MPC와 ACLMS-MPC에서는 TSIUVC 구간을 추출하고 FFT하여 얻은 주파수 정보의 SNR을 고려하여 LMS 보정계수(k), 낮은 주파수 대역의 신호(LF)와 높은 주파수 대역의 신호(HF)의 주파수를 사용하여 음성신호를 재생하도록 하였다. 이때 LMS-MPC의 보정계수($k \leq 5$)는 5개를 사용하였으며, 낮은 주파수 대역의 신호(LF)에 6bit, 높은 주파수 대역의 신호(HF)에 5bit를

(표 1) 부호화 조건

(Table 1) Coding Condition

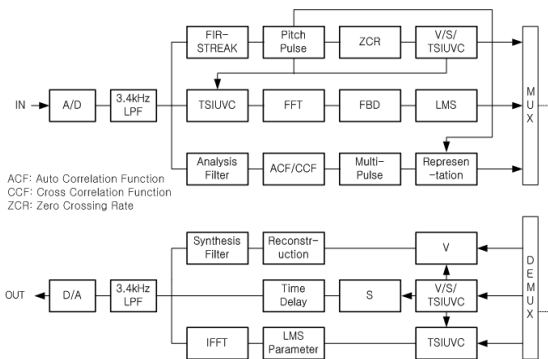
Parameter(bit)	MPC	ACFBD-MPC	LMS-MPC	ACLMS-MPC
V/S/TSIUVC (Voiced)	2	2	2	2
Coefficient of Synthesis Filter	7,6,5,5,4 3,3,3,3,3(42)	7,6,5,5,4 3,3,3,3,3(42)	7,6,5,5,4 3,3,3,3,3(42)	7,6,5,5,4 3,3,3,3,3(42)
g_{max}	10	10	6	6
g_k	8	6	4	4
m_k	8	6	5	5
β_k		3(3*10=30)		3(3*10=30)
γ		2		2
Number of Multi-Pulse	10(160)	10(120)	10(90)	10(90)
Pitch	8	8		
P_0			7	7
I_{AV}			7	7
$DP_i (i = 2 \sim 9)$ (TSIUVC)			3(3*8=24)	3(3*8=24)
$k \leq 5$			6(6*5=30)	6(6*5=30)
Max Amplitude		7		7
Frequency Signal of Low Frequency Band(LF)		6(6*12=72)		6(6*12=72)
Frequency Signal of High Frequency Band(HF)		5(5*15=75)		5(5*15=75)
Total Bit	222	214	208	208
kbps	8.67	8.36	8.13	8.13

할당하였다[8].

5. 실험결과

본 연구에서는 기존의 MPC, ACFBD-MPC와 LMS-MPC를 통합한 ACLMS-MPC의 재생 음성파형의 객관적 평가 척도인 SNR을 비교하고자 한다.

MPC, ACFBD-MPC, LMS-MPC와 ACLMS-MPC는 10kHz 12bit A/D, D/A변환기와 3.4kHz LPF(FIR Digital Filter)를 입출력단에 구비하고, 연속 음성신호에서 TSIUVC 구간을 추출한 다음 유성음/무음(TSIUVC(V/S/TSIUVC) 선택정보에 의하여 각기 다른 음성신호 재생방식을 선택하도록 하였다. 이러한 특징은 다른 음성부호화 방식에서는 찾기 어려운 독특한 방법으로 ACFBD-MPC와 LMS-MPC를 통합한 ACLMS-MPC 시스템을 그림 3에 나타내었다. ACLMS-MPC 시스템에 있어서, 유성음(V)의 경우, 멀티펄스 음원을 사용하여 음성신호를 재생하는데, MPC와 LMS-MPC는 멀티펄스의 진폭보정이 없는 음원을 사용하고 ACFBD-MPC는 멀티펄스의 진폭을 보정하는 보정계수에 의하여 멀티펄스 음원의 진폭을 보정 하고, 진폭이 보정된 멀티펄스 음원으로 합성필터를 구동하여 음성신호를 재생한다. 무음(S)의 경우, 음성신호가 없는 프레임이므로 특별한 신호처리를 하지 않고 25.6ms의 시간 지연 처리를 한다.



(그림 3) ACLMS-MPC 시스템
(Figure. 3) System of ACLMS-MPC

TSIUVC의 경우, TSIUVC 구간의 음성신호를 FFT하여 얻은 스펙트럼 정보에서 TSIUVC 재생에 유효한 주파수 정보를 LMS 근사 값으로 추정하여 TSIUVC를 재생하도록 하였다. TSIUVC를 재생하기 위해서는 우선 FIR-STREAK

필터로부터 추출한 개별 피치와 ZCR(Zero Crossing Rate)를 적용하여 V/S/TSIUVC 선택정보를 얻고, TSIUVC 재생에 FBD(Frequency Band Division)와 LMS(Least Mean Square)를 적용하여 TSIUVC를 재생하는 것이다. 즉, ACLMS-MPC에서는 TSIUVC를 FFT처리한 후 음성합성에 유효한 주파수 대역의 신호에 LMS를 적용하여 근사 합성 처리함으로써 보다 음질을 개선하고자 한다. MPC, ACFBD-MPC, LMS-MPC와 ACLMS-MPC의 SNR 평가에 사용한 음성샘플을 표 2에 나타냈다. 음성신호의 총 길이는 54.4초이며 남녀 4명이 발성한 16문장(유성음145개, 무성자음34개)이다.

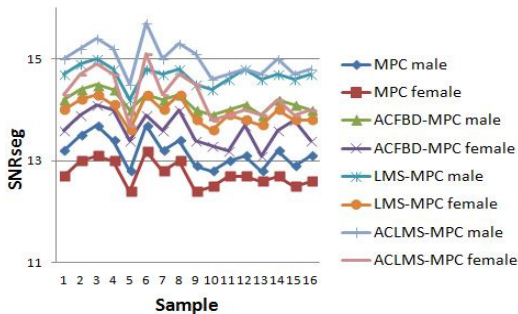
(표 2) 음성샘플
(Table 2) Speech Sample

	male	female
participants	4	4
Time	54.4 sec	54.4 sec
sentences	16	16
Vowel	145	145
Unvoiced Cononant	34	34

MPC, ACFBD-MPC, LMS-MPC와 ACLMS-MPC를 C/C++로 코딩한 후 음성신호를 재생하여 SNR를 측정하는 실험을 하였다. 실험한 결과, 표 3과 그림 4에 나타낸 바와 같이, MPC는 남자음성에서 13.2dB, 여자음성에서 12.7dB를 얻었으며, ACFBD-MPC와 LMS-MPC는 각각 남자음성에서 14.2dB와 14.7dB, 여자음성에서 13.6dB와 14.0dB의 결과를 얻었다. 또한 ACLMS-MPC는 남자와 여자음성에서 각각 15.0dB와 14.3dB를 얻을 수 있었다. 따라서 ACFBD-MPC와 LMS-MPC는 기존의 MPC에 비하여 남자음성에서 1dB와 1.5dB, 여자음성에서 0.9dB와 1.3dB 정도 개선된 것을 알 수 있었다. 한편, ACLMS-MPC는 MPC, ACFBD-MPC, LMS-MPC에 비하여 남자음성에서 각각 1.8dB, 0.8dB, 0.3dB가 개선되었고, 여자음성에서 각각 1.6dB, 0.7dB, 0.3dB가 개선된 것을 확인할 수 있었다. 이러한 SNR 개선의 정도는 음성신호의 물리적 속성인 음성파형의 진폭과 위상을 보상함으로써 음성신호의 심리적 속성인 음량과 음색이 개선된 것으로 볼 수 있다. 개선된 정도가 기존의 MPC에 비하여 ACFBD-MPC의 경우, 남자와 여자음성에서 각각 1.0dB, 0.9dB가 개선되었고, LMS-MPC의 경우는 남자와 여자음성에서 각각 1.5dB, 1.3dB의 개선된 것을 입증한바 있다[7,8]. 이번 실험에서는 ACFBD-MPC와 LMS-MPC를 통합한 방식인

(표 3) MPC, ACFBD-MPC, LMS-MPC, ACLMS-MPC의 SNR_{seg} (Table 3) SNR_{seg} of MPC, ACFBD-MPC, LMS-MPC and ACLMS-MPC

Method [dB]	kbps	male	female
MPC	8	13.2	12.7
ACFBD-MPC	8	14.2	13.6
LMS-MPC	8	14.7	14.0
ACLMS-MPC	8	15.0	14.3



(그림 4) MPC, ACFBD-MPC, LMS-MPC, ACLMS-MPC의 SNR

(Figure 4) SNR of MPC, ACFBD-MPC, LMS-MPC and ACLMS-MPC

ACLMS-MPC의 음질 개선 정도를 확인하는 것이 목표였으며, 기존의 MPC에 비하여 남자음성에서 1.8dB, 여자음성에서 1.6dB 정도 SNR이 개선된 것을 확인할 수 있었다.

6. 결 론

본 연구에서는 셀룰러 폰, 인터넷 통신과 무선통신에서 활용할 수 있는 8kbps ACLMS-MPC를 제안하였으며, 기존의 MPC에 비하여 남자음성에서 1.8dB, 여자음성에서 1.6dB 정도 개선된 것을 확인하였다. 이러한 방식은 음질개선을 필요로 하는 음성통신 환경에 적용할 수 있을 것으로 기대한다. 다만, 신호처리 방법이 다소 복잡하여 연산처리 속도의 제약으로 신호처리 지연시간이 발생할 수 있는 단점이 있다. 이러한 단점을 보완한다면 실질적인 음성통신 시스템에 적용하는 것은 무리가 없어 보이며 향후 심도 있는 연구과정이 필요하다. 이번 연구에서는 8kbps의 전송률에서 음질개선 정도를 확인하였으나, 향후에는 8kbps 보다 낮은 전송률에서 다양한 음질개선방법을 적용하여 양호한 음성신호를 재생할 수 있는 음성통신시스템을 연구하고자 한다.

참고문헌(Reference)

- [1] Selma Ozaydm, Buyurman Baykal: "Matrix quantization and mixed excitation based linear predictive speech coding at very low bit rates", Speech Communication 41, p381-392, 2003, [https://doi.org/10.1016/S0167-6393\(03\)00009-8](https://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, p203-209, 2001, [https://doi.org/10.1016/S0165-1684\(00\)00200-0](https://doi.org/10.1016/S0165-1684(00)00200-0)
- [3] Ghaemmaghami, S., Sridharan, S.: "Very low rate speech coding using temporal decomposition". IEE Electron. Lett. 35(6), p456-457, 1999, <https://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. p373-385, 1993, <https://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, p242-250, 1995, <https://doi.org/10.1109/89.397089>
- [6] Feng Huang, Tan Lee, W.Bastiaan Kleijn, Ying-Yee Kong: "A method of speech periodicity enhancement using transform-domain signal decomposition", Speech Communication, p102-112, 2015, <https://doi.org/10.1016/j.specom.2014.12.001>
- [7] SeeWoo Lee: "A Study on ACFBD-MPC in 8kbps", Journal of the Korea Academia-Industrial cooperation Society, Vol. 17, No.7, p49-53, 2016, <http://dx.doi.org/10.5762/KAIS.2016.17.7.49>
- [8] SeeWoo Lee: "A Study on LMS-MPC Method Considering Low Bit Rate", The Journal of Policy & Management, Vol. 10, No. 5, p233-238, 2012, <http://dx.doi.org/10.14400/JDPM.2012.10.5.233>
- [9] SeeWoo Lee: "A Study on 8kbps IP-MPC Method Considering Cellular Phone", The Journal of Korea Knowledge Information Technology Society, Vol.5, No. 5, p149-154, 2010, http://www.kkits.or.kr/bbs/board.php?bo_table=doc2&wr_id=109

● 저 자 소 개 ●



이 시 우(See-woo Lee)

1990년 日本大學(Nihon Univ) 전자공학과 (공학석사)

1994년 日本大學(Nihon Univ) 전자공학과 (공학박사)

1994년 3월~1998년 2월 (주)삼성전자 통신연구소/멀티미디어 연구소

1998년 3월~현재 상명대학교 스마트정보통신공학과 교수

관심분야 : 신호처리, 전기 및 전자부품, 빅데이터 etc.

E-mail : swlee@smu.ac.kr