

멀티펄스의 위치보정 방법을 이용한 8kbps PC-MPC에 관한 연구

이시우*
상명대학교 정보통신공학과

A Study on 8kbps PC-MPC by Using Position Compensation Method of Multi-Pulse

See-Woo Lee*

Dept. of Information and Telecommunication Engineering, Sangmyung University

요 약 유성음원과 무성음원을 사용하는 멀티펄스 음성부호화 방식에 있어서, 대표구간의 멀티펄스를 사용하는 경우에 유성음의 합성음성파형에서 일그러짐이 나타난다. 이것은 대표구간의 멀티펄스를 피치구간마다 복원하는 과정에서 재생 음성신호가 정규화되는 것이 원인으로 작용한다. 이것을 해결하기 위하여 본 논문에서는 피치구간마다 멀티펄스의 위치를 보정하는 방법(PC-MPC)을 제시하였으며, 기존의 MPC와 멀티펄스 위치를 보정한 PC-MPC의 SNR_{seg} 를 평가한 결과, PC-MPC의 남자음성에서 0.4dB, 여자음성에서 0.5dB 개선된 것을 확인할 수 있었다. 결국, MPC에 비해 PC-MPC의 SNR_{seg} 가 개선되어 음성파형의 일그러짐을 제어할 수 있었으며, 본 방법은 셀룰러폰이나 스마트폰과 같이 Low Bit Rate의 음원을 사용하여 음성신호를 부호화하는 방식에 활용할 수 있을 것으로 기대된다.

주제어 : 셀룰러폰, 스마트폰, 인터넷폰, 음성코딩, 코딩시스템

Abstract In a MPC coding using excitation source of voiced and unvoiced, it would be a distortion of speech waveform. This is caused by normalization of synthesis speech waveform of voiced in the process of restoration the multi-pulses of representation section. To solve this problem, this paper present a method of position compensation(PC-MPC) in a multi-pulses each pitch interval in order to reduce distortion of speech waveform. I was confirmed that the method can be synthesized close to the original speech waveform. And I evaluate the MPC and PC-MPC using multi-pulses position compensation method. As a result, SNR_{seg} of PC-MPC was improved 0.4dB for female voice and 0.5dB for male voice respectively. Compared to the MPC, SNR_{seg} of PC-MPC has been improved that I was able to control the distortion of the speech waveform finally. And so, I expect to be able to this method for cellular phone and smart phone using excitation source of low bit rate.

Key Words : Cellular Phone, Smart Phone, Internet Phone, Speech Coding, Coding System

* 본 논문은 2013년 상명대학교 연구과제로 수행되었음.

Received 5 March 2013, Revised 27 March 2013

Accepted 20 May 2013

Corresponding Author: See-Woo Lee(Dept. of Information and Telecommunication Engineering, Sangmyung University

Email: swlee@smu.ac.kr

1. 서론

최근 스마트폰의 보급과 더불어 다양한 형태의 멀티미디어 데이터 송수신이 급증함에 따라 이동통신 네트워크에 과부하가 걸리는 현상이 나타나고 있다. 이와 더불어 인터넷폰을 이용하는 통신수단이 발달하면서 통신 네트워크의 과부하에 대한 문제점은 더욱 높아지고 있다. 이러한 문제점을 해결하기 위한 수단으로 통신 트래킹을 효율적으로 관리하는 방법이나 데이터 및 음성신호를 효율적으로 압축복원 기술이 주목받을 것으로 생각된다. 특히 음성신호의 압축복원 기술은 일반적으로 Low Bit Rate을 구현하기 위해 음성파형을 원음성에 가깝게 복원하는 신호처리 기술과 적절한 음원을 선정하는 방법을 사용한다 [1]~[3]. 이러한 방식들은 음성신호를 일정한 프레임으로 잘라낸 후 유성음과 무성음을 판독한 다음 유성음원과 무성음원을 사용하여 음성신호를 복원하는 기술을 사용한다. 이러한 유형의 일종인 멀티펄스 음성부호화 방식(MPC)은 Bell연구소의 Atal에 의하여 처음으로 제안되었으며[4], AbS(Analysis by Synthesis)법에 의하여 멀티펄스(Multi-Pulse)를 탐색하고, 멀티펄스에 의하여 합성 필터를 구동하여 음성신호를 재생하는 방식이다. 이 방식에 Ozawa는 자기상관법에 의하여 산출한 정규화된 피치정보를 적용하여 낮은 전송 Bit Rate을 유지할 수 있도록 개선하였다[5]. 또한 이 방식은 정규화된 피치정보를 이용하여 대표구간의 멀티펄스를 송신측에 전송하여 음원을 재구성함으로써 Low Bit Rate을 유지하고 있다. 그러나 대표구간의 멀티펄스를 이용하여 재구성하는 과정에서 송신하기전의 본래의 멀티펄스 진폭과 위치가 바뀔으로써 음성품질을 저하시키는 요인으로 작용한다.

일반적으로 인간이 발생하는 음성에너지는 수시로 시간에 따라 변하며 이러한 변화는 멀티펄스의 진폭과 위치변화로 나타난다. 따라서 생성된 본래의 멀티펄스를 사용하면 원 음성파형을 재현할 수 있다. 그러나 멀티펄스 전부를 사용하면 Low Bit Rate의 구현이 어려운 만큼 피치정보를 이용한 대표구간의 멀티펄스를 수신측에 전송하여 사용하게 된다. 이때 송신측에 전송한 대표구간의 멀티펄스를 피치구간 마다 대표구간의 멀티펄스를 사용하여 멀티펄스 음원을 재구성하게 된다. 이러한 과정에서 본래의 멀티펄스의 진폭과 위치는 대표구간의 멀티펄스 진폭과 위치로 정규화 된다. 이러한 정규화 과정으로 인하여 음질의 객관적 척도인 SNR_{seg} 가 저하되는 요

인으로 작용한다. 따라서 본 연구에서는 멀티펄스의 위치를 보정하여 SNR_{seg} 을 개선하는 PC-MPC (Position Compensation Multi Pulse Coding)을 제안한다.

2. 멀티펄스 탐색

연속 음성신호는 속성을 달리하는 구간별 식별이 가능하며, 음성신호는 폐의 압력으로 빠져나온 공기가 성대와 구강을 빠져나오면서 공기의 진동과 흐름에 따라서 유성음(V: Voiced)과 무성음(UV: Unvoiced)으로 나누어진다. 유성음은 비교적 낮은 주파수대에 존재하는 주기적인 신호인 반면, 무성음은 높은 주파수대에 존재하는 비주기적인 신호의 특징을 갖는다. 이러한 신호의 특징을 이용하여 V/UV 선택신호에 의하여 V와 UV에 적합한 음원을 선택적으로 적용하는 Low Bit Rate의 음성부호화 방식들이 제안되었다[6]~[9].

V/UV 음원을 사용하는 멀티펄스 음성부호화 방식에 있어서, 멀티펄스의 진폭과 위치를 결정하기 위해서는 자기상관함수와 상호상관함수를 사용한다. 원 음성신호에서 구한 자기상관함수와 상호상관함수를 사용하여 재생한 음성신호와의 차가 최소가 되도록 한다. 즉, 원 음성신호 $x(n)$ 와 멀티펄스 $v(n)$ 에 의하여 재생된 음성신호 $\bar{x}(n)$ 로부터 식(1)이 최소가 되도록 멀티펄스의 진폭(g_i)과 위치(m_i)를 결정하게 된다.

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

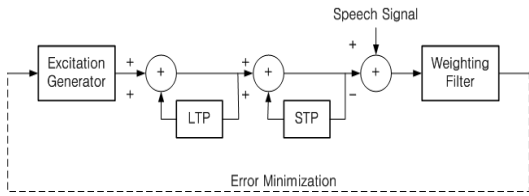
$$\bar{x}(n) = \sum_{i=1}^K g_i \cdot h(n - m_i)$$

$$h(n) = h_p(n) * h_s(n)$$

여기에서 K , g_i , m_i 는 멀티펄스의 수, i 번째의 펄스진폭과 위치를 나타낸다. $h_p(n)$, $h_s(n)$, $h(n)$ 는 각각 피치재생필터(LTP), 음성스펙트럼합성필터(STP), LTP와 STP를 접목한 합성 필터의 임펄스 응답을 나타낸다. 여기에서, $w(n)$ 은 Weighting 필터로서 다음 식과 같은 특성을 갖는다.

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

a_i : 예측계수



[Fig. 1] Multi-Pulse Coding(MPC)

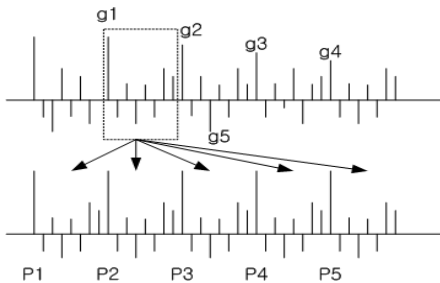
LTP(Long Term Prediction)와 STP(Short Term Prediction)로 구성된 멀티펄스 탐색의 블록도(MPC)를 [그림 1]에 나타내었으며, 이 시스템의 전달함수 $H(z)$ 는 다음 식으로 나타낼 수 있다.

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

여기에서 1항과 2항은 각각 LTP, STP의 전달함수이며, b_i , T 는 자기상관함수와 상호상관함수로부터 구하여진 피치계수, 피치주기를 나타낸다. 또한 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 식에서 R 는 자기상관함수, Φ 는 상호상관함수를 나타낸다. 자기상관함수는 멀티펄스의 탐색은 물론 STP를 구동하여 음성신호를 재생하는 과정에서도 사용되는 파라미터이다.



[Fig. 2] Exciting Source of Multi-Pulse

피치는 음성신호의 기본주파수를 나타내는 것으로 멀티펄스 음원의 대표구간을 [그림 2]와 같이 피치주기마다 복원함으로써 멀티펄스 음원을 생성한다. 음성신호의

기본 주파수인 피치 주파수는 약 80~370Hz이기 때문에 이를 시간 간격으로 나타내면 약 2.7ms~12.5ms이고, 프레임 길이를 25.6ms로 하였을 경우 2개~9개의 피치가 존재하게 된다. 대표구간의 멀티펄스와 피치정보를 수신측에 전송하고, 수신측에서는 피치구간마다 대표구간의 멀티펄스를 재생하여 멀티펄스 음원을 재구성한다.

3. 멀티펄스 보정

본 연구에서 제한한 PC-MPC는 유성음(V)의 음원인 멀티펄스 중에서 피치위치에 있는 멀티펄스의 위치를 보정함으로써 SNR_{seg} 을 개선하고자 하였다. 우선, V/UV 음원을 살펴볼 필요가 있는데, 유성음(V)은 폐에서 나오는 공기압의 변화 즉 음성에너지가 시간에 따라서 주기적으로 변화하며, 무성음(UV)은 음성에너지가 불규칙하여 비주기적으로 변화하는 특성을 갖는다. 음성신호는 V와 UV의 결합에 의하여 발생되는 신호로서 다음과 같은 식으로 나타낼 수 있다.

$$f_{UC}(t) = [S_{UC}(t) \cdot N_{UC}(t)] * h_{UC}(t) \quad (5)$$

여기에서 $f_{UC}(t)$ 는 방사출력, $N_{UC}(t)$ 는 백색잡음, $S_{UC}(t)$ 는 $N_{UC}(t)$ 에 대한 무성자음의 시간길이에 상응하는 시간간격, $h_{UC}(t)$ 는 성대의 임펄스응답을 나타낸다. 여기에서 랜덤 한 특성을 갖는 백색잡음의 경우 위치를 제어하거나 보정하는 것이 어려운 반면, 유성음은 한정된 길이의 프레임에 주기적인 파형이 나타나기 때문에 다음식과 같이 n 개의 파형으로 용이하게 나타낼 수 있다.

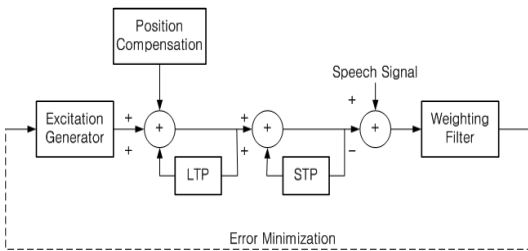
$$f_V(t) = \sum_{k=1}^n A_V(k) e^{(-\pi B_V(k)t)} \cdot \cos(2\pi F_V(k)t + \phi_V(k)) \quad (6)$$

여기에서 $F_V(k)$ 는 공진주파수, $B_V(k)$ 는 대역폭, $A_V(k)$ 와 $\phi_V(k)$ 는 파형의 진폭과 위상을 나타낸다. 윗 식에서 $A_V(k)$ 는 피치구간내의 멀티펄스의 진폭에 상응하는 진폭 값을 사용하고, $\phi_V(k)$ 는 피치위치에 있는 멀티펄스의 위치를 보정하여 식(7)의 멀티펄스 음원을 사용한다.

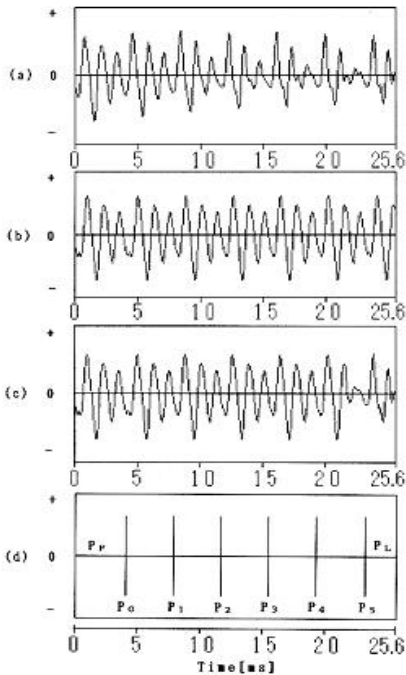
$$V_C(n) = \sum_{k=1}^N g_k \cdot \delta(n - m_k) \quad (7)$$

이 멀티펄스 음원을 피치간격마다 재구성하여 음성합성필터(STP)를 구동하여 음성신호를 재생하게 된다.

정상적인 유성음을 효율적으로 재생하기 위하여 수십 ms의 프레임으로 음성분석을 실행한다. 이 경우 프레임 끝단의 음성파형이 완전한 Sin파형이나 Cos파형이 되지 않는다. 이런 음성파형을 프리에 변환하면, 스펙트럼에 잡음이 급증하게 된다. 이러한 대비책으로 주파수분석을 실행할 때 프레임 경계부의 불연속성 영향을 제어하기 위하여 hamming 함수 등을 사용한다.



[Fig. 3] Position Compensation Multi-Pulse Coding(PC-MPC)



[Fig. 4] (a) Original Speech Waveform, (b) Not apply PC, (c) Apply PC (d) Pitch Pulse

그러나 음성신호를 시간영역에서 분석하는 멀티펄스

부호화 방식에 hamming 함수를 직접 적용하거나 멀티펄스를 탐색하는 것은 어려운 일이다. 또한 프레임 길이가 피치펄스 간격의 정수배와 일치하지 않는 경우에 프레임 끝단부에서 파형 일그러짐이 종종 나타난다.

이러한 프레임 끝단부의 파형 일그러짐은 P_F 구간(프레임 시작부에서 최초 피치가 나타나는 구간) 혹은 P_L 구간(프레임 끝단에서 최종 피치펄스 위치까지의 구간)의 어느쪽 구간에서 나타나는 파형 일그러짐을 제어할 수 있다면, 음질을 향상시킬 수 있다. 그래서 본 연구에서는 멀티펄스 위치보정을 실행하는 구간을 P_L 구간으로 한정하고, 스펙트럼 일그러짐이나 파형 일그러짐에 강한 영향을 미치는 진폭이 큰 멀티펄스의 위치를 보정대상으로 한다. 아울러 보정하는 멀티펄스의 수만큼 통신시스템의 전송비트가 높아지기 때문에 보정하는 멀티펄스의 수는 가능한 적정한 수가 좋다.

멀티펄스의 위치를 보정하는 순서는 우선 P_L 구간에서 진폭이 큰 순서로 멀티펄스를 검출하고, 검출한 멀티펄스의 위치 및 진폭정보를 취득한다. 다음으로 대표구간의 멀티펄스를 피치간격 마다 되풀이 하여 멀티펄스를 재생한다. 단, P_L 구간은 분석측에서 검출한 멀티펄스의 위치 및 진폭 정보를 사용하여 멀티펄스열을 생성하여 음성합성 필터를 구동함으로써 음성신호를 재생한다. 멀티펄스의 위치보정을 적용한 경우와 그렇지 않은 경우의 음성파형을 [그림 4]에 제시하였다. 여기에서 (a)는 원 음성파형이며, (b)는 멀티펄스의 위치를 보정하지 않은 경우, (c)는 멀티펄스의 위치를 보정한 경우, (d)는 피치펄스를 나타낸 것이다. 멀티펄스의 위치를 보정하지 않은 경우(b)에 비하여 멀티펄스의 위치를 보정한 경우(c)가 원 음성파형에 보다 가깝게 재생되어 있음을 알 수 있으며, 원 음성파형에 가까게 재생신호를 얻음으로서 SNR_{seg} 의 개선 효과를 얻을 수 있다.

4. 시스템 구현

기존의 MPC와 PC-MPC의 성능을 비교평가하기 위해서 같은 bit rate이 되도록 부호화한 다음, 재생한 음성신호를 객관적 평가척도인 SNR_{seg} 를 통하여 평가하고자 하였다. MPC와 PC-MPC는 3.4kHz LPF로 대역제한하고 10kHz, 12bit로 표본화 및 양자화 하였으며 시플레이션에는 C 언어를 사용하였다. 또한 향후 FFT를 적용

할 것을 고려하여 프레임 길이는 25.6ms로 하였다. 좀더 구체적인 부호화 조건을 살펴보면 <표 1>에 나타낸바와 같이 V/UV의 선택정보에 각각 2bit를 할당하였으며, V의 경우에는 피치와 멀티펄스, 합성필터 계수를 사용하여 음성신호를 합성하고, UV의 경우에는 white noise를 사용하였다. 합성필터의 차수는 10차를 사용하였으며, 합성필터 계수의 변화가 스펙트럼의 변화에 미치는 영향은 낮은 차수의 계수일수록 영향이 크기 때문에 낮은 차수일수록 많은 bit를 할당하였다[9].

<Table 1> Coding Condition

Parameter[bit]	MPC	PC-MPC
V/UV(1)	2	2
Coefficient of Synthesis Filter	7,6,5,5,4 3,3,3,3,3	7,6,5,5,4 3,3,3,3,3
g_{max}	10	10
g_k	6	6
m_k	8	6
\hat{m}_k		7
Number of Multi-Pulse	10	10
Average Pitch	8	8
kbps	8	8

MPC, PC-MPC 모두 10개의 멀티펄스를 사용하였으며, 평균 피치정보에 8bit를, MPC의 경우, 멀티펄스 진폭 및 위치에 6bit와 8bit를, PC-MPC의 경우는 진폭 및 위치에 각각 6bit, 위치보정에 7bit를 할당하였으며 멀티펄스의 최대진폭에는 10bit를 할당하였다.

<Table 2> Sound Spec

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

이러한 부호화 조건하에서 남녀 4명이 발성한 <표 2>의 음성제원을 사용하여 MPC와 PC-MPC의 SNR_{seg} 를 측정하였다.

실험결과, <표 3>에 나타낸바와 같이 MPC의 경우 남자음성은 13.2dB, 여자음성은 12.7dB를 얻을 수 있었으며, PC-MPC의 경우 남자와 여자음성은 각각 13.6dB,

13.2dB를 얻었다. 결국, PC-MPC의 SNR_{seg} 이 MPC에 비하여 남자와 여자 음성에서 각각 0.4dB, 0.5dB정도 개선된 것을 알 수 있었다.

<Table 3> SNR_{seg} of MPC and PC-MPC

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

5. 결론

V/UV 스위칭 음원절환 음성부호화 방식인 MPC방식을 개선하고자 멀티펄스의 위치를 보정한 PC-MPC방식을 제안하고, 객관적 평가척도인 SNR_{seg} 를 사용하여 비교평가 하였다. 실험결과, 기존의 MPC에 비하여 PC-MPC에서 남자와 여자음성에서 각각 0.4dB, 0.5dB 정도 개선되었으며, 멀티펄스의 위치를 보정함으로써 원 음성에 비교적 가까운 파형을 재생할 수 있었다. 다만, 멀티펄스의 위치보정이 까다로운 반면 음질개선 정도는 미약한 수준으로 멀티펄스의 위치뿐만 아니라 진폭을 함께 보정한다면 보다 음질 개선에 도움이 될 것으로 기대된다. 향후, 본 연구에서 제안한 PC-MPC를 셀룰러폰이나 스마트폰에 적용할 수 있는 Bit Rate으로 낮추는 동시에 음질을 개선할 수 있는 다양한 연구를 진행하고자 한다.

ACKNOWLEDGMENTS

This study was sponsored by the 2013 SANGMYUNG University intramural research fund.

REFERENCES

- [1] Ghaemmaghami, S., Sridharan,S.: "Very low rate speech coding using temporal decomposition".IEE Electron. Lett.35(6), p456-457.1999
- [2] 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
- [3] 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
- [4] B.S. Atal and J.R. Remdo: "A New Medel of LPC Excitation for Producing Natural Sounding Speech at Low Bit Rates", IEEE, ICASSP, p614~617, 1982
- [5] Kazunori OZAWA, Takashi ARASEKI: "Multi-Pulse Excited Speech Coding Utilizing Pitch Information at Rates Between 9.6 and 4.8 kbit/s", IEICE, Vol. J72-D-2, No.8, 1989
- [6] Campbell, J.P, Tremain, T.E: "Voiced/unvoiced classification of speech with applications to the U.S. Government LPC-10e algorithm", Proc.IEEE Int.Conf. on Acoustics, Speech, Signal Processing, p473-476.1986
- [7] Nobuhiko KITAWAKI, Fumitada ITAKURA and Shuzo SAITO: "Optimum Coding of Transmission Parameters in PARCOR Speech Analysis Synthesis System", IEICE, Vol. J61-A No.2, 1978
- [8] K.Krishna, V.L. N.Murty, K.R.Ramakrishnan: "Vector quantization of excitation gains in speech coding", Signal Processing 81, p203-209, 2001
- [9] 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

이 시 우(Lee, See Woo)



- 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 : 음성신호처리, 음주판독시스템, 유무선통신
- E-Mail : swlee@smu.ac.kr