

# Low Bit Rate을 고려한 8kbps FBD-MPC 방식에 관한 연구

이시우  
상명대학교 정보통신공학과

## A Study on 8kbps FBD-MPC Method Considering Low Bit Rate

See-Woo Lee

Dept. of Information and Telecommunication Engineering,  
Sangmyung University

**요약** 유성음원과 무성음원을 사용하는 음성부호화 방식에 있어서, 같은 프레임 안에 모음과 무성자음이 있는 경우에 음질저하현상이 나타난다. 본 연구에서는 연속음성에서 무성자음을 포함한 천이구간을 탐색, 추출하고 주파수대역에서 근사합성하는 8kbps의 멀티펄스 음성부호화 방식(FBD-MPC)을 제안하였다. 기존의 8kbps MPC와 FBD-MPC의 SNRseg를 평가한 결과, FBD-MPC의 남자음성에서 0.5dB, 여자음성에서 0.2dB 개선된 것을 확인할 수 있었다. 결국, MPC에 비해 FBD-MPC의 SNRseg가 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 speech coding system using excitation source of voiced and unvoiced, it would be involved a distortion of speech quality in case coexist with a voiced and unvoiced consonants in a frame. In this paper, I propose a method of 8kbps Multi-Pulse Speech Coding(FBD-MPC: Frequency Band Division MPC) by using TSIUVC(Transition Segment Including Unvoiced Consonant) searching, extraction and approximation-synthesis method in a frequency domain. I evaluate the 8kbps MPC and FBD-MPC. As a result, SNRseg of FBD-MPC was improved 0.5dB for female voice and 0.2dB for male voice respectively. Compared to the MPC, SNRseg of FBD-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

### 1. 서론

스마트폰 등의 이동통신시스템에 적용되는 음성부호

화 방식은 Low Bit Rate을 실현하기 위하여 주로 음원을 구동하여 음성신호를 재생한다. 음원으로 음성합성 필터를 구동하여 음성신호를 재생하는 음성부호화 방식의 일

\* 본 논문은 2014년 상명대학교 연구과제로 수행되었음

Received 2 April 2014, Revised 14 May 2014

Accepted 20 June 2014

Corresponding Author: See-Woo Lee(The Society of Digital Policy)

Email: swlee@smu.ac.kr

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

중인 멀티펄스 음성부호화 방식은[1][2], 피치예측법과 피치보간법을 이용하여 4.8~9.6kbit/s의 멀티펄스 음성부호화 방식(MPC)을 실현하였다[3]. 이 방식은 자기상관법으로 추출한 피치정보에 의하여 V/UV를 선택하고, V/UV 선택정보로 유성음원과 무성음원을 선택하여 음성신호를 재생한다. 그러나 주기적인 특성의 유성음과 비주기적인 무성음이 혼재하는 프레임을 유성음원과 무성음원 어느 한쪽의 음원으로 음성신호를 재생하는 모순이 발견된다. 또한 유성음과 무성음의 중간특성을 갖는 천이구간을 유성음원과 무성음원의 어느 한쪽의 음원을 사용하여 재생함으로써 음질저하의 원인으로 작용하게 된다. 이러한 문제점을 해결하기 위하여 본 논문에서는 전송률 8kbps에서 V/S/TSIUVC의 선택정보에 의하여 유성음(V), 무음(S), TSIUVC(Transition Segment Including UnVoiced Consonant)를 선별하여 각각의 신호특성에 맞는 처리방법으로 음성신호를 재생하는 FBD-MPC 음성부호화 방식(FBD-MPC: Frequency Band Division Multi Pulse Coding)과 기존의 MPC방식과 음질을 비교하였다.

## 2. 멀티펄스 탐색

유성음(V: Voiced)과 무성음(UV: Unvoiced) 신호의 특징을 살펴보면, 유성음은 비교적 낮은 주파수대에 존재하는 주기적인 신호인 반면, 무성음은 높은 주파수대에 존재하는 비주기적인 신호의 특징을 갖는다. 이러한 특징을 이용하여 피치정보로부터 V/UV 선택신호를 생성하고, V와 UV의 음원을 선택적으로 적용한다[4][7]. V/UV 음원을 사용하는 멀티펄스 부호화 방식에 있어서, 멀티펄스의 진폭과 위치는 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 결정한다.

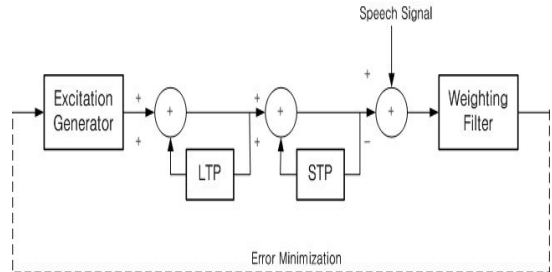
$$E = \sum_{n=0}^{N-1} [x(n) - \bar{x}(n) * w(n)]^2 \quad (1)$$

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

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

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

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



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

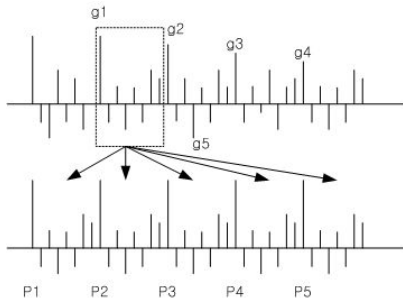
LTP(Long Term Prediction)와 STP(Short Term Prediction)로 구성된 멀티펄스 부호화 방식(MPC)을 [Fig. 1]에 나타내었으며, 이 시스템의 전달함수 $H(z)$ 는 다음 식으로 나타낼 수 있다.

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

여기에서 앞의 분수함수와 다음 이와 곱해지는 함수는 각각 LTP, STP의 전달함수를 나타낸다. 여기에서  $b_i, T, a_i$ 는 자기상관함수와 상호상관함수로부터 구하여진 피치계수, 피치주기, 예측계수를 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  $R$ 은 자기상관함수,  $\Phi$ 는 상호상관함수를 나타낸다.



[Fig. 2] Sound Source of Multi-Pulse

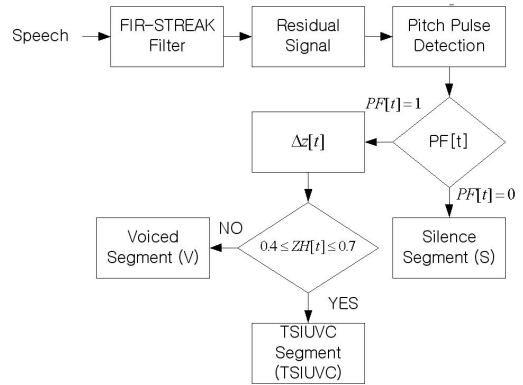
음성신호의 기본주파수인 피치 주기마다 멀티펄스 음원의 대표구간을 [Fig. 2]와 같이 복원하여 멀티펄스 음원을 생성한다. 이때 피치 주파수는 약80~370Hz인 것을 고려하면 피치 간격은 약2.7ms~12.5ms이고, 프레임 길이를 25.6ms로 하였을 경우 대략 2개~9개의 피치가 존재하게 된다.

### 3. FBD 신호처리

#### 3.1 FBD 전처리

음성신호의 특징을 살펴보면 유성음(V)에서는 피치정보와 낮은 zero crossing rate(ZCR)를 나타내며, 무성자음(UVC)에서는 피치정보가 없고 높은 ZCR를 나타낸다. 아울러, 천이구간(TS)에서는 피치정보가 없고 낮은 ZCR를 나타낸다. 이러한 특징들을 고려하여 연속음성에서 V, S, TSIUVC를 탐색·추출하고, FBD 재생에 필요한 유효한 주파수 정보를 전송하기 위한 전처리 방법을 [Fig. 3]에 나타냈다. 이때 음성신호는 3.4kHz LPF를 사용하였고, 표본화 주파수는 10kHz, 양자화는 12bit, 프레임 길이는 25.6ms로 하였다. [Fig. 3]의 처리방법을 살펴보면, 프레임 안에 개별 피치정보가 하나도 존재하지 않으면 (PF[t]=0) 프레임을 S로 판정하였고, 그렇지 않다면 해당 프레임의 ZCR( $Z[t]$ )과 프레임간의 ZCR( $\Delta Z[t] = Z[t] - Z[t-1]$ )차, 천이구간(TS)과 무성자음구간(UVC)의 ZCR( $ZH[t]$ )이  $\Delta Z[t] < 0$ ,  $Z[t-1] \geq 0.4$ ,  $0.4 \leq ZH[t] \leq 0.7$ 인 조건을 만족한 경우에 최초로 나타나는 개별피치( $P_0$ )의 위치에서 25.6ms 이전의 음성신호를 TSIUVC로 판정하고, 그렇지

않으면 V로 판정한다.



[Fig. 3] preprocessing of FBD

남여 9명의 연속음성(73문장, 모음수:609개, 무성자음수:195개)에서 본래 TSIUVC가 존재함에도 불구하고 추출되지 않았을 경우( $b_j$ )와 본래의 TSIUVC가 존재하지 않는데도 불구하고 추출된 경우( $c_j$ )를 TSIUVC추출오류로 규정한 식(5)에 의하여 TSIUVC 추출률을 산출하였다.

$$\mathfrak{R} = \frac{\sum_{j=1}^m \{a_j - (b_j + c_j)\}}{\sum_{j=1}^m a_j} \quad (5)$$

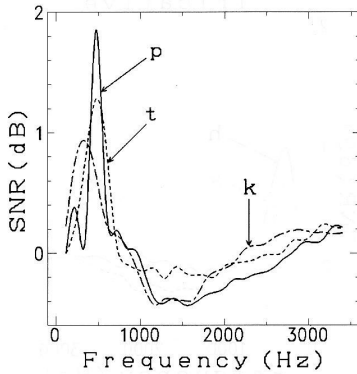
$a_j$ : TSIUVC 관찰 수,  $m$ : 음성샘플 수

실험결과, TSIUVC 추출률은 남자음성에서 96.2%, 여자음성에서 91%의 결과를 얻을 수 있었다. 이러한 추출률의 차이는 여자음성이 남자음성 보다 피치추출 오류가 높기 때문에 발생하는 문제로 생각된다.

#### 3.2 FBD 재생

V/S/TSIUVC의 선택정보에 의하여 V, S, TSIUVC를 추출하여 프레임안에 V, S, TSIUVC가 같이 있지 않도록 프레임을 재구성한다. 다음으로 V의 재생에 필요한 피치와 대표구간의 멀티펄스를 사용하여 V를 재생하고, S는 음성신호가 없는 구간임으로 프레임 길이만큼 시간간격을 갖도록 처리한다. 마지막으로 TSIUVC는 TSIUVC 재생에 유효한 주파수 정보를 사용하여 재생한다. 우선 TSIUVC 재생에 유효한 주파수 정보를 알아보기 위하여 TSIUVC 주파수 대역을 여러 주파수 영역으

로 분할하여 TSIUVC의 SNR 및 스펙트럼을 분석할 필요가 있다.



[Fig. 4] SNR of TSIUVC

일반적으로 음성통신방식에 3.4kHz LPF를 사용하기 때문에 본 연구에서도 3.4kHz LPF를 적용하였으며, 음성신호는 10kHz로 표본화하였다. 이때 25.6ms의 프레임은 29개의 주파수 영역으로 분할할 수 있고 주파수 해상도( $\Delta f$ )은 39.0625Hz 이 된다. 분할된 29개 영역의 주파수 신호를 사용하여 SNR를 측정함으로써 TSIUVC 재생에 유효한 주파수 영역을 알아낼 수 있는데, 나머지 9명의 대화채 음성신호(73문장, 무성 자음수:195개)를 사용하여 TSIUVC의 SNR를 측정된 결과의 대표예로서 "p", "t", "k"의 SNR를 [Fig. 4]에 제시하였다. TSIUVC의 SNR를 분석한 결과에서 주목할 것은 0.547kHz 이하의 낮은 주파수 영역과 2.813kHz 이상의 높은 주파수 영역에서 상대적으로 높은 SNR를 얻을 수 있었는데, 0.547kHz 이하에서는 1.24~1.82dB, 2.813kHz 이상에서는 0.65~0.9dB 이었다. 이것은 TSIUVC의 주요 주파수 정보가 높은 주파수와 중간 주파수 영역으로 양분되어 있는 것을 나타내는 것으로서, 천이구간(TS)과 무성자음(UVC)의 주파수 특성과도 부합하는 실험결과이다. 따라서 0.547kHz 이하의 주파수 정보와 2.813kHz 이상의 주파수 정보를 전송하고, 256 Point IFFT를 사용하여 TSIUVC를 재생하는 방법은 한정된 주파수 영역을 선택적으로 사용한다.

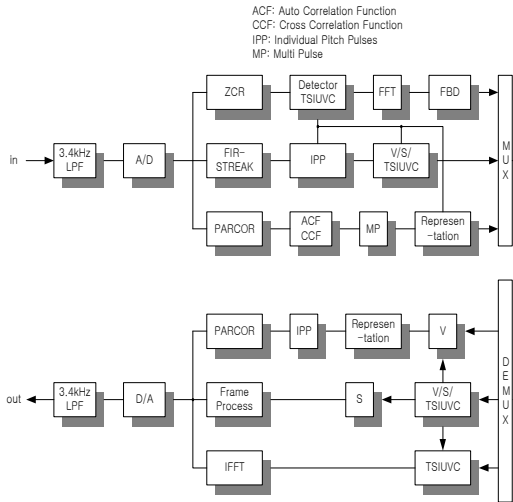
#### 4. FBD-MPC 시스템

일반적으로 음성부호화 방식을 비교평가 하기 위해서는 같은 bit rate에서 객관적 평가척도인  $SNR_{seg}$ 로 평가하여야 한다. 기존의 MPC 방식은 유성음원인 V음원과 무성음원인 UV음원을 V/UV 선택정보에 의하여 선택적으로 사용한다. FBD-MPC가 기존의 MPC와 차이점은 V, S, TSIUVC를 추출하여 같은 프레임에 존재하지 않도록 프레임을 재구성하고, V/S/TSIUVC 선택정보에 의하여 유성음(V) 프레임은 V음원을 사용하고 무성(S) 프레임은 25.6ms 시간간격 처리하며, TSIUVC 프레임은 FFT처리하여 0.547kHz 이하와 2.813kHz 이상의 주파수 정보를 이용하여 음성신호를 재생한다. low bit rate 이동통신의 음성부호화 방식은 일반적으로 4~10kbps으로 부호화 한다. 본 연구에서는 <Table 1>에 나타낸바와 같이 약 8kbps로 MPC와 FBD-MPC를 부호화하고, [Fig. 5]와 같이 FBD-MPC의 시스템을 구성하였다.

<Table 1> Coding Condition

parameter[bit]	MPC	FBD-MPC
V/UV	2	
V/S/TSIUVC		2
[V]		
PARCOR $k_i$	7,6,5,5,4 3,3,3,3,3(42)	7,6,5,5,4 3,3,3,3,3(42)
$g_{max}$	8	8
$g_k$	7(7*10=70)	4(4*10=40)
$m_k$	7(7*10=70)	5(5*10=50)
Multi-Pulse	10	10
$P_{AV}$	8	
$P_0$		7
$I_{AV}$		7
$DP_i, (i=2\sim9)$		3(3*8=24)
[TSIUVC]		
Max Amplitude		7
Frequency Signal of Low Band		6(6*12=72)
Frequency Signal of High Band		5(5*15=75)
bit	200	200
kbit/s	8	8

MPC와 FBD-MPC의 부호화 조건을 살펴보면, MPC와 FBD-MPC의 음원 선택정보인 V/UV와 V/S/TSIUVC에 2bit를 할당하였고 10개의 멀티펄스를 사용하였다.



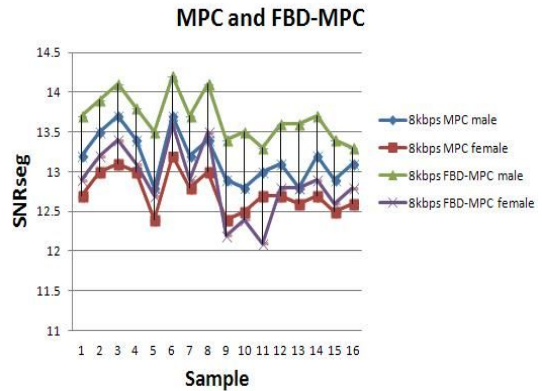
[Fig. 5] System of FBD-MPC

또한 유성음(V)을 재생하는 합성필터로서 10차의 PARCOR 필터를 사용하였으며, 낮은 차수일수록 스펙트럼에 미치는 영향이 크기 때문에 차등적인 bit를 할당하였다. MPC의 경우 멀티펄스 진폭( $g_k$ ) 및 위치( $m_k$ )에 각각 7bit를, FBD-MPC의 경우는 각각 4bit와 5bit를 할당하였으며, 진폭 값이 큰 멀티펄스의 최대 진폭( $g_{max}$ )에는 8bit를 할당하였다.

피치정보에 할당한 bit를 살펴보면 MPC의 경우에는 평균 피치정보( $P_{AV}$ )에 8bit를 할당하였으며, FBD-MPC의 경우에는 최초의 피치 펄스의 위치( $P_0$ )에 7bit, 개별피치 간격의 평균( $I_{AV}$ )에 7bit, 피치 간격의 편차( $DP_i, (i=2\sim9)$ )에 3bit를 할당하였다. 이때 25.6ms의 프레임마다 나타나는 피치의 개수는 최대 9개로 산정하여 계산하였다. 이것은 연속음성신호에서 피치 주파수는 약80~370Hz이고 이를 시간 간격으로 나타내면 약 2.7ms~12.5ms이 되기 때문에 25.6ms에 최대 9개의 피치가 존재하게 된다.

TSIUVC인 경우에 60Hz~0.547kHz의 저역주파수에는 6bit, 2.813kHz~3.4kHz의 고역 주파수에는 5bit를 할당하였다. 결국 MPC 및 FBD-MPC의 프레임마다 총200bit를 사용하였고 전송bit rate은 약8kbps가 된다.

남녀 각각 4명의 32개 단문(모음수:290개, 자음수:68



[Fig. 6]  $SNR_{seg}$  of MPC and FBD-MPC

개)을 사용하여 8kbps MPC와 FBD-MPC의  $SNR_{seg}$ 를 비교한 결과를 [Fig. 6]에 나타내었다. 실험 결과 <Table 2>에 나타낸바와 같이 FBD-MPC의  $SNR_{seg}$ 이 MPC에 비하여 남자 음성에서 0.5dB, 여자 음성에서 0.2dB정도 개선되었다. 이것은 연속음성을 프레임으로 처리하는 경우에 프레임에 모음과 자음이 같이 존재하는 프레임은 기존의 MPC에서는 유성음원 혹은 무성음원 어느 한쪽의 음원으로 음성신호를 재생하는데, 이러한 문제점을 본 연구에서는 특정 주파수대의 신호를 사용하여 음성신호를 재생함으로써  $SNR_{seg}$ 이 개선된 것을 알 수 있었다.

<Table 2>  $SNR_{seg}$  of MPC and FBD-MPC

Method [dB]	kbit/s	male	female
MPC	8	13.2	12.7
FBD-MPC	8	13.7	12.9

### 5. 결론

본 논문에서는 8kbps 동일한 전송량을 갖도록 부호화한 MPC와 FBD-MPC를 객관적 평가 척도인  $SNR_{seg}$ 로 비교 평가 하였다. 실험결과, 기존의 MPC에 비하여 FBD-MPC에서 남자와 여자음성에서 각각 0.5dB, 0.2dB 정도 개선되었다. 이것은 같은 프레임에 모음과 자음이 같이 존재하지 않도록 하고 음성신호 재생에 유효한 주파수 정보를 사용함으로써 음성품질 개선한 주요 원인으

로 작용하였다.

향후, 셀룰러폰이나 스마트폰에 적용할 수 있는 Bit Rate으로 낮추면서 음질을 개선할 수 있는 알고리즘이나 음성부호화 시스템을 주요 추진과제로 진행하고자 한다.

## ACKNOWLEDGMENTS

This paper presents research performed in 2014 under SANGMYUNG University.

## REFERECNES

- [1] 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
- [2] Z.A.Putnins, G.A.Wilson, J.Kumar and R.D. Trupp: "A Multi-Pulse LPC Synthesizer for Telecommunications use",IEEE,ICASSP,Mar,1985
- [3] 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
- [4] 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, Sinal Processing, p473-476.1986
- [5] 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
- [6] K.Krishna, V.L.N.Murty, K.R.Ramakrishnan:"Vector quantization of excitation gains in speech coding", Signal Processing 81,p203-209, 2001
- [7] 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