휴대폰 통신을 고려한 8kbps IP-MPC 방식에 관한 연구

이시우*

요약

본 논문은 피치추출 오류를 줄이고 피치간격의 변위에 적응할 수 있도록 피치간격을 정규화하지 않은 개별피치를 이용한 IP-MPC를 제안한다. 피치 추출률은 자기상관법, Cepstrum법, 개별피치추출법을 평가한 결과, 개별피치추출의 경우에 남자음성에서 96%, 여장음성에서 85%를 얻을 수 있었다. 또한 피치추출률에 있어서 개별피치 추출법이 자기상관법이나 Cepstrum법에 비하여 좋은 결과를 얻을 수 있었다. 휴대폰 통신을 고려한 8kbps에서 MPC와 IP-MPC의 음질평가를 실시한 결과, IP-MPC의 합성음성이 MPC의 합성음성에 비하여 음질이 보다 좋은 것을 알 수 있었다. 음질은 객관적인 척도인 SNRseg와 주관적인 척도인 MOS를 병행하여 평가 하였다. 실험결과, MPC에 비하여 IP-MPC의 남녀 음성에서 SNRseg의 경우에 각각 1.4dB, 1.2dB가 개선되었으며, MOS의 경우에 남녀 음성에서 각각 0.37, 0.22이 개선되었음을 알 수 있었다.

A Study on 8kbps IP-MPC Method Considering Cellular Phone

See-Woo Lee*

ABSTRACT

In this paper, I propose a new method of IP-MPC uses individual pitch in order to accommodate the changes in each pitch interval and reduce pitch errors. The pitch extraction rate was evaluated autocorrelation method, cepstrum method and individual pitch extraction method. As a result, the extraction rate of individual pitch was 85% for female voice and 96% for male voice. Also, I knew that pitch extraction rate was better in individual pitch extraction method than autocorrelation method and cepstrum method. I have evaluated the speech quality of MPC and IP-MPC considering a cellular phone of 8kbps. As a result, I knew that synthesis speech of the IP-MPC was better in speech quality than synthesis speech of the MPC. The speech quality was evaluated both the SNRseg and MOS. As a result, the SNRseg of IP-MPC improved 1.4dB and 1.2dB in female and male voice. In case of MOS, it was 0.37 and 0.22 in female and male voice respectively. In conclusion, I have found out that speech quality of IP-MPC was better than speech quality of MPC.

Key Words : Multi-Pulse Coding, Quality Improvement, Cellular Phone Network, Signal Processing

* 상명대학교 정보통신공학과(✉swlee@smu.ac.kr)
· 제1저자(First Author) : 이시우 · 교신저자(Correspondent Author) : 이시우
· 접수일(2010년 8월 10일), 수정일(1차 : 2010년 9월 10일), 게재확정일(2010년 9월 14일)

Ⅰ. 서 론

무선통신을 기반으로 하는 휴대폰의 경우에 낮은 Bit Rate이면서 음질이 우수한 음성부호화 방식이 요구된다. 낮은 Bit Rate을 실현하기 위하여 제안된 멀티펄스 음성부호화 방식(MPC: Multi Pulse Coding)은 AbS(Analysis by Synthesis)에 의하여 멀티펄스를 탐색하여 LPC 합성 필터를 구동함으로서 음성신호를 재생하는 방식이다[1]. 이 방식을 Putnins는 9.6kbit/s의 Bit Rate으로 음질을 개선하였고[2], Ozawa는 피치예측법과 피치보간법을 이용하여 4.8~9.6kbit/s에서 음질을 개선한 멀티펄스 음성부호화 방식을 제안하였다[3]. 특히 이 방식들은 모음부에서 피치주기를 추출하고 한주기의 피치구간에서 구동음원을 산출하여 송신하고, 수신측에서 피치구간마다 구동음원을 재생함으로서 낮은 Bit Rate을 실현하고 있다.

그러나 이 방식들은 모두 피치추출법에 자기상관법을 사용하고 있으나, 자기상관법은 디지털신호처리의 수학적 기법으로 매우 우수한 처리방법이라 할 수 있으나 실제적인 음성신호처리의 피치추출에서 많은 피치추출 오류가 발생한다. 이러한 피치추출 오류는 송신측에서피치정보에 의하여 유성음(V)/무성음(UV)을 선택하는 과정에서 좋지 않은 영향을 미치며, 수신측에서는 V/UV 선택정보에 의하여 유성음원과 무성음원을 선택하는 과정에서 잘못 선택된 음원에 의하여 음성신호를 재생하는 경우가 발생한다. 이것은 연속된 음성신호를 고정 프레임으로 처리할 경우 프레임마다 음성신호가 모음(V), 자음(C), 무음(S)의 형태로만 존재하는 것이 아니라 S+C 또는 S+V, C+V의 형태로 존재하며, 프레임내에 언어학적으로 상호 특성을 달리하는 신호가 혼재하게 된다. 이러한 프레임의 신호를 자기상관법으로 처리하는 경우에 피추출 오류나 V/UV 판독오류를 동반하게 된다. 이러한 오류는 휴대폰 음성통신과 같이 음원을 사용하여 음성신호를 재생하는 방식에서는 음성통신의 품질을 저하시키는 요인으로 작용한다. 이와 같은 문제점을 해결하기 위하여 본 논문에서는 프레임에 V, C, C의 단독으로 존재하는 경우나 S+C, S+V, C+V의 형태를 갖는 프레임에서 보다 정확한 피치를 추출할 수 있는 방법으로, LPC분석에서 원래의 신호와 재생한 신호의 차를 나타내는 잔차신호(Residual Signal)를 이용한 개별피치(IP) 추출법을 적용한 8kbps의 IP-MPC를 제안하고 기존의 MPC 방식과 음질을 비교 평가한 결과를 제시하고자 한다.

Ⅱ. 시스템 구성

본 논문에서 제시한 8kbps IP-MPC 방식을 그림 1에 나타내었다. 제시한 IP-MPC와 기존의 MPC에서 시스템 구조상 다른 부분은 개별피치(IP-Pulse)를 추출하는 부분이며, 이외에는 기존MPC와 시스템 구조가 같다.

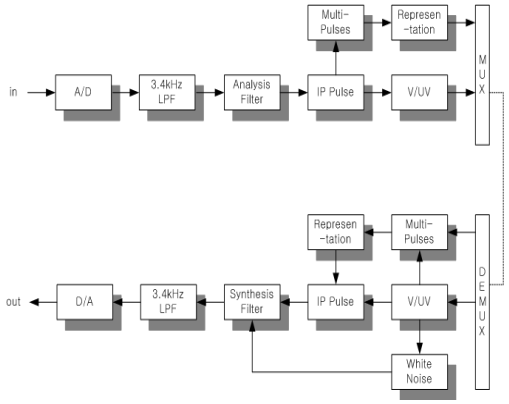

그림 1. 8kbps IP-MPC
Fig. 1 8kbps IP-MPC

시스템의 흐름을 간략히 설명하면, 마이크로폰에 입력된 음성신호는 10kHz, 12bit로 표본화 및 양자화하고 3.4kHz 로 대역제안 하였으며, FIR필터와

STREAK필터를 결합한 FIR-STREAK필터에 의하여 잔차신호를 얻는다[4].

FIR-STREAK 필터에서 출력된 잔차신호는 펄스성 잔차신호(R_P)와 잡음성 잔차신호로 구성되어 있는데, R_P로부터 그림 2의 (d)와 같이 개별피치 위치를 추출한다.

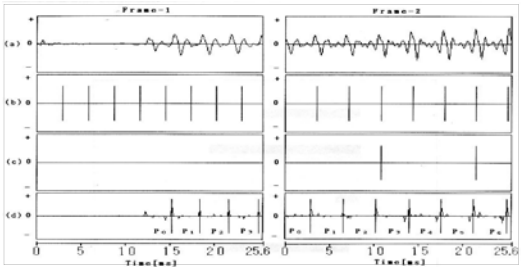

그림 2. 피치추출 예 (a)원음성 (b)자기상관법
(c)Cepstrum법 (d)개별피치 추출법
Fig. 2 Pitch Extraction Sample (a)Original Speech Signal
(b)Auto-Correlation Method (c)Cepstrum Method
(d)Individual Pitch Extraction Method

기존의 MPC에 사용되어온 자기상관법과 Cepstrum법은 그림 2의 (b), (c)에서와 같이 S+V, C+V 프레임과 프레임 경계부에서 피치추출 오류가 발생하는 반면 개별피치 추출법에서는 첫 번째 R_P로부터 음성신호의 시작위치를 추정하기 쉽고 피치위치를 보다 정확히 추출할 수 있다. 피치정보를 이용하여 연속된 음성신호에서 프레임의 V/UV 판정방법을 그림 3에 나타내었다. 일단, 프레임에 개별 피치가 하나라도 존재하면(PF[t]=1) 프레임을 V로 판정하고, 그렇지 않으면(PF[t]=0) 프레임을 UV로 판정하였다.

V로 판정된 프레임의 경우, Zero-Crossing Rate(ZCR)의 차($Z[t]$)와 프레임간의 ZCR($\Delta Z[t]=Z[t]-Z[t-1]$)차, S+V 혹은 C+V인 프레임의 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 처음에 나타난 개별피치(P_0)의 위치로 부터 UV구간을 추출하였다.

한편, 음성신호의 구동음원으로 사용하는 멀티펄스의 진폭과 위치를 g_k, m_k라고 하면, 멀티펄스의 음원 $v(n)$은 다음과 같이 나타낼 수 있다.

$$v(n)=\sum_{k=1}^{N} g_k \cdot \delta(n-m_k) \qquad (1)$$

$$\{\text{if } n=m_k, \delta(n-m_k)=1\}$$
$$\{\text{if } n\neq m_k, \delta(n-m_k)=0\}$$

$v(n)$에 의한 합성신호는

$$y(n)=\sum_{k=1}^{N} g_k \cdot h(n-m_k) \qquad (2)$$

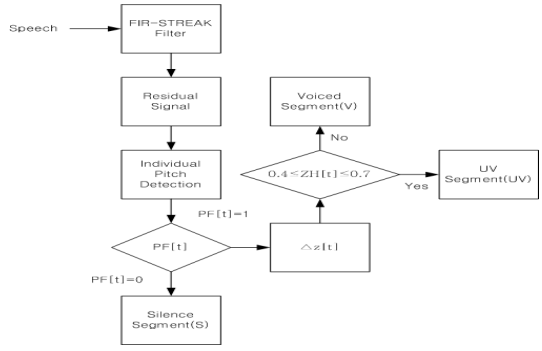

그림 3. V/UV 추출법
Fig. 3 Extraction Method of V/UV

멀티펄스의 진폭(g_k) 및 위치(m_k)는 식 3이 최소가 되도록 식 4에 의하여 구할 수 있다[1][3].

$$E=\sum_{n=1}^{N}[x(n)-y(n)\otimes w(n)]^2 \qquad (3)$$

$x(n)$: 원음성신호, $\otimes$: $convolution$,
N: 샘플수, $w(n)$: $weighting$필터

$$g_k=\max\left\{\Theta_{hx}(m_k-\sum_{j=1}^{k-1} g_j R_{hh}(|m_j-m_k|)\right\} \qquad (4)$$

R_{hh} : 자기상관함수, θ_{hw} : 상호상관함수

여기에서 얻은 멀티펄스를 그림 2의 개별피치 구간마다 분배하여 멀티펄스 구동음원을 만들고, 음성합성필터를 구동하여 음성신호를 재생하게 된다.

III. 부호화 조건

기존의 멀티펄스 음성부호화 방식(MPC)과 개별피치 추출법과 개별피치에 의한 V/UV 정보를 이용한 IP-MPC 방식을 8kbps에서 음질을 비교 평가하고자 한다.

표 1 Bit 할당
Table 1. Bit Allocation

파라미터	bit 할당	
	MPC	IP-MPC
V/UV	2	2
PARCOR계수 $k_i(i=1\sim10)$	7,6,5,5,4 3,3,3,3,3	7,6,5,5,4 3,3,3,3,3
g_{max}	13	7
$g_k(k=1\sim10)$	7	6
$m_k(k=1\sim10)$	7	6
평균 피치간격	8	
P_0		8
I_{AV}		2
$DP_i,(i=2\sim9)$		3
총 bit 수	205	205
kbps	8	8

* g_{max}:멀티펄스의 최대 진폭

음성신호는 3.4kHz LPF로 대역제안하고 10kHz, 12bit로 표본화 및 양자화 하였으며, 프레임 길이는 25.6ms로 하였으며, MPC와 IP-MPC의 bit 할당조건을 표 1에 나타내었다. V/UV의 판정에 의해 음성재생에 적합한 여러 형태의 음원으로 음성신호를 부호화 하는 방법[5][6]의 일종으로 MPC와 IP-MPC 방식은 모두 V/UV 판정에 의해 멀티펄스 음원과 White Noise 음원을 사용하여 PARCOR 합성필터를 구동함으로서 음성신호를 재생한다.

MPC와 IP-MPC 방식에서 사용하는 파라미터에 할당한 bit는 V/UV 판정에 2bit를 할당하였으며, PARCOR 합성 필터의 차수는 10차를 사용하는데, 이때 PARCOR계수의 변화가 스펙트럼의 변화에 미치는 영향은 낮은 차수의 계수일수록 영향이 크기 때문에[7] 낮은 차수일수록 차등적인 bit를 할당 하였다. 음성신호에서 피치 주파수는 대략 80~370Hz 이며 이것은 약2.7ms~12.5ms에 해당된다. 때문에 25.6ms에 최대 9개의 피치가 존재하게 되고, 이 9개의 피치를 최초의 개별피치펄스의 위치(P_0), 개별피치 간격의 평균(I_{AV}) 그리고 개별피치 간격의 편차($DP_i,(i=2\sim9)$)로 나타냄으로서 개별피치에 bit를 일률적으로 할당하는 것보다 할당 bit를 최소화할 수 있는 장점이 있다. 이러한 bit 할당의 최소화는 특히 전송속도를 가변적으로 적용하는 방법[8]에 특히 자주 사용한다.

개별피치와 멀티펄스에 bit 할당을 살펴보면, IP-MPC에서 P_0는 첫 번째 출현하는 개별피치로서 프레임 25.6ms내에 모든 위치를 나타낼 수 있도록 8bit를 할당하였으며, 개별피치 I_{AV}에 2bit, DP_i에 3bit를 할당하였다. 한편, MPC의 평균 피치정보에는 프레임 25.6ms내의 피치정보를 나타낼 수 있도록 8bit를 할당하였다. 사용한 멀티펄스 수는 모두 10개로, 멀티펄스의 최대 진폭(g_{max})에 MPC와 IP-MPC에서 각각 13bit, 7bit를, 멀티펄스 진폭 및 위치에 MPC에서는 각각 7bit를, IP-MPC에서는 각각 6bit 할당 하였다. 멀티펄스의 bit할당에 IP-MPC 보다 MPC에 보다 많은 bit를 할당한 것은 IP-MPC에서 피치정보에 보다 많은 bit를 사용하였기 때문에 MPC의 멀티펄스에 좀 더 많은 bit를 할당한 것이다. 이러한 부호화 조건은 MPC, IP-MPC 모두 사용한 bit 수가 같도록 설정하기 위한 것으로서 총 205 bit를 사용하였으며 전송속도는 8kbps 이다.

Ⅳ. 실험

음성부호화 방식의 음질평가는 일반적으로 객관적인 평가방식과 주관적인 평가방식을 병행하여 실시한다. 이것은 신호의 왜곡과 청각적 효과가 항상 일치하는 결과를 얻는 것은 아니기 때문이다. 일반적으로 객관적인 척도로는 식 5의 SNR_{seg}를 사용하고, 주관적인 척도로는 5단계 MOS(Mean Opinion Score, -2~2점)를 사용한다.

$$SNdR_{seg} = \frac{1}{T}\sum_{i=1}^{T}(SNR)_i f \qquad (5)$$

T: 피치가 존재하는 프레임 수

i: 프레임 번호

SNR_{seg}는 Spectrum의 일그러짐 정도를 알 수 있는 절대적인 평가방식이고, MOS는 음질의 상대적인 평가방식이다. 즉, MPC와 IP-MPC의 재생음성을 평가자에게 들려주고 아주 좋다 +2점, 좋다 +1점, 보통 0점, 나쁘다 -1점, 아주 나쁘다 -2점으로 평가하게 한다. 사용하는 음성제원에 있어서, 발성자의 인원이나 발성한 단문의 수가 많을수록 통신품질의 평가에 좋을 것으로 이해하는 경우가 있으나, 통신품질은 발성자수나 단어 수에 따라서 크게 변하지 않기 때문에 일반적으로 소수명이 발성한 수개의 단어로 시스템의 통신품질 평가가 가능하다.

다만, MOS 평가에 있어서 평가자 수가 극히 적으면 통신품질 평가에 영향을 미칠 수 있기 때문에 실험환경을 고려하여 10명 내외로 결정하는 경우가 많다. 참고로 Ozawa가 제안한 방법에서는 남자 3명, 여자 3명이 발성한 8개의 단문과 평가자 6명을 사용하였다[3]. 이러한 것을 고려할 때 MPC와 IP-MPC를 평가하기 위하여 적용한 표 2의 음성제원과 20명의 평가자수는 통신방식의 품질을 평가하는데 적정한 수준임을 알 수 있다.

표 2. 음성제원
Table 2. Speech Spec

제 원	남자음성	여자음성
발성자	10	10
단문 수	5	5
표본화(kHz)	10	10
양자화(bit)	12	12

표 2의 음성제원을 사용하여 MPC와 IP-MPC의 SNR_{seg}와 MOS를 측정한 결과, SNR_{seg}에서는 MPC의 경우에 남녀 음성에서 각각 13.2dB, 12.7dB의 결과를 얻을 수 있었으며, IP-MPC의 경우에는 각각 14.6dB, 13.9dB이였다. MOS에서는 MPC의 경우에 남녀 음성에서 각각 1.75, 1.77이였으며, IP-MPC의 경우에는 각각 2.12, 1.99의 결과를 얻을 수 있었다.

IP-MPC가 MPC에 비하여 남자 음성에서 1.4dB, 여자 음성에서 1.2dB정도 개선되었고, MOS 평가에서도 IP-MPC가 MPC에 비하여 남자 음성에서 0.37, 여자 음성에서 0.22 정도 개선된 것을 알 수 있었다. 개선된 지표에 대한 평가를 남자의 음성만으로 상대 평가해 보면, 8kbps MPC에서 13.2dB, 피치보간법을 사용한 MPC에서 13.6~14.2dB의 결과를 얻었으며[3], IP-MPC에서 는 14.6dB의 결과를 얻었다. 따라서 MPC에서 비해서는 1.4dB, 피치보간법을 사용한 MPC에 비해 0.4~1dB 정도의 개선효과를 얻을 수 있음을 알 수 있다.

이것은 기존의 MPC에서 피치추출법으로 사용하는 자기상관법은 모음부의 정상부에서도 종종 피치추출오류가 발생하고, 또한 피치의 정확한 위치를 추정하기 곤란한 근본적인 문제점이 음질열화 직접적인 원인으로 이해할 수 있다. 이러한 문제점을 해결할 수 있도록 설계한 개별피치추출법을 사용한 IP-MPC는 상대적으로 피치추출오류가 적고 피치의 위치를 정확히 추정할 수 있어 보다 음질열화를 억제할 수 있었던 것으로 해석할 수 있다.

Ⅴ. 결론

본 논문에서는 기존의 MPC 방식에서 사용하는 피치 추출 방법을 개선한 IP-MPC를 제안하고, 휴대폰 통신에 사용할 수 있는 8kbps에서 MPC와 IP-MPC의 음질평가를 실시하였다. 음질평가는 객관적인 방법인 SNR_{seg}와 주관적인 방법인 MOS를 병행하여 실시한 결과, SNR_{seg}와 MOS에서 음질이 개선된 것을 확인할 수 있었다. 이 정도의 결과는 8kbps 휴대폰 통신에 적용할 수 있는 수준이나 피치추출 방법이 기존의 MPC에 비하여 IP-MPC가 상대적으로 신호처리 방법이 복잡하기 때문에 하드웨어 구성면에서 다소 불리한 점이 있다. 때문에 향후 하드웨어로 구성할 경우나 신호처리 속도를 고려하여 시스템을 보다 단순화 할 필요가 있다. 또한 전송속도를 낮출 경우의 음질저하를 제어할 수 있는 새로운 방법을 고안하여야 하는 것은 향후 연구과제로 남겨두고자 한다.

참고문헌

[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, R.D.Trupp:"*A Multi-Pulse LPC Synthesizer for Telecommunications use*",IEEE,ICASSP,Mar,1985

[3] 小澤 一範, 恙關 卓: "*ピッチ情報を用いる 9.6~4.8 kbit/s マルチパルス 音聲符號化方式*" ,電子情報通信學會論文誌,Vol.J72-D-2, No.8, 1989

[4] 이시우:"*FIR-STREAK 디지털 필터를 사용한 피치추출 방법에 관한 연구*", 한국정보처리학회, 제6권 제1호, pp. 247~252, 1999

[5] McCree, A.V., Barnwell, T.P.,:"*A mixed excitation LPC vocoder model for low bit rate speech coding*", IEEE Trans. Speech Audio Process. 3 (4), 242-250, 1995

[6] 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

[7] 北脇 信彦, 板倉 文忠他:"*PARCOR形音聲分析合成系 における最適符號構成*",電子情報通信學會論文誌,Vol. J61-A No.2, pp.121-130, 1978

[8] Phu Chien Nguyen, Masato Akagi, Binh Phu Nguyen: "*Limited error based event localizing temporal decomposition and its application to variable-rate speech coding*", Speech Communication 49, p292-304, 2007

이시우(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~현재 상명대학교 정보통신공학과 교수
※ 관심분야 : 유무선통신, 음성신호처리, 감성처리