개별 피치정보를 이용한 멀티펄스 음성부호화 방식에 관한 연구

A Study on Multi-Pulse Speech Coding Method by Using Individual Pitch Information

이시우
상명대학교 정보통신공학과

See-Woo Lee(swlee@smu.ac.kr)

요약

본 연구에서는 피치추출 오류를 줄이고 피치간격의 변위에 적응할 수 있도록 피치간격을 정규화하지 않은 개별피치 펄스를 이용한 새로운 멀티펄스 음성부호화 방식(IP-MPC)을 제안하였다. 여기에서, 개별 피치 펄스의 추출률은 남자음성에서 96%, 여자음성에서 85%를 얻을 수 있었으며, 개별피치 펄스를 이용한 IP-MPC와 자기상관법의 피치정보를 이용한 MPC를 평가한 결과, IP-MPC의 음질이 MPC의 음질에 비하여 상당히 개선되었음을 알 수 있었다.

■ 중심어 : | 멀티펄스 | 음성부호화 | 피치 |

Abstract

In this paper, I propose a new method of Multi-Pulse Coding(IP-MPC) use individual pitch pulses in order to accommodate the changes in each pitch interval and reduce pitch errors. The extraction rate of individual pitch pulses was 85% for female voice and 96% for male voice respectively. I evaluate the MPC by using pitch information of autocorrelation method and the IP-MPC by using individual pitch pulses. As a result, I knew that synthesis speech of the IP-MPC was better in speech quality than synthesis speech of the MPC.

■ Keyword : | Multi-Pulse | Speech Coding | Pitch |

I. 서 론

통신망의 회선을 유효하게 사용하고 전송 효율을 높이는 낮은 bit rate의 음성부호화 방식으로 제안된 멀티펄스(Multi-Pulse) 음성부호화 방식은[1][2] 피치예측법과 피치보간법을 이용하여 멀티펄스 음성부호화 방식의 음질을 개선하였다[3]. 이 방식은 "피치 간격의 변동이 청각에 미치는 영향은 무시해도 좋다"라는 생각에 근거하여 프레임당 평균 피치간격을 구하는 자기상관법(Auto Correlation Method)을 사용하고 있다. 그러나 이러한 방법은 모음이 많지 않은 언어에는 적합하다고 볼 수 있으나 우리나라 언어와 같이 모음과 자음이 많은 경우에는 적합한 방법이라 볼 수 없다. 특히, 모음과 자음의 경계부 혹은 모음과 자음의 음소결합이 이루어지는 부분에서의 불안정한 피치 주기의 변화에 적응적으로 피치 위치를 정확히 추출할 필요가 있다.

접수번호 : #051124-001
접수일자 : 2005년 11월 24일
심사완료일 : 2006년 01월 04일
교신저자 : 이시우, e-mail : swlee@smu.ac.kr

본 연구에서는 "피치간격의 변동이 청각에 미치는 영향은 무시할 수 없다"라는 생각에서 프레임당 다수의 피치펄스를 추출하는 개별피치 추출법을 이용한 새로운 멀티펄스 음성부호화 방식을 제안한다.

II. 피치 추출

1. 자기상관법과 Cepstrum법

프레임당 평균 피치 정보를 추출하는 대표적인 방법인 자기 상관법과 Cepstrum법은 널리 알려진 방법이기 때문에 간략하게 기술하고자 한다.

자기 상관법에 의한 피치 추출법은 자기상관 계수 $R(t)$가 1에 근접한 시점을 피치의 개시·종료점으로 규정하여 피치주기를 구하게 된다.

$$R(t)=\sum_{n=0}^{N-1}(x(n)\cdot x(n-t))/\sum_{n=0}^{N-1}x(n)^2 \quad (1)$$

Cepstrum법은 식(2)의 최대 값으로부터 피치주기를 구할 수 있다. 여기에서, $x(k)$, $g(k)$, $h(k)$는 각각 음성신호 $x(n)$, 주기적인 음원 $g(n)$,임펄스 응답 $h(n)$을 FFT하여 얻는다.

$$c(n)=\frac{1}{N}\sum_{k=0}^{N-1}\log\ |x(k)|^2\exp(j(2\pi/N)\cdot k\cdot n)$$
$$=\frac{1}{N}\sum_{k=0}^{N-1}\log\ |g(k)|^2\exp(j(2\pi/N)\cdot k\cdot n)$$
$$+\frac{1}{N}\sum_{k=0}^{N-1}\log\ |h(k)|^2\exp(j(2\pi/N)\cdot k\cdot n) \quad (2)$$

2. FIR-STREAK필터에 의한 개별 피치 추출

일반적으로 자기상관법[4]은 수십ms 프레임 단위로 한개의 정규화된 피치정보를 산출한다. 따라서 음소 상호간의 간섭에 의해 피치간격이 일정하지 않을 경우, 또는 음성의 시작이나 끝부분과 같이 준주기성의 음성파형, 무성음과 유성음 혹은 무성자음과 유성음이 같이 존재하는 프레임에서는 피치추출 오류가 종종 발생한다. 이러한 오류를 억제할 수 있는 방법으로 본 연구에서는 FIR- STREAK 필터의 오차신호에서 추출한 개별피치 펄스를 적용하고자 한다[5].

FIR필터는 LPF의 역할을 하며, STREAK 필터는 전방향 오차신호($f_i(n)$)와 후방향 오차신호($g_i(n)$)의 순시값을 최소화 한다.

$$A_s=\ f_i(n)^2+\ g_i(n)^2$$
$$=-4\ k_i\cdot\ f_{i-1}(n)\cdot\ g_{i-1}(n-1)$$
$$+(1+\ k_i^2)\cdot(\ f_{i-1}(n)^2+\ g_{i-1}(n-1)^2) \quad (3)$$

여기에서 입력된 음성신호를 FIR필터로 처리하여 출력된 신호$y(n)$에 의한 $f_i(n)$과 $g_i(n)$의 초기값은 $f_0(n)=y(n)$,$g_0(n)=y(n-1)$이 된다.

STREAK 계수 k_i는 윗 식을 k_i에 관하여 편미분함으로서 다음과 같이 구할 수 있다.

$$k_i=\frac{2\cdot\ f_{i-1}(n)\cdot\ g_{i-1}(n-1)}{f_{i-1}(n)^2+\ g_{i-1}(n-1)^2} \quad (4)$$

여기에서, i=1,2,...,M이고, n=1,2,...,N이다.

k_i를 사용한 STREAK 필터의 전달함수는 다음과 같다.

$$H_s(z)=\frac{1}{\sum_{i=0}^{M_i}k_i\ z^{-i}} \quad (5)$$

[그림 1]은 프레임 길이가 25.6ms인 연속된 프레임에서 흔히 볼 수 있는 음성파형으로서 자기상관법과 Cepstrum법에 의하여 추출한 피치와 개별 피치추출법에 의하여 추출한 피치를 나타내고 있다. [그림 1]에서 Frame-2와 같이 ①유성음의 경우에는 개별피치 추출법과 자기상관법에서 유효한 피치정보를 얻을 수 있었던 반면 Cepstrum법에서는 피치추출 오류를 볼 수 있다. 또한 Frame-1과 같이 ②무성음과 유성음, 혹은 무성자음

과 유성음이 같이 존재하는 부분, ③음소가 변위하는 부분, ④프레임의 경계 부분, ⑤음성의 시작 부분, ⑥음성의 끝 부분 등에서 자기상관법과 Cepstrum법에서는 피치추출 오류를 볼 수 있는 반면 개별 피치 추출법에서는 이러한 피치추출 오류를 해결할 수 있음을 알 수 있다.

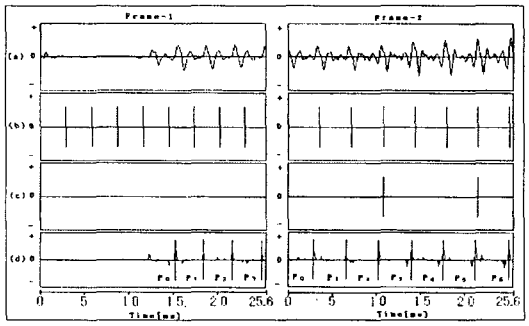

그림 1. 피치추출 (a) 원음성 (b) 자기상관법
(c) Cepstrum법 (d) 개별피치 추출법

3. 피치 추출률

피치 추출률은 합성음성의 품질을 향상시키기 위한 요소이며, 엄정한 규칙과 세심한 관찰력에 의하여 산출되어야 한다.

본 연구에서는 피치추출 오류를 시간상의 음성파형에서 관찰된 실제의 피치간격과 R_P에서 산출한 피치 간격이 일치하는지의 여부를 비교 관찰하여 판정하도록 하였다. 구체적으로는 본래 피치가 존재함에도 불구하고 이를 추출하지 못한 경우(b_{ij}), 또는 피치가 존재하지 않는데 추출된 경우(c_{ij})를 피치추출 오류로 판정하여 피치 추출률(P_R)을 산출하도록 하였다.

$$P_R = \frac{\sum_{i=1}^{T}\sum_{j=1}^{m}[\,a_{ij} - (\,b_{ij} + c_{ij})]}{\sum_{i=1}^{T}\sum_{j=1}^{m} a_{ij}} \qquad (6)$$

위식에서 m, T, a_{ij}, b_{ij}, c_{ij}는 각각 프레임 총수, 총 음성제원 수, 관찰된 피치수, 피치를 추출하지 못한 경우의 오류, 피치가 존재하지 않는데 추출된 경우의 오류를 나타낸다.

[표 1]의 음성 표본과 식(6)을 사용하여 피치 추출률을 산출한 결과, [표 2]와 같은 결과를 얻을 수 있었다. 이때, 자기상관법과 Cepstrum법을 개별 피치펄스 추출법과 비교하기 위하여 자기상관법과 Cepstrum법에 의하여 추출한 피치는 25.6ms의 프레임에 나타낼 수 있는 피치의 수로 환산하였다.

표 1. 음성 표본

제 원	남자음성	여자음성
발성자 및 단문 수	4명, 16개	4명, 16개
모음, 자음 수	145개, 34개	145개, 34개

표 2. 피치 추출률

방 법	남자	여자
개별피치 추출법	96%	85%
자기상관법	89%	80%
Cepstrum법	92%	86%

III. 멀티펄스의 탐색

멀티펄스는 자기상관 함수와 상호상관 함수에 의하여 구할 수 있다[6]. 멀티펄스의 진폭과 위치를 g_k, m_k라고 하면, 멀티펄스의 음원 $v(n)$은 다음과 같이 나타낼 수 있다.

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

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

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

$v(n)$에 의한 합성신호는

$$y(n) = \sum_{k=1}^{N} g_k \cdot h(n - m_k) \qquad (8)$$

여기에서, $h(n)$은 합성필터의 임펄스응답으로 전달함수는 다음과 같다.

$$H(z)=\frac{1}{1-\sum_{i=1}^{M} a_i z^{-i}} \quad (9)$$

a_i: 선형예측계수, M: 필터차수

멀티펄스의 진폭 및 위치는 다음 식의 오차가 최소가 되도록 결정한다.

$$E=\sum_{n=1}^{N}[x(n)-y(n)\otimes w(n)]^2 \quad (10)$$

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

여기에서, $w(n)$은 Noise-Weighting 필터로서 다음과 같다.

$$w(n)=\frac{1-\sum_{i=1}^{M} a_i z^{-1}}{1-\sum_{i=1}^{M} a_i \rho^i z^{-i}} \quad (11)$$

ρ는 계수로서 $0\le\rho\le1$로서 ρ는 멀티펄스 수와 SNR_{seg}의 관계를 고려하면[7] $\rho=0.8$이 적절한 것으로 판단된다.

식(11)에서 선형예측계수 a_i는 식(12)가 최소가 되도록 a_i에 대하여 편미분하여 구할 수 있다.

$$J=\sum_{n=1}^{N} e(n)^2=\sum_{n=1}^{N}[x(n)-y(n)]^2 \quad (12)$$

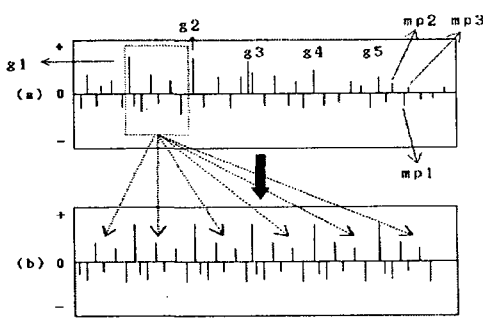

그림 2. 멀티펄스의 진폭과 위치

식(12)를 최소화하는 멀티펄스의 진폭 g_k와 위치 m_k는 다음과 같다.

$$g_k=\max\left\{\Theta_{hx}(m_k-\sum_{j=1}^{k-1} g_j R_{hh}(|m_j-m_k|)\right\} \quad (13)$$

여기에서, $1\le k\le M$이고 장기상관 함수($R_{hh}(m)$)와 상호상관 함수($\Theta_{hx}(m)$)는 다음과 같다.

$$R_{hh}(m)=\sum_{n=1}^{N} h(n)-h(n-m), 1\le m\le N \quad (14)$$

$$\Theta_{hx}(m)=\sum_{n=1}^{N} x(n)h(n-m), 1\le m\le L \quad (15)$$

결국, 이러한 과정에 의하여 얻은 [그림 2](a)와 같은 멀티펄스를 [그림 2](b)에 나타낸바와 같이 개별피치의 구간에 멀티펄스를 재분배함으로서 음성합성필터를 구동할 수 있는 멀티펄스열을 만들게 된다. 여기에서 개별피치구간의 멀티펄스의 진폭과 위치, 개별 피치펄스의 위치정보를 수신측에 전송하여 사용하게 된다.

Ⅳ. 실험결과

1. 부호화 조건

멀티펄스 음성부호화 방식에 있어서 피치 추출법에 자기상관법을 적용한 경우(MPC)와 개별피치 추출법을 적용한 경우(IP-MPC)의 음질을 비교 평가하고자 한다. 우선, 서로 다른 방식의 음질을 비교 평가하기 위해서는 같은 bit rate이 되도록 전송 파라미터에 할당하는 bit를 조절할 필요가 있다.

음성신호는 3.4kHz LPF로 대역제안하고 10kHz, 12bit로 표본화 및 양자화 하였으며, 프레임 길이는 25.6ms, V/UV 음원절환 방식을 채택하였다. V/UV는 프레임에 피치정보가 있으면 V로, 그렇지 않으면 UV로 결정하였으며, V일 때 [표 3]의 부호화 조건을 사용하고, UV일 때는 white noise를 사용하였다. PARCOR 합성 필터의 차

수는 일반적으로 10차를 사용하는데, 이때 PARCOR계수의 변화가 스펙트럼의 변화에 미치는 영향은 낮은 차수의 계수일수록 영향이 크기 때문에[8] 낮은 차수일수록 차등적인 bit를 할당하였다.

표 3. 부호화 조건

parameter[bit]	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}, g_k, m_k, 멀티펄스수	6,6,6 10 (126bit)	6,4,5 10 (96bit)
평균 피치정보	8	
P_0, I_{AV}		7,7
$DP_i,(i=2\sim9)$		2 (3×8)
총 bit 수	178	178
kbps	6.9	6.9

* g_{max}:멀티펄스의 최대 진폭

남녀 음성신호에서 피치 주파수는 약 80~370Hz이기 때문에 이를 시간 간격으로 나타내면 약 2.7ms~12.5ms이고, 25.6ms에 최대 9개의 피치가 존재하게 된다. 이러한 까닭으로 IP-MPC의 경우에 프레임당 최대 9개의 피치가 존재하는 것으로 간주하여 최초의 개별피치 펄스의 위치(P_0)에 7bit, 개별피치 간격의 평균(I_{AV})에 7bit, 개별피치 간격의 편차($DP_i,(i=2\sim9)$)에 3bit를 할당하였다. 이것은 각각의 개별피치 펄스에 bit를 할당하는 것보다 개별피치 정보를 모두 표현할 수 있는 P_0, I_{AV}, $DP_i,(i=2\sim9)$에 bit를 할당함으로서 bit를 절약할 수 있기 때문이며, MPC의 경우에는 평균 피치정보에 8bit 할당하였다.

MPC, IP-MPC 모두 10개의 멀티펄스를 사용하였는데, MPC와 IP-MPC의 bit rate을 같게 하기 위하여 MPC의 멀티펄스 진폭 및 위치에 각각 2bit, 1bit 높게 할당하였으며, 상대적으로 진폭 값이 큰 멀티펄스의 최대 진폭에는 6bit를 할당하였다.

[표 3]의 부호화 조건에 의하여 할당한 총bit수는 MPC와 IP-MPC에서 각각 178bit가 되며, 25.6ms의 프레임당 총 bit rate은 6.9kbps가 된다.

2. 음질의 평가

개별피치 추출법을 사용한 IP-MPC의 시스템 블록도를 [그림 3]에 나타내었다.

[표 1]의 음성표본을 사용하여 MPC와 IP-MPC에 대하여 SNR_{seg}와 MOS(Mean Opinion Score)를 사용하여 음질을 평가하였다. 음질평가는 일반적으로 객관적인 평가와 주관적인 평가를 동시에 수행하게 되는데, 객관적인 평가는 식 (16)의하여 구할 수 있으며, 객관적인 평가는 5단계 MOS(-2~2점, 20명)를 사용하였다. 즉, 아주 좋다 +2점, 좋다 +1점, 보통 0점, 나쁘다 -1점, 아주 나쁘다 -2점으로 평가하게 된다. SNR_{seg}는 Spectrum의 일그러짐 정도를 알아내는 절대평가 방식이고, MOS는 MPC와 IP-MPC의 상대평가를 통하여 MOS평가 점수를 얻게 된다.

$$SNR_{seg}=\frac{1}{T}\sum_{i=1}^{T}(SNR)_i \qquad (16)$$

T: 피치가 존재하는 프레임 수

i: 프레임 번호

$(SNR)_i$: 프레임별 Signal to Noise Ratio

실험결과, IP-MPC가 MPC에 비하여 남자 음성에서 1.2dB, 여자 음성에서 1dB 정도 개선되었고, MOS 평가에서도 IP-MPC가 MPC에 비하여 남자 음성에서 0.26, 여자 음성에서 0.16 정도 개선된 것을 알 수 있었다.

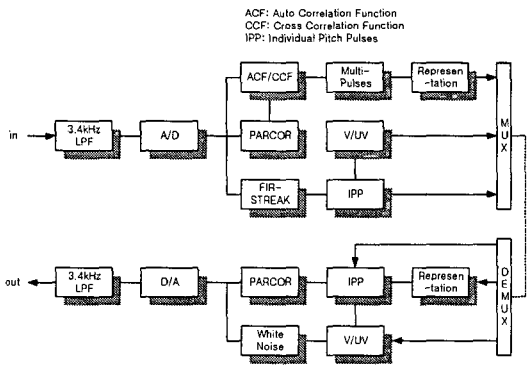

그림 3. IP-MPC 방식의 블록도

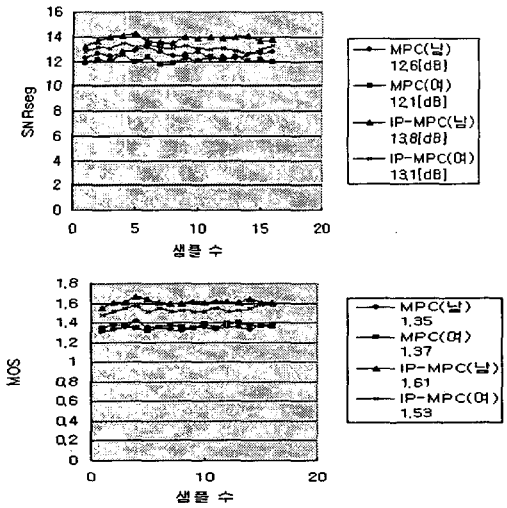

그림 4. SNRseg와 MOS

V. 결 론

본 논문에서는 FIR-STREAK 필터의 오차신호에서 피치를 추출하는 개별 피치펄스 추출법을 적용한 멀티펄스 음성부호화 방식(IP-MPC)을 제안하였다. 실험결과, 기존의 멀티펄스 음성부호화 방식에 자기상관법을 적용한 경우(MPC)에 비하여 피치추출 오류를 줄일 수 있고, 피치의 시간적 변동에 적응할 수 있는 IP-MPC가 상대적으로 높은 SNR_{seg}와 MOS값을 얻을 수 있었다.

향후, 연구과제로는 유성음, 무성음, 자음의 처리를 개별적으로 처리할 수 있는 멀티펄스 음성부호화 방식의 연구가 될 것이다.

참 고 문 헌

[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, pp.614-617, 1982.

[2] Z. A. Putnins, G. A. Wilson, J. Kumar, and R. D. Trupp, "A Multi-Pulse LPC Synthesizer for Telecommun-ications use," IEEE, ICASSP, pp.221-229, 1985.

[3] 小澤 一範, 蒜關 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, pp.159-168, 1989.

[4] 藤井健作:"自己相關法による電話帶域音聲のピッチ 抽出法"電子情報通信學會技術報告書, sp87-65, pp.33-40, 1987.

[5] 이시우, "FIR-STREAK 디지털 필터를 사용한 피치추출 방법에 관한 연구", 한국정보처리학회 논문지, 제6권, 제1호, pp.247-252, 1999.

[6] K. Ozawa, S. Ono, and T. Araseki, "A Study on Pulse Search Algorithms for Multipulse Excited Speech Coder Realization," IEEE, Vol.SAC-4, No1, pp.133-141, Jan, 1986.

[7] 武田 昌一他:"殘差音源利用分析合成方式とマルチルス法の基本特性の比較檢討", 電子情報通信學會論 文誌, Vol.J73-A, No.11, pp.132-140, 1990.

[8] 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.

저 자 소 개

이 시 우(See-Woo Lee) 정회원

- 1987년 : 동국대학교 전자공학과 (공학사)
- 1990년 : 日本大學(Nihon Univ) 전자공학과(공학석사)
- 1994년 : 日本大學(Nihon Univ) 전자공학과(공학박사)
- 1994년~1998년 : (주)삼성전자 통신연구소/멀티미디어 연구소
- 1998년~현재 : 상명대학교 정보통신공학과 교수

<관심분야> : 음성신호처리, 유무선통신, 음주지각