주파수 영역의 선택정보를 이용한 멀티펄스 음성부호화 방식에 관한 연구

A Study on Multi-Pulse Speech Coding Method by using Selected Information in a Frequency Domain

이 시 우*
See-Woo Lee

요 약

본 연구에서는 연속음성에서 무성자음을 포함한 천이구간을 탐색, 추출하고 주파수대역에서 근사합성하는 새로운 멀티펄스 음성부호화 방식(FBD-MPC)를 제안하였다. 실험결과, 여자 음성의 경우 TSIUVC 추출율은 84.8%(파열음), 94.9%(마찰음), 92.3%(파찰음) , 남자 음성의 경우는 88%(파열음), 94.9%(마찰음), 92.3%(파찰음)의 결과를 얻었다. 아울러, 0.547kHz 이하 2.813kHz 이상의 주파수 정보를 사용하여 TSIUVC 음성파형을 양호하게 근사합성할 수 있었으며, 유성음/무성음 선택정보를 이용한 MPC와 유성음/무음/TSIUVC를 이용한 FBD-MPC를 평가한 결과, FBD-MPC의 음질이 MPC의 음질에 비하여 개선되었음을 알 수 있었다.

Abstract

In this paper, I propose a new method of 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. As a result, the extraction rates of TSIUVC are 84.8%(plosive), 94.9%(fricative) and 92.3%(affricative) in female voice, 88%(plosive), 94.9%(fricative) and 92.3%(affricative) in male voice respectively. Also, I obtain a high quality approximation-synthesis waveforms within TSIUVC by using frequency information of 0.547kHz below and 2.813kHz above. I evaluate MPC by using switching information of voiced/unvoiced and FBD-MPC by using switching information of voiced/Silence/TSIUVC. As a result, I knew that synthesis speech of FBD-MPC was better in speech quality than synthesis speech of the MPC.

☞ Keyword : Speech Signal Processing, Speech Coding, Frequency Domain, 음성신호처리, 음성부호화, 주파수영역

1. 서 론

Atal에 의하여 처음으로 제안된 멀티펄스 음성부호화 방식은 음성신호를 유성음(V: Voiced)과 무성음(UV: Unvoiced)의 선택정보에 의하여 유성음(V)은 피치구간마다 유성음원을 사용하고, 무성음(UV)은 White Noise를 사용하여 음성을 재생한다[1]. Ozawa는 Atal이 제안한 멀티펄스 음성부호화 방식에서 멀티펄스를 탐색하는 알고리즘을 적용하여 음성품질을 개선하였다[2]. 이 밖에 음성신호를 압축·복원하는 여러 형태의 음성부호화 방식이 있는데, 대부분이 음성신호를 유성음(V)/무성음(UV) 혹은 유성음(V)/무성음(UV)/무음(S: Silence)과 같은 선택정보에 의하여 음성신호를 재생하는 방식[3-8]이다. 일반적으로 음성처리를 위하여 음성신호를 수십ms의 고정된 프레임으로 분할하여 처리하는데, 음성신호는 음소간의 상호작용에 의하여 연속적으로 변화하는 신호이다. 즉 실제 음성신호에서 프레임내 음성신호가 유성음(V), 무성음(UV), 무음(S)과 같이 각기 독립적으로 존재하는 것이 아니라 무음(S)+무성음(UV) 또는 무음(S)+유성음(V), 유성음(V)+무성음(UV)의 형태로 존재한다. 이러한 형태의 음성신호는 과도기적인 특성

* 정회원 : 상명대학교 정보통신공학과 교수
swlee@smu.ac.kr
[2005/12/23 투고 - 2006/01/19 심사 - 2006/05/04 심사완료]

을 나타내며, 특히, 모음과 자음이 결합하여 유성음(V)과 무성음(UV)의 중간특성을 나타내는 천이구간(TS: Transition Segment)이 존재한다. 이 천이구간(TS)의 음성신호를 유성음원 혹은 무성음원으로 재생하는 것은 문제점이라 볼 수 있다. 이러한 문제점을 해결하는 방법으로 유성음(V)과 무성자음(UVC: Unvoiced Consonant)이 같은 프레임에 존재하지 않도록 프레임의 길이를 동적으로 할당하는 것도 고려해 볼 수 있으나, 이것은 디지털 신호처리의 특성상 상당히 어려운 처리과정이라 할 수 있다. 그래서 본 논문에서는 특성을 달리하는 유성음(V)부, 무음(S)부, TSIUVC(Transition Segment Including Unvoiced Consonant)부를 연속음성에서 추출하고 프레임을 재구성하 방법과 TSIUVC를 근사합성 하는데 유효한 주파수 대역을 선택적으로 사용하는 새로운 멀티펄스 음성부호화 방식을 제안한다. 기존의 멀티펄스 음성부호화 방식은 유성음(V)/무성음(UV)의 선택정보에 의하여 유성음원과 무성음원을 사용하는 것을 특징이라고 한다면, 본 논문에서 제안하는 방식은 유성음(V)/무음(S)/TSIUVC의 선택정보에 의하여 프레임을 자동으로 재구성하는 것과, TSIUVC를 근사합성 하는데 유효한 주파수 정보를 사용하여 음성신호를 부호화하는 것을 특징으로 한다. 이러한 특징은 기존의 음성부호화 방식에서 찾아볼 수 없는 새로운 방식으로서 크게 두가지 문제점을 해결하고 있다. 첫째는 유성음(V)과 무성음(UV)이 같은 프레임에 존재하는 경우에 유성음원 혹은 무성음원 어느 한쪽의 음원만을 사용하는 문제점을 해결하였다. 둘째는 유성음(V)에서 무성음(UV)으로 천이되는 과정에서 존재하는 과도기적인 음성파형인 TSIUVC를 유성음원 혹은 무성음원 어느 한쪽의 음원을 사용하지 않고 재생하는 방법을 구현한 것이다.

본 논문의 전체적인 구성을 살펴보면, Ⅱ장에서는 연속음성에서 유성음(V), 무음(S), TSIUVC를 탐색하고 추출하여 V/S/TSIUVC 선택정보를 만들고, TSIUVC를 근사합성 하는데 유효한 주파수 정보에 대하여 소개하였다. Ⅲ장에서는 유성음(V)의 재생에 사용하는 유성음원인 멀티펄스를 구하는 방법에 대하여 기술하였다. Ⅳ장에서는 기존 방식의 부호화 조건과 본 논문에서 제안한 방식의 부호화 조건을 같은 bit rate이 되도록 한 상태에서 객관적인 평가인 SNR과 주관적인 평가인 MOS 평가의 결과에 대하여 소개하였고, Ⅴ장에서 본 논문의 결론을 맺는다.

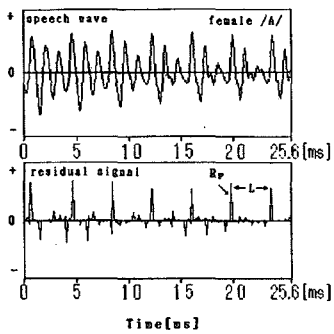
2. V/S/TSIUVC 신호처리

2.1 탐색 및 추출

연속음성을 프레임으로 처리할 때 유성음(V), 무음(S), TSIUVC를 판정하기 위한 유효한 정보로 피치(pitch)와 ZCR(zero crossing rate)이 있다. 일반적으로 유성음(V)에서는 낮은 ZCR과 피치정보를 갖고, 무성자음(UVC: Unvoiced Consonant)에서는 높은 ZCR과 피치정보가 없으며, 천이구간(TS: Transition Segment)에서는 낮은 ZCR과 피치정보가 없는 특징을 나타낸다[9,10]. 따라서 TSIUVC가 무성자음(UVC)+천이구간(TS: Transition Segment)형태인 것을 고려하면 TSIUVC 시작부는 낮은 ZCR을 나타내고, 후반부는 높은 ZCR을 나타내는 것을 유추할 수 있다. 남여 9명의 39문장을 사용하여 연속음성의 지속시간을 관찰한 결과, 연속음성에서 유성음의 지속 시간은 100ms~500ms정도이며 약 2.7ms~12.5ms 간격마다 주기적인 특징을 갖고 있는 반면, 무성자음의 경우는 무성 파열자음, 무성 마찰자음, 무성 파찰자음 별로 약간의 차이는 있으나 대개 20ms 전후이고, 천이구간(TS)의 경우는 약 5ms전후인 지속시간을 갖는다. 따라서 연속음성에서 TSIUVC는 약 25ms전후 길이를 갖고 있음을 알 수 있었다.

이러한 특징들과 TSIUVC의 위치를 효과적으로 추출하기 위해서는 TSIUVC가 끝나는 위

치를 탐색할 필요가 있다. TSIUVC가 끝나는 위치가 모음의 시작위치인 만큼, 모음에 존재하는 피치의 시작위치를 탐색한다면 TSIUVC를 추출할 수 있다. 따라서 시간영역에서 피치위치를 추출하기 위하여 FIR(Finite Impulse Response) 필터와 STREAK(Simplified Technique for Recursively Estimating Autocorrelation K-parameters)필터를 혼합한 형태의 FIR-STREAK 디지털 필터로 음성신호를 처리하여 그림 1과 같은 잡음성 잔차신호와 펄스성 잔차신호(R_P)를 추출하여 피치위치를 탐색할 수 있다[11]. 이러한 잔차신호는 원래의 신호와 예측한 신호의 오차신호로서, 시간영역에서 진폭 값이 큰 파형에서 펄스성 잔차신호(R_P)가 나타나고, 진폭 값이 작은 파형에서 잡음성 잔차신호가 나타난다. 이것은 유성음에서 진폭 값이 큰 파형의 주기성을 근거로 피치위치를 추정하는 것으로서 피치정보가 일반적으로 80Hz~370Hz에 존재하기 때문에 펄스성 잔차신호(R_P)의 간격 L은 $2.7ms \le L \le 12.5ms$을 갖게 된다. 따라서 프레임의 길이가 25.6ms라고 하다면, 펄스성 잔차신호(R_P)는 프레임당 최소 2개에서 최대 9개가 나타나게 된다. 결국, 펄스성 잔차신호(R_P)에서 피치펄스의 위치정보를 얻고, 이러한 피치정보는 프레임 단위로 정규화한 피치정보[12]가 아니라 그림 1과 같이 프레임의 시간상에 여러 개의 피치정보를 갖는 피치펄스로서 TSIUVC를 추출하는데 중요한 정보가 된다.



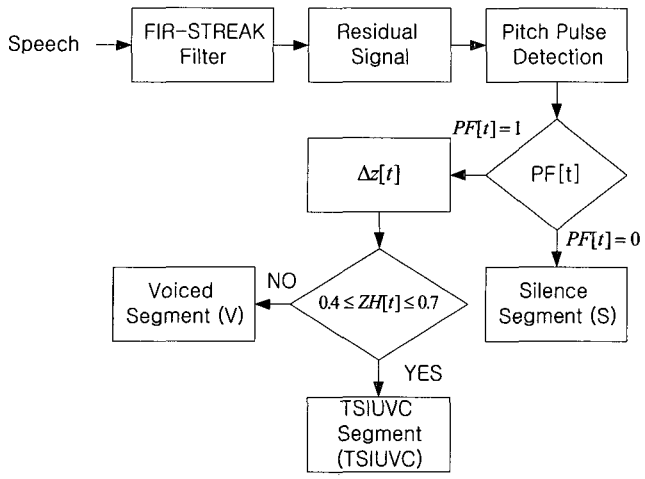
〈그림 1〉 피치펄스

이러한 피치펄스와 식 (1)의 ZCR을 이용하여 TSIUVC를 탐색 추출하는 방법을 그림 2에 제시하였다. 우선, 음성신호는 3.4kHz LPF로 주파수 대역을 제한한 다음 10kHz, 12bit로 표본화 및 양자화하고, FFT 처리를 위하여 프레임의 길이는 25.6ms로 하였다. 다음으로 프레임 안에 피치펄스가 하나라도 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 $\Delta Z[t] < 0, Z[t-1] \ge 0.4$, $0.4 \le ZH[t] \le 0.7$인 조건을 만족한 경우에 최초에 나타나는 피치펄스(P_0)위치에서 25.6ms 이전의 음성신호를 TSIUVC로 판정하였고, 그렇지 않다면 유성음(V)로 판정하였다. 이와 같은 판정조건에서 남여 9명의 39문장의 음성샘플을 사용하여 연속음성에서 TSIUVC를 추출한 결과, 여자 음성의 경우 파열음, 마찰음, 파찰음에서 각각 84.8%, 94.9%, 92.3%의 결과를 얻을 수 있었으며, 남자의 경우는 각각 88%, 94.9%, 92.3%의 결과를 얻었다. TSIUVC 추출율은 남자와 여자음성에서 자음의 종류에 관계없이 거의 같은 결과를 얻을 수 있었으며, 특히 파열음에서 상대적으로 낮은 추출율이 얻어진 이유는 피치펄스에서 최초로 나타나는P_0의 정확성이 낮은 것이 원인으로 추정된다. 따라서 P_0의 위치정보를 보다 정확히 추출할 수 있다면, 보다 좋은 성과를 얻을 수 있을 것으로 생각된다.

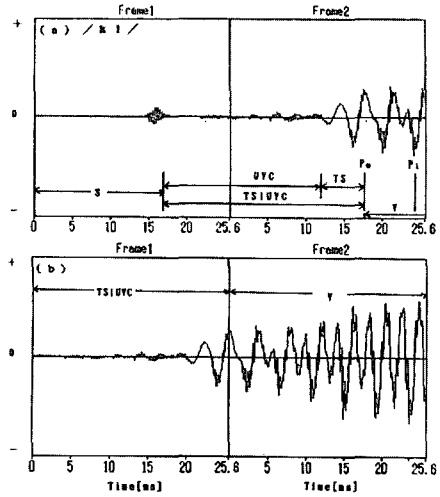
$$Z[t] = \frac{1}{2 \cdot N} \sum_{n=1}^{N} |sgn[x(n)] - sgn[x(n-1)]| \quad (1)$$

if $x(n) \ge 0, sgn[x(n)] = 1$, else if $sgn[x(n)] = -1$, t : 프레임 번호

이와 같은 방법에 의하여 연속음성에서 유성음(V), 무음(S), TSIUVC를 추출하여 프레임을 재구성한 예를 그림 3에 나타내었다. 그림 3(a)는 무음(S), 자음(UVC), 천이구간(TS), 모음(V)



〈그림 2〉 TSIUVC 탐색과 추출



〈그림 3〉 V/S/TSIUVC 프레임 재구성
(a) 본래의 프레임 (b) 재구성한 프레임

가 모두 존재하는 여자음성 /KI/의 연속적인 25.6ms의 두 프레임을 나타낸 것이다. 이와 같은 프레임은 유성음원과 무성음원 어느 한쪽의 음원을 사용하여 음성신호를 재생하는 것이 문제가 된다. 이와 같은 문제점을 해결하기 위해서 본 논문에서 제안한 방법을 사용하여 그림 3(b)와 같이 프레임을 재구성함으로서 프레임 마다 음성신호의 특성에 맞는 신호처리 방법을 선택할 수 있도록 하였다. 즉, V/S/TSIUVC의 정보에 의하여 유성음(V), 무음(S), TSIUVC에 적합한 신호처리 방법을 선택할 수 있도록 하였다.

2.2 주파수 영역의 선택

연속음성에서 탐색 추출한 TSIUVC를 재생하는데 유효한 주파수 영역을 선택하여 근사합성하는 방법(Approximate-Synthesis Method)을 그림 4에 제시하였다. 우선 TSIUVC 추출방법에 의하여 프레임을 재구성함과 동시에 얻어진 V/S/TSIUVC의 선택정보를 음성부호화 방식의 수신 측에 부호화하여 전송하게 된다. 여기에서 TSIUVC 재생에 유효한 주파수 영역을 알아보기 위하여 TSIUVC의 SNR및 스펙트럼을 분석할 필요가 있다. 그러기위해서는 TSIUVC 주파수 대역을 여러 주파수 영역으로 분할하여야 한다. 본 연구에서 제안한 방식을 실제 음성통신방식에 적용하였을 경우를 고려하여 3.4kHz의 LPF를 사용하였기 때문에 음성신호를 10kHz로 표본화한 경우에 주파수 간격이 Δf=39.0625Hz 이 된다. 따라서 최소 3개의 주파수를 사용하면 총 3.4kHz 주파수 대역은 29개의 주파수 영역으로 분할 할 수 있다. 분할된 각 주파수 영역의 신호를 사용하여 재생된 신호와의 SNR를 측정함으로서 TSIUVC의 근사합성에 유효한 주파수 영역을 선별할 수 있을 것으로 기대된다. 결국, TSIUVC 스펙트럼은 29개의 주파수 영역으로 나누어, 각 영역의 주파수 정보를 IFFT하여 재생한 신호와의 SNR를 측정하였다. 실험에 사용한 음성샘플은 남여 9명의 대화체 음성신호(73문장, 무성 자음수:195개)였으며, TSIUVC의 SNR 결과로서 무성자음 "p", "t", "k"의 SNR를 그림 5에 제시하였다. TSIUVC의 SNR을 분석한 결과에서 주목할 것은 0.547kHz 이하의 낮은 주파수 영역과

2.813kHz 이상의 높은 주파수 영역에서 상대적으로 높은 SNR를 얻을 수 있었다는 것이다. 실제로 0.547kHz 이하에서는 1.24~1.82dB, 2.813kHz 이상에서는 0.65~0.9dB를 얻을 수 있었다. 이것은 TSIUVC의 주요 주파수 정보가 높은 주파수와 중간 주파수 영역으로 양분되어 있는 것을 나타내는 것으로서, 천이구간(TS)와 무성자음(UVC)의 주파수 특성과도 부합하는 실험결과이다. 이러한 실험결과를 토대로 TSIUVC의 모든 주파수 정보를 사용하지 않고 특정한 부분의 주파수 정보만을 사용함으로서 한정된 정보량으로 TSIUVC를 근사합성할 수 있다. 이러한 방법에 의하여 0.547kHz 이하의 주파수 정보와 2.813kHz 이상의 주파수 정보를 사용하여 256 Point IFFT를 사용하여 TSIUVC를 근사합성 한다.

3. 멀티펄스의 탐색

멀티펄스 음성부호화 방식에서 유성음을 재생하기 위하여 멀티펄스를 사용하는데, 멀티펄스를 구하는 방법은 자기상관 함수와 상호상관 함수를 사용하여 멀티펄스의 진폭g_k과 위치 m_k를 탐색하는 방법에 의하여 구하였다[2]. 멀티펄스의 진폭g_k과 위치m_k에 의하여 명시된 멀티펄스의 음원 $v(n)$은 다음과 같이 나타낼 수 있다.

$$v(n)=\sum_{k=1}^{N} g_k \cdot \delta(n-m_k) \tag{2}$$

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

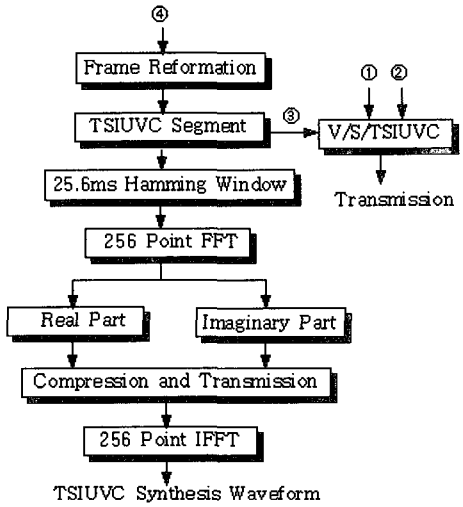
$v(n)$에 의한 합성신호는

$$y(n)=\sum_{k=1}^{N} g_k \cdot h(n-m_k) \tag{3}$$

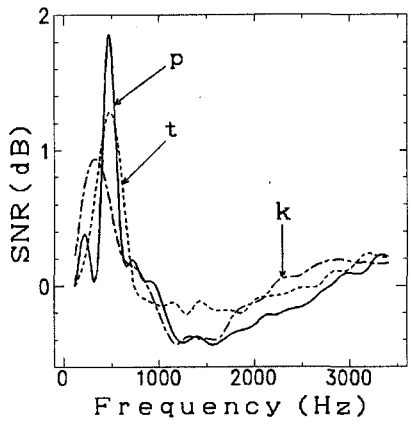
여기에서, $h(n)$은 합성필터의 임펄스응답으로 전달함수는 다음과 같다.

$$H(z)=\frac{1}{1-\sum_{i=1}^{M} a_i z^{-i}} \tag{4}$$

a_i : 선형예측계수, M: 필터차수



〈그림 4〉 주파수 영역의 선택에 의한 TSIUVC 근사합성법



〈그림 5〉 TSIUVC 주파수 영역의 SNR

멀티펄스의 진폭 및 위치는 다음 식의 오차가 최소가 되도록 결정한다.

$$E=\sum_{n=1}^{N}[x(n)-y(n)\otimes w(n)]^2 \quad (5)$$

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

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

$$w(n)=\frac{1-\sum_{i=1}^{M}a_i z^{-1}}{1-\sum_{i=1}^{M}a_i\rho^i z^{-i}} \quad (6)$$

ρ는 $0\leq\rho\leq 1$의 계수로서 멀티펄스 수와 SNR_{seg}의 관계를 고려하면 $\rho=0.8$이 적절한 것으로 밝혀졌다[2]. 식(6)에서 선형예측계수a_i는 식(7)가 최소가 되도록a_i에 대하여 편미분하여 구할 수 있다.

$$J=\sum_{n=1}^{N}e(n)^2=\sum_{n=1}^{N}[x(n)-y(n)]^2 \quad (7)$$

식(7)를 최소화하는 멀티펄스의 진폭g_k와 위치m_k는 다음과 같다.

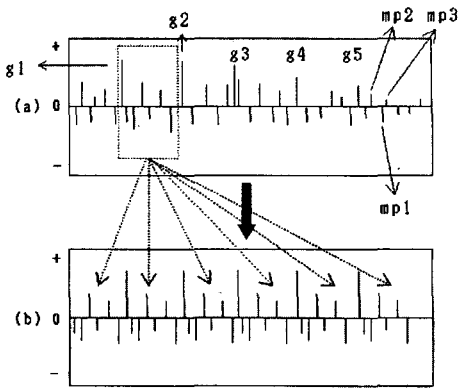
$$g_k=\max\left\{\Theta_{hx}(m_k-\sum_{j=1}^{k-1}g_j R_{hh}(|m_j-m_k|)\right\} \quad (8)$$

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

$$R_{hh}(m)=\sum_{n=1}^{N}h(n)-h(n-m),1\leq m\leq N \quad (9)$$

$$\Theta_{hx}(m)=\sum_{n=1}^{N}x(n)h(n-m),1\leq m\leq L \quad (10)$$

결국, 이러한 과정에 의하여 얻은 멀티펄스의 예를 그림 6에 나타내었다.



〈그림 6〉 멀티펄스의 진폭과 위치

그림 6(a)는 프레임 길이 25.6ms내에 진폭값이 큰 순서대로 순차적으로 구한 멀티펄스의 진폭(g_k=g1~g5)과 위치(m_k=mp1~mp3)를 나타낸 것이다. 이러한 멀티펄스열을 모두 전송하여 유성음원으로 사용하면 원래의 음성파형을 얻을 수 있을 것이다. 그러나 낮은 Bit Rate의 음성부호화 방식을 구현하기 위해서는 정보를 압축할 필요가 있다. 따라서 순차적으로 구한 멀티펄스열을 전부 송신 측에 전송하는 것이 아니라 개별피치구간내의 멀티펄스 만을 전송하고, 수신 측에서는 그림 6(b)와 같이 개별피치구간마다 멀티펄스를 재분배하여 음성합성필터를 구동할 수 있는 유성음원을 만든다. 따라서 개별 피치구간의 멀티펄스의 진폭과 위치, 개별 피치펄스의 위치정보는 반드시 수신 측에 전송하여야한다.

4. 실험결과

4.1 부호화 조건

V/UV의 선택정보를 이용하는 기존의 멀티펄스 음성부호화 방식(MPC)과 V/S/TSIUVC의 선택정보와 TSIUVC 근사합성 방법을 사용하는 새로운 멀티펄스 음성부호화 방식(FBD-MPC)의 부호화 조건을 표 1에 제시하였다. 부호화에 사

용한 음성신호는 3.4kHz LPF(Low Pass Filter)를 사용하여 주파수 대역제안을 하였으며, 프레임 길이는 FFT를 사용하는 것을 고려하여 25.6ms로 하였다. 이때 음성신호의 표본화 및 양자화는 각각 10kHz, 12bit로 하였다. 기존의 MPC 방식에서는 프레임마다 피치정보의 유무에 따라서 V/UV를 결정하고 유성음(V)인 경우에는 멀티펄스를 사용하고 무성음(UV)인 경우에는 white noise를 사용하였다. 반면 FBD-MPC 방식에서는 피치펄스와 ZCR을 이용하여 프레임을 재구성하고 V/S/TSIUVC의 선택정보에 의하여 각 음성신호의 특성에 적합한 방법을 선택하여 합성하게 된다. 즉 유성음(V)인 경우에는 멀티펄스의 유성음원을 사용하고, 무음(S)인 경우에는 신호가 없기 때문에 시간지연처리를 한다. 그리고 TSIUVC인 경우에는 0.547kHz이하 및 2.813kHz이상의 주파수 정보를 사용하여 신호를 근사합성하게 된다. 유성음원을 사용하여 구동할 합성필터가 필요한데, 본 연구에서는 합성 필터로서 PARCOR 필터를 사용하였다. PARCOR 필터의 차수는 10차를 사용하였는데, 낮은 차수일수록 스펙트럼에 미치는 영향이 크기 때문에 차등적인 bit를 할당하였다. 피치정보에 할당한 bit를 보면 MPC의 경우에는 평균 피치정보에 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 약2.7ms~12.5ms이 되기 때문에 25.6ms에 최대 9개의 피치가 존재하게 된다. 유성음(V)의 경우에 사용하는 멀티펄스는 총 10개이며, 그림 6에 나타낸바와 같이 피치간격마다 멀티펄스를 재생하여 사용하게 된다. 멀티펄스에 할당한 bit rate을 MPC의 멀티펄스 진폭(g_k) 및 위치(m_k)에 각각 2bit, 1bit 높게 할당하였으며, 상대적으로 진폭 값이 큰 멀티펄스의 최대 진폭($g_{\max}$)에는 6bit를 할당하였다. TSIUVC인 경우에 60Hz~0.547kHz와 2.813kHz~3.4kHz의 주파수 영역에 존재하는 총 28개의 주파수 정보에 6bit를 할당하였다. 따라서 MPC 및 FBD-MPC의 프레임마다 총178bit를 사용하였고 전송bit rate은 약6.9kbps가 된다.

〈표 1〉 부호화 조건

parameter[bit]	MPC	FBD-MPC
V/UV	2	
V/S/TSIUVC		3
[유성음 구간]		
$k_i(i=1\sim10)$	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)
피치정보	8	
P_0, I_{AV}		7,7
$DP_i,(i=2\sim9)$		2(3×8)
[TSIUVC구간]		
최대 진폭(Re&Im)		7
0.547kHz이하 주파수		6
2.813kHz이상 주파수		6
총 bit 수	178	178
kbps	6.9	6.9

*g_{max}:멀티펄스의 최대 진폭

4.2 음질평가

음성부호화 방식의 음질평가는 일반적으로 단문 20문항 내외의 문장을 사용하여 객관적인 평가척도인 SNR과 주관적인 평가척도인 MOS(Mean Opinion Score)를 동시에 수행한다. SNR은 재생음성의 스펙트럼의 일그러짐 정도를 평가할 수 있고, MOS는 청각적인 효과를

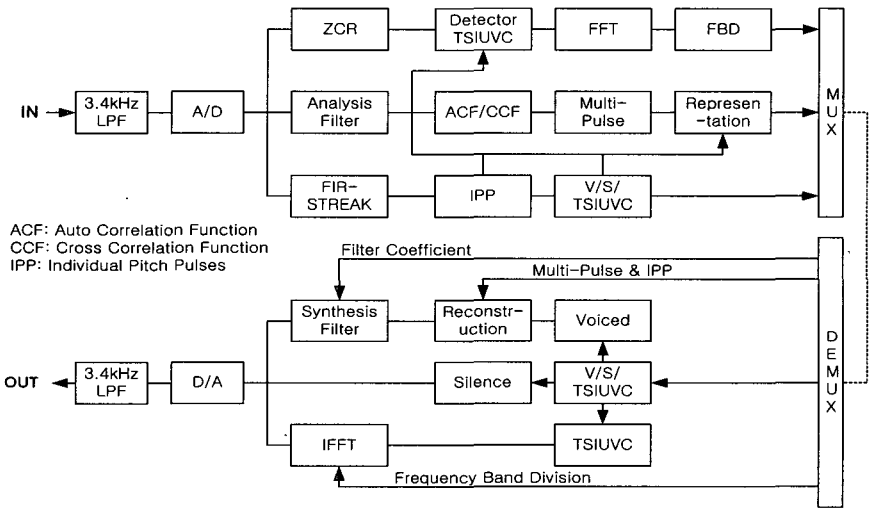
평가할 수 있다. 따라서 객관적이고 주관적인 평가실험을 동시에 수행함으로서 음질평가의 신뢰성을 높일 수 있다. MOS 평가는 5단계 MOS(-2~2점, 20명)를 사용하였으며, 상대평가를 위하여 MPC, FBD-MPC, 4bit~6bit log PCM을 사용하였다.

V/S/TSIUVC의 선택정보에 의하여 음성신호를 재생하는 6.9kbps의 FBD-MPC 방식의 블록도를 그림 7에 나타내었다. 간단히 설명하면 3.4kHz로 대역 제한된 음성신호는 10kHz, 12bit의 A/D변환한다. 변환된 디지털 신호는 ZCR과 FIR-STREAK 필터에서 구한 피치펄스(IPP)를 사용하여 연속음성에서 TSIUVC를 탐색 추출한다. 그리고 TSIUVC 재생에 유효한 주파수 영역의 신호만을 전송하고, 자기상관 함수(ACF)와 상호상관 함수(CCF)에 의하여 멀티펄스를 구하여 피치펄스구간의 멀티펄스만을 부호하여 전송한다. 수신측에서는 V/S/TSIUVC 선택정보에 따라서 유성음(V)일 경우에는 피치펄스구간마다 멀티펄스를 재생하여 합성필터를 구동하여 합성된 음성신호를 얻는다. 무음(S)인 경우에는 25.6ms 시간을 지연시킨다. TSIUVC인 경우에는 TSIUVC의 근사합성에 유효한 주파수 정보를 사용하여 재생하게 된다.

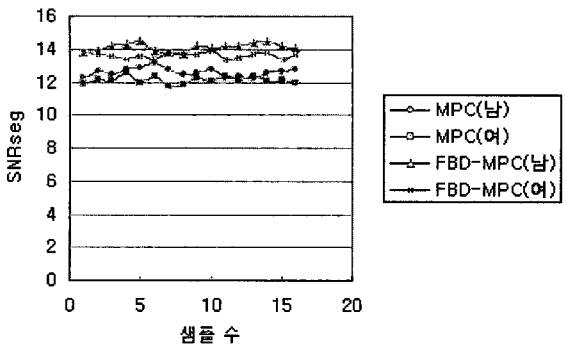
새롭게 제안한 6.9kbps의 FBD-MPC와 기존 MPC의 음질을 표 2의 음성표본을 사용하여 평가한 결과를 그림 8과 표 3, 표 4에 나타내었다. 실험결과, MPC의 경우에 남여 음성의 *SNR*은 각각 12.6dB와 12.1dB를 얻을 수 있었고, MOS에서는 각각 1.35와 1.37를 얻을 수 있었다. 한편, FBD-MPC의 경우에 남여 음성의 *SNR*은 각각 14.1dB와 13.6dB이고, MOS에서는 각각 1.99와 1.75를 얻었다. 결과적으로 FBD-MPC가 MPC에 비하여 *SNR*에서는 약 1.5dB정도, MOS에서는 남여 음성에서 각각 0.64와 0.38정도 개선된 것을 알 수 있었다.

〈표 2〉 음성 표본

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



〈그림 7〉 FBD-MPC 방식의 블록도



〈그림 8〉 MPC와 FBD-MPC의 *SNR*

〈표 3〉 MPC와 FBD-MPC의 *SNR*

Method [dB]	kbit/s	male	female
MPC	6.9	12.6	12.1
FBD-MPC	6.9	14.1	13.6

〈표 4〉 MPC와 FBD-MPC의 MOS

Method	kbit/s	male	female
MPC	6.9	1.35	1.37
FBD-MPC	6.9	1.99	1.75
4bit log PCM	40	1.08	1.09
5bit log PCM	50	1.82	1.83
6bit log PCM	60	2.88	2.90

5. 결 론

본 논문에서는 V/S/TSIUVC 선택정보를 이용한 멀티펄스 음성부호화 방식을 제안하였다. 특히, TSIUVC를 탐색 추출하여 프레임을 재구성함으로서 유성음과 무성자음이 같이 존재하는 프레임을 유성음원 혹은 무성음원 어느 한 쪽의 음원으로 처리하는 문제점을 해결 할 수 있었다. 즉, 프레임을 재구성함으로서 프레임내 음성신호의 특성에 맞는 신호처리방법을 적절히 선택할 수 있게 되었다. 특히 TSIUVC의 근사합성에 필요한 주파수 정보가 0.547kHz 이하의 낮은 주파수 영역과 2.813kHz 이상의 높은 주파수 영역에 분포하고 있다는 것을 알 수 있었다. 새롭게 제안한 V/S/TSIUVC에 의한 FBD-MPC 방식과 기존의 V/UV에 의한 MPC 방식의 음질을 평가한 결과, FBD-MPC가 MPC에 비하여 상대적으로 높은 *SNR*과 MOS값을 얻을 수 있었다. 이러한 연구결과를 토대로 새로운 형태의 음성부호화방식을 연구하는데 전념하고자 한다.

참 고 문 헌

[1] B.S.Atal and J.R.Remde: "A New Model of LPC Excitation for Producing Natural Sounding Speech at Low Bit Rates", ICASSP, p614-617, 1982

[2] Ozawa.K, Ono.S and Araseki.T: "A Study on Pulse Search Algorithms for Multipulse Excited Speech Coder Realization, IEEE, Vol. SAC-4, No.1, 1986

[3] Ergun Ercelebi:"Second Generation Wavelet Transform-based Pitch Period Estimation and Voiced/Unvoiced Decision for Speech Signals" Applied Acoustics, Vol 64, Issue 1, 2003

[4] CHONG KWAN UN and HYEONG HO LEE: "Voiced/Unvoiced/Silence Discrimination of Speech by Delta Modulation", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-28, No.4, 1980

[5] HIDEFUMI KOBATAKE: "Optimization of Voiced/Unvoiced Decisions in Nonstationary Noice Envirionments", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-35, No.1, 1987

[6] 武田 昌一他:"残差音源利用分析合成方式とマルチパルス法の基本特性の比較検討",電子

情報通信學會論文誌, Vol.J73-A,No.11,1990
[7] 眞野 淳,小澤 愼治:"LPC有聲音殘差のピッチ同期メルLSP分析合成方式",電子情報通信學會論文誌, Vol. J71-A, No.3,1988
[8] 武田 昌一他:"殘差音源利用分析合成方式とマルチパ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, 1990
[9] Georg Meyer and Robert Morse: "The Intelligibility of Consonants in Noisy Vowel-Consonant-Vowel Sequences when the Vowels are Selectively Enhanced", Speech Communication, Vol 41, Issue2-3, 2003
[10] Jeremiah Remus and Leslie M. Collins: "Predicting Vowel and Consonant Confusions using Signal Processing Techniques", International Congress Series, Vol 1273, 2004
[11] 이시우: "FIR-STREAK 디지털 필터를 사용한 피치추출 방법에 관한 연구", 한국정보처리학회 논문지, 제6권 제1호, p. 247-252, 1999
[12] Yusuke Hiwasaki, Kazunori Mano and Takao Kaneko: "An LPC vocoder based on phase-equalized pitch waveform" Speech Communication, Vol 40, Issue 3, 2003

◐ 저 자 소 개 ◑

이 시 우(See-Woo Lee)
1987년 동국대학교 전자공학과(학사)
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, 감성처리, 유무선통신
E-mail : swlee@smu.ac.kr