한국산학기술학회논문지
Vol. 7, No. 6, pp. 1180-1184, 2006

V/S/TSIUVC 스위칭을 이용한 음성부호화 방식에 관한 연구

이시우[1*]

A study on Speech Coding Method Using V/S/TSIUVC Switching

See-Woo Lee[1*]

요 약 유성음원과 무성음원을 사용하는 음성부호화 방식에 있어서 모음과 무성자음이 있는 프레임에서 음질저하 현상이 나타난다. 본 논문에서는 음질을 개선하기 위해 V/S/TSIUVC 스위칭과 TSIUVC 근사합성 방법을 사용한 새로운 멀티펄스 음성부호화 방식을 제시한다. TSIUVC는 영교차율과 개별피치 펄스에 의하여 추출되며, TSIUVC의 추출율은 여자와 남자음성에서 각각 91%와 96.2%를 얻었다. 여기에서 중요한 사실은 양질의 TSIUVC 합성 파형을 얻기 위해서는 0.547kHz 이하와 2.813kHz 이상의 주파수 정보를 사용하여야 한다. V/UV를 이용한 MPC와 V/S/TSIUVC를 이용한 FBD-MPC의 비교평가를 하였다. 실험결과, FBD-MPC의 음질이 MPC의 음질에 비하여 상당히 개선되었음을 알 수 있었다.

Abstract In a speech coding system using excitation source of voiced and unvoiced, it would be a distortion of speech quality in a voiced and an unvoiced consonants in a frame. In this paper, I propose a new multi-pulse coding method make use of V/S/TSIUVC switching and TSIUVC approximation-synthesis method in order to restrict a distortion of speech quality. The TSIUVC is extracted by using the zero crossing rate and individual pitch pulse. And the TSIUVC extraction rate was 91% for female voice and 96.2% for male voice. The important thing is that the frequency information of 0.547kHz below and 2.813kHz above can be made with high quality synthesis waveform within TSIUVC. I evaluated the MPC of V/UV and FBD-MPC of V/S/TSIUVC. As a result, the synthesis speech of FBD-MPC was better in speech quality than the MPC.

Key words : Multi-Pulse Speech Coding, 멀티펄스 음성부호화, Speech Signal Processing, 음성신호처리

1. 서 론

Atal은 낮은 bit rate의 음성부호화 방식으로 AbS(Analysis by Synthesis)법에 의하여 멀티펄스(Multi-Pulse)를 탐색하고, 이 멀티펄스에 의하여 합성 필터를 구동함으로서 음성신호를 합성하는 방식을 제안하였다[1]. 이를 Putnins는 9.6kbit/s의 bit rate에서 멀티펄스 음성부호화 방식의 음질을 개선하였다[2]. 그러나 이 방식에서는 bit rate을 더욱 낮추기 위하여 멀티펄스의 수를 감소시켜야 하는 문제점을 가지고 있다. 이러한 문제점을 Ozawa는 멀티펄스의 수를 감소시키지 않고 음질을 개선하기 위한 방법으로서 피치예측법과 피치보간법을 이용하여 4.8~9.6kbit/s의 멀티펄스 음성부호화 방식(MPC)을 제안하였다[3]. 이러한 방식에서는 자기상관법에 의하여 추출한 피치정보에 의하여 유성음/무성음(V/UV)를 선택하고, V/UV에 의하여 유성음원과 무성음원을 선택하여 음성신호를 재생하기 때문에 유성음원 혹은 무성음원 어느 한쪽의 음원을 선택하여 음성신호를 재생하게 된다. 즉, 주기적인 특성의 유성음과 비주기적인 무성자음을 유성음원과 무성음원으로 처리하게 되는데 유성음과 무성음의 중간특성을 갖는 천이구간을 유성음원과 무성음원의 어느 한쪽의 음원을 사용하여 재생함으로서 음질저하의 원인으로 작용한다.

이러한 문제점을 해결하기 위하여 본 논문에서는 개별피치 추출법을 사용하고, 특성을 달리하는 유성음(V), 무음(S), TSIUVC(Transition Segment Includ ing UnVoiced Consonant)의 음성신호를 V/S/TSIU VC의 선택정보에 의하여 음성신호를 재생하는 새로운 멀티펄스 음성부호화 방식(FBD-MPC: Frequency Band Division Multi Pulse Coding)을 제안하고자 한다.

본 연구는 상명대학교 2006년도 교내연구비 지원에 의하여 연구됨

[1]상명대학교 정보통신공학과
*교신저자: 이시우(swlee@smu.ac.kr)

2. 개별 피치 추출

2.1 개별 피치 추출 알고리즘

일반적으로 피치추출에 자주 이용되는 자기상관(Auto-Correlation)법[4]은 수십ms 프레임 단위로 정규화한 하나의 피치정보를 산출하며 음성의 시작이나 끝부분, 무성음과 유성음 혹은 무성자음과 유성음이 같이 존재하는 프레임에서는 피치추출 오류가 종종 발생한다. 이러한 오류를 억제할 수 있는 방법으로 본 연구에서는 그림 1의 FIR-STREAK 필터의 잔차신호에서 개별 피치 펄스를 추출하는 방법을 적용하고자 한다. 실제의 음성신호는 5kHz 주파수 대역에 약 3~4개의 포어먼트 정보가 존재하며 일반적으로 10차의 필터를 사용한다. 따라서 10kHz로 표본화한 음성신호를 10차의 STREAK 필터를 사용하였다.

STREAK 필터에서 k_i를 구하기 위해서는 전 방향 오차신호($f_i(n)$)와 후 방향 오차신호($g_i(n)$)의 순시값을 최소화 한 다음

$$\begin{aligned} A_s &= f_i(n)^2 + g_i(n)^2 \\ &= -4\, k_i \cdot f_{i-1}(n) \cdot g_{i-1}(n-1) \\ &\quad +(1+k_i^2)\cdot(f_{i-1}(n)^2 + g_{i-1}(n-1)^2) \end{aligned} \quad (1)$$

윗식을 k_i에 관하여 편미분함으로서 STREAK계수 k_i를 구할 수 있다.

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

여기에서, i=1,2,...,M 이고, n=1,2,...,N 이다.

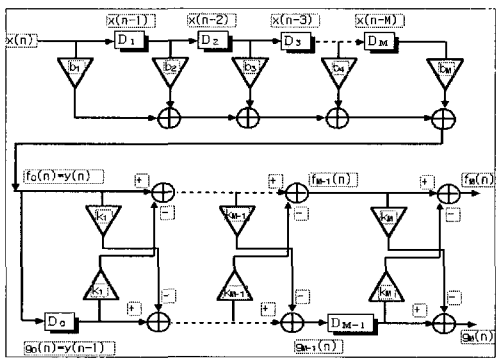


그림 1. FIR-STREAK 필터

k_i를 사용한 STREAK 필터의 전달함수는 다음과 같다.

$$H_s(z) = \frac{1}{\sum_{i=0}^{M_s} k_i\, z^{-i}} \quad (3)$$

프레임 길이가 25.6ms인 두 프레임의 연속된 음성파형을 사용하여 피치를 추출한 결과를 그림 2에 나타내었다. 실험결과, 개별피치 추출법에서는 유효한 피치정보를 추출할 수 있었던 반면, 자기상관법과 Cepstrum법에서는 무성음과 유성음, 혹은 무성자음과 유성음이 같이 존재하는 부분, 음소가 변위하는 부분, 프레임의 경계 부분, 음성의 시작 부분, 음성의 끝 부분에서 피치추출 오류를 볼 수 있다.

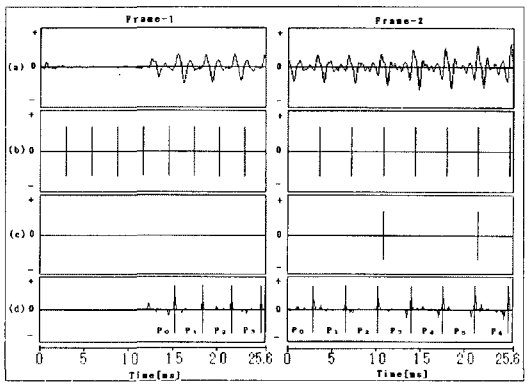


그림 2. 피치추출 (a)원 파형 (b)자기상관 (c)Cepstrum (d) 개별피치 펄스

2.2 개별 피치 추출률

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

$$P_R = \frac{\sum_{i=1}^{T}\sum_{j=1}^{m}[\, a_{ij} - (\, b_{ij} + c_{ij})]}{\sum_{i=1}^{T}\sum_{j=1}^{m} a_{ij}} \quad (4)$$

윗 식에서, m, T, a_{ij}는 각각 프레임 총수, 총 음성제원 수, 관찰된 피치 수를 나타낸다. 표 1의 음성표본과 식(4)를 사용하여 피치 추출률을 산출한 결과, 표 2와 같은 결과를 얻었다. 여기에서, 피치 추출률이

여자음성에서 낮게 산출된 까닭은 여자음성이 남자음성에 비하여 피치주파수가 급격히 변하는 특성 때문으로 판단된다.

표 1. 음성샘플

제 원	남자음성	여자음성
발성자	4	4
발성 시간	54.4초	54.4초
단문 수	16	16
모음 수	145	145
무성자음 수	34	34

표 2. 피치추출율

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

3. V/S/TSIUVC 추출과 근사합성

3.1 V/S/TSIUVC 추출

일반적으로 유성음(V)에서는 낮은 영교차율(ZCR: Zero Crossing Rate)과 피치정보를 갖는 것을 특징으로 하며, 무성자음(UVC)에서는 높은 ZCR과 피치정보가 없는 것이 특징이다. 또한, 천이구간(TS)에서는 낮은 영교차율과 피치정보가 없는 특징을 나타낸다. 이러한 특징들을 고려하여 연속음성에서 유성음(V), 무음(S), TSIUVC를 탐색·추출하는 방법을 그림 3에 나타냈다. 이 방법에 있어서 음성신호는 3.4kHz LPF로 주파수 대역을 제한한 다음 10kHz, 12bit로 표본화 및 양자화고, FFT 처리를 위하여 프레임 길이는 25.6ms로 하였다.

프레임 안에 개별 피치정보가 하나도 존재하지 않으면(PF[t]=0) 프레임을 무음(S)로 판정하였고, 그렇지 않다면 해당 프레임의 영교차율($Z[t]$)과 프레임간의 ZCR($\Delta Z[t]=Z[t]-Z[t-1]$)차, 천이구간(TS)과 무성자음구간(UVC)의 영교차율($ZH[t]$)이 $\Delta Z[t]<0$, $Z[t-1]\geq 0.4$, $0.4\leq ZH[t]\leq 0.7$인 조건을 만족한 경우에 최초에 나타나는 개별피치(P_0)의 위치에서 25.6ms 이전의 음성신호를 TSIUVC로 판정하였고, 그렇지 않다면 유성음(V)로 판정하였다.

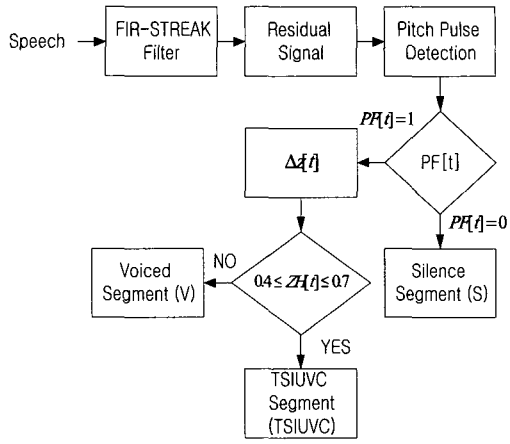


그림 3. TSIUVC 추출법

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 TSIUVC 추출오류로 규정한 식(5)에 의하여 TSIUVC 추출률을 산출하였다.

$$\Re = \sum_{j=1}^{m}\{a_j-(b_j+c_j)\} / \sum_{j=1}^{m} a_j \qquad (5)$$

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 TSIUVC 근사합성

남여 9명의 대화체 음성(73문장, 무성자음 수:195개)신호를 사용하여 추출한 TSIUVC의 SNR를 분석한 결과의 한 예로 무성자음 "p", "t", "k"의 SNR를 그림 4에 나타냈다. 여기에서 주목할 것은 0.547kHz 이하의 낮은 주파수 대역과 2.813kHz 이상의 높은 주파수 대역에서 상대적으로 높은 SNR를 나타내고 있는데, 이것은 TSIUVC의 주요 주파수 정보가 높은 주파수와 중간 주파수대역으로 양분되어 있는 것을 나타낸다. 따라서 본 연구에서는 TSIUVC를 재생에 0.547kHz 이하와 2.813kHz 이상의 주파수 정보를 사용하기로 하였다.

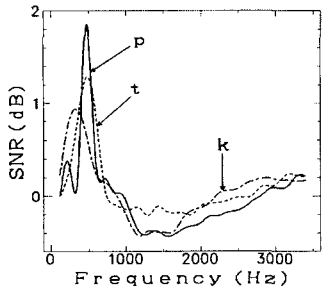


그림 4. TSIUVC의 SNR

3.3 멀티펄스 탐색

본래의 음성신호$x(n)$와 멀티펄스$v(n)$에 의하여 재생된 신호$y(n)$로부터 식(6)이 최소가 되도록 멀티펄스의 진폭(g_k)과 위치(m_k)를 결정하게 된다[5].

$$J=\sum_{n=1}^{N} e(n)^2=\sum_{n=1}^{N}[x(n)-y(n)]^2 \qquad (6)$$

이때, 멀티펄스의 음원 $v(n)$은 다음과 같이 나타낼 수 있다.

$$v(n)=\sum_{k=1}^{N} g_k\cdot\delta(n-m_k) \qquad (7)$$
$$\{\text{if } n=m_k, \delta(n-m_k)=g_k\}$$
$$\{\text{if } n\neq m_k, \delta(n-m_k)=0\}$$

결국, 그림 5와 같이 대표구간의 멀티펄스와 피치정보를 수신측에 전송하고, 수신측에서는 피치구간마다 대표구간의 멀티펄스를 재생하여 멀티펄스 음원을 만든다.

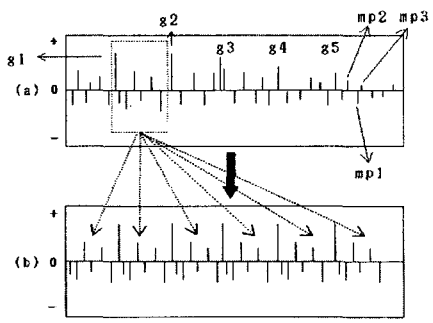


그림 5. 멀티펄스

4. 실험결과

유성음/무성음(V/UV) 형태의 MPC와 V/S/TSIUVC 형태의 FBD-MPC의 음질을 비교 평가할 때 같은 비트 율(bit rate)이 되도록 전송 파라미터에 할당하는 비트(bit)를 조절할 필요가 있고, 음질의 주관적 평가와 객관적 평가가 동시에 이루어져야 한다. 주관적 평가와 객관적 평가의 척도로 자주 사용되는 방법으로 MOS(Mean Opinion Score)와 SNR_{seg}가 있다. 여기에서, MOS는 청각적인 음질을 나타내는 척도이고, SNR_{seg}는 음성 파형의 일그러짐을 나타내는 척도라 할 수 있다. 표 3과 같이 MPC, FBD-MPC 모두 10개의 멀티펄스를 사용하였으며 MPC와 FBD-MPC의 bit rate을 같게 하기 위하여 MPC의 멀티펄스 진폭 및 위치에 각각 2bit, 1bit 높게 할당하였다. 또한 상대적으로 진폭 값이 큰 멀티펄스의 최대 진폭에는 6bit를 할당하였다. V/S/TSIUVC 형태의 FBD-MPC 블록도를 그림 6에 나타내었다. 표 1의 음성표본을 사용하여 MPC와 FBD-MPC에 대하여 SNR_{seg}와 MOS의 음질 평가를 하였다. 그림 8은 MPC와 FBD-MPC의 SNR_{seg} 분포를 나타낸 것인데, 표 4에 나타낸바와 같이 FBD-MPC의 SNR_{seg}이 MPC에 비하여 남자 음성에서 1.5dB, 여자 음성에서 1.5dB정도 개선되었다.

[표 3] 음성부호화

parameter[bit]	MPC	FBD-MPC
V/UV	2	
V/S/TSIUVC		2
[유성음 구간]		
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}(멀티펄스의 최대 진폭)	6	6
g_k(멀티펄스의 진폭)	6	4
m_k(멀티펄스의 위치)	6	5
멀티펄스 수	10	10
평균 피치정보	8	
P_0(최초 개별피치의 위치)		7
I_{AV}(개별피치 간격의 평균)		7
$DP_i,(i=2\sim9)$ (개별피치 간격의 편차)		24 (3×8)
[TSIUVC 구간] Re&Im		
최대 진폭		7
저역의 주파수 신호		3
고역의 주파수 신호		3
총 bit 수	178	178
kbit/s	6.9	6.9

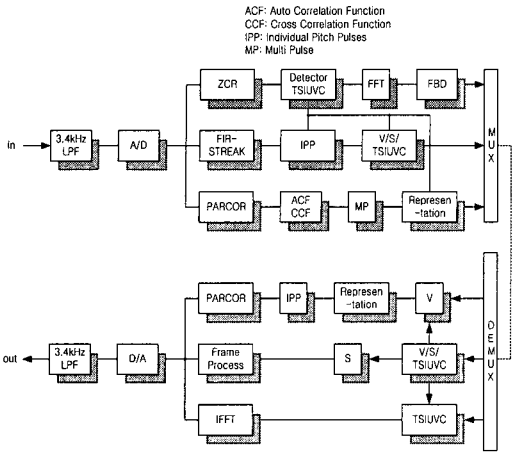


그림 6. FBD-MPC 블록도

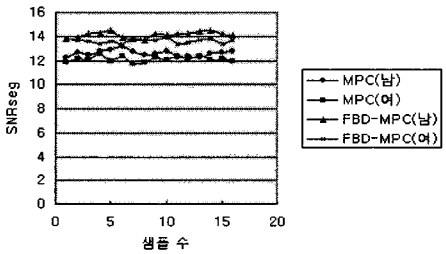


그림 7. MPC와 FBD-MPC의 SNR_{seg}

[표 4] MPC와 FBD-MPC의 SNR_{seg}

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

표 5. 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

한편, 청각적 실험인 MOS 평가에서는 MPC, FBD-MPC, 4~6bit log PCM 방식을 비교 평가한 결과, 표 5에 나타낸바와 같이 FBD-MPC가 MPC에 비하여 남자 음성에서 0.64, 여자 음성에서 0.38 정도의 음질이 개선되었음을 알 수 있었다.

5. 결 론

유성음/무성음(V/UV) 형태의 MPC에 있어서, 유성음과 무성자음이 같이 존재하는 연속된 프레임의 음성신호를 유성음(V) 혹은 무성음(UV)의 음원으로 처리하기 곤란하다. 이러한 문제점을 해결하기 위하여 본 논문에서는 FBD-MPC를 제안하였다. 실험결과, SNR_{seg}와 MOS에서 기존의 방식에 비하여 파형의 일그러짐과 청각적인 음질이 개선되었음을 확인하였다.

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 and R.D. Trupp: "A Multi-Pulse LPC Synthesizer for Telecommun -ications use", IEEE, ICASSP, Mar, 1985

[3] 小澤 一範, 蒜關 卓: "ピッチ情報を用いる 9.6~4.8kbit/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

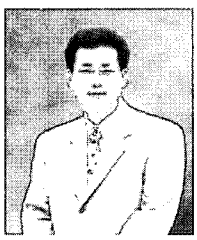
[4] 藤井健作:" 自己相關法による電話帶域音聲の ピッチ抽出法"電子情報通信學會技術報告書,sp87-65, 1987.

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

[6] 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

[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, 1978

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 정보통신학부 교수

<관심분야>
음성신호처리, 음주판독시스템, 유무선통신