

연속음성신호에서 피치와 TSIUVC 추출에 관한 연구

A Study on Extraction of Pitch and TSIUVC in Continuous Speech

이 시 우*

See-Woo Lee

요 약

본 연구에서는 연속음성에서 개별 피치펄스와 TSIUVC를 추출하는 새로운 방법을 제안하고자 한다. TSIUVC 탐색과 추출은 FIR-STREAK 필터를 사용한 개별 피치펄스와 영교차율을 사용한다. 실험결과, 개별 피치펄스의 추출률은 남자음성에서 96%, 여자음성에서 85%를 얻을 수 있었다. 아울러, TSIUVC 추출률은 남자 음성의 경우 88%에서 94.9%, 여자 음성의 경우는 84.8%에서 94.9%의 결과를 얻었다. 제안한 방법은 음성분석, 음성합성, 새로운 Voiced/Silence/TSIUVC의 음성부호화 방식에 활용할 수 있을 것으로 기대된다.

Abstract

In this paper, I propose a new extraction method of Individual Pitch Pulse and TSIUVC in continuous speech. The TSIUVC searching and extraction method is based on a zero-crossing rate and Individual Pitch Pulse extraction method using FIR-STREAK filter. As a result, the extraction rate of individual pitch pulses was 96% for male voice and 85% for female voice respectively. The TSIUVC extraction rates are 94.9% under 88% for male voice and 94.9% under 84.8% for female voice. This method has the capability of being applied to a new speech coding of Voiced/Silence/TSIUVC, speech analysis and speech synthesis.

↳ Keyword : speech signal processing, pitch, TSIUVC

1. 서 론

음원을 사용하는 낮은 bit rate의 음성부호화 방식은 음성신호를 압축복원하기 위하여 음성특징 파라미터인 피치(Pitch)와 유성음원과 무성음원의 이원화된 음원을 사용한다. 이러한 방식은 일반적으로 음성신호를 수십ms의 프레임으로 분할하여 분석하는데, 이때 같은 프레임 안에 유성음과 무성자음의 음성신호가 존재할 수 있다. 이러한 경우에 이원화된 음원을 사용하는 음성부호화 방식에서는 유성음원 혹은 무성음원 어느 한쪽의 음원을 선택하여 음성신호를 재생시키는 문제점이 있다[1]. 또한 무성자음에서 유성음으로 변위하는 과정에서 나타나는 천이구간(TSIUVC: Transition Segment

Including UnVoiced Consonant)은 유성음과 무성자음의 중간특성을 갖고 있기 때문에 유성음원 혹은 무성음원으로 재생하는데 한계가 있다.

음성신호에서 피치를 추출하는 방법은 시간영역, 주파수영역, 시간과 주파수영역을 혼용하여 추출하는 방법으로 나뉠 수 있다. 시간영역에서 추출하는 방법으로 자기상관법[2], 주파수 영역에서 추출하는 방법으로 ALNF를 이용한 방법[3]과 캡스트럼[4], 시간과 주파수영역에서 피치를 추출하는 방법으로 AMDF(Average Magnitude Difference Function)법[5]과 선형예측을 이용한 LPC(Linear Prediction Coding) 역 필터 추정법과 AMDF를 혼합한 방법 등이 있다[6,7]. 이러한 방법들은 음소 상호간섭이 있는 부분이나 음성의 시작이나 끝부분, 무성음과 유성음 혹은 무성자음과 유성음이 같이 존재하는 프레임에서는 피치추출 오류가 종종 발생한다.

* 정 회 원 : 상명대학교 정보통신공학과 교수
swlee@smu.ac.kr(제 1저자)

[2004/04/26 투고 - 2005/03/14 심사 - 2005/05/27 심사완료]

이러한 문제점을 해결하기 위하여 본 연구에서는 피치주파수 영역에 있는 낮은 주파수를 얻기 위하여 LPF(Low Pass Filter) 역할을 하는 FIR (Finite Impulse Response) 필터와 오차신호를 얻기 위한 STREAK (Simplified Technique for Recursively Estimating Autocorrelation K-parameters) 필터를 종속으로 결합한 형태인 새로운 필터(이후 FIR-STREAK 필터로 명명함)를 고안하여 사용하였고, 이 FIR-STREAK 필터의 출력 오차신호로부터 연속적으로 변위하는 피치위치를 추출할 수 있는 개별 피치펄스 추출법과, 개별 피치펄스와 영교차율을 이용하여 유성음과 무성자음이 같은 프레임 안에 존재하지 않도록 유성음, 무성음, TSIUVC를 탐색추출하여 프레임을 재구성하는 방법을 새롭게 제안하고자 한다.

II. 피치 추출

1. 자기상관법과 캡스트럼

본 논문에서 제안한 개별 피치펄스 추출법과 비교평가하기 위하여 프레임마다 평균 피치 정보를 추출하는 일반적인 방법인 자기 상관법과 캡스트럼에 대하여 간략히 기술하고자 한다. 자기상관법은 식 (1)에 나타낸바와 같이 자기상관 계수 $R(t)$ 가 1에 근접한 시점을 피치의 개사종료 지점으로 규정하여 피치주기를 구하게 된다.

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

캡스트럼은 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 나타낸다.

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

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

2. 개별 피치펄스 추출법

프레임 단위로 하나의 규격화된 피치정보를 추출하는 방법은 음성의 시작이나 끝부분과 같이 음소 상호간의 간섭에 의해 피치간격이 일정하지 않거나 무성음과 유성음 혹은 무성자음과 유성음이 같이 존재하는 프레임에서 피치추출 오류가 나타나는데, 이러한 피치추출 오류를 억제하는 방법을 그림 1에 나타내었다. 여기에서 FIR 필터는 주파수 대역을 제한하기 위하여 사용하였으며, STREAK 필터는 오차신호를 출력하는 역할을 한다.

STREAK 필터는 식 (3)에 나타낸바와 같이 전방향 오차신호($f_i(n)$)와 후방향 오차신호($g_i(n)$)의 순시값을 최소화 한다.

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

STREAK계수 k_i 는 식 (3)을 편미분함으로서 구할 수 있으며, 여기에서 $i=1,2,\dots,M$ 이고, $n=1,2,\dots,N$ 이다.

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

Z Transform 영역에서 k_i 를 사용하여 STREAK 필터의 전달함수를 표현하면 다음과 같다.

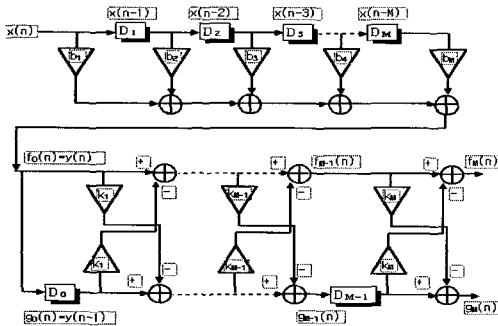
$$H_s(z) = \frac{1}{\sum_{i=0}^{M_i} k_i z^{-i}} \quad (5)$$

FIR-STREAK 필터의 오차신호($E_{PN}(n)$)는 그

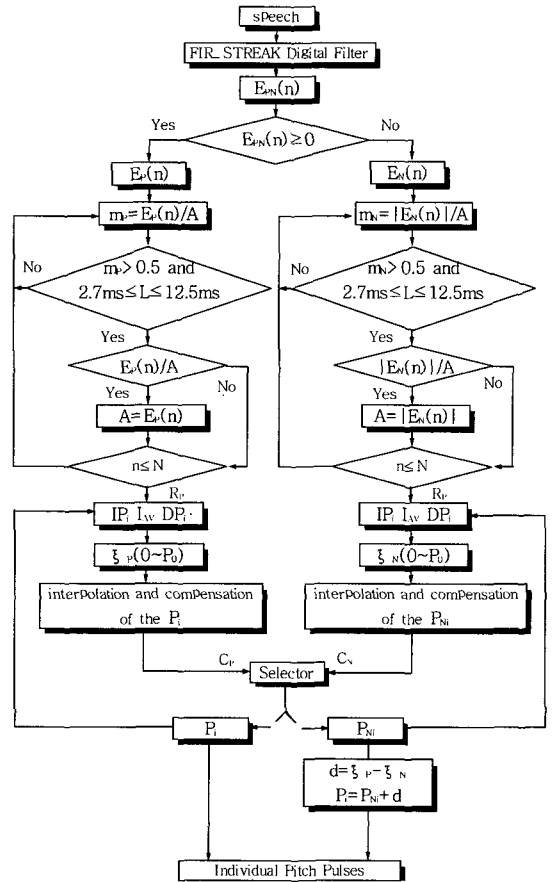
림 2와 같이 시간영역의 (+)측 오차신호($E_p(n)$)와 (-)측 오차신호($E_N(n)$)로 구성되어 있으며, 그림 2의 결과에서 알 수 있듯이 오차신호에는 펄스성 오차신호(R_p)와 잡음성 오차신호가 존재하며, R_p 는 피치주기의 위치에서 발생하고 있는 것을 확인할 수 있다. 따라서 본 논문에서는 잡음성 오차신호에서 R_p 의 위치를 검출하여 개별 피치펄스의 위치를 추정하는 방법을 그림 3에 나타내었다. $E_{PN}(n)$ 은 $E_{PN}(n) \geq 0$ 인 경우에 $E_p(n)$ 으로, $E_{PN}(n) < 0$ 인 경우에는 $E_N(n)$ 으로 하였다. 이때 $E_p(n)$ 과 $E_N(n)$ 을 병렬로 처리하여 검출한 R_p 로부터 개별 피치펄스를 구하는데, $E_p(n), E_N(n), E_p(n)$ 에서 개별 피치펄스를 검출하는 방법은 우선, $E_p(n)$ 를 A 로 정규화한 값이 $m_p > 0.5$ 를 만족하는 동시에 피치의 간격(L)이 피치주파수 범위(80Hz~370Hz)인 $2.7ms \leq L \leq 12.5ms$ 를 만족하는 다음으로 R_p 간의 간격 $IP_i (IP_i = P_i - P_{i-1})$, 평균간격 $I_{AV} (I_{AV} = (P_M - P_0)/M)$ 의 편차 $DP_i (DP_i = I_{AV} - IP_i)$ 를 구하고, $0.5I_{AV} \geq IP_i$ 를 만족하는 경우는 식 (6)으로 R_p 위치를 수정하고, $0.5I_{AV} \geq IP_i$ 및 $|DP_i| > 2.7$ 를 만족하는 경우는 식 (6)으로 R_p 위치를 보완 하도록 하였다.

$$P_i = (P_{i-1} + P_{i+1})/2 \quad (6)$$

일반적으로 유성음은 수십ms 동안 피치주기의

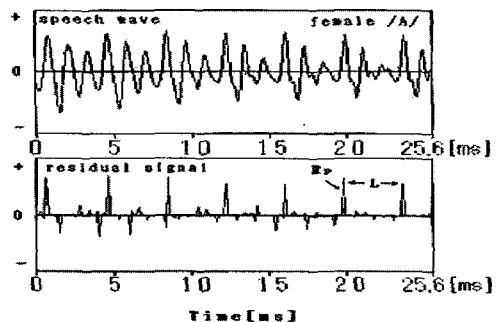


〈그림 1〉 FIR-STREAK 필터의 구성



〈그림 3〉 개별 피치 펄스 추출법

변화가 적기 때문에 식 (7)을 만족하는 경우는 시간영역 (+)측의 P_i 를, 그렇지 않은 경우는 (-)측의 P_i 를 선택하도록 하였다.

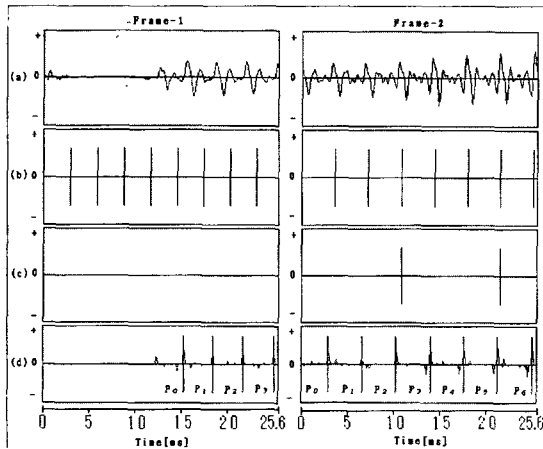


〈그림 2〉 FIR-STREAK 필터의 오차신호

$$\sum_{i=1}^M \frac{IP_i}{I_{AV}} \leq \sum_{i=1}^M \frac{IP'_i}{I'_{AV}} \quad (7)$$

본 논문에서 제안한 방법은 프레임의 경계부분, 무성음과 유성음, 혹은 무성자음과 유성음이 같이 존재하는 부분, 음소가 변위하는 부분, 프레임의 경계 부분, 음성의 시작 부분, 음성의 끝 부분에서 피치의 위치를 보다 적절히 나타낼 수 있는데, 이러한 결과로서 음성의 시작위치에서 무성음과 유성음이 같이 존재하는 두 프레임에서 피치를 추출한 결과를 그림 4에 나타내었다.

그림 4에 나타낸바와 같이 프레임 길이가 25.6ms 인 두 프레임의 연속된 음성파형에서 캡스트럼에서는 Frame-1의 무성음부에서 피치추출 오류를 볼 수 있으며, 자기상관법에서는 Frame-1에서 피치가 추출되어 있지 않으며, Frame-2에서도 피치추출오류가 나타나 있는 것을 알 수 있다. 반면 개별 피치펄스 추출법에서는 R_p 를 검출하여 피치의 위치를 추정함으로써 무성음과 유성음이 같이 존재하는 Frame-1이나 유성음인 Frame-2에서 피치의 위치를 정확하게 추출할 수 있었다. 아울러, Frame-1과 Frame-2 경계부의 피치간격이 적정한 피치간격임을 확인할 수 있다.



〈그림 4〉 피치추출 (a) 원음성 (b) 캡스트럼 (c) 자기상관법 (d) 개별 피치펄스 추출법

III. TSIUVC 추출

1. TSIUVC 추출

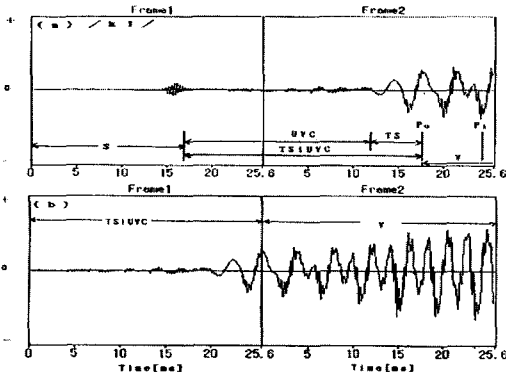
유성음의 경우에는 낮은 영교차율과 피치정보를 갖고 있으며, 무성자음의 경우는 피치정보가 없으며 높은 영교차율(식 (8))을 갖는다. 또한 천이구간의 경우에는 피치정보가 없으며 낮은 영교차율을 나타내는 것이 특징이라 할 수 있다. 아울러, 연속 음성에서 유성음의 지속시간은 100ms~500ms 정도이며 약 2.7ms~12.5ms 간격마다 유사한 음성파형이 주기적으로 반복되며, 무성자음의 경우는 대개 20ms 전후이고, 천이구간의 경우는 약 5ms 전후인 지속시간을 갖는 것으로 알려져 있다. 천이구간은 발생한 음성신호가 주기성을 갖는 유성음의 파형이 나타나기 전의 준주기적인 음성파형으로서 천이구간의 시간적 정보가 음성인지에 중요한 역할을 한다는 지식에 기반을 두고 천이구간을 탐색하는 방법으로 영교차율, 단구간 에너지, LPC 캡스트럼 거리를 이용하는 방법이 제안되었다 [8,9]. 이러한 방법에서는 음성신호의 시작과 끝점을 추출하거나 천이구간을 포함하는 구간을 대략적으로 탐색할 수 있으나 천이구간과 무성자음 구간을 정확히 추출하는데 한계점이 있다. 이러한 한계점의 또 다른 이유로 천이구간의 끝점이 유성음의 시작위치를 고려할 때 유성음의 시작위치를 알아내기 위해서는 유성음과 무성자음 혹은 무성자음을 포함한 천이구간을 분리한 후 유성음의 시작위치를 탐색하여야 하는데, 프레임 단위로 한 개의 정규화된 피치정보를 추출하는 기존의 방법 [2,4-7]으로 유성음의 시작위치를 추정하는 것은 사실적으로 불가능하다.

$$Z[t] = \frac{1}{2 \cdot N} \sum_{n=1}^N | \text{sgn}[x(n)] - \text{sgn}[x(n-1)] | \quad (8)$$

if $x(n) \geq 0, \text{sgn}[x(n)] = 1,$

else if $\text{sgn}[x(n)] = -1,$

t : 프레임 번호

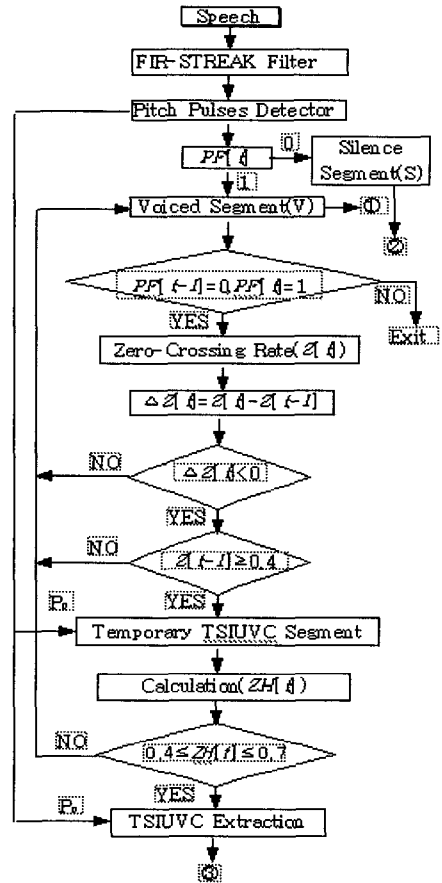


〈그림 6〉 유성음과 TSIUVC가 있는 프레임의 재구성
(a) 본래의 프레임 (b) 재구성한 프레임

그러나 본 논문에서 제안한 개별 피치펄스 추출법에서는 유성음만 존재하는 프레임이나 혹은 같은 프레임 안에 유성음과 무성자음이 존재하더라도 최초로 나타나는 개별 피치펄스의 위치로부터 유성음의 시작위치를 알 수 있다. 즉 최초로 나타나는 개별 피치펄스의 위치가 천이구간의 끝점인 동시에 유성음의 시작위치이기 때문에 개별 피치펄스의 위치로부터 천이구간과 유성음구간을 용이하게 추출할 수 있다.

개별 피치정보와 영교차율을 이용하여 연속음성에서 TSIUVC를 탐색추출하는 방법을 그림 5에 나타냈다. 그림 5의 방법을 간단히 설명하자면 FIR-STREAK 필터의 오차신호에서 추출한 개별 피치펄스가 프레임 안에 한 개라도 존재하면 프레임 안의 음성신호를 유성음($V:PF[t]=1$)으로 간주하고, 그렇지 않은 경우에는 무음($S:PF[t]=0$)으로 간주하였다. 그리고 유성음으로 판단된 프레임과 다음 프레임간의 영교차율($Z[t]$)의 차($\Delta Z[t]$)가 $\Delta Z[t] < 0$ 인 경우에는 프레임에 TSIUVC가 존재하는 것으로 간주하였다.

다음으로, 유성음의 시작위치가 결국 TSIUVC가 끝나는 위치이기 때문에 최초의 개별 피치펄스의 위치(P_0)를 유성음의 시작위치인 동시에 TSIUVC가 끝나는 위치로 간주하고 P_0 에서 약 25.6ms 이전의 지점을 무성자음의 시작위치로 간주하여 TSIUVC



〈그림 5〉 TSIUVC의 탐색과 추출

구간을 검출하였다. 다만, 검출한 TSIUVC 구간에 확실히 무성자음을 포함하고 있는가를 재차 검증하고자 천이구간을 제외한 무성자음 구간에 해당하는 0~12.8ms 구간의 영교차율($ZH[t]$)가 $0.4 \leq ZH[t] \leq 0.7$ 의 조건을 만족하는지를 평가하도록 하였다. 실험에 의하면 모음에서의 $Z[t]$ 는 약 0.1 부근에 분포하고 있으며, 무성자음의 경우는 0.4~0.7 정도이고, 천이구간의 경우는 모음과 무성자음의 중간 값을 갖는다. 이와 같은 방법으로 유성음(V)과 TSIUVC 구간을 탐색추출하여 재구성한 프레임을 그림 6에 나타내었다. 이와 같이 특성을 달리하는 무성음, 유성음, TSIUVC가 같은 프레임에 존재하지 않도록 프레임을 재구성함으로써 무성음, 유성음, TSIUVC 특성에 맞는 신호처리

방법을 능동적으로 선택할 수 있다.

IV. 추출물

1. 피치 추출물

피치 추출 오류의 검증은 보다 세심한 관찰력과 엄정한 규칙에 의하여 평가되어야 한다. 본 연구에서는 음성파형에서 관찰된 실제의 피치 위치와 R_p 에서 추출한 피치 위치가 일치하는지의 여부를 비교 관찰하여 피치가 존재함에도 불구하고 이를 추출하지 못한 경우(b_{ij}), 또는 피치가 존재하지 않는데 추출된 경우(C_{ij})를 피치추출 오류로 판정하여 식(9)와 같이 피치 추출률(P_R)을 계산하였다.

$$P_R = \frac{\sum_{i=1}^T \sum_{j=1}^m [a_{ij} - (b_{ij} + C_{ij})]}{\sum_{i=1}^T \sum_{j=1}^m a_{ij}} \quad (9)$$

위식에서 $T, m, a_{ij}, b_{ij}, c_{ij}$ 는 각각 프레임 총수, 총 음성샘플 수, 관찰된 피치수, 피치를 추출하지 못한 경우의 오류, 한개 이상 피치를 잘못 추출한 경우의 오류를 나타낸다. 본 연구에서 사용한 음성 표본은 표 1에 나타낸바와 같이 남녀 각각 4명의 아나운서가 발성한 32 종류의 단문을 사용하였다. 표 1의 음성 표본과 식 (9)을 사용하여 피치 추출률을 계산한 결과, 남자 음성의 경우에 자기상관법과 캡스트럼, 개별피치 펄스추출법에서 각각 89%, 92%, 96%, 여자의 경우 80%, 86%, 85%의 결과를 얻었다. 남자음성의 경우에는 피치추출률이 상당히 높은 것으로 평가되나 여자음성의 경우에는 캡스트럼과 거의 같은 결과를 얻었다. RMSE(Root Mean-Square Energy), LPC(Linear Predictive Coding), AC(Autocorrelation Coefficient), CEP(Cepstrum) 등을 사용하여 추출한 결과인 약 66.3%~87.8% 정도의 추출률[10]과 비교하면 개별피

〈표 1〉 음성 표본

| 제 원 | 남자음성 | 여자음성 |
|-----------|---------|---------|
| 발성자릿 단문 수 | 4, 16 | 4, 16 |
| 모음, 자음 수 | 145, 34 | 145, 34 |

치 펄스 추출법이 기존의 방법에 비하여 성능이 우수한 것으로 평가된다. 피치 추출률은 음성부호화, 음성합성 등의 음질향상 또는 피치 추출 알고리즘의 성능개선을 위한 지표로 활용할 수 있다.

2. TSIUVC 추출물

비선형적인 신호에 가까운 자음과 선형적인 신호에 가까운 모음이 같은 프레임에 존재할 경우에 이것을 신호처리하기란 상당히 곤란할 것이다. 이러한 프레임을 적절히 처리하지 않을 경우에 음질을 저하시키거나 음성인식률을 저하시키는 요인으로 작용할 수 있다. 때문에, 자음과 모음이 결합하는 부분에서 관찰할 수 있는 천이구간의 정보가 음성인지에 중요한 역할을 하는 것으로 간주하여 단구간 에너지나 LPC 캡스트럼을 사용하여 추출하고 분석하는 연구가 진행되었는데[8,9], 이 연구에서는 천이구간을 추출하는 방법론을 다루고 있는 관계로 천이구간의 추출률이 어느 정도인지는 확인할 수 없었다. 다만, 모음과 무성음이 존재하는 음성구간을 RMSE, LPC, AC, CEP 등을 사용하여 추출한 경우에 약 66.3%~87.8% 추출률[10]이 산출된 것을 고려하면 본 논문에서 제안한 방법이 보다 안정적으로 TSIUVC를 추출할 수 있는 것으로 평가된다. 보다 중요한 것은 개별피치 펄스를 사용함으로써 모음의 시작위치를 정확히 탐색추출할 수 있을 뿐만 아니라 연속음성에서 자연스럽게 TSIUVC가 포함되어 있는 프레임의 탐색추출이 가능하다는 것이다.

TSIUVC는 무음과 유성음 또는 무성자음과 유성음이 같이 존재하는 프레임에 한정하여 식 (10)에 의하여 TSIUVC를 추출하였으며, TSIUVC 추출률도 피치 추출률과 마찬가지로 본래 TSIUVC가 존재

함에도 불구하고 추출되지 않았을 경우(b_i)와 본래의 TSIUVC가 존재하지 않는데도 불구하고 추출된 경우(c_i)를 TSIUVC 추출 오류로 규정하였다.

$$\mathfrak{R} = \frac{\sum_{j=1}^m \{a_j - (b_j + c_j)\}}{\sum_{j=1}^m a_j} \quad (10)$$

aj:TSIUVC 관찰수, m:음성샘플 수

표 1의 음성표본을 사용하여 실험한 결과, 여자 음성의 경우 84.8%(파열음), 94.9%(마찰음), 92.3%(파찰음), 남자 음성의 경우는 88%(파열음), 94.9%(마찰음), 92.3%(파찰음)의 결과를 얻었다. 이때, 여자 음성에서 TSIUVC 추출률이 낮게 산출되었는데, 이것은 여자 음성이 남자 음성에 비해 피치 주파수가 급격히 변화하기 때문에 여자음성에서 피치 추출률이 상대적으로 낮게 산출되는 것이 원인으로 해석된다. 아울러, 자음과 모음이 같이 존재하는 음성구간을 RMSE, LPC, AC등을 이용하여 추출하는 방법에 비해 추출률의 측면에서도 향상된 방법임을 알 수 있었으며, 무엇보다 자음과 모음이 같이 존재하는 음성구간에서 TSIUVC 구간을 정확히 추출할 수 있었다는 것에 그 의미가 있다고 할 수 있다.

V. 결 론

유성음원과 무성음원을 사용하는 음성부호화 방식에 있어서, 유성음과 무성음이 같은 프레임에 존재할 경우, 이 프레임의 음성신호를 재생하기 위하여 유성음원 혹은 무성음원의 어느 한쪽의 음원을 획일적으로 적용하는 문제점이 발생한다. 아울러, 프레임 경계 부분, 무성음 혹은 무성자음과 유성음이 같이 존재하는 프레임, 음성의 시작이나 끝부분의 경우에 프레임 단위로 정규화한 피치를 얻는 방법에서는 피치추출 오류가 발생한다. 이러한 문제점을 해결하기 위하여 본 논문에서는 FIR-STREAK 필터의 오차신호로부터 연속적으로 변위하는 피치 정보를 능동적으로 추출할 수 있는 개별 피치펄스 추출법을 제안하였으며, 영교차율과 개별 피치펄

스를 사용하여 연속음성에서 TSIUVC를 탐색추출한 다음 무음/TSIUVC/유성음의 형태로 프레임을 재구성하는 방법을 제안하였다.

실험결과, 프레임 경계 부분, 무성음 혹은 무성자음과 유성음이 같이 존재하는 프레임, 음성의 시작이나 끝부분에서 발생하는 피치추출 오류를 억제할 수 있었으며 피치추출률의 측면에서도 자기상관법, 캡스트럼에 비하여 양호한 결과를 얻을 수 있었다. 또한, 연속음성에서 TSIUVC를 탐색추출하여 프레임을 재구성함으로써 특성을 달리하는 무음/TSIUVC/유성음에 적합한 음원 또는 신호처리 방법을 적절하게 선택할 수 있게 되었다.

본 논문에서 제안한 방법은 특성을 달리하는 음성신호에 적합한 신호처리 방법을 적절하게 선택하여 음성신호를 재생하는 Voiced/Silence/TSIUVC 절환 음성부호화 방식에 적용할 수 있을 것으로 기대된다.

향후 연구과제로서 개별 피치펄스 추출법을 음성부호화 방식에 활용하기 위하여 개별 피치펄스에 적절한 bit를 할당하는 연구가 이루어져야하며, TSIUVC를 낮은 bit rate의 음성부호화 방식에 적용함으로써 개선되는 음질을 정량적으로 측정하는 청각적 실험이 필요하다.

참 고 문 헌

- [1] Yusuke Hiwasaki, Kazunori Mano and Takao Kaneko: "An LPC vocoder based on phase-equalized pitch waveform" *Speech Communication*, Vol 40, Issue 3, May 2003.
- [2] 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*, sp 87-65.1987.
- [3] Ergun Ercelesi: "Second generation wavelet transform-based pitch period estimation and voiced/unvoiced decision for speech signals" *Applied Acoustics*, Vol 64, Issue 1, January 2003.

- [4] L.Hodgson, M.E.Jernigan, and B.L.Wills: "Nonlinear Multiplicative Cepstral Analysis for Pitch Extraction in Speech" IEEE, S4b. 11.1990.
- [5] Lawrence R, Rabiner, Michael J.Cheng, Aaron. Rosenberg, and Carol A.McGonegal: "A Comparative Performance Study of Several Pitch Detection Algorithms", IEEE, Vol. ASSP-24, Oct 1976.
- [6] Chong Kwan Un and Shin-Chien Yang: "A Pitch Extraction Algorithm Based on LPC Inverse Filtering and AMDF" IEEE, Vol. ASSP-39, Feb 1991.
- [7] Carol A.McGonegal, Lawrence R.Rabiner and Aaron E.Rosenberg: "Subjective Evaluation of Pitch Detection Methods Using LPC Synthesized Speech", IEEE. Vol. ASSP-25, June 1997.
- [8] 최일홍, 장승관, 차태호, 최응세, 김창석, "변곡점치 단구간 에너지 평가에 의한 음성의 천이구간 특징분석", 한국음성과학회, 제3권, 1998.
- [9] 이성주, 김희동, 김형순, "LPC 쉘스트림 거리 기반의 천이구간 정보를 이용한 음성의 가변적인 시간축 변환", 한국음성과학회지, 제3권, 1999.
- [10] LEAH J. SIEGEL and ALAN C. BESSEY: "Voiced/Unvoiced/Mixed Excitation Classification of Speech", IEEE. Vol. ASSP-30, June 1982.

○ 저자 소개 ○



이 시 우

1987년 동국대학교 전자공학과 졸업(학사)
 1990년 日本大學(Nihon Univ) 대학원 전자공학과 졸업(석사)
 1994년 日本大學(Nihon Univ) 대학원 전자공학과 졸업(박사)
 1994년 ~ 1998년 (주)삼성전자 통신연구소/멀티미디어 연구소
 1998년 ~ 현재 상명대학교 정보통신공학과 교수
 관심분야 : 음성신호처리, 유무선통신, 음주지각시스템
 E-mail : swlee@smu.ac.kr