음성부호화 방식에 있어서 FIR-STREAK 필터를 사용한 개별 피치펄스에 관한 연구

A Study on Individual Pitch Pulse using FIR-STREAK Filter in Speech Coding Method

이시우
상명대학교 컴퓨터정보통신공학부

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

요약

본 연구에서는 음성부호화 방식에서 피치추출 오류를 줄이고 피치간격의 변위에 적응할 수 있도록 피치간격을 정규화하지 않은 개별 피치펄스 추출법을 제안하였다. 개별피치 펄스의 추출률은 남자음성에서 96%, 여자음성에서 85%를 얻을 수 있었으며, 이 방법은 음성부호화방식, 음성분석, 음성합성, 음성인식 등에 활용할 수 있을 것으로 기대된다.

■ 중심어 : | 음성신호 | 피치 |

Abstract

In this paper, I propose a new extraction method of Individual Pitch Pulse in order to accommodate the changes in each pitch interval and reduce pitch errors in Speech Coding. The extraction rate of individual pitch pulses was 96% for male voice and 85% for female voice respectively. This method has the capability of being applied to many fields, such as speech coding, speech analysis, speech synthesis and speech recognition.

■ keyword : | Speech Signal | Pitch |

I. 서 론

이동통신, 유선통신의 채널을 효과적으로 사용하기 위하여 낮은 bit rate의 음성부호화 방식을 사용한다. 이러한 낮은 bit rate의 음성부호화 방식에서는 음성신호를 효율적으로 압축/복원하기 위하여 피치(Pitch)정보를 종종 사용한다. 피치정보는 음성부호화 방식의 음질을 향상시키는 중요한 요소로 작용할 뿐만 아니라 남자와 여자음성을 판별하거나 모음과 자음을 구분지울 수 있는 상당히 유용한 파라미터이다.

피치를 추출하는 방법은 시간영역, 주파수영역, 시간과 주파수영역을 혼용하여 추출하는 방법으로 나뉠 수 있다. 시간영역에서 추출하는 방법으로 자기상관법[1], 주파수 영역에서 추출하는 방법으로 Cepstrum법[2], 시간과 주파수영역에서 피치를 추출하는 방법으로 AMDF(Average Magnitude Difference Function)법[3]과 LPC와 AMDF를 혼합한 방법[4][5] 등이 있다.

접수번호 : #040419-001
접수일자 : 2004년 4월 19일
심사완료일 : 2004년 5월 7일
교신저자 : 이시우, e-mail : swlee@smu.ac.kr

이러한 방법들은 한 프레임에 하나의 규격화된 피치정보를 얻을 수 있는 방법으로 음소 상호간섭이 있는 부분, 음성의 시작이나 끝부분, 무성음과 유성음 혹은 무성자음과 유성음이 같이 존재하는 프레임에서는 피치추출 오류가 종종 발생한다. 이러한 오류들은 연속적인 음성신호에서 볼 수 있는 피치정보의 연속적인 변위를 하나의 피치정보로 규격화 하는데 그 원인이 있다고 볼 수 있다. 이러한 오류를 억제할 수 있는 방법으로 본 연구에서는 시간영역에서 연속적으로 변위하는 피치위치를 추출할 수 있도록 FIR 필터와 STREAK(Simplified Technique for Recursively Estimating Autocorrelation K-parameters) 필터를 조합한 FIR-STREAK 필터의 오차신호로부터 피치펄스를 추출하는 방법을 제안하고자 한다. 이 방법에 의하여 추출된 피치펄스는 프레임마다 여러 개의 피치정보를 추출할 수 있기 때문에 변위하는 음성신호의 분석, 합성, 인식 등에 유용하게 활용할 수 있을 것으로 기대된다.

II. 피치 추출 알고리즘

1. 자기상관법과 Cepstrum법

프레임마다 평균 피치 정보를 추출하는 대표적인 방법인 자기상관법과 Cepstrum법은 널리 알려진 방법이기 때문에 간략하게 기술하고자 한다.

자기상관법에 의한 피치 추출법은 자기상관 계수 $R(t)$가 1에 근접한 시점을 피치의 개시 · 종료 지점으로 규정하여 피치주기를 구하게 된다.

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

Cepstrum법은 식(2)의 최대값으로부터 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.

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

2. 개별 피치펄스 추출법

일반적으로 자기상관법은 프레임 단위로 하나의 규격화된 피치정보를 추출하기 때문에 음소 상호간의 간섭에 의해 피치간격이 일정하지 않거나 음성의 시작이나 끝부분과 같이 준주기성의 음성파형, 무성음과 유성음 혹은 무성자음과 유성음이 같이 존재하는 프레임에서는 피치추출 오류가 종종 발생한다. 이러한 오류를 억제하고, 변화하는 음성파형에 따라서 능동적으로 피치의 위치를 추정하는 방법으로 본 연구에서는 FIR-STREAK 필터의 오차신호에서 개별피치 펄스를 추출하고자 한다.

그림1에 나타낸바와 같이 FIR-STREAK는 FIR 필터와 STREAK 필터를 조합한 형태로서 FIR 필터는 주파수 대역을 제한하기 위한 LPF(Low Pass Filter)의 역할을 하며, STREAK 필터는 오차신호를 출력하는 역할을 한다.

STREAK 필터는 전방향 오차신호($f_i(n)$)와 후방향 오차신호($g_i(n)$)의 순시값을 최소화 한다.

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

윗식을 k_i에 관하여 편미분함으로서 STREAK계수 k_i를 구할 수 있다.

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

여기에서, i=1,2,...,M 이고, n=1,2,...,N이다.

k_i를 사용한 STREAK 필터의 전달함수는 다음과 같다.

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

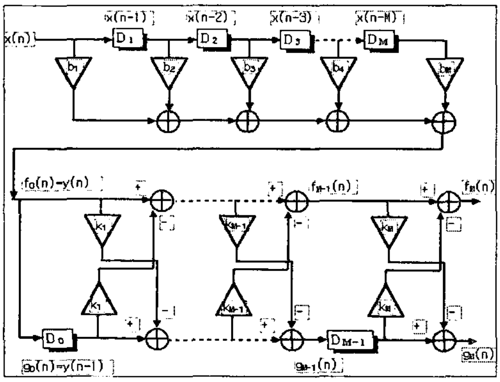


그림 1. FIR-STREAK 필터의 구성

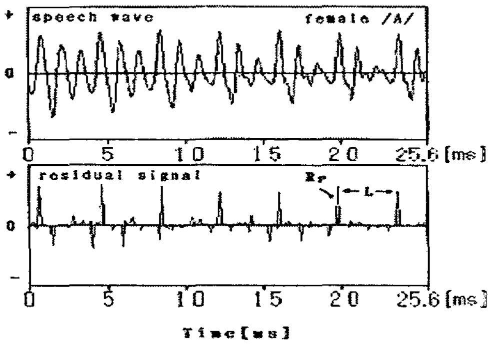


그림 2. 개별 피치 펄스

FIR-STREAK 필터의 오차신호($E_{PN}(n)$는 그림2와 같이 시간영역의 +측 오차신호($E_P(n)$)와 -측 오차신호($E_N(n)$)로 구성되어 있으며, 잡음성 오차신호에서 펄스성 오차신호(R_P)를 검출하여 개별 피치펄스의 위치를 추정할 수 있다.

FIR-STREAK 필터의 오차신호로부터 개별 피치펄스를 추출하는 방법을 그림3에 나타내었다. FIR-STREAK 필터의 오차신호 $E_{PN}(n)$은 $E_{PN}(n) \geq 0$ 인 경우에 $E_P(n)$으로, $E_{PN}(n) < 0$인 경우에는 $E_N(n)$으로 결정하였다.

이때 $E_P(n)$와 $E_N(n)$을 병렬로 처리하여 검출한 R_P로부터 개별 피치펄스를 구하는데, $E_P(n)$과 $E_N(n)$에서 개별 피치펄스를 구하는 방법은 모두 같기 때문에 본 논문에서는 $E_P(n)$에서 개별 피치펄스를 검출하는 방법에 관해서 언급하기로 한다. 우선, $E_P(n)$를 A로 정

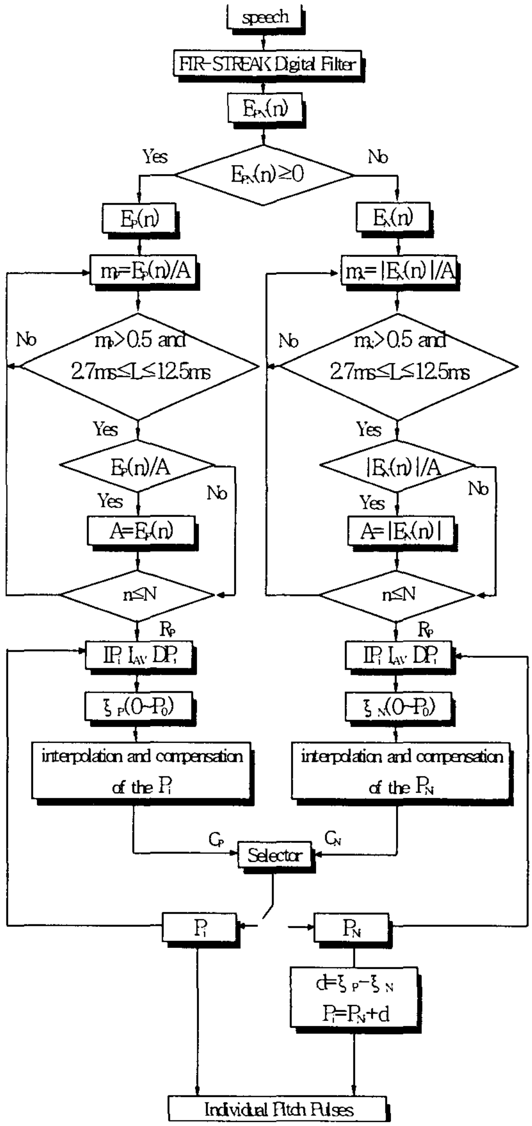


그림 3. 개별 피치 펄스 추출법

규화한 값이 $m_p > 0.5$를 만족하는 동시에 $2.7ms \leq L \leq 12.5ms$ (피치주파수 범위:80~370Hz)를 만족하는 오차신호를 검출하도록 하였다. 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 $DP_i(DP_i = I_{AV} - IP)$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 만

족하는 경우는 식(6)으로 R_P위치를 보완 하도록 하였다.

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

일반적으로 유성음은 수십ms 동안 피치주기의 변화가 적기 때문에 식(7)을 만족하는 경우는 시간영역 +측의 P_i를, 그렇지 않은 경우는 -측의 P_i'를 선택하도록 하였다.

$$\sum_{i=1}^{M} \frac{IP_i}{I_{AV}} \leq \sum_{i=1}^{M} \frac{IP_i'}{I_{AV}'} \tag{7}$$

Ⅲ. 음성샘플과 피치 추출률

피치정보는 V/UV의 판독, 남여음성의 판독 등에 사용되는 유용한 파라미터이며, 유성음원과 무성음원에 의하여 음성신호를 부호화하고 합성하는 음성부호화 방식의 음질을 향상시키기 위한 중요한 요소이기도 하다. 비주기적인 특성을 나타내는 자음 혹은 자음에서 모음으로 천이하는 구간의 음성신호처리는 주기적인 특성을 나타내는 모음의 음성신호에 비하여 신호처리하기가 어렵고 피치추출 오류가 종종 발생한다. 이러한 오류를 개선하기 위한 방법을 보다 엄정하게 평가하기 위해서는 자음이 많은 음성샘플이 바람직하며 엄정한 규칙과 평가방법이 필요하다.

본 연구에서는 표1에 나타낸바와 같이 남자(4명), 여자(4명)이 발성한 단문 32개의 음성샘플을 사용하였으며, 이 음성샘플의 모음과 자음의 수는 총 290개, 68개이다. 이 음성샘플을 사용하여 본 연구에서 제안한 개별 피치펄스 추출법과 일반적으로 널리 사용되고 있는 자기상관법과 Cepstrum법을 피치 추출률의 측면에서 평가하고자 한다. 우선, 프레임마다 규격화된 하나의 피치정보를 추출하는 자기상관법 혹은 Cepstrum법과 프레임마다 여러 개의 피치정보를 추출하는 방법을 피치 추출 오류 측면에서 비교하기 위해서 자기상관법과 Cepstrum법에 의하여 추출한 규격화된 하나의 피치주기를 한 프레임에 나타낼 수 있는 여러 개의 피치주기로 표현한 후 피치추출 오류를 판독하도록 하였다. 즉, 피치추출 오류를 시간상의 음성파형에서 관찰된 실제의 피치간격과 R_P에서 검출한 피치 간격이 일치하는지의 여부를 비교 관찰하여 판정하도록 하였다. 결과적으로 본래 피치가 존재함에도 불구하고 이를 추출하지 못한 경우(b_{ij}), 또는 피치가 존재하지 않는 위치에서 추출된 경우(C_{ij})를 피치추출 오류로 판정하여 피치 추출률(P_R)을 계산하도록 하였다.

$$P_R = \frac{\sum_{i=1}^{T} \sum_{j=1}^{m} [\, a_{ij} - (\, b_{ij} + C_{ij})]}{\sum_{i=1}^{T} \sum_{j=1}^{m} a_{ij}} \tag{8}$$

위식에서 m, T, a_{ij}, b_{ij}, c_{ij}는 각각 프레임 총수, 총 음성샘플 수, 관찰된 피치 수, 피치를 추출하지 못한 경우의 오류, 한개 이상의 피치를 잘못 추출한 경우의 오류를 나타낸다.

표 1. 음성샘플

제 원	남자음성	여자음성
발성자및 단문 수	4, 16	4, 16
모음, 자음 수	145, 34	145, 34

Ⅳ. 실험결과

본 연구에서 제안한 FIR-STREAK 필터의 오차신호에서 추출한 R_P로부터 개별 피치추출법을 통하여 피치펄스를 추출한 후, 연속적인 프레임 경계면, 음소 상호간섭이 있는 부분, 무성음과 유성음이 같이 존재하는 프레임에서 종종 발생하는 피치추출 오류를 해결하였는지 여부와 피치추출률의 측면에서 비교 평가하는 실험을 하였다.

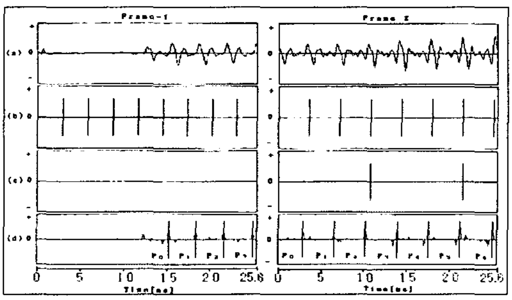
본 연구에서 피치주파수는 억제하지 않고 높은 주파수인 잡음성 오차신호만 억제하기 위하여 FIR 필터의 계수는 복소수[6]를 사용하였으며, 필터의 차수와 대역제한 주파수는 각각 800Hz, 40차로 하였다. 이러한 조건은 피치주파수가 400Hz 이하에 존재하며 잡음성 오차신호를 효과적으로 억제하기 위하여 설정된 조건으로

서 본 연구결과에 중대한 영향을 미치지는 않는다. 다만, 40차의 필터를 사용함으로서 3.4초 동안에 약 0.3ms의 신호처리 지연시간이 발생하는데 이 정도의 지연시간은 인간의 청각으로 느낄 수 없는 정도라 할 수 있다.

다음으로 자기상관법과 Cepstrum법을 개별 피치펄스 추출법과 피치추출의 정확성과 추출률 측면에서 비교하기 위하여 자기상관법과 Cepstrum법에 의하여 추출한 피치를 25.6ms의 프레임에 나타낼 수 있는 피치의 수로 환산하여 나타내었다. 표1의 음성샘플과 식(8)을 사용하여 피치 추출률을 계산한 결과, 표2와 같은 결과를 얻을 수 있었다. 개별 피치추출법의 경우에 남녀 4명의 32문장에서 a_{ij}는 남자와 여자음성에서 각각 3483개, 5374개 이며, b_{ij}, C_{ij}의 오류 없이 추출된 피치는 남자와 여자음성에서 각각 3343개, 4566개였다. 따라서 개별 피치펄스 추출법에 의한 피치추출률은 남자와 여자 음성에서 각각 96%, 85%를 얻었으며, 자기상관법에서는 89%, 80%를, Cepstrum법에서는 92%, 86%를 얻었다. 이때, 피치추출률이 여자음성에서 낮게 얻어진 이유는 여자음성이 남자음성에 비하여 피치주파수가 급격히 변하는 특성 때문으로 해석된다. 이러한 까닭에 피치 추출률은 여자의 경우가 일반적으로 낮게 평가 된다.

표 2. 피치 추출률

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



(a) 원음성 (b) 자기상관법 (c)Cepstrum법 (d)개별피치 추출법

그림 4. 피치추출

프레임의 경계부분, 무성음과 유성음, 혹은 무성자음과 유성음이 같이 존재하는 부분, 음소가 변위하는 부분, 프레임의 경계 부분, 음성의 시작 부분, 음성의 끝 부분에서 자주 발생하는 피치추출 오류를 본 연구에 의하여 해결한 예를 그림4에 나타내었다. 그림4에 나타낸 바와 같이 프레임 길이가 25.6ms인 두 프레임의 연속된 음성파형에서 FIR-STREAK 필터의 오차신호가 펄스성 오차신호(R_P)와 잡음성 오차신호로 구성되어 있으며, Frame-2와 같이 유성음의 경우에는 개별피치 추출법과 자기상관법에서 유용한 피치정보를 추출할 수 있었던 반면 Cepstrum법에서는 피치추출 오류를 볼 수 있다.

한편 Frame-1과 같이 무성음과 유성음, 혹은 무성자음과 유성음이 같이 존재하는 부분, 음소가 변위하는 부분, 프레임의 경계 부분, 음성의 시작 부분, 음성의 끝 부분에서는 개별피치 추출법이 보다 안정된 피치정보를 얻을 수 있다.

V. 결 론

본 논문에서는 FIR-STREAK필터의 오차신호를 처리하여 연속적으로 변위하는 피치정보를 능동적으로 추출할 수 있는 개별피치 추출법을 제안하였다. 실험결과, 일반적으로 사용하는 자기상관법, Cepstrum법에서는 프레임 경계 부분, 무성음 혹은 무성자음과 유성음이 같이 존재하는 프레임, 음성의 시작이나 끝부분에서 발생하는 피치추출 오류가 발생한다. 그러나 본 논문에서 제안한 개별 피치추출법에서는 프레임 경계 부분, 무성음 혹은 무성자음과 유성음이 같이 존재하는 프레임, 음성의 시작이나 끝부분에서 피치추출 오류를 억제할 수 있었으며 피치추출률의 측면에서도 자기상관법, Cepstrum법에 비하여 양호한 결과를 얻을 수 있었다.

본 연구에 의하여 제안된 방법은 음성부호화 방식, 음성분석, 음성인식 등에 활용할 수 있을 것으로 기대된다.

향후, 연구과제로는 개별 피치펄스 추출법을 음성부호화 방식에 활용하기 위하여 적정한 bit할당과 전송량에 관한 연구가 이루어져야 하겠다. 이것은 본 연구에서

제안한 방법이 음성신호를 좀더 섬세하게 분석, 합성할 수 있는 요소를 제공하는 측면에서는 유리하지만 프레임마다 여러 개의 피치정보를 사용하여야 함으로 정보량의 측면에서 최적화할 필요가 있기 때문이다.

참 고 문 헌

[1] 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, pp. 87-65.1987.

[2] L. Hodgson, M. E. Jernigan, B. L. Wills, "Nonlinear Multiplicative Cepstral Analysis for Pitch Extraction in Speech," IEEE, S4b.11. 1990.

[3] Lawrence R, Rabiner, Michael J. Cheng, Aarone. Rosenberg, Carol A. McGonegal, "A Comparative Performance Study of Several Pitch DetectionAlgorithms," IEEE, Vol. ASSP-24, Oct. 1976.

[4] Chong Kwan Un, Shin-Chien Yang, "A Pitch Extraction Algorithm Based on LPC Inverse Filtering and AMDF," IEEE, Vol.ASSP-39, Feb. 1991.

[5] Carol A. McGonegal, Lawrence R. Rabiner, Aaron E. Rosenberg, "Subjective Evaluation of Pitch Detection Methods Using LPC Synthesized Speech," IEEE. Vol.ASSP-25, June. 1997.

[6] T. H. Crystal and L. Ehrman, "The design and application of digital filter with complex coefficients," IEEE Trans. Audio & Electroacoust, AU-16.3, 1968.

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 정보통신전공 교수

<관심분야> : 음성신호처리, 유무선통신, 음주지각시스템