제15회 음성통신 및 신호처리 워크샵(KSCSP '98 15권1호)

피치검색기법과 3-Level Clipping을 이용한 음성 파형부호화법에 관한 연구

김규홍, 정형교, 장금영, 배명진
숭실대학교 정보통신공학과
156-743 서울시 동작구 상도동 1-1

On a Waveform Coding Technique Using Pitch Searching and 3-Level Clipping

KyuHong Kim, HyungGoue Chung, KeumYoung Jang, MyungJin Bae
Dept. of Telecomm., Engr., Soongsil Univ.
1-1 Sangdo-5Dong, Dongjak-Ku, Seoul 156-743, KOREA
khkim@ifcom.soongsil.ac.kr, mjbae@saint.soongsil.ac.kr

*이 논문은 산학연 공동기술개발사업 연구지원비에 의해 이루어졌습니다.

요 약 문

본 논문에서는 피치검색과 3-Level Clipping을 이용한 새로운 파형부호화법을 제안하고자 한다. 제안한 방법에서는 우선 피치를 검출하여 기준피치파형과 인근피치파형을 검색한다. 그 후 유사도 측정시 기준피치파형과 인근피치파형에 대해 3-Level Clipping을 수행한다. 클리핑된 기준피치파형과 인근피치파형간의 유사도를 측정하여 유사성이 크다면 피치정보와 에너지 정보만을 전송하거나 저장하여 압축을 하고, 유사성이 적다면 인근피치파형을 압축을 하지않고 저장한다. 그 후에 저장된 파형을 기준피치파형으로 재정의하여 다시 반복적으로 압축을 수행한다. 압축된 음성신호를 다시 복원할 때에는 수신 또는 저장된 음성신호를 이용하여 PSOLA방식으로 합성을 수행한다. 평균압축율이 약 65%일 경우에도, MOS값이 4이상을 유지하였다.

1. 서 론

음성신호를 메모리에 저장하거나 전송하기 위한 음성부호화법에는 크게 파형부호화법, 신호원부호화법, 혼성부호화법으로 나눌 수 있다[1]. 파형부호화법은 음성정보를 발성모델에 따라 분리하지 않고 파형 자체의 잉여성분만을 제거한 후 부호화하여 저장 또는 전송하고 필요에 따라 다시 원래의 신호 파형으로 합성하는 방법이다. 일반적으로 양자화된 음성표본을 B bit를 사용하여 Fs 율로 표본화하면 전송이나 저장에 필요한 정보의 용량 I는 B bit와 Fs의 곱으로 나타낸다. 일반적으로 부호화법에서는 정해진 음질을 유지하는 상태에서 I를 낮출 필요가 있다. 그러나 파형 부호화법에서는 음성신호의 파형 형태를 보존하기 위해 음성신호의 표본화율이 Nyquist 표본화 이론으로 이미 정해져 있기 때문에 표본당 양자화 비트의 수를 줄이는데 주로 연구의 초점을 맞추고 있다. 이러한 비트 수를 줄이는 방법에 따라 PCM(Pulse Code Modulation), ADPCM (Adaptive Differential PCM), ADM(Adaptive Delta Modulation) 등이 제안되어 있다[1]. 파형부호화법의 특징은 고음질과 개성을 유지할 수 있으나 방대한 데이터량 때문에 메모리가 많이 필요하게 된다는 단점이 있다[2]. 최근에는 디지틀 신호처리 전용칩의 제조기술의 발달로 인하여 파형부호화법의 분석 및 합성 알고리즘이 잘 개발되어 32kbps, 16kbps 등의 전송율을 갖는 ADPCM의 표준화가 실현되어졌다. 그러나 ADPCM을 이용할 경우에 별

도의 DSP칩을 사용해야 한다는 문제점이 있다.

본 논문에서는 음성파형의 대부분을 차지하고 있는 유성음의 피치를 검색하며 유사도를 측정하고, 이 유사도를 이용하여 음성신호의 반복성 즉, 잉여성분을 제거하는 기법으로 음성을 압축하였다. 이렇게 측정된 유사도 값을 문턱값과 비교하여, 유사도 값이 문턱값 이상이면, 기준피치구간과 비교해 변경된 정보만을 전송하거나 저장하는 방법을 이용하여 음성을 압축할 수 있는 새로운 파형 부호화방법을 제안하였다. 제안한 방법을 이용할 경우 알고리즘이 간단하여 저렴한 범용칩을 사용하여 구현할 수 있기 때문에 위에서 설명하였던 ADPCM의 문제점을 해결할 수 있다. 따라서 저가의 음성부호화기 구현이 가능하게 된다.

Ⅱ. 피치검색에 의한 유사도 측정

음성신호에서 대부분을 차지하고 있는 유성음을 압축하기 위해서는 여러 가지 방법이 있지만, 고음질을 유지하는데 적합한 방법은 파형부호화법이라 할 수 있다. 일반적으로 음성신호는 무성음과 유성음으로 나뉘어질 수 있다. 특히 유성음은 음성신호에서 대부분을 차지하고 있다. 그림 1(a)는 유성음 /나/에 대한 파형을 나타내고 있다. 그림에서 알 수 있듯이 유성음은 피치단위로 유사한 파형이 반복되는 형태를 띄고 있는데 이러한 유사성을 이용하여 파형을 압축할 수 있다. 이러한 유사성 정보를 추출하기 위하여 본 논문에서는 피치를 검색하여, 시간영역에서 피치 단위별로 유사도를 추출할 수 있는 방법을 제안하고자 한다.

우선 AMDF법을 사용하여 피치를 사전에 계산하여 피치단위로 음성 압축 및 부호화 처리를 하였다. 식 (1)을 이용하여 기준피치구간과 인근피치구간의 스케일링 정보를 구한 후 식 (2)를 이용하여 인근피치구간의 크기를 스케일링하여 진폭을 조절한다.

$$a=\sqrt{\frac{E_X}{E_Y}} \qquad (1)$$

$$S_Y'(n)=a\,S_Y(n) \qquad (2)$$

이렇게 식 (1), (2)을 이용하여 에너지를 일치시킨 후 3-Level Clipping을 시킨다. 3-Level Clipping은 식 (3)과 같다.

$$f(x)=\begin{cases} 1, & C_L \le x \\ 0, & -C_L \le x < C_L \\ -1, & x < -C_L \end{cases} \qquad (3)$$

여기서 C_L은 Clipping되는 임계치이다. 이 값은 기준피치구간에서 최대값의 20%를 사용하였다. 3 Level Clipping된 기준피치구간과 인근피치구간에 대해서 유사도를 측정하게되는데, 유사도의 정의는 다음 식 (4)와 같다.

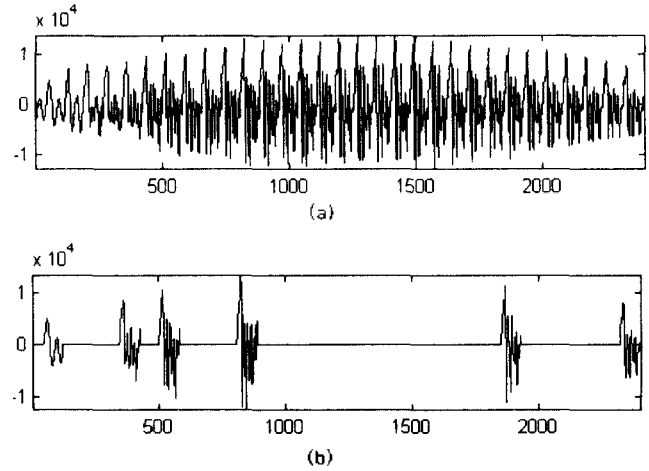

그림 1. 유성음 /나/에 대한 유사도 측정 예

$$S = \frac{\sum_{n=0}^{P}\{f(S_X(n)\, f(S_Y'))\}}{P} \times 100 \qquad (4)$$

$$P = \min(L_1, L_2) \qquad (5)$$

식 (4)에서 P값의 정의는 식(5)에 나타내었으며, 식 (5)에서 L_1, L_2는 각각 기준피치구간의 길이와 인근피치구간의 길이이다. 식(1)-(5)를 이용하여 측정한 유사도 값이 크면 클수록 기준피치구간과 인근피치구간의 유사성이 높다고 할 수 있다. 이러한 사실을 이용하여 이 유사도 값을 미리 정의한 유사도 문턱값과 비교하여 기준피치구간과 인근피치구간이 유사하다고 판단되면 인근피치정보 L_2와 에너지정보 a만을 저장하거나 전송한다. 그 후에 다음 피치구간을 검색하여 인근피치구간으로 정의한 후 처음의 과정을 반복한다. 만약 유사도 값이 문턱값보다 작아서 파형의 모양이 유사하지 않다고 판단되면, 파형을 그대로 저장하거나 전송하고 기준피치구간을 방금 저장하거나 전송한 파형으로 대체하고, 처

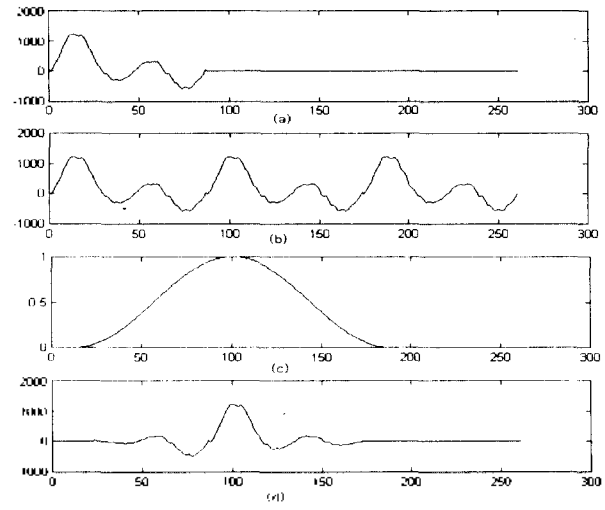

그림 2. PSOLA 합성을 위한 기준파형

음의 과정을 반복한다.

그림 1(b)는 유성음 /나/에 대한 파형에 대하여 유사도를 측정하여 그 유사도 값이 문턱값(60%)이하로 되었을 경우 파형을 표시한 그림이다.

Ⅲ. PSOLA를 이용한 음성합성

본 논문에서는 음성신호를 복원할 때 스펙트럼 왜곡률과 복잡성이 적은 PSOLA 방식의 합성법을 적용하였다[7,8]. PSOLA 합성을 위한 기준파형을 구성하기 위하여 저장 또는 수신된 기준피치구간을 3번 반복시킨 후 윈도우를 적용하여 기준파형을 만든다. 그림 2는 기준파형을 만드는 과정을 보여준다. 그림 2(a)는 저장 또는 수신된 기준 피치파형이며 그림 2(b)는 기준피치구간을 3번 반복시킨 파형이고 그림 2(c)는 사용된 윈도우의 모양이다. 최종적으로 그림 2(d)와 같은 파형을 합성에 이용하게 된다. 이렇게 구성된 파형과 저장된 피치정보 및 진폭정보를 이용하여 스케일링하며 피치구간만큼 간격을 두면서 누적 시켜가는 방식(PSOLA)으로 합성을 하여 음성을 복원한다[7,8].

Ⅳ.실험 및 결과

본 논문에서 제안한 방법을 시뮬레이션하기 위해 Pentium(150Mhz)에 마이크 입력이 가능한 16비트 A/D 변환기를 인터페이스하여 5명의 남성과 2명의 여성화자를 통해 다음 음성시료를 발성하게 하고 이를 11kHz의 표본화율로 16비트 양자화하여 저장하였다.

발성 1: /인수네 꼬마는 천재소년을 좋아한다./
발성 2: /예수님께서 천지창조의 고운을 말씀하셨다./.
발성 3: /숭실대 정보통신과 음성통신연구팀이다./
발성 4: /창공을 헤쳐나가는 인간의 도전은 끝이 없다./
발성 5: /MAY I HELP YOU?/

그림 3은 본 논문에서 제안한 방법의 블럭도이다. 시뮬레이션시 사전 피치검색 구간을 256 샘플로 정하였으며 음성압축처리를 수행할 때에는 피치단위로 처리하였다. 부호화단에서는 먼저 AMDF법을 사용하여 한 피치구간의 음성표본을 피치단위로 자른 다음 기준피치파형으로 저장한다. 그리고 인근피치파형을 조사한 다음 두 파형의 유사도 값을 측정한다. 그 후에 문턱값과 유사도 값을 조사하여 유사도 값이 문턱값보다 작으면 파형을 전송하거나 저장하고 그렇지 않으면 기준파형과 비교하여 늘거나 줄어든 피치정보와 진폭정보를 저장하거나 전송한다. 수신단에서는 수신된 비트 스트림을 받아서 헤더를 체크하여 기준피치파형일 경우 PSOLA 합성에 적합한 형태의 파형을 만들고 다시 헤더를 체크하여

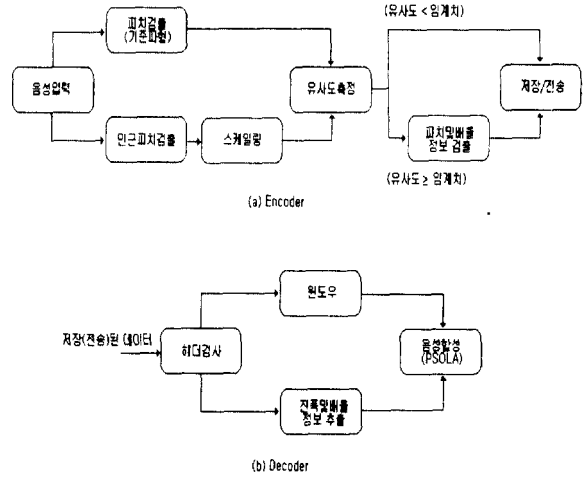

그림 3. 제안한 방법의 블록도

압축된 데이터일 경우에는 피치정보와 진폭정보를 이용하여 기준 파형을 누적시켜 가면서 디하는 과정으로 음성을 복원하였다.

부호화 단계에 있어서 유사도 값과 문턱값 비교 과정에서 문턱값을 변화시킴으로써 압축율을 조정할 수 있다. 이렇게 하여 음성을 압축하였을 경우 압축율에 따른 결과를 <표 1>에 나타내었다. <표 1>에서 볼 수 있듯이 60% 압축 수행결과 약 4.2의 MOS Score를 얻을 수 있었고 65%, 70%, 75%일 때 각각 4.1, 3.8, 3.5의 MOS Score를 얻을 수 있었다. 압축율을 증가시킴에 따라 MOS Score 값이 상대적으로 낮아지는 것을 볼 수 있다. 제안한 방법은 기존의 범용 칩으로 구현이 가능하지 못한 10MIPS정도를 차지하는 ADPCM같은 파형 부호화 방법을 저가의 칩에 본 논문에서 제안한 방법을 적

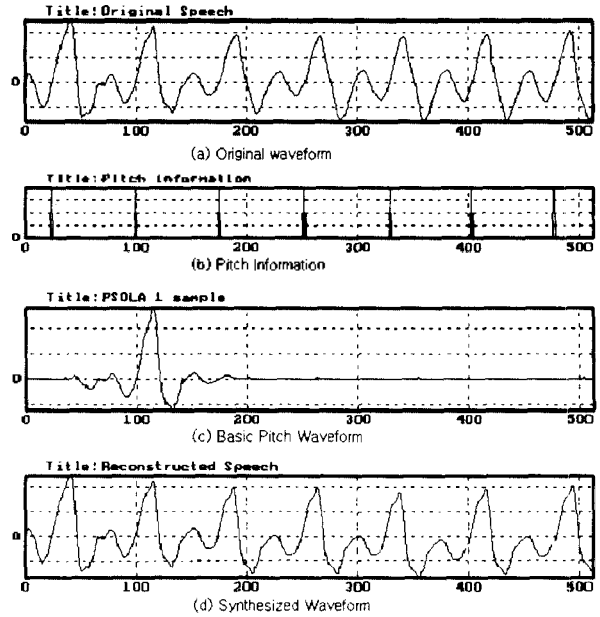

그림 4. 피치단위의 처리과정 예

용하여 2MIPS성도의 한계를 갖고 있는 범용칩에 구현할 수 있다. 그림 4는 피치단위의 처리과정 예를 보여주는 그림이다. 그림 4(a)는 원래의 음성파형이고, 그림 4(b)는 피치정보이고, 그림 4(c)는 음성을 합성할 때 PSOLA 합성에 적합한 형태로 만들어 준 파형이다. 그림 4(d)는 PSOLA 합성법을 이용하여 복원한 파형이다.

<표 1> 압축율에 따른 MOS Score

압축율	MOS Score
60%	4.2
65%	4.1
70%	3.8
75%	3.5

V. 결 론

파형부호화법의 대표적인 방법이라고 할 수 있는 ADPCM을 이용하여 음성을 처리하는 시제품에 적용할 경우 DSP칩을 사용해야 한다는 문제점이 있다. 이것은 제품의 가격경쟁력을 약화시키게 된다. 따라서 본 논문에서는 기존의 파형 압축방법과는 전혀 다른 피치단위로 파형을 부호화하여 2MIPS 정도의 범용칩으로도 음성부호화가 가능한 새로운 방법을 제안하였다.

피치를 검색하며 기준피치구간을 얻고 인근피치구간을 검색한 다음 유사도를 측정한다. 문턱값과 유사도의 관계에 의해서 파형을 압축할 것인가를 결정한다. 압축할 경우에는 기준 파형과 비교하여 늘어나거나 줄어든 진폭정보와 피치정보만을 저장하거나 전송한다. 결과 60%이상의 압축율에도 MOS Score가 4 이상을 유지하는 것을 볼 수 있었다. 본 논문에서 제안한 음성 부호화 알고리즘은 계산량이 적기 때문에 기기의 사용시간을 늘리거나 무게를 줄일 수 있다는 장점을 갖는다.

따라서 음성부호화 방법을 이용하여 상품화하려는 분야에 본 논문에서 제안한 방법을 이용하여 음성데이터를 압축하여 전송하거나 저장할 경우 저가의 범용칩을 이용하여 상품화할 수 있으므로 대외 경쟁력을 가질 수 있다.

VI. 참고 논문

[1] N. S. Jayant and P. Noll, Digital Coding of Waveforms-Principles and Applicants to Speech and Video, pp. 220-221, Prentice-Hall, 1978.

[2] M. J. Bae, D. S. Kim, H. Y. Jeon and S. G. Ann, "On a new predictor for the waveform coding of speech signal by using the dual autocorrelation and the sigma-delta technique," *IEEE Proc. of ISCAS'94*, vol.6, No. 3, pp.261-264, June 1994.

[3] D. Chung, M. BAE, S. ANN, "On Detecting the Steady State Segments of Phonemes by using the Magnitude Distribution of Speech Signals", J., Acoust., Soc., Korea, Vol.10, No.6, pp.5-11, Dec. 1991.

[4] A.M. Kondoz *"Digital Speech"*, John Wiley & Sons Ltd, Baffins Lane, Chichester, England, 1994

[5] L.R. Rabiner, and R.W. Schafer *"Digital Processing of Speech Signals"*, Prentice-Hall, Englewood Cliffs, New Jersey, 1978.

[6] A. varga and F. Fallside, "A Technique for Using Multipulse Linear Predictive Speech Synthesis in Text-to-speech Type System", IEEE signal processing, Vol.ASSP-35, No.4, pp.586-587, APRIL 1987.

[7] F. Chapentier, M. G. Stella, "Diphone Synthesis Using Overlap-add Technique for Speech Waveforms Concatination", ICASSP 86, pp.2015-2018, 1986

[8] E. Moulines and F. Charpentier, "Pitch-synchronous waveform processing techniques for test to-speech synthesis using diphones", Speech Comm., vol.9, pp. 453-467, 1990