포만트 유사도 측정에 의한 PSOLA 음성 부호화에 관한 연구

나 덕 수 , 이 희 원 , 김 규 홍 , 배 명 진
숭실대학교 정보통신공학과
156-743 서울시 동작구 상도동 1 1
dsna@assp.soongsil.ac.kr mjbac@saint.soongsil.ac.kr

On A Study of PSOLA Coding Technique based on the measurement of Formant similarity

Duck Su Na , Hee Won Lee , Kyu Hong Kim , Myung Jin Bae
Dept. Telecommunication Engr., SoongSil University
1-1 Sangdo-5Dong, Dongjak-Ku, Seoul 156-743, KOREA
dsna@assp.soongsil.ac.kr mjbac@saint.soongsil.ac.kr

* 이 논문은 산학연 공동기술개발사업 연구지원비에 의해 이루어졌습니다.

Abstract

The major objectives of speech coding include high compression ratio for transmission in the band limited channel, high synthesized speech quality in terms of the intelligibility and the naturalness and fast processing speed. In general, speech coding methods are classified into the following three categories: the waveform coding, the source coding and the hybrid coding.

In this paper, we proposed a new waveform coding method using PSOLA(Pitch-Synchronous Overlap Add) technique. First, we fixed one basic waveform per pitch and measured the formant similarity between basic and neighbor waveform. Second, if the similarity satisfied threshold values, we compress the neighbor waveform per pitch and then store or transmit. When the compression is about 45%, we obtained about 4 in MOS.

Ⅰ. 서론

일반적으로 양자화된 음성표본을 B bit 부호어를 사용하여 Fs율로 표본화하면 전송이나 저장에 필요한 정보의 용량 I는 B bit와 Fs의 곱으로 나타낸다. 일반 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 ADPCM (Adaptive Differential PCM), ADM(Adaptive Delta Modulation) 등이 제안되어 있다[1]. 파형부호화법은 고음질과 개성을 유지할 수 있으나 방대한 데이터량 때문에 메모리가 많이 필요하게 된다[2]. 최근에는 디지털 신호처리 전용 칩의 제조기술과 파형부호화법의 분석 및 합성 알고리즘이 잘 개발되어 32kbps, 16kbps 등의 전송률을 갖는 ADPCM의 표준화가 실현되어졌다. 그러나 ADPCM을 이용할 경우에 별도의 DSP칩을 사용해야 한다는 문제점이 있다.

본 논문에서는 피치단위로 기준 피치 파형과 인근 피치파형의 포만트 유사도를 측정하여 유사도가 높은 경우 피치정보와 진폭정보만을 전송하거나 저장하는 방

법을 이용하여 음성을 압축하는 새로운 파형부호화 방법을 제안하였다. 압축시에는 Cross Normalized AMDF 파형의 면적으로 유사도를 측정하여 압축하였고 합성을 할 때에는 PSOLA 기법을 사용하였다.

제안한 방법을 이용할 경우 범용칩을 사용하여 합성할 수 있기 때문에 위에서 설명하였던 ADPCM의 문제점을 해결할 수 있다.

Ⅱ. NAMDF (Normalized AMDF)

현재 프레임의 피치를 측정하는 방법으로는 다음과 같이 NAMDF를 정의하여 사용할 수 있다[3].

$$NAMDF(d) = \frac{\sum_{n=1}^{N} |s(n) - s(n-d)|}{\sum_{n=1}^{N} |s(n)| + |s(n-d)|} \quad (2.1)$$

여기서 s(n)은 음성신호이고 N은 NAMDF를 구하려는 윈도우 구간이다. 지연인자 d를 점차 증가시키면서 NAMDF를 구해보면, 지연인자가 프레임내 음성피치에 정수배가 될 때마다 NAMDF는 거의 영이 된다.

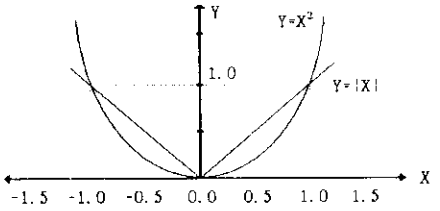

그림. 2-1 1차와 2차 함수의 에러 함수 비교

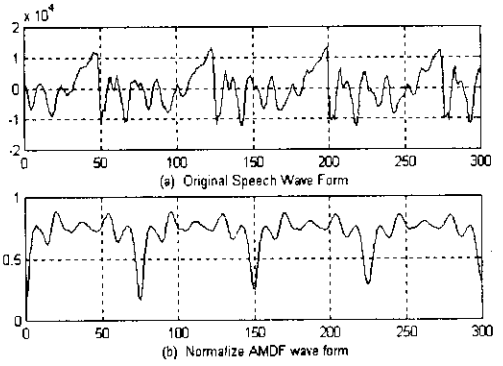

그림. 2-2 (a)음성파형
(b) NAMDF 파형

그림 2-1에서 보면 $Y=X^2$ 그래프와 $Y=|X|$ 그래프는 자기상관함수와 AMDF를 취했을 경우 피크점에서의 그래프를 나타내고 있다[4]. 영점 위치를 살펴보면 자기상관함수의 정확한 피크 값을 찾는 것이 AMDF의 피크 값을 찾는 것 보다 더 어렵다는 것을 볼 수 있다. 이러한 이유 때문에 피치검색시에 잘못된 피크 값을 얻게 됨으로써 피치검색시 오차를 발생시킬 수 있는 문제를 내포하고 있어 AMDF가 자기상관함수 대신에 주기성을 강조하는데 오랫동안 적용되어 왔다[5].

또한 AMDF는 곱셈을 사용하지 않는 장점이 있다. 단 규준화시 한 번의 나눗셈은 전체 계산량에 커다란 영향을 주지 않기 때문에 NAMDF의 장점을 유지할 수 있다.

본 논문에서는 NAMDF를 이용하여 피치를 검색하고 유사도 측정 구간을 정하였다. 그리고, 한 구간 안의 피크들의 변화는 Cross NAMDF법을 이용하여 측정할 수 있다. 본 논문에서는 Cross NAMDF법을 이용하여 포만트 유사도 측정에 적용하였다.

Ⅲ. 포만트의 유사도 측정

유성음 구간을 관찰하면 피치가 일정하게 유지되는 구간에서도 포만트는 조금씩 변화하는 것을 알수 있다. 이러한 포만트의 정보는 한 피치주기 사이에 나타나는 피크의 수와 모양, 크기, 위치 등에 좌우된다. 따라서 포만트의 유사도를 측정하기 위하여 기준 피치와 인근 피치 주기내에 나타나는 피크들의 특성을 비교하였다.

한 주기안에 나타나는 피크들의 특성을 비교하기 위하여 기준피치와 인근피치 한 주기 파형에 대해 Cross NAMDF를 수행하였다. Cross NAMDF는 식 (3.1)과 같다.

$$NAMDF_{Cross}(d) = \frac{\sum_{n=1}^{N} |S_{ref}(n) - S_p(n-d)|}{\sum_{n=1}^{N} |S_{ref}(n)| + |S_p(n-d)|} \quad (3.1)$$

여기서 S_{ref} 는 기준이되는 피치주기의 파형이고 S_p는 p번째 주기의 파형이다. N은 윈도우 크기이고 S_{ref}와 S_p 길이중 작은값이다. d는 지연인자이다.

그림 3-1은 기준 피치주기와 p번째 주기의 파형과의 Cross NAMDF를 수행한 결과이다. 단 같은 피치를 반복하여 두 주기로 만든 후에 수행하여 대칭적이다. 구해진 파형에 대한 면적은 식 (3.2)와 같이 구해진다.

$$A(p) = \frac{1}{N} \sum_{d=1}^{N} NAMDF_{Cross}(d) \quad (3.2)$$

여기서 $A(p)$는 p번째 파형의 Cross NAMDF 파형의 면적이다. 구해진 면적과 기준 피치주기의 NAMDF 파형의 면적을 비교하여 유사도를 측정한다.

유사도 측정은 식 (3.3)과 같다.

$$D(p) = \frac{|A_{ref} - A_p|}{A_{ref}} \times 100 \quad (\%) \quad (3.3)$$

A_{ref}는 기준파형의 NAMDF 파형의 면적이고, A_p는 식 (3.2)와 같이 구한 기준파형과 인접 파형의 Cross NAMDF 파형의 면적이다. D(p)는 p번째 파형의 포만트 유사도를 나타내며, 값이 작을수록 p번째 파형은 기준 파형과 유사하다. 그림 3-2는 D의 문턱값의 변화에 따른 압축률 변화를 보여주고 있다.

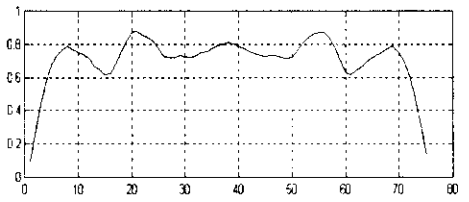

그림. 3-1 Cross NAMDF 파형

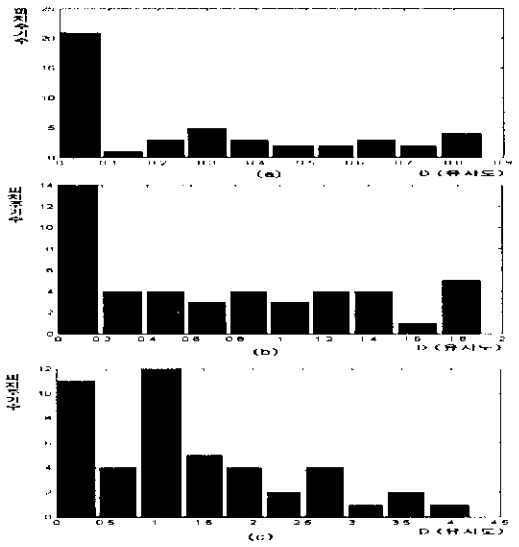

그림. 3-2 문턱값에 따른 피치주기의 압축률
첫막대(D=0) -> 전송되는 피치 주기 수
그외 (D>0) -> 압축되는 피치 주기 수
(a) D = 1를 문턱값으로 했을 때 (45.6%)
(b) D = 2를 문턱값으로 했을 때 (30.4%)
(c) D = 5를 문턱값으로 했을 때 (23.9%)

IV. PSOLA 기법에 의한 음성합성

본 논문에서는 음성신호를 복원할 때 스펙트럼 왜곡율과 복잡성이 적은 PSOLA 방법이 적합하다[6].

전송 또는 압축된 파형과 진폭정보와 피치정보를 이용하여 PSOLA합성을 수행한다[7].

그림 4-1은 PSOLA 기법으로 합성하는 과정을 나타내었다. 그림 4-1(a)은 원래 음성 파형이고, (b)는 피치정보를 표시한 그림이고, (c)는 합성을 위한 한 주기 파형이며, (d)는 PSOLA 방법을 이용하여 합성한 파형이다.

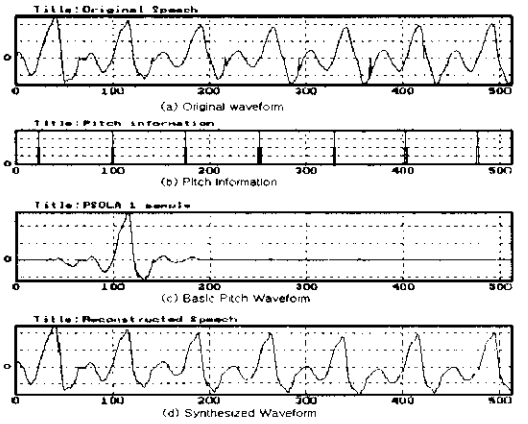

그림 4-1. 피치단위의 처리과정 예

V.실험 및 결과

본 논문에서 제안한 방법을 시뮬레이션하기 위해 IBM-PC/Pentium-150MHz에 마이크 입력이 가능한 16비트 A/D변환기를 인터페이스하여 11kHz의 표본화율로 16비트 양자화하여 저장하였다. 시뮬레이션시 피치분석 프레임단위를 256표본으로 사용하였으며, 피치주기 단위로 부호화 하였다.

그림 5-1은 본 논문에서 제안한 방법의 블럭도이다. 송신단에서는 먼저 한프레임에 대한 NAMDF법을 사용하여 피치를 구한다. 피치는 그림 2-2 (b)에서 가장 먼저 0점에 가까워지는 Valley 까지의 간격으로 정한다. 이렇게 구해진 피치에 일치하는 한 주기를 기준파형으로 정하고 저장하거나 전송한다. (제일 처음 기준파형은 처음 한 피치 주기에 해당되는 파형이다.) 기준파형의 진폭정보를 추출하고 기준 파형만의 NAMDF를 수행하여 기준면적을 구한다. 기준면적은 유사도가 문턱값을 넘어 기준 파형이 달라질 때 새로이 구해진다.

기준파형의 면적이 구해지면 처리된 파형의 피치만큼 전진하여 새로운 프레임을 잡고 NAMDF를 수행하여 피치를 구하고 진폭정보를 추출한다. 그 후 구해진 피치만큼의 파형을 기준 파형과 식(3.1)처럼 Cross NAMDF 수행하여 식(3.2)로 면적 A(p)를 구한다. 구해진 면적과 기준면적으로 식(3.3)처럼 유사도 D(p)를 측정한다.

유사도가 문턱값 보다 작다면 압축하고 위와 같은 과정을 반복한다. 만일 유사도가 문턱값 보다 크다면 그 주기를 기준 파형으로 하여 기준면적을 다시 구한 후 위와 같은 과정을 반복한다.

합성단에서는 전송된 파형과 피치정보, 진폭정보를 이용하여 PSOLA 방법으로 복원해낸다.

송신단에서 문턱값을 변화시킴으로써 압축율을 조정할 수 있다. 이렇게 하여 음성을 압축하였을 경우 압축율에 따른 결과를 표 5-1에 나타내었다. 표 5-1에서 볼 수 있듯이 전체 음성의 45%로 압축 수행결과 약 4.1의 MOS를 얻었고 38.8%, 30.4%, 23.9%일 때 각각 3.9, 3.7, 3.1의 MOS를 얻을 수 있었다.

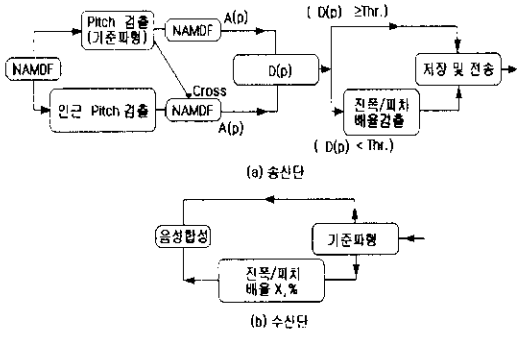

그림 5-1. 제안한 방법의 블록다이어그램

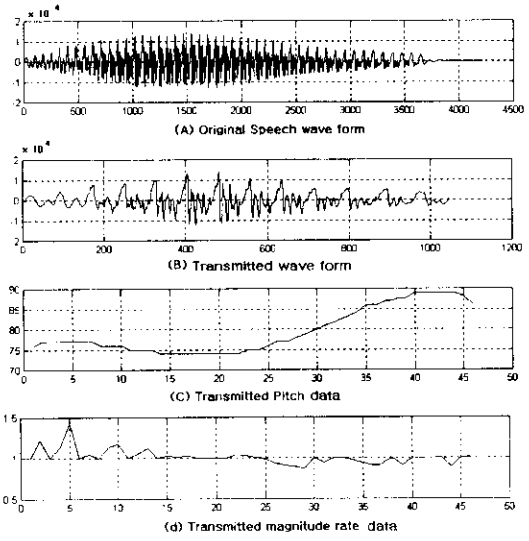

그림.5-2 '아' 음성에 대한 부호화
(A)음성파형 (B)전송되는 파형
(C)전송되는 피치정보 (C)전송되는 진폭(변화)정보

표.5-1 압축율에 따른 MOS

압축률 (부호화된 음성/전체 음성) * 100	MOS
45.6%	4.1
38.8%	3.9
30.4%	3.7
23.9%	3.1

VI. 결론

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 범용칩으로도 합성이 가능한 새로운 방법을 제안하였다.

우선 NAMDF로 피치를 검색하여 기준 파형을 얻고 각 피치구간 별로 유사도를 측정한다. 유사도의 문턱값을 정하여 파형을 압축할 것인가를 결정한다. 압축할 경우에는 진폭과 피치정보만을 저장하거나 전송한다. 결과 전체 음성의 45%정도로 압축하여도 MOS 4.1을 유지하는 것을 볼 수 있었다.

본 논문에서 제안한 음성부호화법은 유성음만 압축을 수행하고 있으나, 무성음 및 묵음에 대해서도 압축을 수행한다면, 좀더 높은 압축률을 얻을 수 있다. 본 논문에서 제안하는 음성부호화법의 특징은 알고리즘이 매우 간단하다는 특징이 있다. 따라서 음성부호화법을 이용하여 상품화하려는 분야에 본 논문에서 제안한 방법을 이용하여 음성데이터를 압축하여 전송하거나 저장할 경우 저가의 범용칩을 이용하여 상품화할 수 있으므로 대외경쟁력을 가질 수 있다.

VII. 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] 배명진 외 1인, "On Detecting the Steady State Segments of Speech Waveform by using the Normalized AMDF", 대한전자공학회지, Vol.14, No.1, pp.600-603, Jun., 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] F. Chapentier, M. G. Stella, "Diphone Synthesis Using Overlap-add Technique for Speech Waveforms Concatination", ICASSP 86, pp.2015-2018, 1986

[7] E. Moulines and F. Charpentier, "Pitch- synchronous waveform processing techniques for test to-speech synthesis using diphones," Speech Comm., vol. 9 no. r, pp. 453-467, 1990