

음성전보 시스템 구현에 적합한 음성 부호화 알고리즘

A Vocoder Algorithm for Implementation of Voice Telegram System

정 찬 중, 배 명 진
(Chan Joong Jung, Myung Jin Bae)

요 약

전보(Telegram)는 우리의 생활에서 필요한 정보통신 시스템이다. 지금까지 전보의 기능은 긴급하거나 짧은 정보의 내용을 전화국의 메시지를 통해 문자위주로 변환하여 수신자에게 전달하고있다. 이와는 달리 음성 전보 시스템(Voice Telegram System)은 기존의 문자전달방식에 음성을 가미한 것으로서 발신자의 목소리도 함께 기록하여 수신자에게 전달하는 방식이다.

본 논문에서 음성 전보에 적합한 알고리즘을 새로이 제안한다. 유/무성음 판별 후 유성음에 대해 피치주 기별로 유사도를 측정하여 유사도가 낮은 파형은 전송하고 유사도가 높은 파형은 피치와 에너지 정보만 전송하는 방법이다. 실험결과 전체 음성에 대해서 41%의 압축율과 MOS 테스트 결과 4.1을 얻을 수 있었다.

1. 서 론

전보는 오랫동안 우리 생활과 밀접한 관계를 가지면서 이용되어 왔으며, 만나서 전할 수 없는 진술하고 다양한 내용을 짧은 메시지에 담아서 상대방에게 전달하는 매체로서 전보가 사용된다.

전보수신자에게 지금까지 전달되는 문자위주의 전보전달과 함께 발신자의 목소리도 함께 전송하는 목소리 전보시스템을 음성전보 시스템이라고 한다.

음성 전보 시스템에 필요한 보코더는 계산량이 적어야 한다. 그런면에서는 CELP나 IMBE보다 시간영역 부호화법이 적합할 수 있다. 하지만 시간영역 부호화법의 문제점은 압축을 많이 할수록 음질의 열화가 생긴다는 단점이 있다.

전송형 부호화기법을 음성 보코더 시스템에 응용하려면 분석 및 부호화된 데이터를 메모리에 저장하고, 합성 시에 메모리에 있는 데이터를 복호화하여 분석의 역과정으로 음성신호를 합성하면 된다. 따라서 전송형 부호화법은 기본적으로 분석과정과 합성과정의 처리속도와 데이터율이 거의 비슷해야만 일정한 전송률로 송수신을 할 수 있게 된다.

메모리형 부호화 시스템을 구성하려고 할 때는 합성을 위해 작성되는 데이터베이스를 실시간으로 구할 필요가 없다. 차라리 그 분석과정은 복잡하게 처리하여도 복호화 하는 시간을 줄여주는 것이 메모리형 부호화시스템에서는 바람직하게 된다. 그리고 메모리에 저장되는 합성용 데이터는 저장률이 정해져 있지 않으므로 합성에 불필요한 성분을 제

거한 후에 저장시킬 수 있기 때문에 같은 음질에서도 메모리의 양을 감소시킬 수 있다.

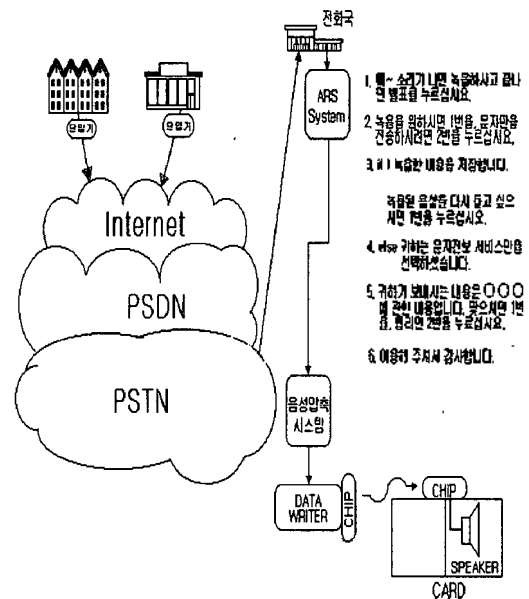


그림 1-1. 보이스 전보 시스템 구성도

그림 1-1은 음성 전보 시스템의 구성도이다. 먼저 사용자가 인터넷이나 전화 또는 직접 방문하여 전보를 신청하여 전하고자 하는 음성 메시지를 녹음하면 음성 압축을 수행하여 데이터량을 줄이고 이미 만들어진 프로그램 된 칩속에 사용자의 데이터만 기록하게 된다. 그런 후 카드에 부착하고 문자 메시지와 함께 수신자에게 전달된다. 수신자가

카드를 펼치면 기록된 음성 메시지는 칩속에 프로그램 된 음성 복호화 알고리즘에 의해 다시 복원되어 보낸 사람의 목소리로 메시지를 전달하게 된다. 이러한 과정에서 사용자는 압축하고 복원되어진 자신의 메시지를 보내기 전에 다시 들어봄으로써 음질에 대한 송신 여부를 결정하게 된다.

2. 제안한 음성 부호화 알고리즘

2-1. 피치분석법

음성신호 중 유성음 구간을 관찰하면 피치가 일정하게 유지되는 것을 알 수 있다. 따라서 유사도를 측정하기 위하여 인접한 피치 주기에 나타나는 피크들의 특성을 비교하였다.

한 주기 안에 나타나는 피크들의 특성을 비교하기 위하여 기준피치 구간을 정한다. 그리고 인근피치 주기에 대해 Cross NAMDF를 수행하였다. Cross NAMDF는 식 (2-1)와 같다.

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

여기서 $s(n)$ 은 음성신호이고 N 은 NAMDF를 구하려는 윈도우 구간이다. 지연인자 d 를 점차 증가시키면서 NAMDF를 구해보면, 지연인자가 프레임 내 음성피치에 정수배가 될 때마다 NAMDF는 거의 영이 된다.

NAMDF는 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 파형의 유사도 측정에 적용하였다.

2-2. 파형의 유사도 측정에 의한 압축 방법

한 주기 안에 나타나는 피크들의 특성을 비교하기 위하여 기준피치 구간을 정한다. 그리고 인근피치 주기에 대해 Cross NAMDF를 수행하였다. Cross NAMDF는 식 (2.2)와 같다.

여기서 S_{ref} 는 기준이 되는 피치주기의 파형이고 S_p 는 p 번째 주기의 파형이다. N 은 윈도우 크기이고 S_{ref} 와 S_p 길이 중 작은 값이다. d 는 지연인

자이다. 구해진 파형에 대한 면적은 식 (2.3)과 같이 구해진다.

$$NAMDF_{Cross}(d) = \frac{\sum_{n=1}^N |S_{ref}(n) - S_p(n-d)|}{\sum_{n=1}^N |S_{ref}(n)| + |S_p(n-d)|} \quad (2.2)$$

$$A(p) = \frac{1}{N} \sum_{d=1}^N NAMDF_{Cross}(d) \quad (2.3)$$

여기서 $A(p)$ 는 p 번째 주기의 Cross NAMDF 파형의 면적이다. 구해진 면적과 기준 피치주기의 NAMDF 파형의 면적을 비교하여 유사도를 측정한다. 유사도 측정은 식 (2.4)와 같다.

$$D(p) = \frac{|A_{ref} - A_p|}{A_{ref}} \times 100 (\%) \quad (2.4)$$

A_{ref} 는 기준이 되는 피치주기를 NAMDF하여 구한 파형의 면적이고, A_p 는 식 (2.3)과 같이 구한 기준 피치주기와 인근 피치주기의 Cross NAMDF 파형의 면적이다. $D(p)$ 는 p 번째 주기의 유사도를 나타내며, 값이 작을수록 p 번째 주기는 기준 피치구간과 유사하다.

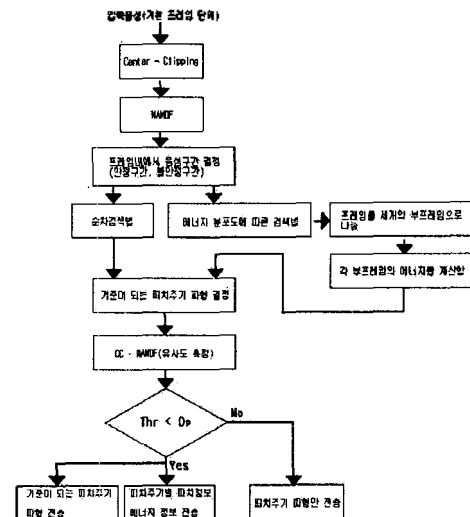


그림 2-1. 부호화단

그림 2-1은 제안한 부호화기의 블록다이어그램이다. 부호화단에서 한 프레임에 대한 NAMDF법을 사용하여 피치를 구한다. 피치는 가장 먼저 0점에 가까워지는 Valley까지의 간격으로 정한다. 이렇게 구해진 피치에 일치하는 한 주기를 기준 파형으로 정하고 저장한다. 기준 파형의 진폭정보를 추출하고 기준 피치주기만의 NAMDF를 수행하여 기준면적을 구한다. 기준면적은 유사도가 문턱 값

이하로 되어 기준 파형이 달라질 때 새로이 구해진다.

기준 면적이 구해지면 처리된 파형의 피치만큼 전진하여 새로운 프레임을 잡고 NAMDF를 수행하여 피치를 구하고 진폭정보를 추출한다. 구해진 피치만큼의 파형을 기준 파형과 식(2.2)처럼 Cross NAMDF 수행하여 식(2.3)로 면적 A(p)를 구한다. 구해진 면적과 기준면적으로 식(2.4)처럼 유사도 D(p)를 측정한다.

D(p)가 문턱 값 보다 작으면 유사하므로 압축하고, D(p)가 문턱 값 보다 크다면 그 주기를 기준 파형으로 하여 기준면적을 다시 구한 후 위와 같은 과정을 반복한다.

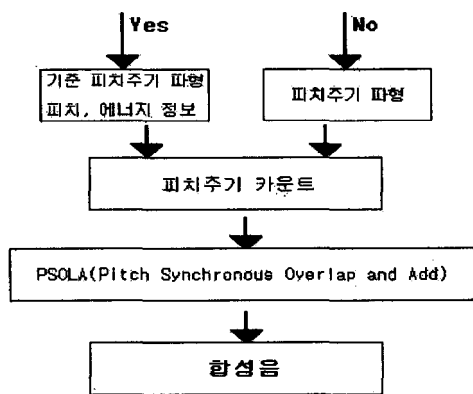


그림 2-2. 복호화단

음성정보에서는 복잡도가 낮은 합성법을 사용해야 한다. 그래서 본 논문에서 사용한 합성 방법은 PSOLA(pitch synchronous overlab add) 방법이다. 이 방법은 음성신호를 복원할 때 스펙트럼 왜곡률과 복잡성이 낮은 장점이 있다[7][8].

3. 실험 및 결과

본 논문에서 제안한 방법을 시뮬레이션하기 위해 IBM-PC/Pentium-150MHz에 마이크 입력이 가능한 16비트 A/D변환기를 인터페이스하여 8kHz의 표본화율로 16비트 양자화하여 저장하였다. 시뮬레이션 시 피치분석 프레임단위를 256표본으로 사용하였으며, 피치주기 단위로 부호화 하였다. 실험에 사용한 음성 데이터는 남, 여 화자가 발성하여 수집하였다.

음성시료

1. 순산을 축하하며 산모의 건강을 기원합니다.
2. 개업을 축하합니다.
3. 합격을 진심으로 축하합니다.

4. 결혼을 진심으로 축하하며 복된 가정 이루길 기원합니다.

부호화단에서 문턱 값을 변화시킴으로써 압축율을 조정할 수 있다. 이렇게 음성을 압축하였을 경우 압축율에 따른 결과를 표 3-1에 나타내었다. 유/무성을 비율은 67/33이고 전체음성 길이는 14,080 샘플이다.

표 3-2는 시료5에 결과를 나타내며, 유/무성비율은 72/23이고 전체음성 길이는 26,880 샘플이다.

표 3-1. 시료1에 의한 결과표

시료	유사도 (%)	압축 샘플수	전송 샘플수	전송율	MOS
시료 1	5	2,143	11937	84.78	4.2
	10	2,591	11489	81.60	4.1
	15	2,991	10707	78.76	4.0
	20	3,373	10239	76.04	3.8
	30	3,841	10099	72.72	3.7
	40	3,981	9457	71.73	3.5
	50	4,623	9228	67.17	3.5

표 3-2. 시료5에 의한 결과표

시료	유사도 (%)	압축 샘플수	전송 샘플수	전송율	MOS
시료 5	5	4,323	22557	83.92	4.2
	10	5,887	20993	78.10	4.0
	15	6,914	19966	74.28	3.8
	20	7,536	19344	71.96	3.7
	30	8,443	18437	68.59	3.7
	40	9,080	17800	66.22	3.6
	50	10,006	16874	62.78	3.5

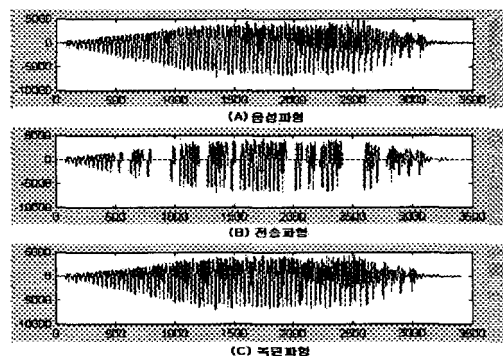


그림 3-1. '우'발음에 대한 부호화,복호화

시료 1에 대한 결과 파형

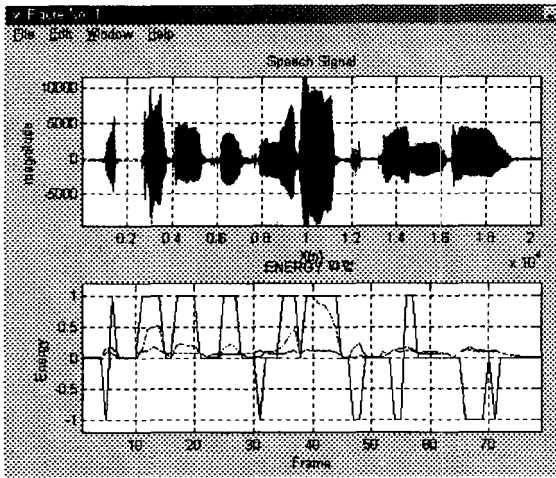


그림 3-2. 음성파형과 유/무성음 판별한 결과

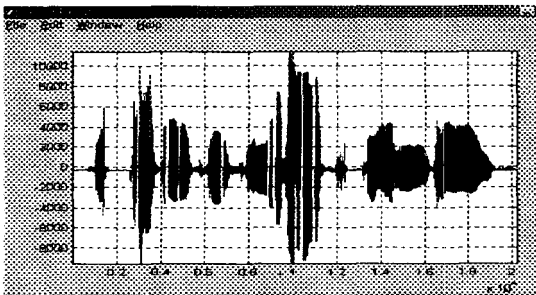


그림 3-3. 부호화한 결과

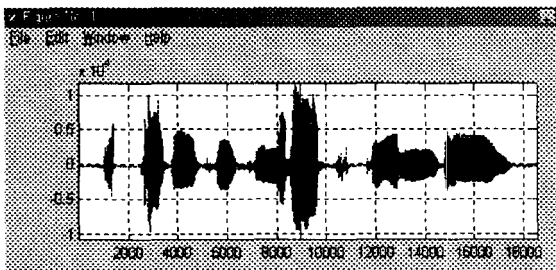


그림 3-4. 복호화한 결과

4. 결론

본 논문에서는 보이스 전보 시스템에 필요한 간단하고 압축율이 높은 새로운 부호화 방법을 제안하였다. 먼저 NAMDF로 피치를 검색하여 기준 파형을 얻고 각 피치구간 별로 유사도를 측정한다. 유사도의 문턱 값을 정하여 파형의 압축 여부를 결정한다. 압축할 경우에는 진폭과 피치정보만을 저장한다. 실험결과 전체 음성의 50%정도로 압축하여도 MOS 3.3을 유지하는 것을 볼 수 있었다.

제안한 음성부호화법은 유성음압축은 물론 무성음의 일부도 압축을 하였다. 본 논문에서 제안하는 음성 부호화법의 특징은 알고리즘이 매우 간단하다는 특징이 있다. 따라서 보이스 전보와 같이 음성 부호화법을 이용하여 상품화하려는 분야에 제안한 방법을 이용할 경우 저가의 범용칩을 이용하여 상품화할 수 있으므로 대외 경쟁력을 가질 수 있다.

참고 문헌

- [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.
- [4] A.M. Kondozi "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] 배명진, "디지털 음성부호화", 동영출판사, 1996
- [7] F. Charpentier, 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 no. 1, pp. 453-467, 1990
- [9] 나덕수 외 2인, "보이스 전보 시스템 구현을 위한 저가형 음성 파형 부호화 알고리즘", *음향학회지*, 19권 2호, pp 101-105, 2000. 2월
- [10] 배명진, "디지털 음성합성", 동영출판사, 1998
- [11] 배명진, "디지털 음성분석", 동영출판사, 1998