2000年度 第13回 信號處理合同學術大會 論文集 第13卷 1號

CELP 보코더 전송률 감소를 위한 발성속도 측정 방법

장 경 아, 나 덕 수, 배 명 진
숭실대학교 정보통신공학과
전화 : (02) 824-0906 / 팩스: (02) 820-0018

On a Study of Measurement Method of Utterance Velocity for the Reduction of Transmission Rate in CELP Vocoder.

KyungA JANG, MyungJin BAE
Dept. Information and Telecommunication Engr., Soongsil University
E-mail : kajang@assp.ssu.ac.kr

Abstract

음성의 발성속도가 빠른 경우에는 발성속도가 느린 경우보다 적은 정보만으로도 부호화가 가능하다. 음성의 발성속도가 빠른 경우에는 청취시 낮은 주파수 대역의 정보가 높은 주파수 대역의 정보보다 중요하게 된다. 음성 부호화 기술은 전송률과 복잡도를 줄이고 음질을 향상시키는 방향으로 진행되고 있다. 현재 상용화되고 있는 CELP형 보코더는 낮은 전송률에 비해 우수한 음질을 제공하지만, 기존 방식은 음성의 발성속도에 대해서 처리를 달리하지 않고 사용하고 있다. 음성의 발성속도를 측정하여 발성속도가 빠를 경우에, 발성속도가 느린 경우보다 낮은 대역의 정보만 전송한다면 전송율을 감소시킬 수 있다. 본 논문에서는 CELP 부호화기의 전송률 감소를 위해 발성속도를 측정하는 방법을 제안한다. LSP 파라미터가 가지고 있는 정보로 음소의 변화율을 측정하였다. 각각 다른 발성속도를 갖는 음성시료에 대하여 음소 변화율을 구한 결과 발성속도가 다른 경우, 뚜렷하게 다른 음소 변화율을 갖는 것을 알 수 있었고, 빠르게 발성한 경우가 느리게 발성한 경우보다 42.8%가 높게 나왔다.

I. 서 론

정보통신 문화의 발달에 따라 디지털 이동통신이나 멀티미디어, 음성우편 시스템 등 음성을 통한 여러 가지 새로운 산업들이 급속히 성장하고 있다. 이중에서도 특히 디지털 이동통신분야와 인터넷 기반 멀티미디어 전송 분야에 적용하기 위한 음성신호의 디지털 변환과 전송 데이터 량을 줄이기 위한 여러 음성 부호화 기술이 사용되어지고 있다. 최근 CELP형 보코더는 낮은 전송률에 비해서 우수한 음질을 제공하고 있다[1].

기존의 CELP 보코더에서 음성의 발성 속도에 상관없이 음성 신호 프레임을 부호화 하였다. 본 논문에서는 음성 신호의 인접 프레임간의 LSP 파라미터의 차를 구해, 이를 이용하여 낮은 계산량으로 발성 속도를 나타내는 음소변화율을 구하는 방법을 제안한다. 먼저 II장에서는 현재 상용되고 있는 G.723.1에서 LSP를 처리하는 과정을, III장에서는 발성속도에 따른 음소변화율을 측정하는 방법을 제안한다. IV장에서는 실험 및 결과, V장에서는 결론을 맺겠다.

II. CELP 부호화기

현재까지 발표된 음성부호화기 중 가장 많은 연구가 이루어지고 있는 방식은 그림 2-1의 구조를 가지고 있는 CELP(Code Excited Linear Prediction)구조이다. 이 방식은 4.8kbps내외의 전송율에서 양호한 음질을 얻을 수 있으며 ITU-T, TIA/EIA 등을 다양한 응용분야에서 표준화가 이루어지고 있다. 특히 한국에서는 PCS 및 전화기 라인상에서의 인터넷을 통한 화상회의를 위하여 낮은 전송률에서 고음질을 가지는 코덱이 많은 주목을 받고 있다[1]. CELP 계열 보코더들 중에서 G.723.1은 멀티미디어 통신 환경하의 음성 전송 표준 보코더로 개발되었다[2].

G.723.1은 5.3/6.3kbps의 이중 전송률을 갖는 구조로 현재 별정 통신으로 상용화되는 인터넷폰과 그 외의 이동 통신용 보코더로 사용되어지고 있으며 낮은 전송률에 비해서 우수한 음질을 제공하고 있다. 더불어 최적의 전송 환경을 위하여 두 개의 전송률을 사용하기 때문에 다른 보코더 표준안들에 비해서 더욱 응용성이 높다[3]-[7]. 현재 사용되는 음성 코덱(codec)이나 인식기에서 음성신호를 분석하여 전송형이나 저장형 파라미터로 변환하는데 LSP 파라미터를 사용한다. LSP 파라미터는 양자화 에러에 강하고 시스템의 안정성이 보장되고 일정한 스펙트럼 민

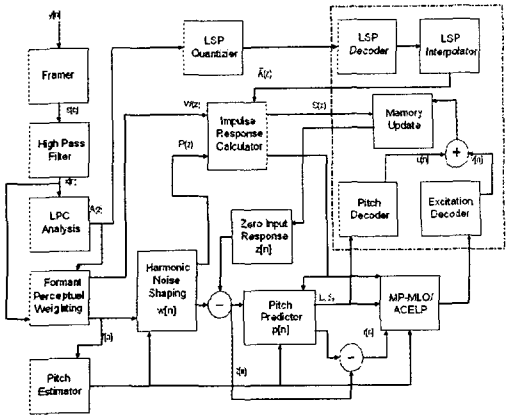

그림 2-1. G723.1 보코더 블록도

감도와 저전송률 부호화에서 낮은 스펙트럼 왜곡, 좋은 선형보간 특성을 가진다. G.723.1에서는 음성신호를 전송하기 위해 LPC분석을 거쳐 LSP를 구해 양자화 하여 전송한다. 이 과정은 다음과 같다.

LSP 양자화 과정

첫 번째로, 약간의 부가적인 대역폭 확장(Bandwidth Expansion) (7.5Hz)이 수행된다. 그리고 그 결과로 $A_3(Z)$ LP 필터는 예측 분산 벡터 양자화기(Predictive Split Vector Quantizer)를 이용하여 양자화 되어진다. 양자화는 다음의 과정에 따라 이루어진다.:

1. LP 계수 $\{a_j\}_{j=1..10}$ 는 단위원과 영교차에 의한 인터폴레이션 동안의 검색에 의해 LSP 계수 $\{p'_j\}_{j=1..10}$로 변환된다.

2. DC 성분 p_{DC} 는 LSP 계수 p'로부터 제거되고 그러면 LSP 벡터 p가 제거된 새로운 DC가 얻어진다.

3. 1차 제한 예측기 b=(12/23)는 시간(프레임) n에서 DC 제거 예측된 LSP 벡터 $\overline{p_n}$과 잔여 LSP 에러 벡터 값을 얻기 위해서 이전에 복호화된 LSP 벡터 $\tilde{p}_{n-1}$에 적용된다.

$$P^T{}_n=[p_{1,n}p_{2,n}\cdots\cdots p_{10,n}] \tag{2.1}$$

$$\overline{P}^T{}_n=[\overline{P}_{1,n}\overline{P}_{2,n}\cdots\cdots\overline{P}_{10,n}] \tag{2.2}$$

$$\overline{P}_n=b[\overline{P}_{n-1}-P_{DC}] \tag{2.3}$$

$$e_n=P_n-\overline{P}_n \tag{2.4}$$

4. 양자화 되지 않은 LSP 벡터 p_n, 양자화 된 LSP 벡터 $\tilde{p}_n$, 잔여 LSP 에러 벡터 e_n 은 각각 3, 3, 4의 크기로 3개의 부프렘임으로 각각 나누어진다. 각 m번째 부벡터는 8비트 코드북을 사용하여 양자화된 벡터이다. 에러 표준치 $E_{l,m}$이 최소가 되는 적당한 부벡터 코드북 기록의 인덱스 l은 선택되어진다.

$$P'^T{}_m=[p'_{1+3m}p'_{2+3m}\cdots p'_{K_m+3m}],\ K_m=\begin{cases}3, m=0\\3, m=1\\4, m=2\end{cases} \tag{2.5}$$

$$\tilde{P}^T{}_{l,m}=[\tilde{p}_{1,l,m}\ \tilde{p}_{2,l,m}\cdots\tilde{p}_{K_m,l,m}],\ \begin{cases}0\le m\le 2\\1\le l\le 256\end{cases} \tag{2.6}$$

$$P'=P+P_{DC} \tag{2.7}$$

$$\hat{P}_{l,m}=\overline{P}_{m+}P_{DC_m}+e_{l,m},\ \begin{cases}0\le m\le 2\\1\le l\le 256\end{cases} \tag{2.8}$$

$$E_{l,m}=(p'_m-\tilde{p}_{l,m})^T W_m(p'_m-\tilde{p}_{l,m}),\ \begin{cases}0\le m\le 2\\1\le l\le 256\end{cases} \tag{2.9}$$

여기서 $e_{l,m}$은 m째로로 분산된 장여 LSP 코드북의 l번째 기록이고, W_m은 직교 가중 행렬식이다. 이것은 양자화 되지 않은 LSP 계수 벡터 p로부터 결정되고 가중치는 다음과 같이 정의된다.

$$W_{j,j}=\frac{1}{\min\{p'_j-p'_{j-1},p'_{j+1}-p'_j\}},\ 2\le j\le 9 \tag{2-10}$$

$$W_{1,1}=\frac{1}{p'_2-p'_1} \tag{2-11}$$

$$W_{10,10}=\frac{1}{p'_{10}-p'_9} \tag{2-12}$$

선택된 인덱스들은 채널로 전송된다.

III. 발성속도에 따른 음소변화율 측정

본 논문에서 제안하고자 하는 알고리즘은 CELP형 부호화기에서 음성신호를 고려하지 않고 전송하는 점을 감안하여 음성신호의 발성속도에 따라 전송

파라미터를 달리 보내려고 할 때, 이 파라미터를 달리 보내려면, 발성속도를 측정을 해야한다.부호화기를 위한 낮은 계산량으로 구현 가능한 발성속도 측정은 시간에 따라 음소 변화율을 구하는 것이다. 음소 변화율을 측정하는 순서 개요도는 다음과 같다.

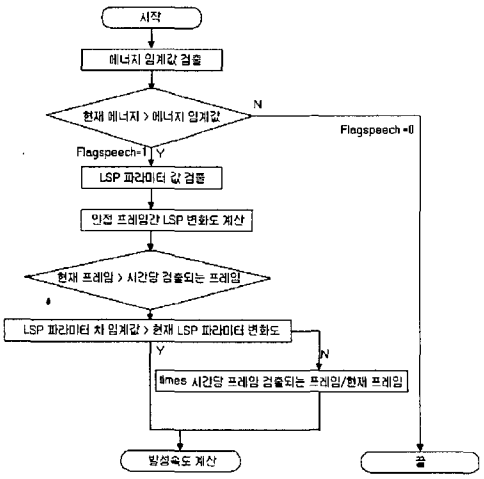

그림 3-1. 음소변화율 측정 과정 순서도

먼저 입력된 음성신호를 묵음과 음성이 있는 구간을 판정하기 위해 에너지 문턱값을 결정하고 이 문턱값이 넘는 음성 프레임만을 가지고 음소 변화율을 구한다.

$$E'(n)=\begin{pmatrix} 1, & E(n)>Thres. \\ 0, & E(n)<Thres. \end{pmatrix} \quad (3.1)$$

N은 윈도우 크기이고 E(n)은 단구간 에너지이고 Thres.는 음성구간을 검출하기 위한 문턱값이다. Thres.는 처음 5 프레임 정도의 에너지 평균이다. 이 값을 이용하여 단구간에너지를 0과 1로 단순화하여 E'(n)을 얻는다. E'(n)의 값이 1로 판정된 프레임은 순차적으로 각 프레임에 대하여 LSP 파라미터를 구한다. 프레임에서 구해진 각 LSP 파라미터 값을 가지고, 인접 프레임간의 같은 차수의 LSP 파라미터의 차이를 구한다. 이것을 프레임간의 LSP 파라미터 변화도로 본다. LSP 파라미터의 변화도인 $LSPdiff$는 다음과 같이 구해진다.

$$LSPdiff(i)=\sum_{n=0}^{N}|LSP_{n+1}(i)-LSP_n(i)|,\ i=1,2\ldots10 \quad (3.2)$$

인접 프레임간의 LSP 파라미터의 변화도를 측정한 결과를 가지고, 변화가 두드러진 구간에 대한 문턱값을 측정한다. LSP 파라미터는 연속되는 음성구간에서는 인접 프레임간 그다지 크게 변화하지 않는다는 특성을 감안하여 얻어진 LSP 파라미터의 변화도로 각 3프레임씩 순차적으로 변화도를 다시 측정한다. 이 변화도를 전체 프레임동안 구한 식은 다음과 같다.

$$diff_{mean}=\sum_{m=1}^{N}\left\{\frac{1}{3}\sum_{n=1}^{n+2}diff(n)\right\},\ \begin{cases} m=the\ number\ of\ frame, \\ n=nth\ FRMAE \end{cases} \quad (3.3)$$

FrPerSec은 시간당 샘플링 되는 프레임 수이다. $diff_{mean}$ 는 한정된 인접 프레임간의 변화도의 평균값이다. 음소의 변화율은 시간에 따라 변화되는 음소의 비율을 측정한다. 따라서 현재 프레임이 FrPerSec을 넘는 경우와 넘지 않은 경우를 달리하여 처리하여야 한다.

$$FrPerSec=Fs/N \quad (3.4)$$

i) $n<FrPerSec,\ diff_n(i)>thDiff$

$$SpRate(n)=\frac{diff(n)}{diff_{mean}},\ n=nth\ FRAME \quad (3.5)$$
$$SpRate'=FrSecPer/n\times SpRate(n)$$

ii) $n>FrPerSec,\ diff_n(i)>thDiff$

$$SpRate'(n)=\frac{diff(n-FrPerSec)}{diff_{mean}},\ n=nth\ FRAME \quad (3.6)$$

$thDiff$ 은 측정된 $LSPdiff$ 에서 두들어진 변화를 보이는 값을 문턱값으로 결정된 것이다. 각 프레임간의 LSP파라미터 변화도를 나타내는 $diff(i)$값이 이 문턱값을 넘었을 경우, 음소율을 계산하게 된다. 따라서 $SpRate$ 는 시간당 변화하는 음소율, 발성속도율이라 한다.

Ⅳ. 실험 및 결과

제안한 방법을 실험하기 위해서 먼저 IBM PC(233 MHz)에 마이크 입력이 가능한 A/D 변환기를 인터페이스 하였다. 음성시료는 남자와 여자가 연구실 환경(30dB의 SNR)에서 발성한 음성을 8kHz로 표본화하고 16bit로 양자화하여 사용하였다. 발성한 문장은 다음과 같다.

발성1) "아야어여우유으이"
발성2) "여기는 음성통신 연구실입니다"
발성3) "일이삼사오육칠팔구십"
발성4) "아름다운 가을입니다"

음성시료는 발성속도를 각각 다르게 하여 같은 문장을 발성하였다.

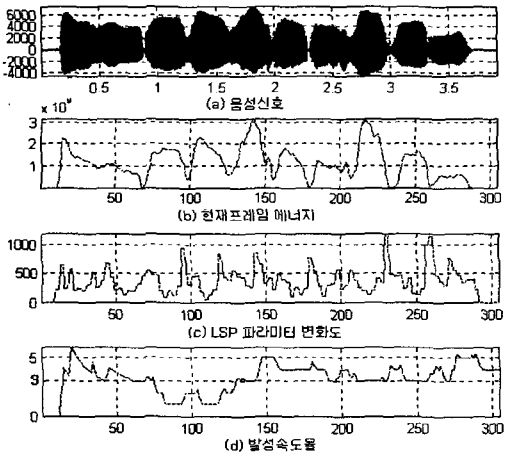

그림4-1. 빠르게 발성한 경우의 음소변화율

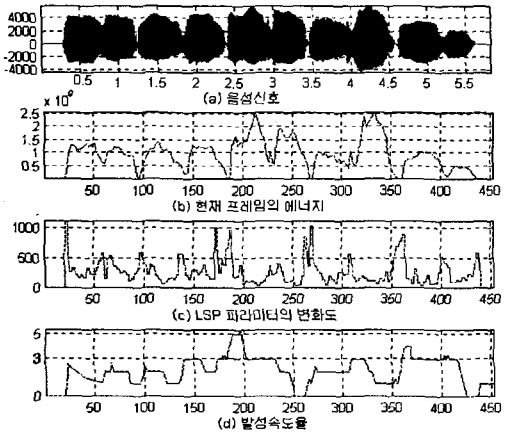

그림4-2. 느리게 발성한 경우의 음소변화율

제안한 음소변화율 알고리즘은 C언어로 구현하여 각각 걸리는 시간과 각 발성문장에 대한 음소변화율을 측정하였다. 그림은 음소변화율을 구하기 위해 각 파라미터들을 구하는 과정이다. (b)는 입력된 음성에 대해 각각의 프레임마다 현재 에너지를 구한 그림이다. (c)는 인접 프레임간 각 차수의 LSP 파라미터의 변화도를 나타낸 것이다. 음소가 지속되는 경우에는 LSP 변화도가 적은 반면에, 음소가 변화는 경우에는 LSP의 변화도의 값이 큰 것을 볼 수 있다. (d)는 음소 변화율을 나타낸 것이다. 빠르게 발성한 경우와 느리게 발성한 경우에 음소변화율의 값이 뚜렷하게 달리 나타나는 것을 볼 수 있다. 이 음소변화율의 측정치에서 빠르게 발성한 경우가 느리게 발성한 경우보다 약 42.8%가 높게 나왔다. 이 결과는 발성속도에 따라 다른 변화율을 가진다는 것을 알 수 있고, 발성 속도에 따라 빠르게 발성한 경우에는 느린 발성보다 변화율이 높다는 것을 알 수 있었다.

Ⅳ. 결 론

음성을 부호화 하는 기술은 전송률과 복잡도를 줄이고 음질을 향상시키는 방향으로 진행되고 있다. 현재 상용화되고 있는 CELP형 보코더는 낮은 전송률에 비해 우수한 음질을 제공하지만, 음성의 발성속도에 대해서는 처리를 달리하지 않고 사용하고 있다. 음성신호 부호화시 빠른 경우, 발성속도가 느린 경우보다 적은 정보만으로 부호화가 가능하다. 음성이 빠르게 발성된 경우에는 듣는 이로 하여금 낮은 주파수 대역의 정보가 높은 주파수 대역의 정보보다 중요하게 작용한다. 따라서 음성의 발성속도를 측정하여 발성속도가 빠를 경우에는 발성속도가 느린 경우보다 낮은 대역의 정보만을 전송한다면 전송률 감소를 시킬 수 있다. 본 논문에서는 HELP 부호화기의 전송률 감소를 위해 발성속도를 측정하는 방법을 제안한다. ESP 파라미터가 가지고 있는 정보를 이용하여 음성의 발성속도에 따른 음소변화율을 구한 결과 빠르게 발성한 경우가 느리게 발성한 경우보다 42.8%가 높게 나왔다.

표 4-1. 시간당 음소변화율

	발성 시간		음성 발성속도 변화율	
	Fast	Slow	Fast	Slow
발성(1)	4.62	7.10	3.879	2.264
발성(2)	3.21	6.52	4.149	3.187
발성(3)	4.2	6.53	3.472	2.129
발성(4)	2.3	4.53	3.432	1.103
평균	3.58	6.10	3.733	2.170

참고 문헌

[1] 배역진, "디지털 음성분석", pp.95-120, 동영출판사, 1998. 4

[2] L R. Rabiner, R.W Schafer, " Digital Processing of Speech Signal", pp.38-115, Prentice Hall, 1978

[3] A. M. Kondoz, "Digital Speech", pp. 84-92, John Wiley & Sons Ltd, 1994.

[4] ITU-T Recommendation G.723.1, March, 1996

[5] John R. Deller, Jr., John G. Proakis, John H.L. Hansen, "Discrete-Time Processing of Speech Signals", pp.124-125, Maxwell Macmillan International, 1993.

[6] Sadaoki Furui, "Digital Speech Processing, Synthesis, and Recognition", pp129, MARCEL DEKKER, INC. 1991.