각국 언어 특성에 독립적인 CELP 계열 보코더에서의 계산량 단축 알고리즘

주상규*
*(주)주성시스템
e-mail:dkmadl@naver.com

The Computation Reduction Algorithm Independent of the Language for CELP Vocoders

Sang-Gyu Ju*
*JuSung System Co., Ltd.

요 약

.In this paper, we propose the computation reduction methods of LSP(Line spectrum pairs) transformation that is mainly used in CELP vocoders. In order to decrease the computational time in real root method the characteristic of four proposed algorithms is as the following. First, scheme to reduce the LSP transformation time uses mel scale. Developed the second scheme is the control of searching order by the distribution characteristic of LSP parameters. Third, scheme to reduce the LSP transformation time uses voice characteristics. Developed the fourth scheme is the control of searching interval and order by the distribution characteristic of LSP parameters. As a result of searching time, computational amount, transformed LSP parameters, SNR, MOS test, waveform of synthesized speech, spectrogram analysis, searching time is reduced about 37.5%, 46.21%, 46.3%, 51.29% in average, computational amount is reduced about 44.76%, 49.44%, 47.03%, 57.40%. But the transformed LSP parameters of the proposed methods were the same as those of real root method.

1. 서 론

과거의 음성 정보를 전송하는 방법은 한정된 일부 사용자와 특정 분야에 적용되었기 때문에 전송률은 크게 고려되지 않았으며 고음질의 음성을 수신 단에 보낼 수 있으면 되었다. 그러나 과거에 비해 정보통신 문화의 발달에 따라 디지털 이동통신이나 멀티미디어, 음성우편 시스템 등 음성을 이용한 여러 가지 새로운 산업들이 급속히 성장하고 특정 집단이나 개인에게 제공되었던 서비스들이 다수의 대중에게도 제공되어지고 그 숫자 역시 기하 급수적으로 늘어나게 되었다. 이중에서도 특히 디지털 이동통신분야와 인터넷 기반 멀티미디어 전송 분야에 적용하기 위한 음성신호의 디지털 변환과 전송 데이터량을 줄이기 위한 여러 음성 부호화 기술의 연구가 진행되고 있다. 음성부호화란 음성 신호를 디지털 부호로 변환시키는 것을 말하며 음성의 전송 또는 저장 등에 주로 사용된다. 음성부호화의 주된 기능은 음성 신호를 분석하여 잉여성분을 제거한 후에 제거되지 않는 나머지 성분들을 타당한 방법을 이용하여 부호화하는 것이다. 현재 이동통신 분야에서 이용되고 있는 디지털 셀룰러 음성 부호화기의 표준화 경향을 각 알고리즘이 이용되는 응용분야로 구분하여 표 1에 나타내었다[1,2,3]. 표 2에서는 표준화된 음성부호화 기법들의 종류와 특성에 대하여 간단히 비교를 하였다. 현재까지 발표된 음성부호화기 중 가장 많은 연구가 이루어지고 있는 방식은 CELP 계열이다. 이 방식은 4.8kbps 내외의 전송률에서 양호한 음질을 얻을 수 있으며 ITU-T, TIA/EIA 등 여러 국제 표준화 기구를 통해 다양한 응용분야에서 표준화가 이루어지고 있다.

[표 1] 디지털 이동통신용 음성 부호화기의 비교

알고리즘	RPE-LPT	VSELP	QCELP	LD-CELP	CS-ACELP	ACELP
전송률	8k bps	8k bps	8,4,2,1k bps	16k bps	8k bps	5.3k bps
프레임 길이	20m sec	20m sec	20m sec	0.625m sec	10m sec	20m sec
MOS Score	3.37	3.56	3.64	3.7	4.0 이상	3.98
주된 응용 분야	GSM	IS-54	CDMA	PSTN	PCS IMT-2000 DCME	Inter-net Phone

[표 2] 음성 부호화 ITU 국제표준

표준 번호	부호화 방식	내용	권고년도
G.711	PCM	3.4KHz 대역 64kbps 부호화	1972
G.712	ADPCM	3.4KHz 대역 32kbps 부호화	1988
G.722	SB-ADPCM	7KHz 대역 48/56/64kbps 부호화	1988
G.723.1	ACELP (5.3kbps) MP-MLQ (6.3)	멀티미디어 전송을 위한 이중 속도(5.3/6.3kbps) 부호화	1996
G.726	ADPCM	16/24/32/40kbps 부호화(G.721보완)	1990
G.727	Embeded ADPCM	2,3,4,5 비트 샘플 할당이 포함된 ADPCM	1990
G.728	LD-CELP	16kbps 부호화	1992
G.729	CS-ACELP	8kbps 부호화	1996

2. LSP 변환시간 단축알고리즘

PARCOR 필터에서 $k_{p+1}=\pm 1$인 전달함수를 $P_{p+1}(z)$와 $Q_{p+1}(z)$로 나타내면 식(1), (2)로 나타낼 수가 있다[1,3,4].

$$\begin{array}{l} k_{p+1}=1 \text{ 일때, } P_{p+1}(z)=A_p(z)-B_p(z) \\ k_{p+1}=-1 \text{ 일때, } P_{p+1}(z)=A_p(z)+B_p(z) \end{array} \quad (1)$$

$$\Rightarrow A_p(z)=\frac{1}{2}[P_{p+1}(z)+Q_{p+1}(z)] \quad (2)$$

두 개의 근($k_{p+1}=\pm 1$)을 알고 있으므로 $P_{p+1}(z)$의 $Q_{p+1}(z)$의 차수를 줄일 수 있다. 즉,

$$\begin{aligned} P'(z) &= \frac{P_{p+1}(z)}{(1-z)} \\ &= A_0 z^p + A_1 z^{(p-1)} + \cdots + A_p \end{aligned} \quad (3)$$

$$\begin{aligned} Q'(z) &= \frac{Q_{p+1}(z)}{(1-z)} \\ &= B_0 z^p + B_1 z^{(p-1)} + \cdots + B_p \end{aligned} \quad (4)$$

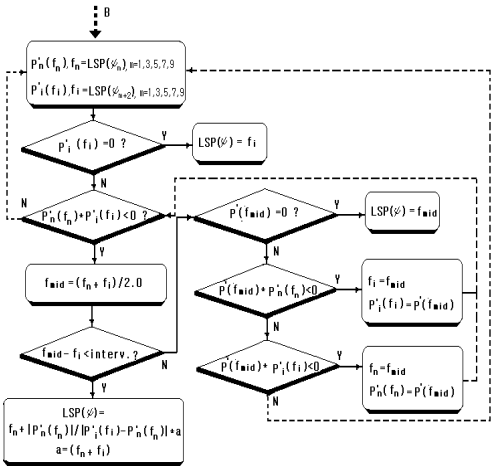
조건: $A_0=1, B_0=1$ (5)

$$A_k=(\alpha_k-\alpha_{p+1-k})+A_{k-1}$$

$$B_k=(\alpha_k-\alpha_{p+1-k})-A_{k-1} \quad \text{for} \quad k=1,\cdots p$$

2-1. 불균등 간격정보를 이용한 계산시간 단축 알고리즘

실근 방법에서는 홀수 번째 근을 먼저 검색하고 이미 찾아진 홀수 번째 근사이에서 짝수번째 근을 검색한다. 그러므로 LSP 파라미터를 찾기위해 주파수 대역을 검색함에 있어서 전체 변환시간은 홀수 번째 근을 검색하는 시간에 의해 좌우가 된다. 제안하는 첫 번째 방식은 멜 스케일을 적용하여 주파수 대역을 불균등하게 조절하여 검색시간의 단축을 가져왔다. 즉, 식 (4)를 사용하여 홀수 번째 근을 검색할 때 불균등하게 조절된 대역에서 근을 찾게 된다.



[그림 1] 제안한 방법 1의 순서도

2-2. LSP 파라미터의 분포특성을 이용한 계산시간 단축 알고리즘

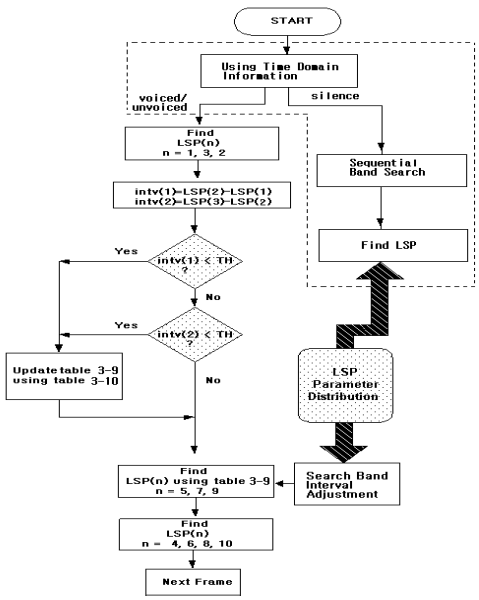
임의의 음성시료에 대하여 홀수 번째 LSP 파라미터

의 분포특성을 조사해본 결과 LSP 파라미터는 어느 특정 주파수 대역에 편중되어 나타난다. 기존의 변환 방식인 실근 방식에서는 홀수 번째 파라미터를 검색함에 있어서 주파수 대역을 순차적으로 검색한다. 제안한 방법 2에서는 홀수 번째 LSP 파라미터가 주로 분포하는 대역을 먼저 검색함으로써 계산시간의 단축을 가져온 방식이다. 제안한 알고리즘에서 홀수 번째 근을 검색할 때 사용한 부분 순서도가 그림 2에 나타나있다. 즉, 그림 2는 제안한 방법에서 첫 번째 근과 3번째 근을 검색할 때 사용한 순서도이다[6].

2-3. 음성신호 특성을 이용한 검색구간 조절 방법

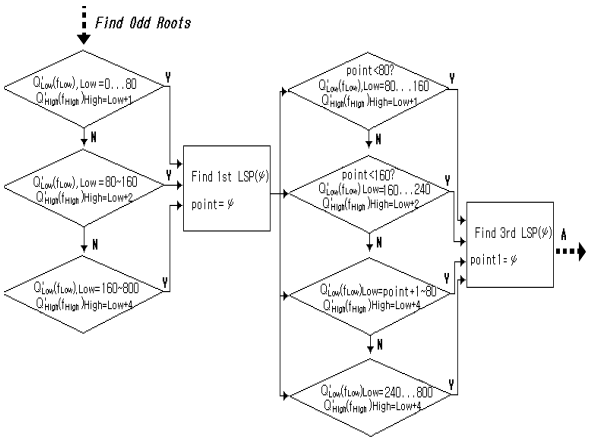
제안한 3번째 방식에서는 묵음구간에서 일정한 LSP 파라미터의 분포특성을 이용하여 근을 검색하고 유/무성음 구간에서는 제안한 방법 2에서와 같이 LSP 파라미터의 분포도를 이용하여 검색순서를 조절한다. 특히 모음에 대해서는 제 1 포만트와 제 2 포만트의 관계를 이용하여 검색구간을 조절함으로써 LSP 파라미터의 변환시간 단축을 도모하였다. 이때 유/무성음과 묵음구간을 판별하기 위한 시간영역정보로는 에너지를 이용하였다. 즉, 입력된 음성신호의 5 프레임까지의 에너지를 구한 다음 평균을 구하여 문턱 값으로 이용한다. 또한 각 프레임 별로 에너지를 구하여 앞에서 구한 문턱 값 에너지의 두 배 이상이 되면 유/무성음구간이라고 판단을 하고, 아니면 묵음구간으로 판단을 하여 검색구간을 조절하게 된다.

2-4. 분포도에 의한 해상도 조절 방법



[그림 3] 제안한 방법 3의 순서도

연구되어진 4번째 방법에서는 두 번째 방법과 마찬가지로 LSP 파라미터의 분포도를 적용한 방식이다. 즉, 비교적 긴 음성시료를 사용하여 LSP 파라미터의 홀수 번째 분포도를 조사하여 분포도가 높은 곳에서는 해상도를 높여서 검색하고 분포도가 낮은 대역을 검색할때는 해상도를 낮추어서 검색함으로써 LSP 변환시간의 단축을 도모한 방식이다. 제안한 방식에서 첫 번째 근과 세 번째 근을 찾을 때 사용한 알고리즘의 부분 순서도를 그림 4에서 나타내었다[5].



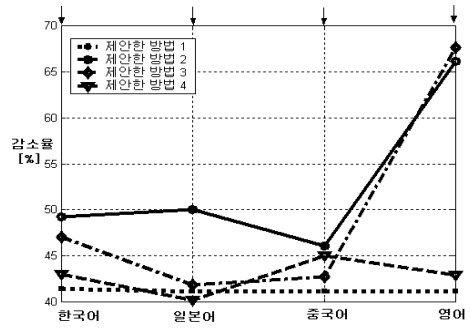
[그림 4] 제안한 방법 4의 부분순서도

3. 실험결과 및 결론

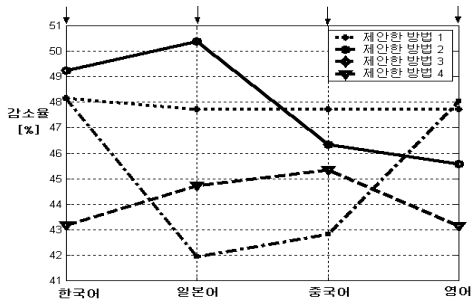
실험을 위해 IBM PC에 마이크 입력이 가능한 A/D 변환기를 인터페이스 하였다. 음성시료는 남성과 여성이 실험실 환경(30dB의 SNR)에서 발성한 음성을 8kHz로 표본화하고 16bit로 양자화하여 사용하였다. 실험과정에서 사용한 언어에 따른 계산시간 단축률의 변동을 조사하기 위하여 한국어, 일본어, 중국어, 영어 등을 사용하였다.

[표 3] clean speech인 경우 전체 통계(단위:[times, %])

	평균 계산량 (프레임당)		평균 계산량 감소율	
	+ , −	* , /	+ , −	* , /
제안한 방법 1	9217.3	14992.4	41.22	**47.82**
제안한 방법 2	7383.5	14932.0	**52.87**	**47.89**
제안한 방법 3	7871.1	15722.0	49.81	45.24
제안한 방법 4	8965.5	16020.1	44.77	44.10



[그림 5] (+, -) 계산량 감소율 비교(I)



[그림 6] (*, /) 계산량 감소율 비교(II)

표 3에서는 4개국어를 사용한 경우 기존의 방법에 비하여 제안한 방법들의 계산량 단축을 나타내고 있다. 표 3에서 보여지듯이 제안한 방법들은 기존의 방법에 비하여 모두 40%가 넘는 계산량 단축 정도를 나타낸다. 그림 5와 6은 사용한 언어에 따라 계산량의 단축정도가 어느 정도인지를 나타낸 결과로써 제안한 방법 1의 계산량 단축정도의 안정성이 가장 우수하다는 결론을 내릴 수가 있다. 즉, 그림 5, 6에서는 4개 국어 음성시료를 사용함에 있어서 제안한 방법 2, 3, 4는 기존의 방법과 비교하여 계산량이 단축정도가 크지만 계산량 단축정도가 경우에 따라 변화함을 알 수가 있다. 이에 반해 제안한 방법 1은 어느 언어를 사용하여도 비교적 일정한 계산량 감소율을 지님을 알 수가 있다. 본 논문에서 기존의 방법과 비교한 제안한 방법 4가지 중에서 검색간격을 불균등하게 조절한 방법이 계산량 감소률의 안정도 측면에서 가장 우수함을 알수가 있었다. 즉, LPC 계수에서 LSP 파라미터로 변환하는 과정에 있어서 어떤 음성신호에 대해서도 또, 분석하기 어려운 특성을 지닌 외국어가 입력신호로 사용되어도 언제나 같은 계산량 감소율을 지니는 점에서 안정되고, clean speech에 대해서도, 잡음 환경 하에서도 기존의 방법과 비교하여 음질 면에서 열하가 발생하지 않기 때문이다. 앞에서 언급하였듯이 제안하는 방법들은 현재 저 전송률 음성 부호화 표준안의 주류를 이루고 있는 CELP 계열의 부호화기에 적용되어질 수가 있다.

참고문헌

1. 배명진, *디지털 음성부호화*, 동영출판사, 1996.
2. 배명진, *디지탈 음성분석*, 동영출판사, 1998.4.
3. A. M. Kondoz, *Digital Speech*, John Wiley & Sons Ltd, 1994.
4. L. R. Rabiner, R.W. Schafer, *Digital Processing of Speech signal*, Prentice Hall, 1978.