1999 年度 大韓電子工學會 夏季綜合學術大會 論文集 第 22 卷 第 1 號 99/6

LSP 파라미터를 이용한 음성신호의 성분분리에 관한 연구

이 희 원, 나 덕 수, 정 찬 중, 배 명 진
숭실대학교 정보통신공학과
전화 : (02) 824-0906 / 팩스 : (02) 820-0018

A Study on a Method of U/V Decision by Using The LSP Parameter in The Speech Signal

HeeWon Lee, DuckSu Na, ChanJoong Jung, MyungJin Bae
Dept. Infomation and Telecommunication Engr., Soongsil University
E-mail : hwlee@ifcom.soongsil.ac.kr, mjbae@saint.soongsil.ac.kr

Abstract

In speech signal processing, the accurate decision of the voiced/unvoiced sound is important for robust word recognition and analysis and a high coding efficiency.

In this paper, we propose the mehod of the voiced/unvoiced decision using the LSP parameter which represents the spectrum characteristics of the speech signal. The voiced sound has many more LSP parameters in low frequency region. To the contrary, the unvoiced sound has many more LSP parameters in high frequency region. That is, the LSP parameter distribution of the voiced sound is different to that of the unvoiced sound. Also, the voiced sound has the minimun value of sequantial intervals of the LSP parameters in low frequency region. The unvoiced sound has it in high frequency region. we decide the voiced/unvoiced sound by using this charateristics.

We used the proposed method to some continuous speech and then achieved good performance.

Ⅰ. 서 론

음성신호는 발성 모델에 따라 유성음, 무성음, 묵음으로 분류될 수 있다. 유성음은 준주기적인 성대 펄스가 성도를 통해감으로써 발생되기 때문에 각 원소마다 성대에서 고유한 공명이 일어난다. 따라서 유성음의 스펙트럼은 음소마다 고유한 공명 봉우리를 갖게 된다. 이러한 공명 봉우리를 포만트라 하며 낮은쪽 주파수에서부터 두드러진 포만트를 차례로 제 1, 제 2, 제 3 포만트 등으로 부른다. 유성음의 스펙트럼에서는 보통 제 1포만트 주파수가 250-750Hz에 존재하며, 또한 공명현상 때문에 무성음에 비해 에너지가 크고, 성대의 진동에 의해 준주기성을 띠게 된다.

무성음은 불규칙한 잡음이 성대를 자극하는 입력으로 성대를 통과하는 동안 성대의 협착점에서 공명이 발생하게 된다. 따라서 무성음의 스펙트럼에서는 2500Hz 근처에서 주된 공명 봉우리가 존재하게 된다[7].

지금까지 제안된 알고리즘은 이러한 분석특성을 이용하여 유/무성음을 분류하였다. 유성음구간을 검출하기 위해서는 음성의 주기적 성질을 이용하거나 에너지 펄스를 사용하였다. 그러나 파열음이나 천이구간에서는 안정된 주기가 구해지지 않아 에러가 발생하고, 유성 자음 구간에서는 배경잡음과 에너지의 구분이 어려워서 유성음 구간을 검출하는데 문제점이 있다[4].

본 논문에서는 음성신호의 LSP 파라미터의 분포와 간격정보를 이용하여 음성신호의 성분분리하는 방법을 새로이 제안하고자 한다. 현재 사용되는 음성 코덱(codec)이나 인식기에서 음성신호를 분석하여 전송형이나 저장형 파라미터로 변환하는데 사용되는 것이 LSP 파라미터이다. LSP 파라미터는 양자화 에러에 강하고 시스템의 안정성과 선형 보간성이 뛰어나 많이 사용되고 있다[3]. 이러한 시스템에서 유성음, 무성음, 묵음을 분리하는데 분석시 추출되는 LSP 파라미터를 사용한다면 별도로 다른 파라미터를 도입하지 않아도 된다는 장점이 있다.

먼저 Ⅱ 장에서는 음성신호의 LSP 파라미터 추출방법을, Ⅲ 장에서는 LSP 파라미터를 이용한 음성신호의 성분분리방법에 대해 설명하고, Ⅳ장에서는 실험 및 결과, Ⅴ 장에서는 결론의 단계로 서술하였다.

Ⅱ. LSP 파라미터 추출

LSP 파라미터를 추출하기 위해서 먼저 LPC(Linear Predictive Coding)분석이 이루어져야 한다[3].

$$H(z) = 1/A_p(z) \qquad (2.1)$$

where $$A_p(z) = 1 + \sum_{k=1}^{p} a_k z^{-k} \qquad (2.2)$$

$H(z)$는 LPC 필터이고 p는 필터의 차수이다.

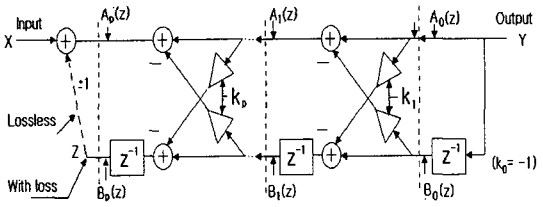
LSP 파라미터를 유도하기 위해서 PARCOR(Partial Correlation) 필터를 이용해서 식(2.1)과 식(2.2)를 표현하면 다음과 같다.

$$A_{p-1}(z)=A_p(z)+k_pB_{p-1}(z)$$
$$B_p(z)=z^{-1}[B_{p-1}(z)-k_pA_{p-1}(z)] \quad (2.3)$$

여기서 $A_0(z)=1$고 $B_0(z)=z^{-1}$ 이고

$$B_p(z)=z^{-(p+1)}A_p(z^{-1}) \quad (2.4)$$

그림<2-1>에 보이는 것처럼 PARCOR 구조는 손실이 없는 음파관에서 음파의 전달로 이해된다. 시스템은 단지 역방향(backward) 에너지 모양에서 Z 종점에서 손실이 있다. 이러한 음관은 Z 종점의 출력이 $k_{p+1}=\pm1$의 경로를 통해 입력의 종점으로 귀환될 때 완전한 무손실이 된다. 각각의 공명의 값인 Q는 무한해지고 에너지 분포 스펙트럼은 몇 개의 선 스펙트럼에 집중된다. $k_{p+1}=-1$ 조건의 귀환은 입력종점에서 완전히 폐쇄되고 $k_{p+1}=+1$ 은 무한 자유공간상으로 개방된다[3].



그림<2-1> PARCOR structure of LPC synthesis

그림<2-1>에서 $k_{p+1}=\pm1$인 전달함수를 $P_{p+1}(z)$와 $Q_{p+1}(z)$로 나타내면:

$k_{p+1}=1$ 일때, $P_{p+1}(z)=A_p(z)-B_p(z)$ (2.5)

$k_{p+1}=-1$ 일때, $P_{p+1}(z)=A_p(z)+B_p(z)$

$$\Rightarrow A_p(z)=\frac{1}{2}[P_{p+1}(z)+Q_{p+1}(z)] \quad (2.6)$$

두 개의 근($k_{p+1}=\pm1$)을 알고 있으므로 $P_{p+1}(z)$의 $Q_{p+1}(z)$의 차수를 줄일 수 있다. 즉,

$$P'(z)=\frac{P_{p+1}(z)}{(1-z)}=A_0z^p+A_1z^{(p-1)}+\cdots+A_p \quad (2.7)$$

그리고

$$Q'(z)=\frac{Q_{p+1}(z)}{(1-z)}=B_0z^p+B_1z^{(p-1)}+\cdots+B_p \quad (2.8)$$

조건 : $A_0=1,\ B_0=1$ (2.9)

$$A_k=(a_k-a_{p+1-k})+A_{k-1} \quad (2.10)$$

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

LSP는 $0\le\omega_i\le\pi$인 범위에서 $P'(z)$와 $Q'(z)$을 통해 얻어진 근의 각(angular) 위치를 나타낸다.

$P'(z)$와 $Q'(z)$의 다차 방정식을 풀기 위해서 실근 방법(real root method)을 사용하였다. $P'(z)$와 $Q'(z)$의 계수는 대칭적이기 때문에 식(2.7)의 차수는 $p/2$로 줄어든다.

$$P'(z)=A_0z^p+A_1^{p-1}+\cdots+A_1z^1+A_0$$
$$=z^{p/2}[A_0(z^{p/2}+z^{-p/2})+A_1(z^{(p/2-1)}+z^{-(p/2-1)})+\cdots+A_{p/2}] \quad (2.11)$$

$$Q'(z)=B_0z^p+B_1^{p-1}+\cdots+B_1z^1+B_0$$
$$=z^{p/2}[B_0(z^{p/2}+z^{-p/2})+B_1(z^{(p/2-1)}+z^{-(p/2-1)})+\cdots+B_{p/2}] \quad (2.12)$$

모든 근이 단위원 상에 있기 때문에, 단지 아래와 같이 정의하고 단위원 상에서 식(2.11)의 값을 구할 수 있다.

$$Let\ \ z=e^{j\omega}\ \ then\ \ z^1+z^{-1}=2\cos(\omega) \quad (2.13)$$

$$P'(z)=2e^{jp\omega/2}[A_0\cos(\frac{p}{2}\omega)+A_1\cos(\frac{p-2}{2}\omega)+\cdots+\frac{1}{2}A_{p/2}] \quad (2.14)$$

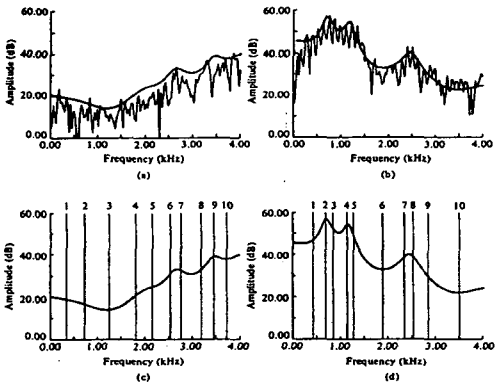
$$Q'(z)=2e^{jp\omega/2}[B_0\cos(\frac{p}{2}\omega)+B_1\cos(\frac{p-2}{2}\omega)+\cdots+\frac{1}{2}B_{p/2}] \quad (2.15)$$

$x=\cos\omega$를 대입해서 식(2.14)와 식(2.15)을 x에 대해서 풀 수 있다. 예를 들어서 $p=10$이면 다음과 같이 얻어진다.

$$P'_{10}(x)=16A_0x^5+8A_1x^4+(4A_2-20A_0)x^3+(2A_3-8A_1)x^2+(5A_0-3A_2+A_4)x+(A_1-A_3+0.5A_5) \quad (2.16)$$

유사하게,

$$Q'_{10}(x)=16B_0x^5+8B_1x^4+(4B_2-20B_0)x^3+(2B_3-8B_1)x^2+(5B_0-3B_2+B_4)x+(B_1-B_3+0.5B_5) \quad (2.17)$$

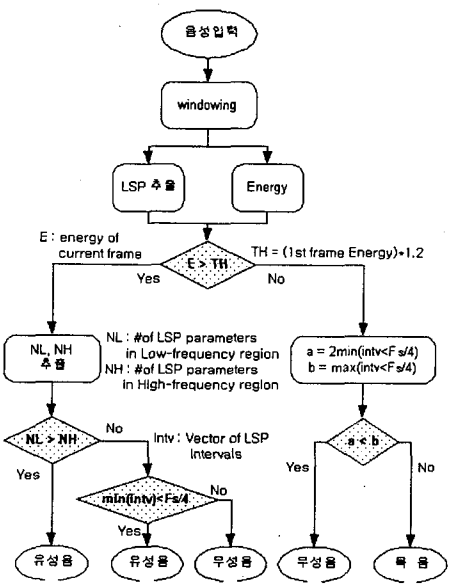


그림<2-2>선형스펙트럼쌍의 예
(a) 자음 /s/ (b) 모음 /a/
(c),(d) /s/와 /a/에 대한 LPC 분석과 LSP

Ⅲ. 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 8kHz, 11kHz로 표본화하고 16bit로 양자화하여 사용하였다. 발성한 문장은 다음과 같다.

발성1) "인수네 꼬마는 천재소년을 좋아한다."
발성2) "창공을 날으는 인간의 도전은 끝이 없다."
발성3) "예수님께서 천지창조의 교훈을 말씀하셨다."



그림<3-1>제안한 성분분리 방법에 대한 순서도

그림<3-1>은 제안한 성분분리 방법에 대한 순서도이다. 먼저 음성신호에 30ms 윈도우를 사용하여 에너지와 LP분석을 한다. LP분석을 통해 LSP 파라미터를 얻을 수 있다. 이때 LSP는 10차를 사용하였다. 그리고 첫 프레임의 에너지의 1.2배를 문턱값으로 하여 두가지 경우로 나눈다. 첫 번째 경우는 에너지가 문턱값 보다 크므로 유성음 또는 무성음이 존재하는 경우이고, 두 번째는 문턱값 보다 작으므로 무성음 또는 묵음인 경우로 간주한다.

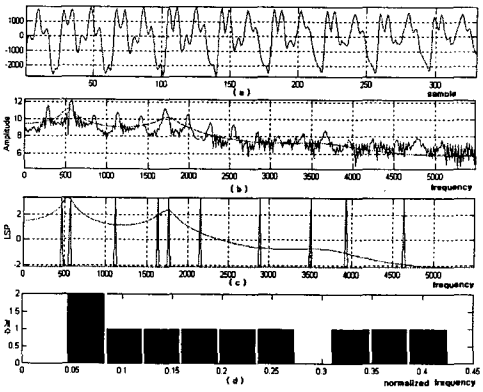
먼저 첫 번째 경우 NL과 NH의 값을 결정한다. NL은 샘플링 주파수를 Fs라고 할 때 $Fs/4$ 이하의 주파수영역에 존재하는 LSP 개수이고 NH는 $Fs/4$ 이상의 주파수영역에 존재하는 LSP 개수이다. NL이 NH보다 큰 경우는 음성신호의 스펙트럼이 저주파 쪽에서 봉우리가 많이 나타나는 모양이어서 유성음의 스펙트럼 특징을 나타낸다고 간주하게 된다. 즉 유성음의 제 1 포만트와 제 2 포만트가 주로 저주파수 영역에 존재하기 때문이다.

이와 반대로 NH가 NL보다 큰 경우는 무성음을 나타낸다고 결정한다. 즉 무성음의 스펙트럼은 주된 포만트가 고주파영역에 나타나기 때문이다. 하지만 /i/, /I/, /ε/, /æ/ 와 같은 유성음은 제 2 포만트, 제 3 포만트 또는 제 4 포만트가 고주파쪽에 존재하여 NH가 NL보다 크게 나타난다. 이와 같은 경우에는 제 1 포만트의 존재 여부로써 무성음인지 유성음인지를 결정하게 된다. 즉, LSP 파라미터들의 간격을 조사하여 $Fs/4$ 이하의 영역에서 좁은 간격을 나타내는 LSP들이 존재하면 유성음으로 간주 하게 된다. 여기서 그림<3-1>에 intv는 다음과 같다.

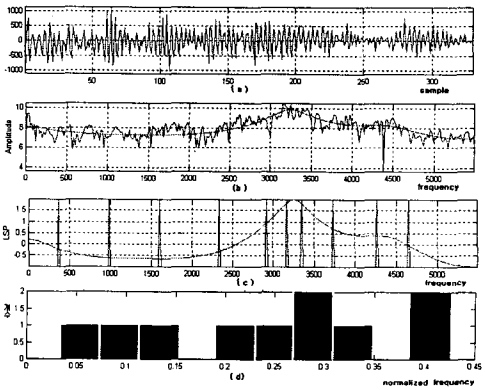
$$P=[p_1, p_2, p_3, \cdots, p_{10}]$$
$$intv(i)=|p_{i+1}-p_i| \quad , \quad 1\le i\le 9 \qquad (3.\ 1)$$

P는 LSP 파라미터이고, intv는 간격정보를 나타낸다.

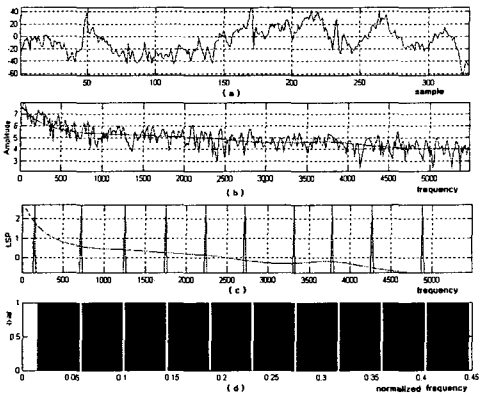
두 번째 경우인 에너지가 문턱값보다 작을 때는 묵음이나 무성음인 경우로 간주한다. 묵음과 무성음의 차이점은 묵음에서는 LSP들이 비교적 일정한 간격으로 나타나고 무성음에서는 고주파 영역에서 좁은 간격을 나타내는 LSP들이 존재하게 된다. 이러한 차이점을 이용하기 위해 $Fs/4$이상의 LSP들의 간격 중 최소값과 최대값이 2배 이상 차이가 있는지를 조사한다. 만일 차이가 2배 이상 존재한다면 무성음이고 그렇지 않다면 묵음으로 간주한다.



그림<3-2> 유성음에 대한 처리예
(a) 유성음파형 (b) 음성파형의 스펙트럼
(c) LPC 분석과 LSP (d) LSP 분포



그림<3-3> 무성음에 대한 처리예
(a) 무성음 파형 (b) 음성파형의 스펙트럼
(c) LPC 분석과 LSP (d) LSP 분포



그림<3-4> 묵음에 대한 처리예
(a) 묵음파형 (b) 음성파형의 스펙트럼
(c) LPC 분석과 LSP (d) LSP 분포

Ⅳ. 결 론

음성신호는 유성음, 무성음, 묵음으로 분류할 수 있으며 각각에 따라 그 성질이 다르게 나타난다. 따라서 음성신호에서 현재 처리하고자 하는 음성이 어떤 성분인지 사전에 알수 있다면 보다 효율적이고 정확한 분석을 할 수 있다. 음성인식에서는 유성음인지 무성음인지에 따라 비교해야할 대상을 줄일 수 있을 뿐 아니라 정확도를 높일수 있다. 또한 음성부호화시 처리될 음성이 유성음, 무성음 또는 묵음부분인지의 정보를 통해 중요도를 달리하거나 각 성분에 보다 적합한 파라미터를 사용하는 방법을 통해 비트를 줄이거나 음질을 향상시킬 수 있다. 따라서 음성신호처리에서 성분분리의 전처리과정은 매우 중요하게 적용될 수 있다.

본 논문에서는 인식 또는 전송형 코덱에 주로 사용되는 LSP 파라미터를 사용하여 유성음, 무성음 또는 묵음을 결정할 수 있는 방법을 제안하였다. 저주파수 영역과 고주파수 영역의 LSP 분포와 포만트의 존재여부를 결정하는 LSP 간격정보를 이용하여 음성신호의 성분을 분리하였다.

제안한 방법을 연속음에 적용하였을 때 좋은 성능을 보였으며 앞으로는 인식기와 부호화기에 적용하여 전체 시스템의 성능향상에 대한 연구가 계속되어야 된다.

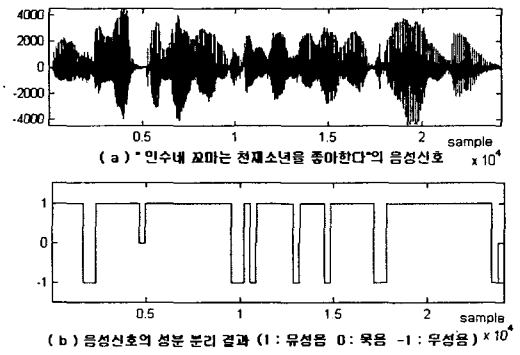


그림 <4-1> 발성1)에 대한 처리결과

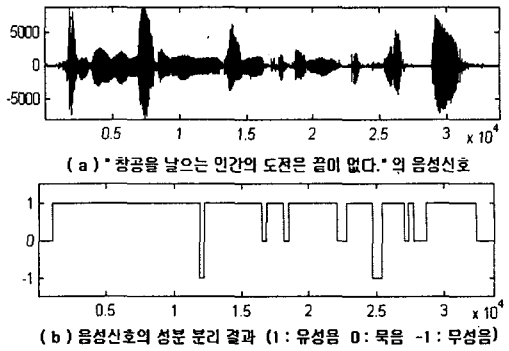


그림 <4-2> 발성2)에 대한 처리결과

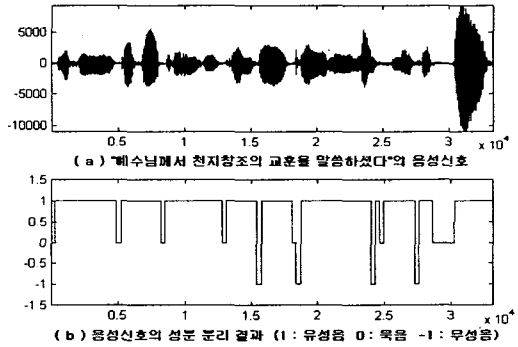


그림 <4-3> 발성3)에 대한 처리결과

Ⅴ. 참고 문헌

[1] Mabo Robert Ito, Robert W. Donaldson, "Zero-Crossing Measurements for Analysis and Recognition of Speech Sounds", IEEE Trans. on A.A.E. Vol. AU-19, No. 3 pp. 235-242, Sep., 1971.

[2] B.S Atal, L.R.Rabiner, "A Pattern Recognition Approach to Voiced-Unvoiced-Silence Classificatio -n with Application to Speech Recognition", IEEE Trans. on ASSP, Vol. ASSP-24, No. 3. June, 1976.

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

[4] 배성근, 백금란, 배명진, 안수길, "음성신호의 진폭분포를 이용한 유/무성음 검출에 대한", 한국음향학회. 학술논문발표회 논문집, 제 12호, 제 1(s)호, 1993.

[5] H. Kobatake, "Optimization of Voiced/Unvoiced Decisions in Noise Environments", IEEE Trans. on ASSP, Vol. ASSP-35, No. 1, pp. 9-18, Jan., 1987.

[6] S.G. Knorr, "Reliable Voiced/Unvoiced Decision", IEEE Trans. on ASSP. Vol. ASSP-27, No. 3, June, 1979

[7] L.R. Rabiner and R.W. Schafer, "Digital processing of Speech Signals Englewood Cliffs", New Jersey : Prentice-Hall, 1978.