제15회 음성통신 및 신호처리 워크샵(KSCSP '98 15권1호)

피치동기에 의한 음성신호의 전이구간 검출

나 덕 수 , 노 원 석 , 함 명 규 , 배 명 진
숭실대학교 정보통신공학과
156-743 서울시 동작구 상도동 1-1
dsna@assp.soongsil.ac.kr mjbae@saint.soongsil.ac.kr

On Detecting the Transition Regions of Speech Signal by Pitch Synchronization

DuckSu Na , WonSuck No , MyungKyu Ham , MyungJin Bae
Dept.infomation and Telecommunication Engr., Soongsil University
1-1 Sangdo-5Dong, Dongjak-Ku, Seoul 156-743, KOREA
dsna@assp.soongsil.ac.kr mjbae@saint.soongsil.ac.kr

요 약

연속된 음성의 인식을 위해서는 음성신호를 음성학적인 단위인 단어, 음절, 음소 등으로 분할하여야 한다. 이러한 분할을 위해서는 전이구간의 검출이 선행되어야 한다.

본 논문에서는 음성신호에서 전이구간을 검출하기 위해 피치동기된 상관관계 계수의 변화를 나타내는 파라미터를 새로이 제안하였다. 이 파라미터는 음성신호의 안정구간에서는 매우 작은 값을 나타내지만 음성의 시작이나 유성음과 무성음의 경계에서는 큰 값을 나타내어 전이구간검출용 파라미터로 매우 용이하다.

1. 서 론

음성신호(혼합음)의 전이구간은 무성음에서 유성음으로 또는 유성음에서 무성음으로 연결되는 혼합영역이며 유/무성음의 성질이 동시에 나타나게 된다. 여기서 유성음이란 성대에서 적절한 장력을 주어서 공기 압력을 주기적으로 변화시키게 하고, 공명을 여기하여 생성하며, 무성음은 성대에서 조음기관들이 아주 좁은 관으로 이루어지게 되어 생성되는 것을 말한다[1].

연속음이나 연결음에서는 이 음들이 시간에 따라 변화하게 되며, 이것은 프레임당 평균진폭의 변화 형태로 나타나게 된다. 이때 평균진폭의 변화도는 음소나 음절의 변화를 근사적으로 대별하게 된다.

음성신호의 전이구간을 검출하는 연구는 특징파라미터를 추출하는 영역에 따라 크게 시간영역법, 스펙트럼 영역법, 혼성영역법으로 나눌 수 있다.

시간영역법은 시간영역에서 계산의 간편성을 취할 수 있으며, VOT(Voice Onset Time)의 연속성이나 진폭의 증감을 이용하는 방법들이 제안되어 졌다.

스펙트럼영역법은 음성신호의 음소의 변화에 따른 포만트의 전이특성이나 주파수성분별 에너지를 이용하는 것 등이 제안되어져 있다. 또한 혼성영역법은 변환영역에서의 특징 파라미터들을 이용하는 것으로 LPC (Linear Prediction Coefficient)의 전이특성, LPC에러의 변화 특성 등을 이용하고 있다.

시간영역법에서 파라미터 검출은 비교적 쉬우나 그 변화정도를 정확히 파악하기 위한 결정논리가 상대적으로 어렵다. 반면 스펙트럼영역법이나 혼성영역법은 비교적 정확하지만 계산의 정밀도나 변환차수등의 영향을 받게 되고, 전처리과정으로 보기에는 계산량의 부담이 시간영역법에 비해 큰 편이다[2]. 따라서 본 논문에서는 시간영역법 전이구간 검출용 파라미터들 중에서 음성에너지 파라미터가 갖는 부정확성과 결정논리의 복잡성을 보완할 수 있는 새로운 파라미터를 제안하고자 한다.

2. 음성신호의 전이구간 검출의 중요성

연속음에서 전이구간을 사전에 판별하는 것은 음성인식에서 인식률 향상과 피치검색시 소요되는 시간 단축에 매우 중요한 역할을 한다.

연속된 음성의 인식을 위해서는 전기신호로 표시된 음성신호를 음성학적 단위인 단어, 음절, 음소 등으로 분할하여야 한다. 연속음을 이러한 단위로 분류하면 분석의 반복을 줄일 수 있고, 음성인식과정에서는 고립단어

의 인식기법을 연속음 인식에 쉽게 연장시켜 적용할 수 있게 된다[3].

시간영역에서 피치를 찾는 경우에는 AMDF, ACM, Parallel Processing법 등을 이용하여 파형의 주기성을 강조한 뒤 결정논리를 적용한다. 그러나 음소가 전이구간에 걸쳐있는 경우에는 프레임내의 음소변화가 심하고 기본주파수의 주기가 일정치 않으므로 피치검출에 어려움이 따른다. 이때에 음소의 전이구간을 미리 검출하여 피치검출구간에서 제외시키면 검출시간을 줄일 수 있으며, 검출오류의 증가도 막을 수 있다[4].

3. 피치동기된 상관계수

본 논문에서 제안한 상관계수는 음성신호의 분석시에 사용되는 피치필터의 피치이득을 변환하여 사용하였다.

$$P(z)=\frac{1}{1-\beta z^{-(j+\tau)}} \tag{3-1}$$

$$R(\tau+i,0)=\sum_{m=0}^{N-1} r(m-\tau-i)r(m)$$
$$V(i,j)=\sum_{m=0}^{N-1} r(m-\tau-i)r(m-\tau-j) \tag{3-2}$$

$$\beta = \frac{R(\tau,0)}{V(0,0)} = \frac{\sum_{m=0}^{N-1} r(m)r(m+\tau)}{\sum_{m=0}^{N-1} r^2(m)} \tag{3-3}$$

식 (3-1)은 일반적인 1-tap 피치필터를 나타낸다. τ는 피치-래그(lag)이고 β는 피치이득 계수이다. 식 (3-3)에서 β는 τ 만큼 떨어진 샘플들의 유사도를 나타낸다[5].

본 논문에서는 먼저 규준화된 AMDF법으로 피치를 찾은 후 피치 단위로 신호를 분절한다. 이렇게 피치단위로 신호를 분절하는 이유는 피치에 동기시켜 각 피치 구간의 상관관계로 음성신호의 전이구간을 검색하기 위해서이다.

$$NAMDF(d)=\frac{\sum_{n=1}^{N}|s(n)-s(n-d)|}{\sum_{n=1}^{N}[|s(n)|+|s(n-d)|]} \tag{3-4}$$

식 (3-4)는 피치 검색시 사용된 식이다. s(n)은 음성신호이고 N은 윈도우 크기, d는 지연인자이다[4].

식 (3-4)를 이용하여 현재 피치주기와 미래 피치주기를 분절한다. 분절된 음성신호로 피치동기된 상관관계를 적용한다.

현재 피치주기의 상관관계 계수는 다음과 같이 정의된다.

$$\beta(i) = \frac{E_{ij}}{E_{ii}} = \frac{\sum_{n=1}^{\min(\tau_i,\tau_j)} s_i(n)s_j(n)}{\sum_{n=1}^{\tau_i} s_i^2(n)} \tag{3-5}$$

여기서 β(i)는 i번째 피치주기의 상관관계 계수이고 $s_i(n)$와 $s_j(n)$는 각각의 피치주기인 τ_i, τ_j에 의해 현재 미래 피치주기로 분절된 음성신호이다.

그림 <3-1>에서처럼 식 (3-5)를 사용하여 피치동기된 상관관계 계수는 두 피치주기의 상관관계가 클수록 1에 가까운 값을 가지게 된다.

$$VR(i) = |1-\beta(i)|^2 \tag{3-6}$$

본 논문에서는 전이구간에 보다 민감하게 적용하기 위해서 식 (3-6)과 같이 상관관계 계수를 변화시켜 전이구간 검출에 필요한 파라미터를 얻어낸다.

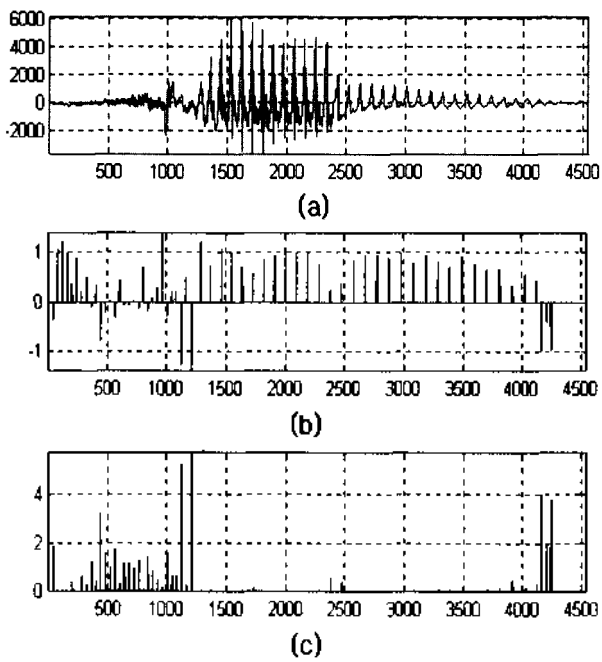

그림 < 3-1 > 음성신호의 피치동기된 상관계수
(a) 음성파형 ('삼')
(b) 상관관계 계수 (β)
(c) VR 파라미터

4. VR(Variable Rate)파라미터에 의한 전이구간의 검출

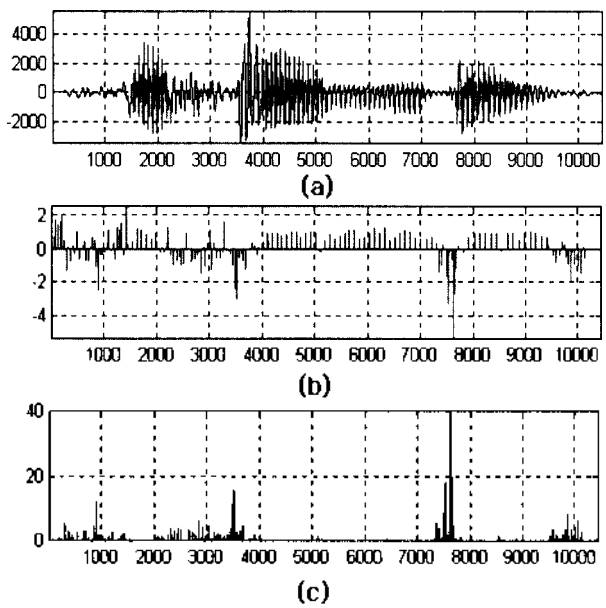

그림 < 4-1 > 음성신호의 피치동기된 상관계수
(a) 음성파형 ('감사합니다.')
(b) 상관관계 계수 (β)
(c) VR 파라미터

그림 <4-1>은 20대 남성화자가 발음한 연속음 '감사합니다'에 대해 상관관계 계수와 VR 파라미터를 구한 것이다. 상관관계 계수에서 음의 값이 나오는 부분은 음절이 바뀌는 부분과 무성음에서 유성음으로 바뀌는 부분이다. 그리고 유성모음에서 비음자음으로 바뀌는 부분은 계수 값이 비록 음(minus)이 아니더라도 값의 변화가 안정구간에 비해 크게 나타난다. 이러한 상관관계 계수 값에서 식 (3-6)에 의해 VR 파라미터를 얻어내면 음의 상관관계가 나타나는 부분은 그 값이 1이상의 매우 큰 값이 나타나고 그 외 전이구간은 1이하의 값이 나타나지만 안정구간보다 크게 나타난다. 그리고 안정구간은 상관관계 계수 값이 1에 가깝게 나타나기 때문에 VR 파라미터 값은 거의 0에 가깝게 나타난다.

VR 파라미터의 값으로 음절이 변하는 부분은 물론 음소가 변하는 전이구간까지 검출하기 위해서 먼저 그 값이 1보다 크게 나타나는 부분을 검출하여 검출된 부분의 파라미터 값을 제외한 나머지 값들만의 평균을 취한다. 그리고 구해진 평균으로 문턱값을 정하여 나머지 전이구간을 검출한다.

5. 실험 및 결과

시뮬레이션을 위해 PC(Pentium 150MHz)에 마이크로폰의 입력이 가능하도록 16비트 A/D 변환기를 인터페이스 시켰다. 화자는 20대 남성화자 3명과 여성화자 1명을 통해 다음 음성을 발성케하고 11kHz로 표본화 하였다.

발성 1) " 삼 "
발성 2) " 감사합니다. "
발성 3) " 안녕하십니까. "
발성 4) " 숭실대학교 "

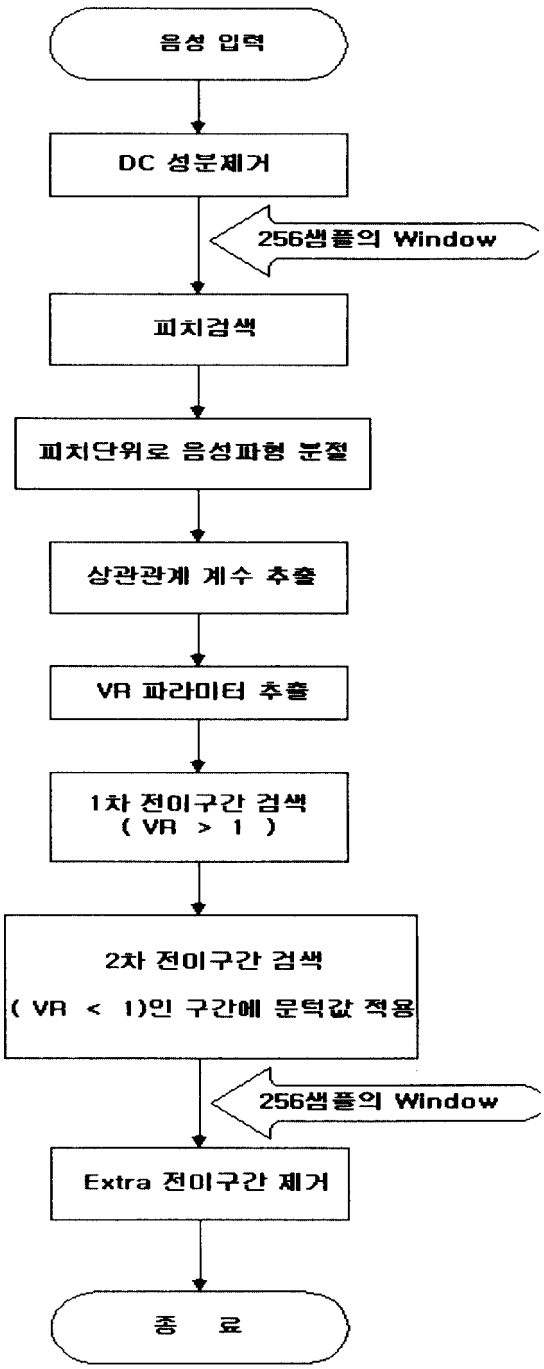

그림 < 5-1 > 전이구간 검출 과정 순서도

각 음성시료에 대해 피치검색 프레임의 크기는 256샘플로 하였고 [프레임 크기 - 검색된 피치] 만큼 겹치게 하였다. 피치 검색법은 식 (3-4)와 같이 규준화된 AMDF법을 사용하였다. 만일 피치검색 결과 무성음이나 묵음으로 간주될 때는 40샘플을 임의의 피치로 결정하였다.

이렇게 구해진 피치로 음성파형을 분절하여 식 (3-5)에 의해 상관관계 계수 값을 구하고 식 (3-6)에 의해 상관관계 계수 값에서 VR 파라미터 값을 얻는다.

이렇게 얻어진 VR 파라미터 값으로 4장에서 설명한 바와 같이 먼저 1보다 큰 값이 나타나는 부분을 검출한다. 그리고 음소변화가 일어나는 구간을 검출하기 위해 VR 파라미터 값이 크게 나타나는 부분을 제외한 나머지 부분만의 VR 파라미터 값의 평균을 취한다. 이 평균값을 문턱값으로 하여 이 값보다 큰 VR 파라미터 값이 나타나는 부분을 전이구간으로 검출한다.

이렇게 구해진 전이구간들은 피치동기된 VR 파라미터에 의해 결정되었기 때문에 실제의 전이구간 보다 많이 나타나게 된다. 따라서 실제보다 많이 검출된 Extra 전이구간을 줄이기 위해 256샘플 크기의 윈도우를 사용하여 그 구간 내에 VR 파라미터 값이 최대인 부분만 남긴다.

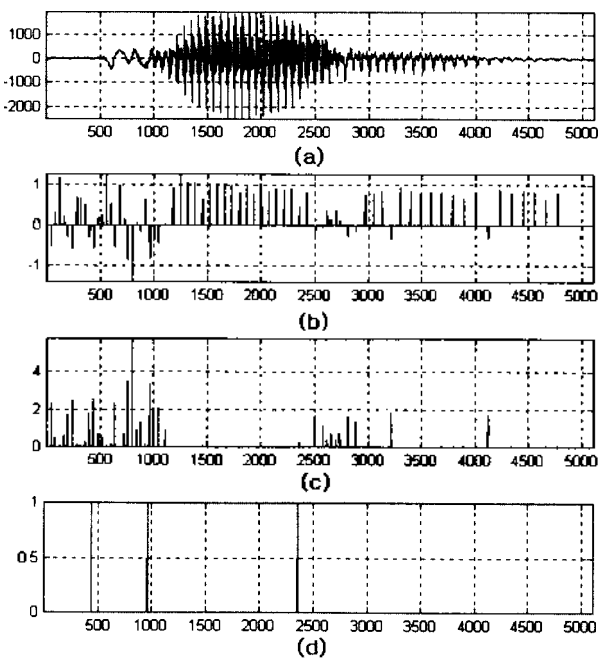

그림 < 5-1 > (발성 1)에 대한 처리결과
(a) 음성파형 (화자 1) (b) 상관관계 계수 (β)
(c) VR 파라미터 (d) 검출된 전이구간

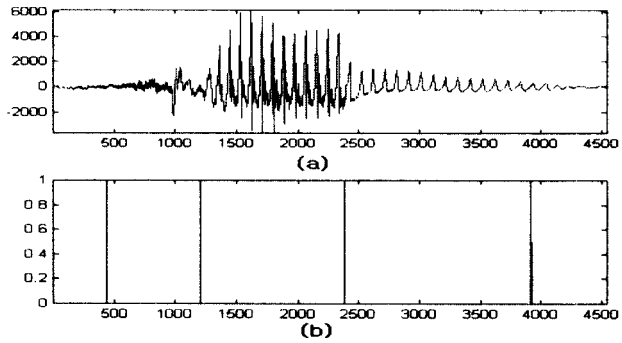

그림 < 5-2 > (발성 1)에 대한 처리결과 (계속)
(a) 음성파형 (화자 2) (b) 검출된 전이구간

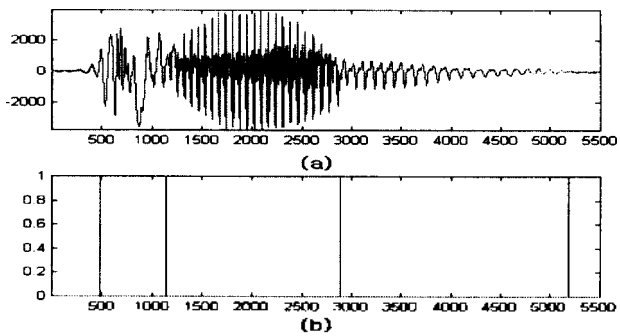

그림 < 5-3 > (발성 1)에 대한 처리결과 (계속)
(a) 음성파형 (화자 3) (b) 검출된 전이구간

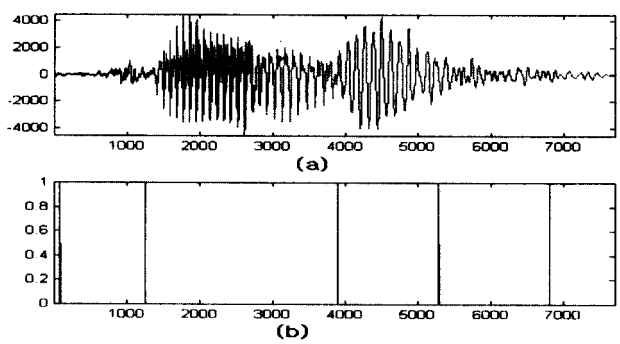

그림 < 5-4 > (발성 1)에 대한 처리결과 (계속)
(a) 음성파형 (화자 4) (b) 검출된 전이구간

그림 <5-1>에서 <5-4>까지는 화자 1, 화자 2, 화자 3, 화자 4의 (발성1)에 대한 처리 결과이다. 그림 <5-1>의 (d)에서 1이 나타나는 부분에서 전이가 일어난다. /삼/에서 무성음 /ㅅ/이 시작되는 부분, /ㅅ/에서 /ㅏ/부분으로 넘어가는 부분 그리고 /ㅏ/부분에서 비음 /ㅁ/으로 넘어가는 부분이 정확히 검출되었다.

그림 <5-1>의 (c)에서 볼 수 있듯이 그 값이 큰 부분이 전이구간의 변화 특성을 나타내고 있어 이 파라미터를 이용하여 전이구간을 검출한 (d)에서도 정확한 결과

를 보여준다.

6. 결 론

연속음 인식을 위해서는 음성신호의 분할과정이 필요하다. 음절단위의 분할이 잘 이루어지면 음성분석시나 인식시에 고립단어의 분석 및 인식 기법들을 쉽게 적용할 수 있게 된다.

본 논문에서는 먼저 음성신호의 피치를 검출하여 피치필터에서 피치이득으로 사용되는 상관관계 계수를 구하였다. 그리고 이 상관관계 계수의 변화특성을 나타내는 VR 파라미터를 이용하여 전이구간을 검출하였다.

본 논문에서 사용된 VR 파라미터는 피치검색 후 추출된 파라미터이므로 음성의 피치변화와 에너지 변화를 잘 반영 할 수 있어 음성의 끝점은 물론 음소의 전이구간도 정확하고 쉽게 검출 할 수 있다.

7. 참고 문헌

[1] L.R. Rabiner, and R.W. Schafer *"Digital Processing of Speech Signal"*, Prentice-Hall, Englewood Cliffs, New Jersey, 1978.

[2] C.J. Weinstein, S.S McCandiess, L.F. Mondshein, and V.W. Zue, *"A System for Acoustic-Phonetic Analysis of Continuous Speech"*, IEEE Trans. on ASSP, Vol ASSP-23, No.1, pp 54-67.

[3] 배명진, 이을재, 안수길, "음성파형의 비대칭을을 이용한 음소의 전이구간 검출", 한국음향학회지, Vol.9, No.4, pp. 55-65, August 1990

[4] 배명진 외 1인 , "On Detecting the Steady State Segments of Speech Waveform by using the Normalized AMDF", 대한전자공학회지, Vol.14, No.1, pp.660-603, Jun.,1991

[5] A.M. Kondoz, *"Digital Speech"*, John Wiley & Sons Ltd, Baffins Lane, Chichester, England, 1994.

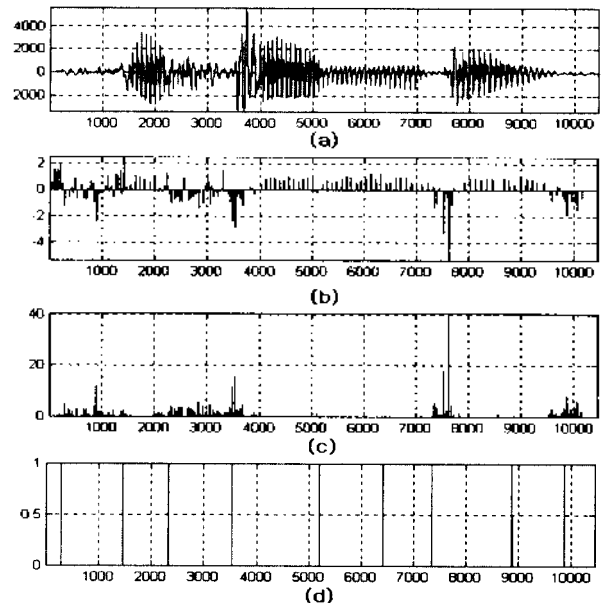

그림 < 5-5 > (발성 2)에 대한 처리결과
(a) 음성파형 (화자 1) (b) 상관관계 계수 (β)
(c) VR 파라미터 (d) 검출된 전이구간

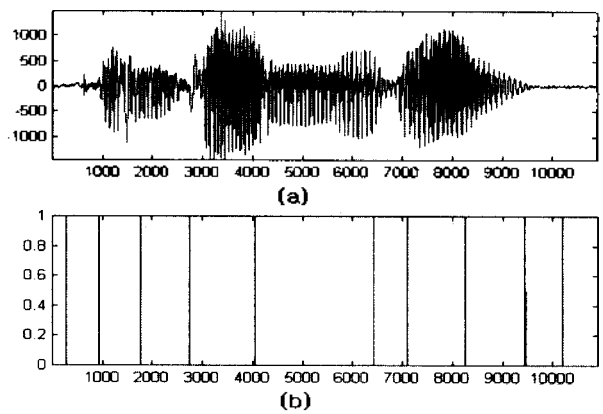

그림 < 5-6 > (발성 2)에 대한 처리결과 (계속)
(a) 음성파형 (화자 2) (b) 검출된 전이구간

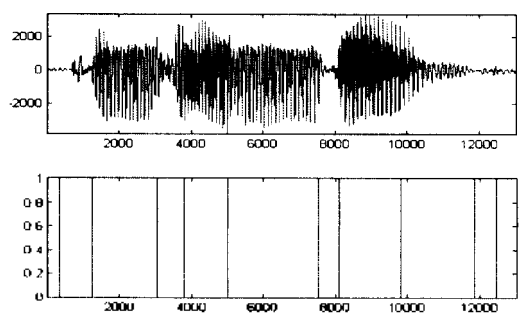

그림 < 5-7 > (발성 2)에 대한 처리결과 (계속)
(a) 음성파형 (화자 3) (b) 검출된 전이구간

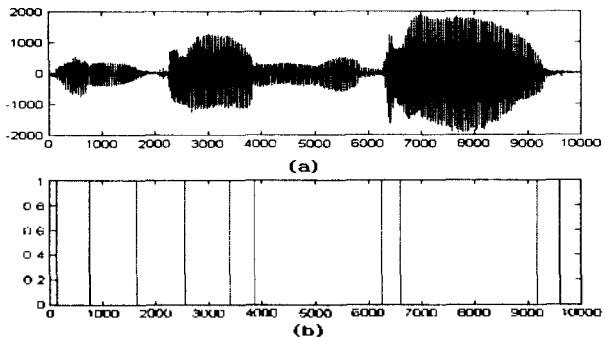

그림 < 5-8 > (발성 2)에 대한 처리결과 (계속)
(a) 음성파형 (화자 4) (b) 검출된 전이구간

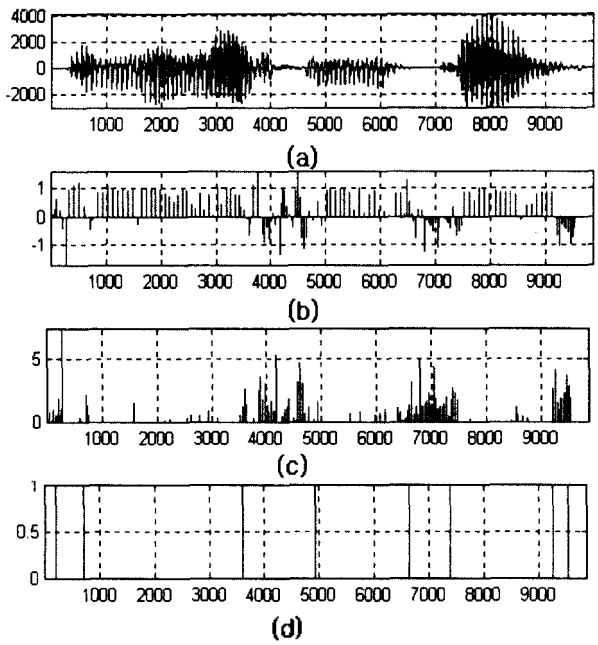

그림 < 5-9 > (발성 3)에 대한 처리결과
(a) 음성파형 (화자 1) (b) 상관관계 계수 (β)
(c) VR 파라미터 (d) 검출된 전이구간

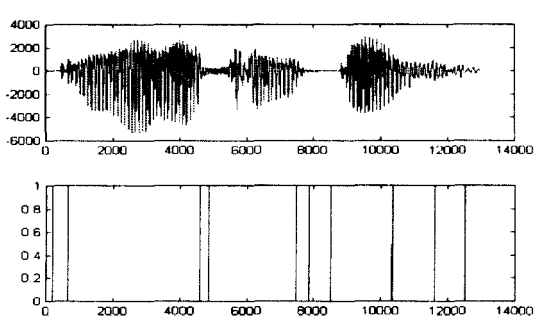

그림 < 5-10 > (발성 3)에 대한 처리결과
(a) 음성파형 (화자 2) (b) 검출된 전이구간

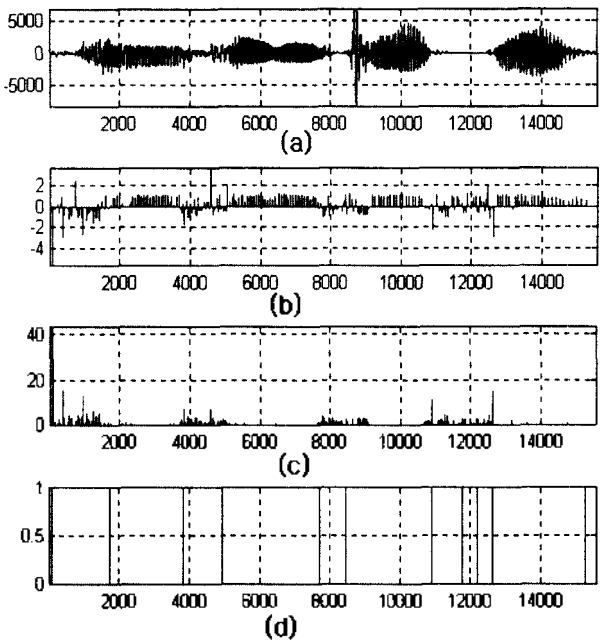

그림 < 5-11 > (발성 4)에 대한 처리결과
(a) 음성파형 (화자 3) (b) 상관관계 계수 (β)
(c) VR 파라미터 (d) 검출된 전이구간

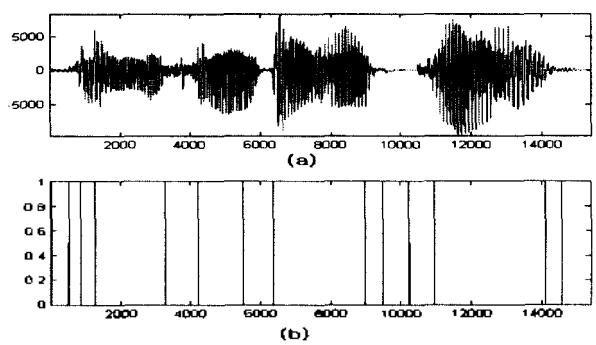

그림 < 5-12 > (발성 4)에 대한 처리결과
(a) 음성파형 (화자 4) (b) 검출된 전이구간