MBE 부호화용 스펙트럼 V-UV 구간 검출에 관한 연구

(*) 김 을재 (*) 김 형태 (**) 한 창문 (**) 배 명진
(*) 호서대학교 (**) 숭실대학교

(On a Detection of V-UV Segments of Speech Spectrum for the MBE Coding)

(*) Ul-je KIM (*) Hyoungtae KIM (**) Changmun HAN (**) Myungjin BAE
(*) HOSEO University, (**) SUNGSIL University

ABSTRACT

In the area of speech vocoder systems, the MBE vocoder allows the high quality and low bit rate. In the MBE parameters detection, the dicision methods of V/UV region proposed until now are dependent highly to the other parameters, fundamental frequency and formant information. In this paper, thus, we propose a new V/UV detection method that uses a zero-crossing rate of flatten harmonices spectrum. This method can reduce the influences of the other parameters for the V/UV regions detection.

Ⅰ. 서 론

음성합성 시스템은 여러 가지 분류기준에 따라 구분될 수 있다. 먼저, 단순히 음성을 분석한 후에 이를 그대로 다시 합성에 이용하는 분석/합성(synthesis by analysis)방식과 한정된 음성 데이터를 바탕으로 법칙에 의해 음성을 발생시키는 법칙합성(synthesis by rule)방식으로 구분될 수 있다. 또한, 신호를 분석하거나 합성하는 부호화기법에 따라, 파형부호화방식, 신호원부호화방식, 혼성부호화방식으로 분류된다[1-2].

분석/합성방식에서는 메모리 효율성 때문에, 신호원 부호화방식이 주로 사용되고 있으며, 낮은 전송율과 고음질을 구현하기 위한 많은 연구가 활발히 진행되고 있다. 이러한 부류의 시스템들은 한 구간의 음성을 여기성분과 성도의 시스템 스펙트럼으로 재표현하고 있다. 일반적으로 이를 구현하기 위하여 음성 발생 모델을 그림 1-1과 같이 가정한다. 먼저, 음성이 유/무성음임에 따라 스위치가 준 주기적인 임펄스 열과 잡음 시퀀스에 의한 여기를 선택하게 된다. 무성음의 경우 잡음 시퀀스에 의해 여기되고, 유성음의 경우 준 주기적인 임펄스 열에 의해 여기된다. 선택된 신호는 성도 여파기, H(Z)를 거쳐 음성정보를 갖게 된다. 이러한 발생 모델을 기본으로 하는 분석/합성 시스템(보코더)은 합성음의 명료성과 자연성을 포함하는 전체적인 음질이 떨어지며, 주위 환경의 잡음에 민감하다는 단점이 있다.

이러한 단점을 극복하기 위하여 여기원에 그 문제에 대한 초점이 맞추어 졌다. MBE(Multi-Band Excitation) 음성 발생모델은 V/UV (Voiced/UnVoiced) 분류를 갖는 스펙트럼 구간을 이용하여 기존의 음성생성 모델의 여기원 부분을 대치함으로써 음질의 개선을 이룰 수 있었다. 그림 1-2는 MBE 음성 발생모델에 대한 블럭도이다. 먼저 어떤 음성에 대한 스펙트럼의 한 부분은 유성화 구간과 무성화 구간으로 다시 분류된다. 이러한 분류에 의해 유성화 구간은 기본주파수의 간격을 갖는 고조파가 여기원으로서 사용되고, 무성화 구간은 잡음 시퀀스에 의한 스펙트럼으로 대치된다. 이렇게 발생된 여기원은 성도 여파기를 통하여 음성 정보를 갖게 되고, 부분적으로 유성음이면서도 부분적으로 무성음인 합성음이 만들어 진다. 이러한 보코더 시스템은 복잡한 계산 없이도 주위 환경에 영향을 적게 받는 고음질의 음성을 발생할 수 있다[3-4].

본 논문은 MBE법에 의한 음성 분석/합성 시의 V/UV 구간 검출에 관한 것으로써, 스펙트럼 평탄화 기법을 이용하여 구해진 고조파에 대한 영교차율을 이용한다. 먼저 2절에서 MBE 부호화법과 기존의 파라미타 결정법을 간단히 알아보고 V/UV 구간 검출에 대한 문제점을 살펴본 후, 3 절에서는 본 논문에서 사용되는 스펙트럼 평탄화 기법과 이를 이용한 스펙트럼 V/UV 결정에 대하여 알아 본다. 4절에서는 제안한 검출법의 실험 결과를 보이고, 마지막으로 결론짓는다.

Ⅱ. 기존의 MBE 부호화법

음성을 분석하다 보면 유성음이라고 단정지을 수 있는 구간 조차도 진폭 스펙트럼의 일부분은 잡음에 의한 에너지로 채워져 있다는 것을 알 수 있다. 더욱이 잡음이 섞인 음성이나, 잡음이 섞이지 않은 혼합구간(mixed voicing segment)일 경우에는 기본주파수에 의한 주기적인 고조파와 잡음 에너지에 의한 스펙트럼이 동시에 존재한다는 것을 알 수 있다. 그러므로 보코더 시스템에 어떤 음성구간에 대하여 단지 2진 V/UV 결정을 전체 음성에 적용하는 것으로는 합성음질을 보장받을 수 없다.

이러한 이유로 MBE법에서는 스펙트럼을 여러 구간으로 나누고, 이 구간에 대한 V/UV를 결정한다. 그러므로 MBE법에 대한 파라미타는 기본주파수, 스펙트럼 포락과 스펙트럼의 각 구간에 대한 V/UV 결정치로 이루어 진다. 그림 2-1은 일반적인 MBE 알고리듬의 한 블럭도이다. 먼저 분석 단에서 창함수가 적용된 음성 $S_W(w)$에 대한 피치와 스펙트럼 포락을 구한 후에 이를 이용하여 각 구간의 V/UV를 결정하고, 이를 그 음성 구간에 대한 시스템 파라미타로 결정한 뒤, 합성단에서 이들 파라미타를 이용하여 음성을 합성, 출력한다.

음성 분석시에 원 음성에 대한 스펙트럼과 합성음성에 대한 스펙트럼 사이의 오차가 최소가 되도록 이들 파라미타를 결정해야 한다. 원 음성과 합성음 사이의 스펙트럼의 차이는 식 2-1에 의해 얻어진다.

$$E = \frac{1}{2\pi} \int_{w=-\pi}^{\pi} G_W(w) \, |S_W(w) - S'(w)|^2 \, dw \qquad (2\text{-}1)$$

여기서 $S'(w)$는 합성 스펙트럼이고 $G(w)$는 주파수에 따른 가중치이다. 이에 대한 최소 에너지를 갖는 파라미타를 검출하려면 고도의 비선형 최적화 문제를 해결해야 한다. 이런 이유로 먼저 유성음이라는 가정하에서 기본주파수와 스펙트럼의 포락을 찾고 V/UV 정보를 최적화하는 근사법을 사용해야 한다.

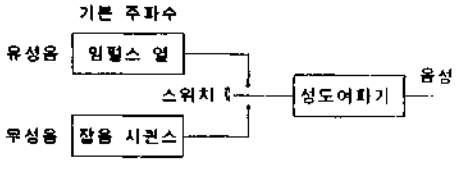

그림 1-1. 일반적인 음성 발생모델

FIG. 1-1. General speech generation model

먼저 원음성의 정확한 기본주파수가 얻어져야 한다. 다음에는 얻어진 기본주파수를 이용하여, 스펙트럼 포락선을 복소 하모닉스 계수(complex harmonic coefficients)의 조합으로 대치시킨다. 이것은 기본주파수에 대한 고조파 스펙트럼 포락선의 값에 대응한다. 기본주파수를 알고 있다면 하모닉스 계수에 대한 에러를 최소화 할 수 있는 계수는 다른 파라미타의 영향이 없는 선형 방정식이며 쉽게 풀 수 있다. 식 2-1에 의해서 기본주파수와 스펙트럼 하모닉스 계수에 대한 최소 에러를 검출한다. V/UV 결정은 이 최소 에러를 갖는 스펙트럼에서 얻어진다. 먼저, 최소 에러에너지를 갖는 $S'(w)$와 원 스펙트럼 사이의 에러 스펙트럼을 구하고, 기본주파수의 3배 길이로 각 구간을 나눈 뒤에 이에 대한 평균 에러를 계산한다. 이 에러 값이 정해진 문턱값을 초과한다면 무성화 구간이라 하고, 초과하지 않으면 유성화 구간이라 결정한다.

상술한 법을 성공적으로 수행하기 위해선 우선 오차율 ±1Hz 정도의 정확한 피치검출이 이루어 져야하고, 이를 이용하여 스펙트럼 포락을 검출하여야 한다. 또한 V/UV 결정은 이 두가지 파라미타를 바탕으로 이루어 지므로, 피치의 검출이 잘못 되었다면 오차가 누적되어 V/UV 검출에 큰 영향을 미치게 된다.

Ⅲ. 고조파 스펙트럼 평탄화에 의한 스펙트럼 V/UV 결정

3.1 고조파 스펙트럼 평탄화 기법

먼저 그 구간에 대한 기본주파수를 검출하기 위하여 SAMDF 함수를 이용한다. 이 방법은 배경 잡음과 포만트의 영향을 제거하기 위하여, 스펙트럼에 AMDF 함수를 적용하는 것이며 식 3-1과 같이 나타낼 수 있다.

$$SAMDF(w) = \sum_{k=1}^{F_{MAX}} |S_W(w) - S_W(w-k)| \qquad (3\text{-}1)$$

$$(w = 1, 2, \cdots\cdots, size)$$

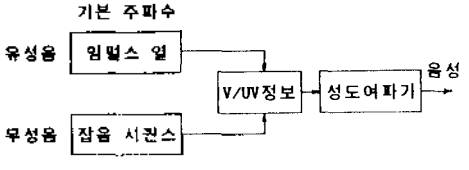

그림 1-2. MBE 음성 발생모델

FIG. 1-2. MBE speech generation model

여기서 F_{MAX}는 제 1포만트에 의한 스펙트럼 최대 에너지 위치이고, size는 한 프레임의 길이이다. 식 3-1에 의해 고조파 스펙트럼의 골이 인근한 골과 겹치는 경우 최소의 값이 된다. 이 함수는 포만트에 의해 원 스펙트럼이 평탄하지 않더라도 이에 비하여 SAMDF(w)의 기울기가 두드러지므로 이들을 서로 보상하면 평탄한 스펙트럼을 얻을 수 있다. 이렇게 강조된 고조파의 간격은 최소값 검출에 의한 간단한 분석값에 의해 얻어질 수 있다. 측정된 고조파의 갯수가 N이라 할때 F_0는 식 (3-2)에 의해 얻어진다[5].

$$F_0 = F_{MAX} / N \qquad (3\text{-}2)$$

SAMDF 함수에 의해 기본주파수 F_0를 얻으면 원 음성에 대한 진폭 스펙트럼 S(w)에 대한 포만트 스펙트럼 F(w)는 다음과 같은 lifer함수를 통과시킴으로써 얻어질 수 있다.

$$F(w) = \frac{1}{F_0} \sum_{L=1}^{F_0} S(w - L) \qquad (3\text{-}3)$$

$$(w = 1, 2, \cdots\cdots, size)$$

이렇게 포만트 스펙트럼이 구해지면 원 스펙트럼과의 차이를 구한다. 평탄화된 고조파 스펙트럼 E(w)은 식 3-4와 같이 나타낼 수 있다[6].

$$E(w) = S_W(w) - F(w) \qquad (3\text{-}4)$$

$$(w = 1, 2, \cdots\cdots, size)$$

3-2. 평탄화기법에 의한 스펙트럼 V/UV 결정

그림 3-1(a)는 유성음에 대하여 해밍창을 적용한 원 음성파형이며, (b)는 (a)에 FFT를 적용한 후 로그 스펙트럼을 구한 것이다. (c)는 스펙트럼 평탄화기법을 이용하여 구한 포만트성분이고, (d)는 (b)와 (c)에 대한 차를 나타낸 것으로써 평탄화된 고조파성분이다.

사용된 원 음성이 유성음이고, 또한 고조파 성분이 잘 찾아졌다면, F_0 길이의 한 구간에서 영을 통과하는 점은 두 개가 존재하여야 한다. 이 말은 또한 어떤 일정한 구간의 길이가 L이고, 기본주파수를 F_0이라 하면, 유성음의 경우에 $2*(L/F_0)$개의 영교차점이 발생하게 된다는 의미이다. E(w)에 대한 영교차율은 식 3-5에 의해 얻어진다.

$$Z(w) = \frac{1}{2\times L} \times \sum_{w=1}^{L} |SGN[E(w)]-SGN[E(w+1)]| \qquad (3\text{-}5)$$

여기서

$$SGN[E(w)] = 1 \qquad E(w) \geq 0$$
$$= -1 \qquad E(w) < 0$$

Z(w)를 구간 단위로 사용하기 위하여, 구간 영교차값 $Z_I(w)$으로 대치하며, 이는 다음과 같다.

$$Z_I(w) = L \times Z(INT[w/L] \times L) \qquad (3\text{-}6)$$

여기서 $INT[\cdot]$는 $[\cdot]$속의 값을 정수값으로 반올림하는 함수이다.

MBE 부호화시에 구간 영교차값, $Z_I(w)$과 이에 대한 문턱값을 이용하여 V/UV 구간을 결정한다. 그림 3-2(a)는 원 음성이고, (b)는 이에 대한 스펙트럼이며, (c)는 고조파가 평탄화된 결과이며 (d)는 (c)의 구간 대표 영교차값을 표시한 것이다.

Ⅳ. 실험 및 결과

시뮬레이션하기 위해 IBM PC/AT에 마이크 입력이 가능하도록 12-비트 A/D 변환기를 인터페이스시켰다. 다음과 같은 음성시료를 발성하고 8KHz의 표본화율로 양자화 하면서 메모리에 표본화된 데이타를 저장하였다.

발성 1) 25세 남성 : /인수내 꼬마가 천재소년을 좋아한다/

발성 2) 23세 여성 : /지금 거신 전화는/

발성 3) 28세 남성 : /호서대 전자공학과 음성 신호처리 연구/

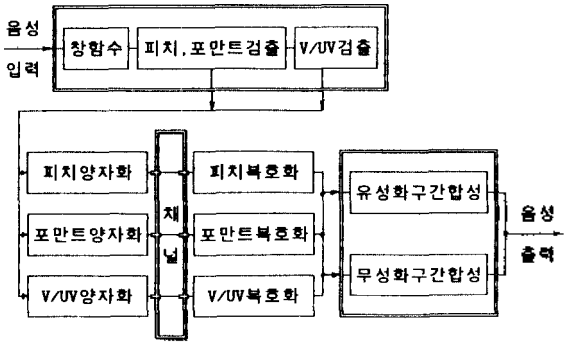

그림 2-1. MBE알고리듬에 대한 블럭도.

FIG. 2-1. Block diagram for MBE algorithm

전체 실험과정을 통하여 한 프레임의 길이를 512샘플로 하였다. 그림 4-1은 본 논문에서 제안한 알고리듬에 대한 블럭도이다. 먼저 음성에 해밍창을 취하고, 이에 FFT를 적용한 다음 LOG를 취하여 진폭 스펙트럼을 얻는다. 이에 SAMDF를 적용하여 기본주파수 F_0를 얻고 하모닉스 스펙트럼 평탄화기법으로 포만트 정보와 평탄화된 하모닉스를 얻는다. 하모닉스에 대한 구간대표 영교차값을 구하여 각 스펙트럼 구간에 대한 V/UV를 결정한다. 스펙트럼의 부분 구간 길이를 16 주파수 표본으로 하였다. 구간 영교차값에 대한 문턱값은 $3*(L/F_0)$로 하여 이를 초과하지 않으면 유성화구간이라 하고, 그렇지 않으면 무성화 구간이라 하였다.

피치검출 오류에 의한 영향을 알아보기 위하여 정확한 피치검출에 의한 결과와 인위적으로 피치를 부정확하게 검출한 결과와 비교하였다. 그림 4-2는 정확한 피치에 의한 결과와 피치를 0.8배, 1.2배 길이로 잘못 검출한 결과를 나타낸다. 영교차를 나타내는 그림에서 가로로 그어진 직선은 문턱값을 나타낸다. 그림에서도 알 수 있드시 3가지 결과가 유사하다. 따라서 본 연구에서 제안한 분류구간은 검출한 피치의 정확도에 둔감하다는 것을 알 수 있다.

그림 4-3, 4는 각각 음성 구간에 대한 결과이다. 각 그림에서 (a)는 원 음성 시료이고, (b)는 평탄화된 고조파 스펙트럼이다. (c)는 구간별 영교차율과 이에 대한 문턱값을 나타내었고, (d)는 (c)에 결정논리를 적용하여 구한 V/UV 구간 결정값이다. 여기서 문턱값을 초과하면 유성음 스펙트럼을 초과하지 않으면 잡음성 스펙트럼 구간으로 부호화한다.

V. 결론

음성을 분석한 후에 합성하는 보코더 시스템중에서 MBE 부호화는 스펙트럼 상에서 기본주파수의 3배 길이에 대한 V/UV 구간 판별방법을 사용하고 있다. 스펙트럼의 어떤 구간이 유성화 구간이라면 여기원으로써 F_0간격의 고조파를 사용하고 무성화 구간이라면 백색 잡음에 대한 스펙트럼을 여기원으로 사용한다. 이렇게 함으로써 합성음의 음질을 높이고 주위 잡음에 대한 영향을 줄일 수 있었다. 그러나, 사전에 검출되는 기본주파수와 스펙트럼 포락이 잘못 검출된다면, V/UV결정시 오차가 누적되어 검출 오류가 크게 발생된다는 문제점이 있다.

이러한 문제를 극복하기 위하여 스펙트럼 평탄화법에 의한 고조파 스펙트럼의 영교차율을 이용하는 방법을 제안하였다. 먼저 기본 주파수를 검출한 후에 이를 이용하여 스펙트럼 포락선을 구하고, 원 스펙트럼과의 차를 구하여 평탄화된 고조파 스펙트럼을 얻는다. 얻어진 스펙트럼에 대한 영교차율을 이용하여 정해진 스펙트럼 구간의 V/UV를 결정한다.

본 논문에서 제안된 방법은 MBE 부호화시 기본주파수와 스펙트럼 포락선의 검출오차에 의한 영향이 적고, 결

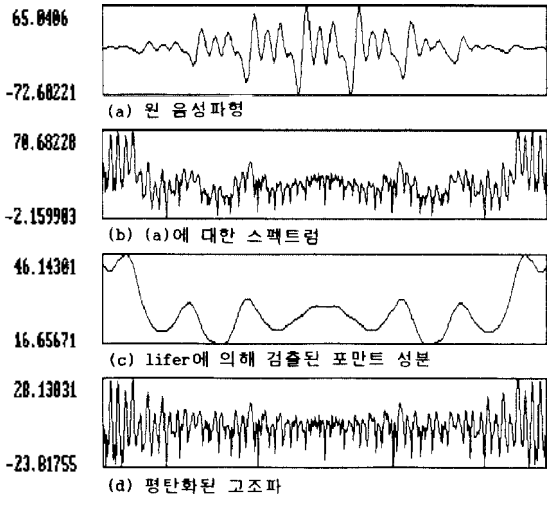

그림 3-1. 평탄화 기법에 의한 스펙트럼 포락선과 고조파의 검출

FIG. 3-1. Detection of Spectral envelop and harmonices by using flattening method

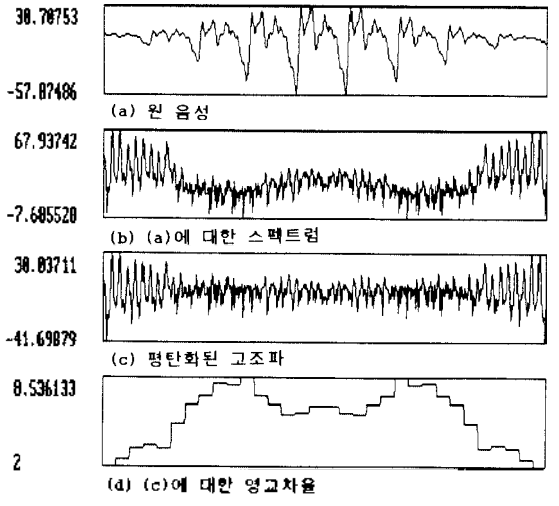

그림 3-2. 평탄화된 고조파 스펙트럼의 영교차값

FIG. 3-2. Zero-crossing rate of flatten harmonices

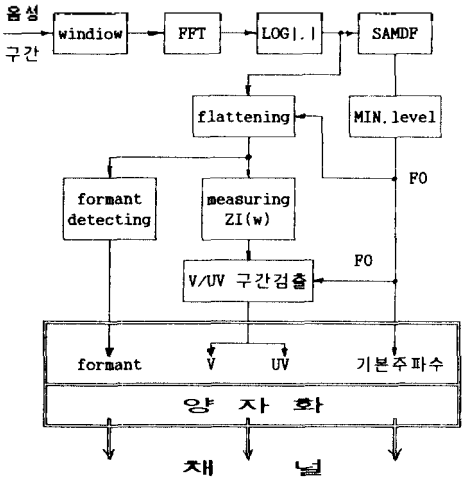

그림 4-1. 제안된 V/UV 구간 검출에 대한 알고리듬

FIG. 4-1. The proposed detection algorithm of V/UV segmentation for the MBE Vocoder

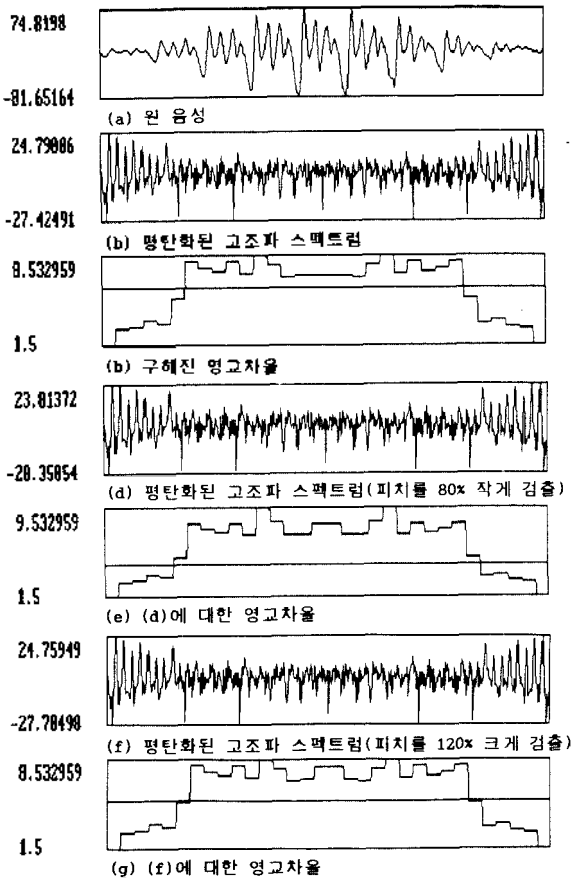

그림 4-2. 영교차값과 문턱값의 변화 (문턱값 3*(L/F0))

FIG. 4-2. Variation of Zero-crossing rate and threshold

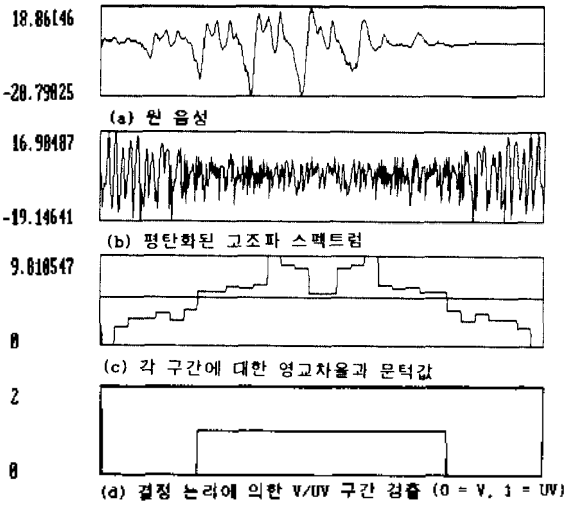

그림 4-3. 얻어진 V/UV 구간 검출 (문턱값 3*(L/F0))

FIG. 4-3. Result of V/UV segment detection

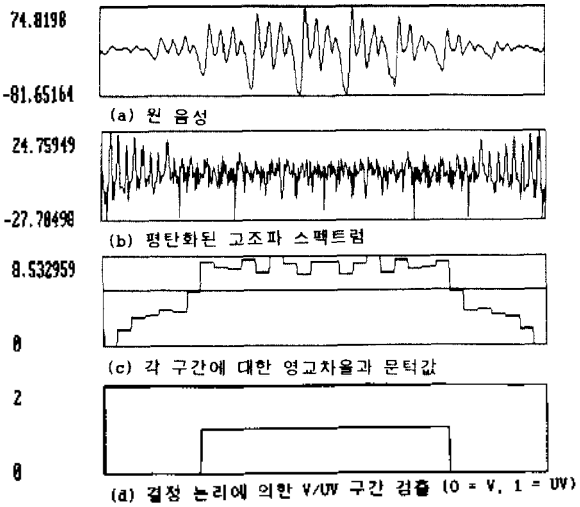

그림 4-4. 얻어진 V/UV 구간 검출 (문턱값 3*(L/F0))

FIG. 4-4. Result of V/UV segment detection

정논리가 간단하다는 이점이 있다. 본 논문에서 사용되는 스펙트럼 평탄화 기법은 스펙트럼의 V/UV 구간 검출뿐만 아니라 피치주기와 스펙트럼 포락선을 검출할 수 있으므로, MBE법에서 사용되는 모든 파라미타를 검출할 수 있다.

[REFERENCE]

[1] L.R.Rabiner and R.W.Schafer, *Digital Processing of Speech Signals*, Prentice-hall, 1978.

[2] S.Saito and K.Nakata, *Fundamentals of Speech Signal Processing*, Academic Press, 1985.

[3] P.E.Papamichalis, *Practical Approaches To Speech Coding*, Prentice-Hall, 1987.

[4] Bishnu S. Atal, V. Cuperman and A. Gersho, *Advances in Speech Coding*, Kluwer Academic Publishers, 1991

[5] 배명진, 박찬수, 안수길, "On the frequency Domain Pitch Detection of Noise Corrupted Speech Signals - Minimizing the Effects of the F1 by the Spectral AMDF -", 한국 음향학회 논문집, 10권 4호, pp12-18, 1991

[6] 이해군, 배명진, 안수길, "음성 고조파의 Flattening에 의한 시간-주파수 혼성형 피치검출법", 전자공학회 하계 학술발표 논문집, 제 15권 1호, pp 681- 684, 1992년 6월.