한국음향학회지 제22권 제4호 pp. 286~292 (2003)

잡음환경에서 우리말 연속음성의 무성자음 구간 추출 방법

Extraction of Unvoiced Consonant Regions from Fluent Korean Speech in Noisy Environments

박 정 임*, 하 동 경*, 신 옥 근*
(Jeong-Im Park*, Dong-Gyung Ha*, Ok-Keun Shin*)

*한국해양대학교 컴퓨터공학과

(접수일자: 2002년 1월 24일; 수정일자: 2002년 8월 19일; 채택일자: 2003년 3월 12일)

음성 구간 추출이란 입력된 음성신호를 음성 구간과 묵음, 또는 잡음 구간으로 구분하는 과정이다. 잡음이 섞여 있는 음성신호의 무성자음신호는 잡음신호와 매우 유사하다. 따라서 음성 구간을 추출하거나 잡음을 제거 또는 감소시킬 때 무성자음에 특별히 주의하지 않으면 무성자음을 손상시키거나 잘못된 잡음 추정으로 이어질 수 있다. 본 논문에서는 잡음 환경에서 연속 음성신호의 음성 구간을 정확하게 추출하기 위해 잡음과 무성자음 사이의 경계를 명시적으로 검출함으로써 무성자음의 구간을 추출하는 방법을 제안한다. 제안하는 추출방법은 Hirsch가 잡음 추정을 위해 사용한 히스토그램 방법과 연속된 프레임 사이의 주파수 성분의 유사성을 나타내는 파라미터들을 이용하였다. 제안한 방법의 성능을 평가하기 위해 음성신호에 SNR이 각각 10 dB와 15 dB인 7가지의 잡음을 첨가하여 무성자음신호의 추출 실험을 수행하였다.

핵심용어: 음성구간추출, 히스토그램, 잡음 제거, 무성자음과 잡음 경계 추출, 잡음첨가 음성 신호
투고분야: 음향처리 분야 (2,3)

Voice activity detection (VAD) is a process that separates the voice region from silence or noise region of input speech signal. Since unvoiced consonant signals have very similar characteristics to those of noise signals, it may result in serious distortion of unvoiced consonants, or in erroneous noise estimation to carry out VAD without paying special attention on unvoiced consonants. In this paper, we propose a method to extract in an explicit way the boundaries between unvoiced consonant and noise in fluent speech so that more exact VAD could be performed. The proposed method is based on histogram in frequency domain which was successfully used by Hirsch for noise estimation, and also on similarity measure of frequency components between adjacent frames. To evaluate the performance of the proposed method, experiments on unvoiced consonant boundary extraction was performed on seven kinds of noisy speech signals of 10 dB and 15 dB SNR respectively.

Keywords: *Voice activity detection, Histogram, Noise estimation, Unvoiced consonant boundary extraction, Noisy speech signal*
ASK subject classification: *Speech signal processing (2,3)*

I. 서 론

음성인식 시스템에서 필수적인 전처리 (preprocessing) 중의 하나가 음성 구간을 추출하는 과정이다. 음성 구간 추출이란 입력된 음성신호를 음성 구간과 묵음, 또는 잡음 구간으로 구분하는 과정으로 음성 구간 추출의 정확도는 인식률에 상당한 영향을 미치게 된다[1]. 잡음이 음성신호에 첨가되면 원래의 음성신호를 왜곡시키며 특히 무성 음성신호는 잡음신호와 유사하기 때문에 이들 사이의 구별은 더욱 어려워진다.

Junqua[3]가 제안한 방법에서는 잡음이 있는 연속된 숫자음 발성신호에 대해 첫 번째 나타나는 유성음과 마지막에 나타나는 유성음의 위치를 찾아 IORB (island of

교신저자: 박정임 (jipark@kmaritime.ac.kr)
606-791 부산광역시 영도구 동삼동
한국해양대학교 컴퓨터공학과
전화: 051-410-4928; 팩스: 051-404-3986)

reliability boundary)를 먼저 결정하였다. 결정된 IORB의 앞부분과 뒷부분을 조사하여 음성 구간의 시작점과 끝점을 추출함으로써 인식률을 향상시킬 수 있었다.

본 논문에서는 잡음이 선형적으로 첨가된 연속 음성신호의 음성 구간을 정확히 추출하기 위해 상호간의 특성이 비슷한 잡음과 무성자음 사이의 경계를 명시적으로 추출하는 방법을 제안한다. 이를 위해 먼저 하동경 등이 제안한 방법[5,6]을 이용하여 유성음 구간을 추출한 다음, 추출된 유성음 구간을 제외한 나머지 비유성음 구간에서 잡음과 무성자음 사이의 경계를 추출한다. 먼저 유성음 구간을 제외함으로써 대상 범위를 줄이고, 본고에서는 구분이 어려운 무성자음과 잡음의 경계를 추출하는데 초점을 두었다. 제안하는 방법은 Hirsch[4]가 잡음 추정을 위해 사용한 히스토그램 방법과 추정한 기준 잡음 프레임(frame)과 각 프레임 사이의 주파수 성분의 유사성을 나타내는 파라미터들을 이용한다. 제안한 방법의 성능을 평가하기 위해 잡음이 없는 음성신호에 신호 대 잡음비(SNR: signal to noise ratio)가 10 dB와 15 dB인 백색잡음과 6개의 유색잡음 등 모두 7가지의 잡음을 첨가하여 잡음과 무성자음 사이의 경계 추출 실험을 수행하였다. 제안하는 방법은 음성인식 및 음성코딩은 물론, 정확한 잡음 추정 및 잡음 제거에도 이용될 수 있을 것으로 기대된다. 본 논문의 II에서 잡음이 섞인 음성신호의 특징을 간략하게 기술하고, III에서는 무성자음 특성과 잡음의 경계 검출을 위한 파라미터 추출 방법, 추출된 파라미터를 이용하여 경계 검출을 하는 방법을 IV에서 설명하고, 제안한 방법에 대한 실험과 성능 평가를 V에 기술한 다음, VI에서 결론을 맺는다.

II. 잡음이 섞인 음성신호의 특징

유성음은 신호가 주기적이며 무성음에 비해 단구간(short-term) 에너지가 크고, 저주파 영역에 많은 에너지가 분포되는 특성을 가지므로 잡음과 비교적 구분하기 쉽다. 그러나 무성음은 신호의 주기성이 없으며 단구간 에너지가 상대적으로 작고 잡음신호와 매우 유사하다.

그림 1은 순수한 음성 신호와 잡음이 섞인 음성신호의 파형과 스펙트로그램 (spectrogram)의 예이다. 그림 1에서 타원으로 표시된 부분은 무성자음 구간이며 가로실선으로 표시된 부분은 잡음 구간, 그리고 아무 표시도 하지 않은 부분은 유성음 구간이다. 그림 1(a)와 같이 잡음이 섞이지 않은 경우에는 무성자음 구간과 묵음 구간을 비교적 쉽게 구분할 수 있지만, 공장 기계 잡음이 첨가된 그림 1(b)의 경우에는 무성자음 구간과 잡음 구간을 구분하기 어려우며 SNR이 낮아질수록 더욱 구분하기 어려워진다.

그러나 그림 1(c)에 보인 것과 같이 잡음 구간과 무성자음 구간의 주파수 스펙트럼이 서로 다른 분포를 띄는 경우에는 구분이 가능할 수도 있다. 본고에서는 이점에 착안하여 잡음 구간과 무성 자음 구간을 구별하는 방법을 제안한다.

III과 IV에서는 각각 경계 검출에 사용되는 파라미터를 추출하는 방법과 비유성음 구간에서 무성자음 구간과 잡음 구간을 검출하는 방법에 대하여 기술한다.

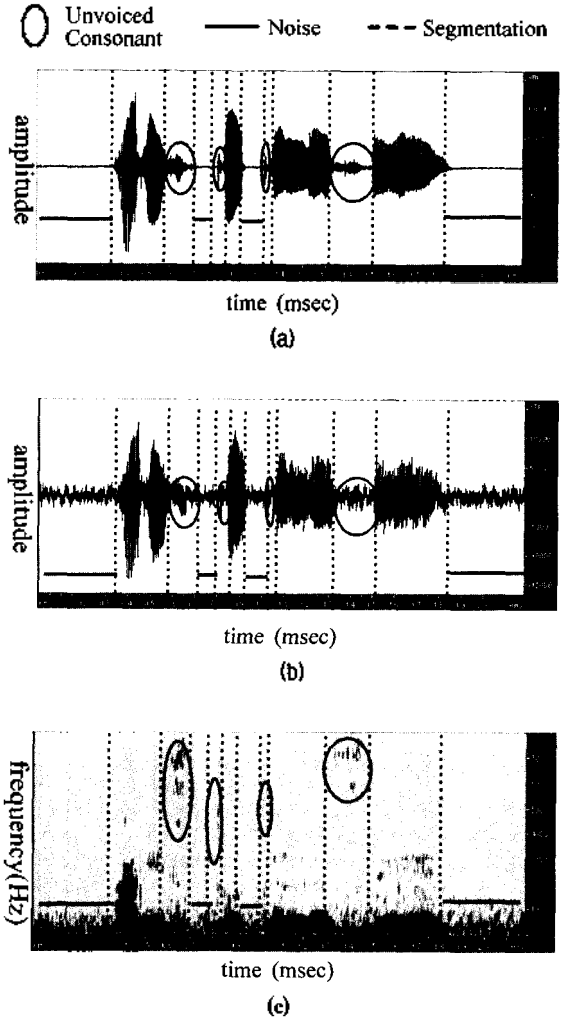

그림 1. 여성화자가 발화한 /아리스토텔레스에/의 (a) 음성 신호 파형, (b) (a)에 10 dB의 공장 잡음이 첨가된 파형, (c) (b)의 스펙트로그램

Fig. 1. (a) Clean speech waveform of a female speaker's utterance/ɑ-li-s-tʰo-tʰel-le-s-e/, (b) factory noise added speech waveform of (a) (SNR=10 dB), (c) spectrogram of (b).

III. 경계 검출을 위한 파라미터 추출

무성자음 구간과 잡음 구간의 경계를 검출하기 위하여 본 논문에서는 세 가지 파라미터와 잡음 모델을 사용한다. 먼저, 입력신호의 모든 프레임의 저주파 영역 평균 에너지 E^{lpf}를 구한다. 다음으로 잡음으로 확인된 구간(입력신호의 초기부분 포함)에 대하여 주파수 영역의 에너지 히스토그램의 최빈값과 편차로 구성되는 잡음 모델 λ를 설정한다.

파라미터 추출을 위해 비유성음 구간 내에 있는 각 프레임의 밴드들과 잡음 모델 λ의 밴드들을 서로 비교하여 파워의 차이가 큰 '특징밴드'를 먼저 찾아낸다. 프레임별 특징밴드의 수, 그리고 연속된 특징밴드들 중 가장 긴 것의 밴드 수를 파라미터로 추출한다. 아래에 이들 파라미터 추출과 잡음 모델 설정에 대하여 자세히 설명한다.

3.1. 저주파 영역의 평균 에너지(E^{lpf})

무성자음과 잡음의 경계 검출에 앞서 정확한 유성음 구간을 검출하여 고려 대상에서 제외함으로써 무성음과 잡음 경계 추출에만 초점을 맞출 수 있다. 이를 위해 각 프레임의 저주파 영역 평균 에너지 E^{lpf}를 식 (2)와 같이 구한 다음 이를 이용하여 유성음 구간을 재설정한다.

$$p_{i,k} = |X_i(k)|^2 = \sum_{n=0}^{N-1} \left| x_i(n)\, h(n) e^{-j\frac{2\pi nk}{N}} \right|^2 \quad (1)$$

$$E_i^{lpf} = \frac{1}{L} \sum_{k=0}^{L-1} p_{i,k} \quad (2)$$

여기서, N은 프레임 크기, $x(n)$은 음성신호, $h(n)$은 해닝 창 함수 (hanning window function), i는 프레임 인덱스, k는 샘플링 주파수 f_s를 FFT 포인터의 수로 균일하게 나눈 주파수 밴드의 인덱스이다. L은 저역통과 필터의 차단 주파수 (cutoff frequency)에 해당하는 밴드의 인덱스이다. 음성 신호의 파워 스펙트럼에서 유성음 구간의 파워는 600 Hz 이하의 주파수 영역에 밀집되어 나타나는 경우가 많으므로[10], FFT를 이용한 저역통과 필터(low-pass filter)의 차단 주파수, L을 600 Hz로 설정하였다.

유성음을 제외한 구간의 각 프레임 에너지 E_i^{lpf}의 히스토그램에서 최빈값 $E^{lpf(peak)}$를 구한 다음, 유성음 구간 배제 설정에 사용할 문턱값 $E_{N_{TH}}$를 아래의 식과 같이 결정한다.

$$E_{N_{TH}} = E^{lpf(peak)} \cdot (1 + \alpha_E) \quad (3)$$

$$\alpha_E = E_{\max}\, dB - E^{lpf(peak)}\, dB \quad (4)$$

여기서, $E_{\max} dB$는 신호가 갖는 최대 파워를 dB로 나타낸 것이며 E^{lpf}가 $E_{N_{TH}}$보다 크면 유성음, 아니면 비유성음 구간으로 결정한다.

3.2. 잡음 모델 설정 $\lambda(P_k^{peak}, \hat{\sigma}_k)$

잡음 구간에서 주파수 밴드별로 파워 스펙트럼의 히스토그램을 구하고 이들의 최빈값 (P_k^{peak})을 각 주파수 밴드의 대표값으로 정한다. 그리고 밴드별로 구해진 최빈값을 중심으로 하는 편차 $\hat{\sigma}$을 식 (5)와 같이 구한다.

$$\hat{\sigma}_k = \sqrt{\frac{1}{M} \sum_{i=1}^{M} (p_{i,k} - P_k^{peak})^2} \quad (5)$$

여기서, M은 잡음 구간의 프레임의 수이다. P_k^{peak}와 $\hat{\sigma}_k$를 파라미터로 하여 잡음 모델 $\lambda(P_k^{peak}, \hat{\sigma}_k)$를 결정한다.

3.3. 특징밴드의 수 *BandCount*와 *BandCount*의 문턱값

식 (6)과 같이 잡음 모델의 P^{peak}와 비유성음 구간의 프레임의 파워 p_i를 밴드별로 비교해 차가 식 (5)에서 구한 $\hat{\sigma}_k$보다 큰 밴드를 '특징밴드'라 정의하고, 식 (7)과 같이 한 프레임에 대한 특징밴드의 갯수, *BandCount*를 구한다.

$$flag_{i,k} = \begin{cases} 1, & |p_{i,k} - P_k^{peak}| > \hat{\sigma}_k \\ 0, & other\ wise \end{cases} \quad (6)$$

$$BandCount_i = \sum_{k=0}^{N/2} flag_{i,k} \quad (7)$$

여기서, i는 프레임의 인덱스이고 k는 밴드의 인덱스이다. $flag_{i,k}$ 값이 1이면 무성자음 특성의 밴드임을 의미하고, 0이면 잡음의 특성을 갖는 밴드를 뜻한다.

잡음 구간 프레임들의 *BandCount*를 가지고 프레임의 특성을 결정하기 위한 문턱값 Cnt_{TH}^{nois}와 Cnt_{TH}^{unv}를 식 (8), (9)과 같이 구한다.

$$Cnt_{TH}^{nois} = 2\alpha + \max(1, BandCount^{(peak)}) \quad (8)$$

$$Cnt_{TH}^{unv} = \frac{3}{2} \cdot Cnt_{TH}^{nois} \quad (9)$$

여기서, $BandCount^{(peak)}$는 잡음 구간에 속하는 프레임들의 *BandCount* 값들로 작성한 히스토그램에서의 최빈값이며, α는 $BandCount^{(peak)}$를 평균으로 하여 구한

*BandCount*의 표준 편차이다. 프레임의 *BandCount*가 Cnt_{TH}^{nois}보다 작으면 잡음 특성의 프레임이고, Cnt_{TH}^{unv}보다 크면 무성자음 특성의 프레임으로 결정한다.

3.4. 연속된 특징밴드의 최대 갯이

그림 1(c)의 예에서 볼 수 있는 것처럼 무성자음 구간의 주파수 스펙트럼은 일반적으로 잡음 구간에 비해 특징밴드가 일정한 대역에 연속적으로 나타나는 것을 알 수 있다. *BandWidth*는 이러한 특성을 나타낸 것으로 한 프레임 안에서 무성자음 특성을 띈 밴드 ($flag_{i,k}$가 1)가 최대로 연속된 경우의 밴드들의 수이다.

IV. 비유성음 구간에서의 무성자음과 잡음의 경계 추출 방법

비유성음 구간 (NVR: non-voiced region)에서 무성자음 영역 (UVR: unvoiced region)과 잡음 영역 (NR: noise region)의 경계는 그림 2의 과정을 거쳐 추출한다.

먼저, [5]에서 제안한 방법으로 구한 유성음구간의 경계 프레임의 인덱스 n을 다음의 알고리즘을 이용하여 재설정한다.

do while on the left/right edge of NVR
if $E_{n+k}^{lpf} \le E_n^{lpf} < E_{N_{TH}}$ ***then***
$n = n$ and *finish*
else
$n = n + k$ and ***continue***
end-if
end-do , $\begin{cases} k=1 & \text{for left} \\ k=-1 & \text{for right} \end{cases}$

여기서, E_n^{lpf}와 $E_{N_{TH}}$는 각각 식 (2)와 식 (3)에서 정의한 것이다.

다음으로 재설정된 유성음 구간 경계 정보를 이용하여 유성음 구간을 제거시키고 남은 비유성음 구간 각각의 프레임들에 대해 III장에서 설명한 파라미터 *BandCount*와 *BandWidth*를 구한다. 주어진 NVR의 각 프레임의 파라미터들을 문턱값과 비교하여 무성자음과 잡음 사이의 경계를 추출한다. 잡음 구간으로 판명된 구간이 있으면 이 잡음 구간의 신호로부터 잡음 모델 λ와 문턱값 Cnt_{TH}^{unv}와 Cnt_{TH}^{nois}를 갱신하여 다음 NVR에 적용한다. 입력신호의 첫 NVR에 적용할 잡음 모델과 문턱값들은 음성신호가 포함되지 않은 첫 800 ms 구간으로부터 구한다.

비유성음 구간에 나타날 수 있는 무성자음과 잡음의 구성 형태는 모든 발화가 모음을 중심으로 앞뒤에 자음이 위치하는 우리말의 특징에 의해 크게 다섯 가지로 나뉘어질 수 있다: 무성자음–잡음 (UN: unvoiced-noise), 잡음–무성자음 (NU: noise-unvoiced), 무성자음–잡음–무성자음 (UNU: unvoiced-noise-unvoiced), 무성자음 (U: all unvoice), 잡음 (N: all noise). 무성자음 구간은 위 다섯 가지 경우 모두에서 유성음 구간에 접하여 위치하고 있으므로, 유성음 구간과 접해 있는 NVR의 양쪽 가장자리에서 안쪽으로 잡음 특성의 프레임이 시작되는 지점을 찾음으로써 무성자음과 잡음의 경계를 추출할 수 있다.

무성자음 구간과 잡음 구간의 경계를 찾기 위한 과정은 다음과 같다. 먼저 3.2절과 3.3절에서 기술한 바와 같이

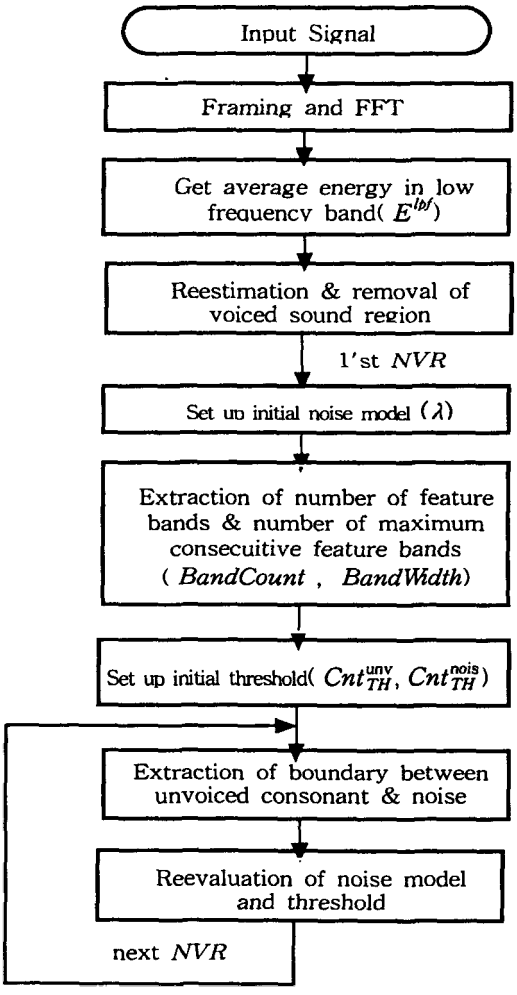

그림 2. 무성자음 구간과 잡음 구간 경계 검출 방법
Fig. 2. Proposed method for boundary detection between unvoiced consonant and noise regions.

표 1. 유성음이 무성음화된 구간의 지속 기간 분포
Table 1. Distribution of the duration of devocalized region.

Duration (ms)	Frequency	Duration (ms)	Frequency
70~80	2	140~150	6
80~90	8	150~160	5
90~100	10	160~170	1
100~110	10	170~180	0
110~120	26	180~190	1
120~130	17	190~200	0
130~140	11	200~210	1

$BandCount < Cnt_{TH}^{nois}$와 $E^{lpf} < E_{N_{TH}}$를 만족하는 프레임의 시작점 (UVR-Noise) 또는 끝점 (Noise-UVR)들을 잡음구간의 경계로 설정한다. 이때, 경계에서부터 잡음 구간 방향으로 $BandCount > Cnt_{TH}^{nois}$이거나 $BandWidth < 5$인 프레임이 두 프레임 이상 지속되면, 이것은 무성자음에서 잡음으로 천이되는 구간 중 무성자음의 특성을 많이 포함한 프레임으로 보고, 이 프레임들을 무성자음에 포함시켜 경계를 재설정한다. 끝으로 위의 과정을 통해 예측된 잡음 구간에 대하여 유성음이 무성음화된 구간의 최소 지속 기간 조건을 검사한다. 유성음 /ㅗ, ㅜ, ㅡ, ㅣ/ 등이 특정한 자음 /ㅆ, ㅊ, ㅋ, ㅌ, ㅍ/ 등과 결합되면 유성음이 무성음화되어 발성되는 경우가 발생한다. 표 1은 본 연구에서 사용한 한국어 음성 데이터베이스인 POW를 이용하여 구한 무성음화된 구간의 지속 기간 분포이다.

표 1에서 볼 수 있는 바와 같이 대부분 80 ms에서 160 ms의 지속 기간을 가지므로 무성음화된 구간의 최소 지속 기간을 80 ms로 설정하고, 이를 예측된 잡음 구간에 대하여 적용시킨다. 잡음 구간 안에 $BandCount > Cnt_{TH}^{nois}$이거나 $BandWidth \geq 5$인 구간이 80 ms이상 지속되면 무성음화된 구간으로 설정해 잡음 구간에서 제외시킨다.

위와 같은 방법으로 첫 번째 NVR에 대하여 무성자음과 잡음 구간의 경계를 추출한 다음, 잡음 구간에 해당하는 프레임들만을 이용해 잡음 모델 λ와 파라미터의 문턱값인 Cnt_{TH}^{unv}와 Cnt_{TH}^{nois}를 갱신한다.

갱신한 잡음 모델 λ와 파라미터 문턱값들을 다음 NVR의 무성자음과 잡음 구간 경계 추출에 사용한다.

V. 실험 및 결과 분석

제안한 무성자음과 잡음의 경계 추출 방법에 관한 실험을 위해서 한국어 음성 데이터 베이스인 POW (phonetically optimized word) 코퍼스(corpus)에서 남녀 각각 5명의 화자의 5음절 이상 비교적 길게 발성된 음성 데이터 240개를 발췌하여 사용하였다. 잡음 데이터는 NoiseX-92에서 자동차 잡음 (Leopard와 Volvo), 비행기 조종석의 잡음 (F16), 여러 사람들의 잡담 (Babble), 공장의 기계 잡음 (Factory), 핑크 잡음 등의 6개 의 유색 잡음과 백색 잡음, 그리고 잡음이 섞이지 않은 음성 데이터를 사용하였다. 음성 데이터는 16 kHz로 샘플링되었고 잡음 데이터는 20 kHz로 샘플링되었으며, 두 가지 데이터 모두 16 bit로 양자화되었다. SNR은 10 dB와 15 dB의 두 가지 경우를 고려하였다.

표 2. NVR에서 제안한 방법으로 무성자음과 잡음 경계 추출 결과
Table 2. Result of boundary detection between unvoiced consonant and noise in NVR by the proposed method.

Type	Insertion (%)		Deletion (%)		Detection (%)	
	15 dB	10 dB	15 dB	10 dB	15 dB	10 dB
Clean	7		5		95	
Babble	25	31	14	26	86	74
F16	23	27	10	15	90	85
Factory	27	33	11	16	89	84
Leopard	18	23	9	19	91	81
Pink	23	36	10	17	90	83
Volvo	15	19	8	13	92	87
white	19	39	12	17	88	83
Average	21.43	29.71	10.57	17.57	89.43	82.43

한 프레임의 크기는 16 ms으로 하였으며, 프레임 간격은 8 ms로 하였다. 먼저 프레임 단위의 데이터에 해닝창 (Hanning Window)을 적용한 뒤 256크기로 FFT를 한 다음 주파수 밴드별 파워를 구하였다. 다음으로 Ⅳ에서 설명한 방법으로 NVR에서 무성자음과 잡음의 경계를 검출하였다.

표 2는 잡음의 종류에 따른 경계 추출 결과이며, 그림 3과 그림 4는 기준 경계와 제안한 방법으로 찾은 경계의 거리 (distance)를 잡음 종류별로 나타낸 것이다.

검출 허용오차 범위는 프레임 간격의 두 배에 해당하는 ±16 ms로 설정하였다. 모든 잡음종류에 대하여 SNR 10 dB에서는 평균 82.4%, 15 dB인 경우는 평균 89.4%의 추출률을 나타내었다. 그림 3과 4에서 제안한 방법으로 찾은 경계와 기준 경계 사이의 오차가 대부분 ±16 ms 미만인 것으로 나타나 양호한 결과를 보였다. 다른 잡음환경에 비해 babble 잡음에서 추출률이 낮은 것은 여러 사람들이 웅성거리는 소리인 babble 잡음의 특성이 음성 신호

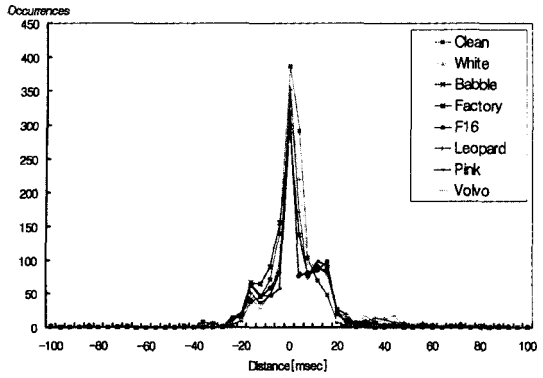

그림 3. 기준경계와 제안한 방법으로 찾은 경계의 거리 (SNR=15 dB)
Fig. 3. Distances between reference and detected boundaries (SNR=15 dB).

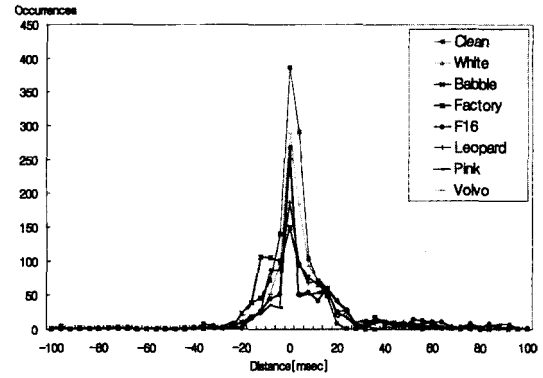

그림 4. 기준경계와 제안한 방법으로 찾은 경계의 거리 (SNR=10 dB)
Fig. 4. Distances between reference and detected boundaries (SNR=10 dB).

와 유사하기 때문이다.

오검출의 대부분이 발화가 끝나는 경계 지점을 잘못 결정한 것으로 나타났다. 발화 끝점 검출에 대한 오류 문제는 잡음이 없는 환경에서도 나타나는 문제이다.

발화의 대부분은 모음으로 끝나며, 이 때의 신호는 크기가 서서히 감소하고 매우 낮은 주파수 성분이 대부분으로 수작업 끝점 검출의 기준도 정확히 정의하기 어렵다.

발화 끝점의 신호 성분은 인식에 큰 영향을 미치지 않는 것으로 연구되어지고 있는 점을 감안하여 발화 끝점에 대한 검출 허용 오차를 다른 지점보다 완화시키면 추출률이 많이 향상될 것으로 기대된다.

VI. 결 론

본 논문에서는 잡음신호가 섞인 3음절이상의 연속 어절에서 무성자음과 잡음의 경계를 추출하여 음성의 시작점과 끝점은 물론, 발화 중간에 포함된 잡음 및 묵음 구간의 경계를 추출하는 방법을 제안하였다. 제안한 방법에서는 잡음이 섞인 음성신호에서 먼저, 유성음 구간을 찾아 제외시킴으로써 문제를 보다 단순화하였으며 신호의 특성상 잡음 성분과 유사하여 추출이 어려운 무성자음신호와 잡음신호의 경계 추출에 중점을 두었다. 경계를 찾기 위한 파라미터 추출을 위해서 히스토그램을 이용하였으며 추출된 파라미터는 비유성음 구간의 프레임의 특성을 무성자음신호에 가까운 프레임과 잡음신호에 가까운 프레임으로 결정한다. 프레임 각각에 대해 특성을 결정한 후 최종 경계 추출을 위해 무성자음신호와 잡음신호가 갖는 특징이 반영된 파라미터들을 생성하여 최종 경계를 찾는 방법을 제안하였다. 제안한 방법의 성능을 평가하기 위해 음성신호에 백색잡음과 6가지의 유색잡음을 SNR을 10 dB와 15 dB로 달리하여 첨가하였을 때의 추출률을 확인하였다. 잡음의 종류에 따라 다소 차이가 있으나 평균적으로 10 dB에서는 82.4%, 15 dB에서는 89.%4의 추출률을 나타냈다.

추출된 잡음 구간에서 잡음신호를 추정하여 제거하는 간단한 실험을 통해 무성자음신호 구간의 손실이 없이 잡음 구간에서 잡음 성분이 제거된 결과를 확인하였다. 그러나 음성 구간에 잔류 잡음 (musical tone)이 남아 있기 때문에 이러한 잔류 잡음을 제거하는 방법에 대해서는 보다 연구가 더 필요하다. 본 연구의 결과는 음성인식 및 음성코딩은 물론, 보다 정확한 잡음 구간 추정이 가능하므로 정확한 잡음 추정 및 잡음 제거를 위해서도 이용될 수 있을 것으로 기대된다.

감사의 글

본 연구는 한국해양대학교 BK21사업의 부분적인 지원을 받아 수행되었습니다.

참 고 문 헌

1. J. G. and Wilpon, L. R. Rabiner and T. B. Martin, "An improved word-detection algorithm for telephone-quality speech incorporating both syntactic and semantic constraints," *AT&T tech. J.*, 63 (3), 479-798, March 1984.
2. L. R. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, 1993.
3. J. C. Junqua, "A robust algorithm for word boundary detection in the presence in of noise," *IEEE, Transactions*

on Speech and Audio Processing, **2** (3), July 1994.
4. H. G. Hirsch, "Estimation of noise spectrum and its application to SNR estimation and speech enhancement," *Technical Report TR-93-012, International Computer Science Institute*, Berkeley, USA, 1993.
5. 하동경, 피치 정보를 이용한 모음의 특징 벡터 변별력 향상에 관한 연구, 한국해양대학교, 컴퓨터공학과 석사논문, 2000.
6. D.-G. Ha and O.-K. Shin, "Adaptation of pitch information in vowel feature extraction for speech recognition," EALPIT2000, 324-329, 2000.
7. 유건수, 김건명, 배명진, "쌍 자기 상관관계에 의한 음성 신호의 끝점검출," 제9회 음성통신 및 신호처리 워크샵 논문집, SCAS-9 (1), 133-137, 1992.
8. R. Chengalvarayan, "Robust energy normalization using speech/nonspeech discri-minator for german connected digit recogniti-on," *EUROSPEECH*, 61-63, 1999.
9. 박정임, 히스토그램을 이용한 무성자음과 잡음의 경계 추출, 한국해양대학교, 컴퓨터공학과 석사논문, 2001.
10. S. A. Liu, "Landmark detection for distinctive feature- based speech recognition," *J. Acoust. Soc. Am.*, **100** (5), 3417-3430, November, 1996.

저자 약력

● 박 정 임 (Jeong-Im Park)

1999년: 동의대학교 컴퓨터공학과 졸업 (학사)
2001년: 한국해양대학교 대학원 컴퓨터공학과 (공학석사)
2001년~현재: 한국해양대학교 대학원 박사과정

● 하 동 경 (Dong-Gyung Ha)

1997년: 한국해양대학교 컴퓨터공학과 졸업 (학사)
2001년: 한국해양대학교 대학원 컴퓨터공학과 (공학석사)
2001년~현재: 한국해양대학교 대학원 박사과정

● 신 옥 근 (Ok-Keun Shin)

1981년: 서강대학교 전자공학과 졸업 (학사)
1983년: 부산대학교 전자공학과 (공학석사)
1989년: 프랑스 Université de Franche-Comté (공학박사)
1983년~1995년: 한국전자통신연구소 선임연구원
1995년~현재: 한국해양대학교 자동화정보공학부 부교수
※ 주관심분야: 신호처리, 음성신호처리, 음성인식