한국음향학회지 제21권 제8호 pp. 727~734 (2002)

유성음 구간 검출을 위한 간단한 알고리즘에 관한 연구

A Study on the Simple Algorithm for Discrimination of Voiced Sounds

장 규 철*, 우 수 영*, 박 용 규*, 유 창 동*
(Gyu-Cheol Jang*, Soo-Young Woo*, Yong-Kyu Park*, Chang D. Yoo*)

*한국과학기술원 전자전산학과

(접수일자: 2002년 4월 23일; 수정일자: 2002년 10월 7일; 채택일자: 2002년 11월 4일)

본 논문에서는 유·무성음 구간을 검출하기 위한 간단한 알고리즘을 제안한다. 제안된 방법은 음성의 유·무성음의 주기성에 대한 특성을 보완할 수 있는 저대역 에너지와 영교차율, 그리고 주기성의 안정성을 판단하기 위한 피치 변화량을 파라미터로 사용하였다. 유·무성음의 구간검출을 음소단위의 검출이라는 측면에서 접근하여 음소군의 검출율과 음소군 내의 음소의 검출율을 얻었다. TIMIT 코퍼스 (corpus)를 데이터 베이스로 사용하여 실험했을 때 유성음 음소 검출율이 약 13% 향상되었다.

핵심용어: 유성음 검출, 피치 변화량
투고분야: 음성처리 분야 (2.4)

A simple algorithm for discriminating voiced sounds in a speech is proposed in this paper. In addition to low-frequency energy and zero-crossing rate (ZCR), both of which have been widely used in the past for identifying voiced sounds, the proposed algorithm incorporates pitch variation to improve the discrimination rate. Based on TIMIT corpus, evaluation result shows an improvement of 13% in the discrimination of voiced phonemes over that of the traditional algorithm using only energy and ZCR.

Keywords: *Voiced sound discrimination, Pitch variation*
ASK subject classification: *Speech signal processing (2.4)*

I. 서 론

사람의 음성은 크게 유성음과 무성음으로 나눌 수 있다. 이는 성대의 진동 유, 무에 의해서 구분된다. 공기가 성문을 통과할 때 성문이 거의 닫힌 상태에서 얇은 막의 성대가 진동하며 만들어지는 음을 유성음 (有聲音: voiced sounds)이라 하며, 반대로 성문이 열려져서 성대가 진동하지 않으며 만들어지는 음을 무성음 (無聲音: unvoiced sounds)이라 한다. 이러한 발성원리에 의해서, 유성음과 무성음은 각각 주기신호와 고대역 랜덤 신호로 표현할 수 있다[1].

사람의 음성을 유·무성음으로 구별할 수 있다면 음성정보적 측면에서 여러 가지 음성가공이나 처리에 많은 도움이 된다. 특히 음성인식의 관점에서 유·무성음의 검출은 연속음 인식과 더불어 사용되었을 때 검색 공간 감소와 인식률의 향상을 기대할 수 있다.

음성의 음향학적 특징을 이용해서 신호처리기법을 통해 구별하는 시도가 많이 있어 왔다[4,11]. 이러한 시도들은 유성음의 피치주기뿐만 아니라 여러 가지 특징 파라미터들을 이용하여 주기성 판단에 의한 유·무성음 검출을 향상시키려고 하였다. 그리고 패턴인식 기법을 이용한 방법[3,10]이나 신경회로망을 이용한 방법도 많이 시도 되어왔다[5-8].

본 논문에서는 유성음의 근사적인 주기성을 보완하기 위해 피치를 정확히 판단하기보다는 피치의 분포를 이용한 유무성음 판단에 대해 연구해 보았다. 앞에서 언급한 기존에 제시된 복잡한 방법 대신에 에너지, 영교차율, 그리고 피치 주기의 변화량에 관한 분포를 이용한 유성음 음소군 구간 판단을 위한 간단한 알고리즘을 제

교신저자: 장규철 (nic@kaist.ac.kr)
(305-701 대전시 유성구 구성동 373-1
한국과학기술원 전기 및 전자공학과 Multimedia Processing Lab.
(전화: 042-869-5470; 팩스: 042-862-0559)

안한다.

II장에서 본 알고리즘에 사용된 특징 파라미터들에 대해서 알아보고, III장에서는 유성음 검출을 위한 알고리즘을 제안한다. IV장에서 실험결과를 보이고 그에 대한 고찰을 한다. 그리고 V장에서 결론을 맺는다.

II. 특징 파라미터

2.1. 저대역 에너지 (low-frequency energy)

전통적으로 저대역 에너지는 유·무성음 구별, 음성구간 구별 등 음성의 특징으로 널리 사용되는 파라미터이다. 유·무성음의 에너지 분포를 살펴보았을 때 유성음은 저대역에 에너지가 높은 반면, 고대역 랜덤 신호로 모델링 할 수 있는 무성음은 그렇지 못하다. n번째 프레임 에너지 $e[\cdot]$은 다음과 같이 표현할 수 있다.

$$e[n] = \sum_{i=1+nM}^{N+nM} s_{LP}[i] \tag{1}$$

여기서 N은 프레임의 길이 M은 프레임의 이동 (shift) 길이를 나타낸다. $s_{LP}[\cdot]$는 저대역 통과된 음성 신호이다. 본 논문에서는 에너지의 단위로서 dB 단위를 사용하였으며, N, M은 모두 160으로 사용하였으며 저대역 필터로는 1 kHz를 차단 주파수로 하는 14차 butterworth 필터를 사용하였다.

2.2. 영교차율 (zero-crossing rate)

에너지와 마찬가지로 영교차율은 유·무성음 구별 등 음성을 구별하기 위한 특징으로 전통적으로 사용되는 파라미터이다. 유성음 구간에서는 일정 크기를 주기를 가지는 큰 진폭의 파형으로 표현되므로 영교차율이 낮고, 반대로 무성음의 구간에서는 영교차율은 높게 나타난다. n번째 프레임의 영교차율 $z[\cdot]$은 다음과 같이 표현할 수 있다.

$$z[n] = \sum_{i=1+nM}^{N+nM} \frac{1}{2N} abs[\, sgn(s[i]) - sgn(s[i-1]) \,]$$

$$where,\ sgn(y) = \begin{cases} -1 & , y<0 \\ 1 & , y \geq 0 \end{cases} \tag{2}$$

여기서 N은 프레임의 크기, M은 프레임의 이동 (shift) 길이를 나타내며 $s[\cdot]$는 음성신호이다.

2.3. 피치 주기 (pitch period)

음성의 피치 주기 정보는 음성의 주기성분의 정보를 나타내는 좋은 파라미터로 널리 사용되어 왔다. 음성신호의 피치 주기를 구하는 방법으로는 선형 예측 (LP: linear prediction) 오차의 자기상관을 이용해서 구한다[1,9]. 자기상관 $R[i]$로부터 얻게 되는 n번째 프레임의 피치 주기 $T[n]$은,

$$T[n] = \arg\max_i R[i] \quad , \quad T_{\min} \leq i \leq T_{\max} \tag{3}$$

이다. 본 논문에서 T_{min}은 20, T_{max}는 140으로 사용하였다. 인간의 유성음의 피치 주기는 약 60~400 Hz 사이에 존재하고, 이는 8 kHz 샘플링에서는 20~140 샘플에 해당한다.

2.4. Delta 피치 주기와 Delta-delta 피치 주기

프레임과 프레임간의 피치 주기의 시간에 따른 변화량은 유성음과 무성음을 구분할 수 있는 중요한 요인이다. delta 및 delta-delta 주기 ($\Delta T[n]$, $\Delta\Delta T[n]$)를 다음과 같이 정의한다.

$$\Delta T[n] = T[n] - T[n-1] \tag{4}$$

$$\Delta\Delta T[n] = \Delta T[n] - \Delta T[n-1] \tag{5}$$

그림 1은 TIMIT 코퍼스[2]에서의, 피치 주기와 피치 주기변화량에 대한 분포를 나타낸 히스토그램이다. 피치 주기의 변화를 나타내는 $\Delta T[n]$와 $\Delta\Delta T[n]$을 관찰하면 유성음의 분포가 무성음에 비해서 더욱 원점에 밀집되어 있다. 이것은 유성음의 구간에서 안정된 피치 주기를 가지는 구간이 많다는 것을 보여주며, 반대로 무성음의 경우에는 피치 주기의 값이 고르지 않다는 것을 나타낸다. 유성음과 무성음 사이의 $T[n]$의 분포는 두드러진 차이점이 관찰되지 않으므로 이를 이용하기는 힘들다.

III. 유성음 구간의 검출

3.1. 선형함수를 이용한 판별율과 피치의 시간적 변화량

저대역 에너지와 영교차율은 유·무성음의 검출에 있어서 유용한 정보들이다. 그림 2는 TIMIT 코퍼스로부터 얻은 약 2000개의 프레임에 대해서 유·무성음 구간에서

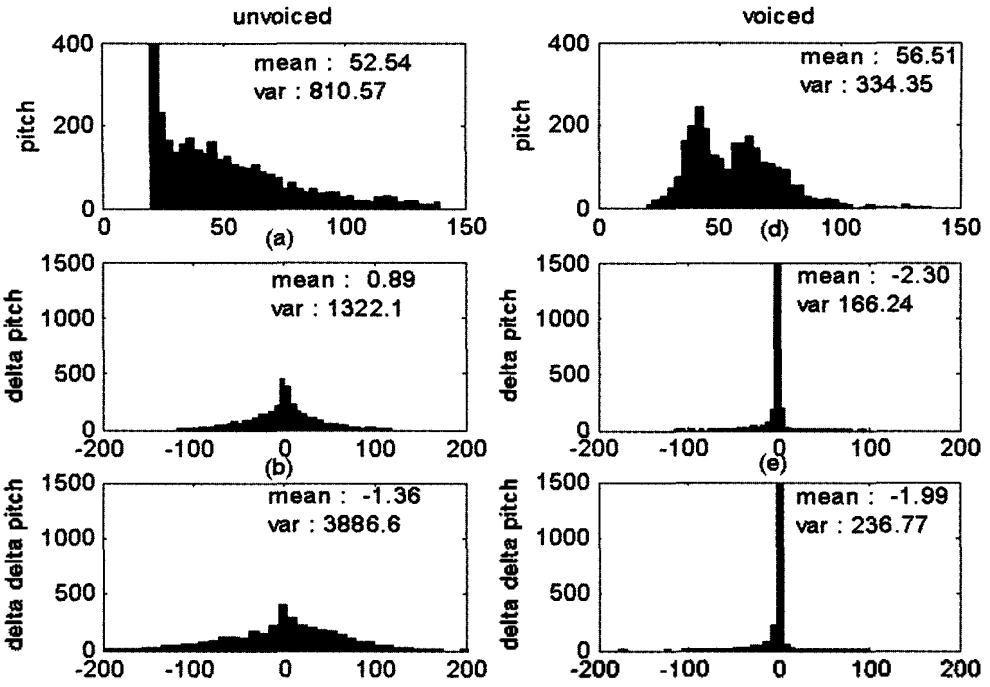


그림 1. TIMIT 코퍼스를 이용한 피치 주기의 분포, (a,b,c) 무성음 구간에 대한 $T[n]$, $\Delta T[n]$, $\Delta\Delta T[n]$ 의 분포, (d,e,f) 유성음 구간에 대한 $T[n]$, $\Delta T[n]$, $\Delta\Delta T[n]$ 의 분포

Fig. 1. Using TIMIT corpus, (a,b,c) distribution of $T[n]$, $\Delta T[n]$, $\Delta\Delta T[n]$ in unvoiced segments respectively (d,e,f) distribution of $T[n]$, $\Delta T[n]$, $\Delta\Delta T[n]$ in voiced segments respectively.

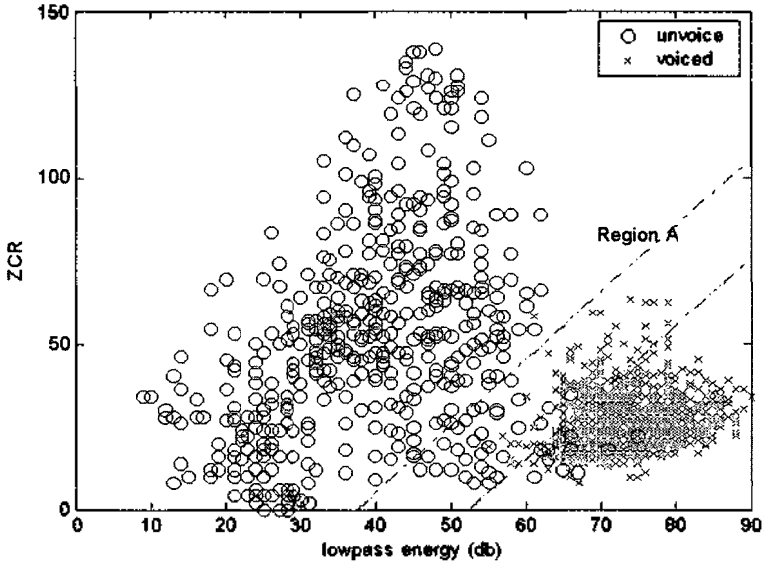


그림 2. 저대역 에너지와 영교차율과의 관계

Fig. 2. Relationship between low-frequency energy and ZCR.

의 에너지와 영교차율의 관계를 나타낸 그래프이다. 유성음과 무성음의 구간이 비교적 뚜렷이 나뉘어지는 것을 관찰할 수 있다. 일반적으로 무성음은 유성음보다 에너지가 더 작고, 영교차율의 분포가 유성음보다 더 넓게 퍼져서 분포하는 경향이 있다[1].

유·무성음을 구분하기 위해서 프레임의 저대역 에너지 e와 영교차율 z의 선형함수 $f(x)=ax+b$를 사용하였다. $f(e)-z<0$이면 무성음, 그렇지 않을 경우에는 유성음으로 판별했을 때의 결과를 표 1에서 확인할 수 있다. 이때 함수의 기울기 $a=2$는 가장 높은 판별율을 얻기 위해 실험적으로 구한 값이다.

피치 주기의 변화량 파라미터를 유·무성음의 판단에 이

표 1. 한 프레임의 길이가 20 ms인 프레임 단위의 유성음/무성음 구간 판별율 (TIMIT corpus에서 임의로 뽑은 400문장을 이용)
Table 1. Discrimination rate of voiced/unvoiced sounds according to frame of length 20 ms. (400 sentences randomly chosen from TIMIT corpus).

[illegible]	유성음 구간 판별율(%)	무성음 구간 판별율(%)
b=-70	99.51	85.13
b=-90	96.74	93.23
b=-110	82.77	98.67
b=-130	39.14	99.90

표 2. 영역 A에 대해서 ΔT와 $\Delta\Delta T$를 사용하여 판단하였을 때 전체 음성 구간에 대한 판별율
Table 2. Discrimination rate of voiced/unvoiced segments according to frame in the region A using the distributions of ΔT and $\Delta\Delta T$.

$f(x)=2x+b_1$ and $f(x)=2x+b_2$ (Region A)	유성음 구간 판별율(%)	무성음 구간 판별율(%)
b_1=-70, b_2=-90	98.89	89.98
b_1=-80, b_2=-100	97.79	93.44
b_1=-90, b_2=-110	94.39	96.16
b_1=-100, b_2=-120	88.14	98.18
b_1=-110, b_2=-130	77.54	99.30

용하면 유성음과 무성음이 섞여 있는 영역의 유·무성음 검출성능을 향상시킬 수 있다. 그림 2에서 $f(x)=2x+b_1$와 $f(x)=2x+b_2$의 사이에 있는 영역 A는 유성음과 무성음 구간이 섞여 있는 영역이다. 이 영역에 존재하는 유·무성음 구간에 대한 ΔT와 $\Delta\Delta T$의 분포는 그림 3과 같다.

그림 3에서 가로축은 ΔT, $\Delta\Delta T$의 크기를 나타내고, 세로축은 발생빈도를 나타낸다. 판별율을 향상시키기 위해서 유성음 또는 무성음의 판별율이 99% 이하인 영역에 대해서 ΔT와 $\Delta\Delta T$의 분포를 적용하였다. 이 영역은 그림 2에 나타난 영역 A이다. 유성음의 ΔT와 $\Delta\Delta T$의 분포의 분산이 무성음 구간에 대한 분포보다 각각 8배, 17배 작다. ΔT와 $\Delta\Delta T$의 분포를 바탕으로 영역 A내의 유·무성음을 구별하였을 때 전체 유성음의 판별율은 표 2와 같다.

이와 같이 에너지, 영교차율보다 안정적이고 밀집된 분포를 보여주는 ΔT와 $\Delta\Delta T$를 위와 같이 기존의 파라미터와 상호 보완적으로 사용한다면 유·무성음 추출에 있어서 검출효과를 증대시킬 수 있을 것이다. 따라서 유성음과 무성음이 섞여 있는 영역에서 신뢰성 있는 판단을 기대할 수 있다.

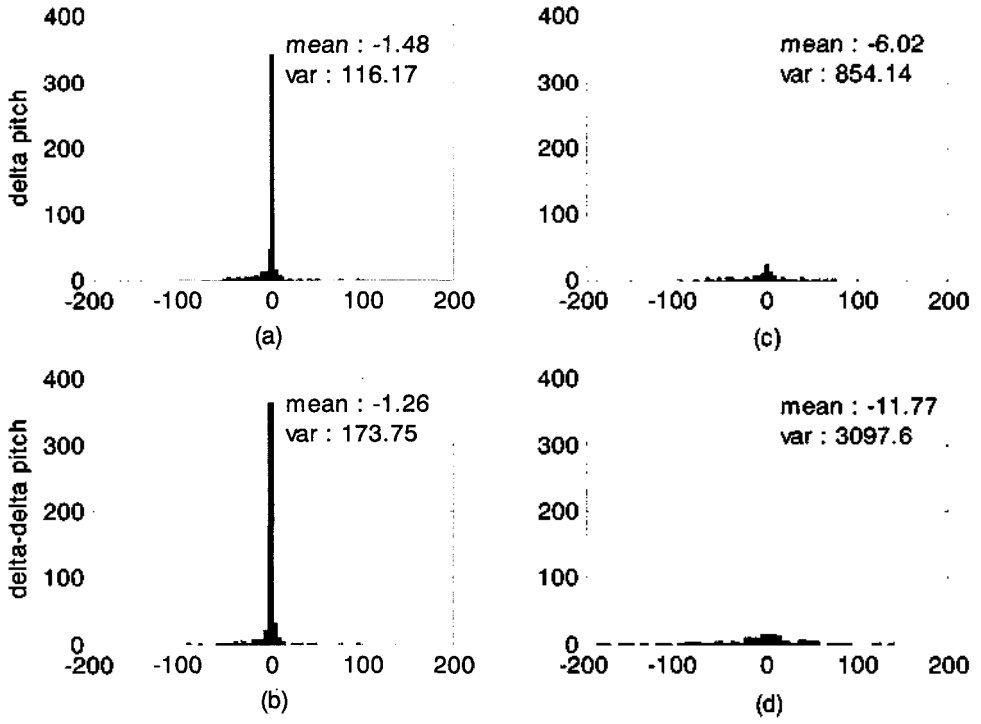
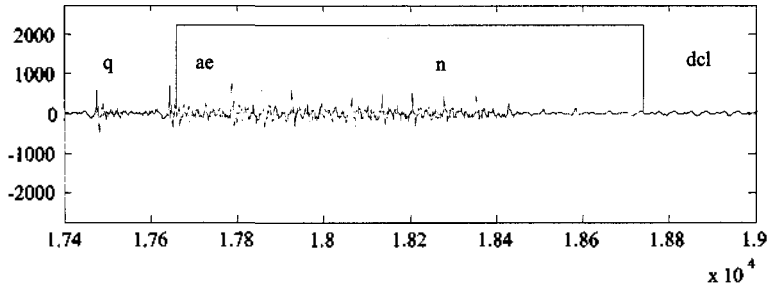


그림 3. 그림 2에서 2e[n]-70<z[n]<2e[n]-130인 영역 A에 존재하는 유성음인 음성 구간에 대한 (a) ΔT와 (b) $\Delta\Delta T$에 대한 분포 그래프, 무성음 구간에 대한 (c) ΔT와 (d) $\Delta\Delta T$에 대한 분포 그래프
Fig. 3. Distributions of ΔT and $\Delta\Delta T$ in the region A that was commented in Fig. 2 (a,b) for voiced sounds and (c,d) unvoiced sounds.

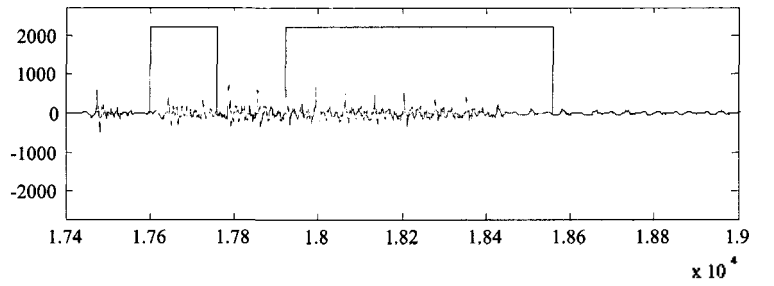
3.2. 유성음 구간을 잘못 검출했을 때의 분석

ΔT와 $\Delta\Delta T$를 사용한 유성음 구간의 검출을 위해 고려해야 할 문제는 유성음 음소가 여러 개 연결되어 있는 유성음 음소군을 판별해야 할 경우가 많다는 것이다. 이러한 경우 개인적인 발음 성향이나 서로 다른 유성음 음소간의 급격한 조음상의 변화가 피치주기 검출 성능을 떨어뜨릴 수 있다.

그림 5는 그림 2의 영역 A에 속해 있는 유성음구간 중에서 무성음으로 잘못 판단된 유성음 프레임의 유성음 음소 구간내의 분포그래프를 그린 것이다. 가로축 값은

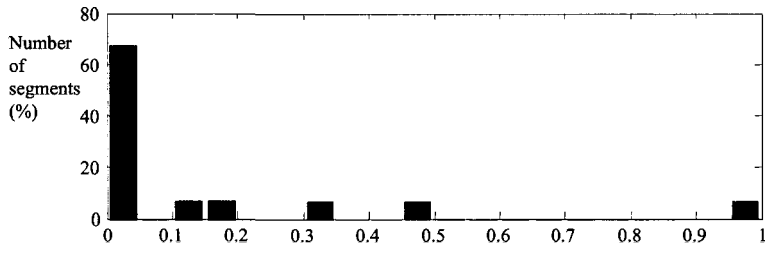


(a) 올바른 검출
(a) correct decision

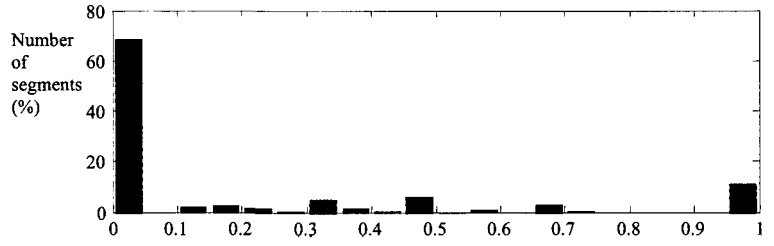


(b) 잘못된 검출
(b) wrong decision

그림 4. 유성음군 구간내의 잘못된 판단의 예 (점선은 유성음 구간)
Fig. 4. An example of the segmentation of voiced group (dotted line represents detected voiced region).



(a) 유성음 음소 사이에 위치하는 유성음 음소에 대한 분포
(a) between voiced phonemes



(b) 그 외의 유성음 음소에 대한 분포
(b) not between voiced phonemes

그림 5. 그림 2의 영역 A 내의 잘못 판단된 유성음 구간들의 해당 음소구간내의 분포
Fig. 5. Distribution of the relative position in the phoneme, when the voiced segments are not detected correctly in the region A shown in Fig. 2.

0에서 1 사이의 값을 가지며 0은 음소의 시작, 1은 음소의 끝을 나타낸다. 세로축은 해당 위치에 나타난 프레임의 개수를 나타낸다. 유성음 검출에 불리한 파라미터 값을 가지는 파라미터 값이 나오는 빈도는 음소구간은 시작 경계에서 가장 두드러진다는 것을 나타내고 있다.

그림 5의 결과에 따르면 음소구간을 검출한 후에 음소군의 검출을 위해서 다음과 같은 과정을 거칠 필요가 있다. 음소군 내의 음소간의 경계에서 유성음 구간임에도 불구하고 무성음 구간으로 잘못 판단된 짧은 구간에 대해서 유성음 구간인지 아닌지 다시 판단해야 한다. 이 때의 짧은 구간의 길이는 30 ms로서, 이는 유성음과 유성음 사이의 최소 길이를 기준으로 실험적으로 결정한 값이다.

3.3. 검출 알고리즘

제안된 유성음 구간 검출 알고리즘은 그림 6의 순서도에 소개가 되어 있다. n번째 프레임의 저대역 에너지 $e[n]$, 영교차율 $z[n]$, 피치주기의 변화량 $\Delta T[n]$, $\Delta\Delta T[n]$에 대해서 $d(\underline{x}_1, \underline{x}_2)$는 $\underline{x}_1$와 $\underline{x}_2$ 사이의 거리척도, $\underline{y}[n]=[2e[n]\ z[n]]^T$, $\underline{o}[n]=[\Delta T[n]\ \Delta\Delta T[n]]^T$라고 정의하고, λ_o를 임계값이라고 할 때 다음과 같이 알고리즘을 표현할 수 있다. 단, $\underline{y}_v$는 유성음 음성구간들에 대한 $\underline{y}[n]$값의 대표값, $\underline{y}_u$는 무성음 음성구간들에 대한 $\underline{y}[n]$값의 대표값, $\underline{y}_m$는 유성음인지 아닌지 구별하기 어려운 영역에 대한 $\underline{y}[n]$값의 대표값, $\underline{o}_v$는 유성음 음성구간들에 대한 $\underline{o}[n]$값의 대표값이다. 거리척도로는 $d(\underline{x}_1, \underline{x}_2)=|\underline{x}_1-\underline{x}_2|$이 사용되었다. (대표값은 임계구간에 따라서 결정된다.)

프레임 n_1과 n_2 사이를 무성음으로 판별된 구간이라고 할 때 $n_1-n_2<L$ 일 경우에, 모든 $n_1<k<n_2$에 대하여 다음이 성립하면 프레임 k를 유성음으로 판단한다. 이 때, 2차 에러가 아닌 1차 에러를 사용한 이유는 이웃한 프레임 간의 피치 변화를 실시간으로 적용하기 위해서이다.

$$d(\underline{y}[k], \underline{y}_v')<d(\underline{y}[k], \underline{y}_u),\ d(\underline{o}[n], \underline{o}_v)<\lambda_o$$

이 때, λ_o'은 λ_o보다 큰 값이며 $\underline{y}_v'$는 $d(\underline{y}_u, \underline{y}_v')<d(\underline{y}_u, \underline{y}_v)$를 만족하는 새로운 대표값으로서 짧은 무성음 구간이 유성음 구간으로 포함하지 않게 하고 피치 주기가 변하는 유성음 음소 경계면의 유성음을 검출해 내기 위한 설정이다.

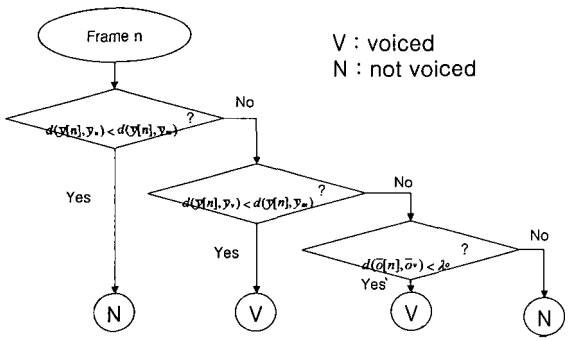


그림 6. 유성음 구간 검출을 위한 순서도
Fig. 6. The flowchart of the proposed algorithm for discrimination of voiced segments.

IV. 실험 및 결과고찰

유·무성음 음소의 구간을 검출해 내고자 할 때 다음과 같은 어려움이 있을 수 있다. 첫 번째로 유성음 음소가 연속적 나타날 때 유성음 세분류의 어려움이다. 유성음 음소 여러 개가 연속적으로 발음되었을 때 연속되는 유성음 음소를 모두 포함하는 구간으로 검출은 가능하나, 해당하는 구간 내에서의 각 유성음을 세분류하는데에는 어려움이 따른다. 두 번째로는 앞서 논의한 바가 있는 유성음군 사이의 짧은 무성음 음소가 존재하는 경우의 검출의 어려움이다. 이러한 경우에는 그림 7과 같이 짧은 무성음 음소의 구간이 유성음화 되는 경향을 보이게 되므로 같이 유성음 군으로 판별되는 경우가 많다.

유·무성음 음소의 구간 검출실험을 TIMIT 코퍼스 중에서 임의로 약 200개의 데이터를 추출하여 실험하였다. 20 ms의 오차를 가지고 TIMIT의 음소 구간정보를 검출해 내는 실험을 프레임 길이를 20 ms로 하여 실험하였다. 사용된 TIMIT 데이터들은 모두 8 kHz로 decimation해서 사용하였다.

$$\text{(유성음 검출율)} = \frac{20ms\text{의 오차로 검출된 유성음의 총 개수}}{\text{해당 유성음의 총 개수}} \quad (6)$$

표 3, 4는 제안된 알고리즘을 사용하여 유성음이 연속하는 경우 몇 개의 연속적인 유성음 군을 검출해내는 실험의 결과이다. 유성음 검출율을 계산하기 위해서 식 (6)을 사용하였다. 연속하는 음소의 개수란 검출된 음소군 내에 총 몇 개의 음소가 존재하는지를 의미한다. 본 실험 결과는 영어 문장 발음상에 있어서 연속하는 유성음 음소의 발음이 얼마나 많이 존재하는가를 보여주는 동시에 음소군 내의 음소 검출의 필요성 또한 제시하고 있다. 연속하

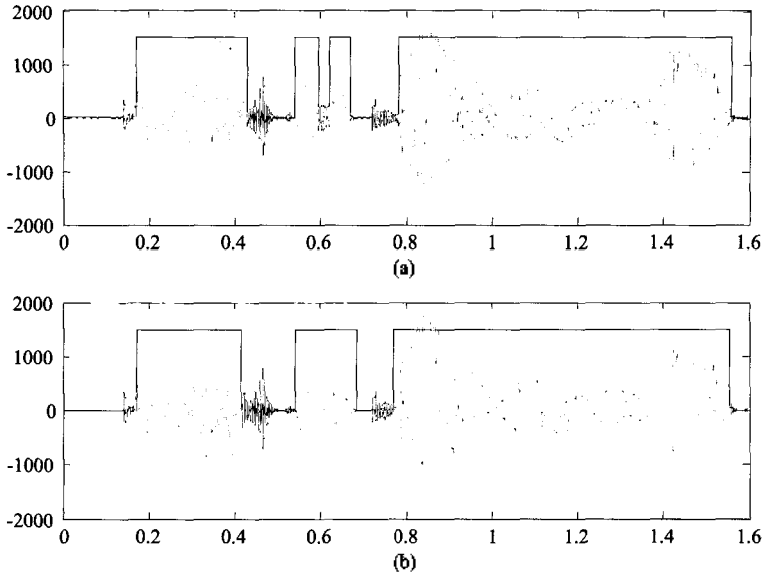


그림 7. 발음 'Don't ask me to carry an oily rag'에 대한 (a) TIMIT 레이블 데이터를 이용한 음소 구간과 (b) 제시된 알고리즘을 통한 유성음 음소구간 검출 결과 (점선으로 표시된 구간이 유성음 구간)

Fig. 7. For a speech, "Don't ask me to carry an oily rag", (a) segmentation with labeling data from TIMIT corpus, and (b) segmentation with proposed algorithm (dotted lines for voiced segments).

표 3. 에너지와 영교차율을 이용한 음소군에 대한 판별율 (6개 연속인 것 까지만 표시)

Table 3. Results of discriminating a group of successive voiced phonemes using only energy and ZCR.

연속하는 음소의 개수	검출된 유성음 음소의 개수	유성음 음소의 개수	%
1	177	309	57.28
2	554	794	69.77
3	425	525	80.95
4	15	24	62.50
5	150	595	25.21
6	38	108	35.19

표 4. 제안된 파라미터와 알고리즘을 이용한 음소군에 대한 판별율 (6개 연속인 것 까지만 표시)

Table 4. Results of discriminating a group of successive voiced phonemes using the proposed algorithm.

연속하는 음소의 개수	검출된 유성음 음소의 개수	유성음 음소의 개수	%
1	158	309	51.13
2	571	794	71.91
3	430	525	81.91
4	18	24	75.00
5	419	595	70.42
6	49	108	45.37

표 5. 저대역 에너지와 영교차율을 이용한 유성음 음소의 종류에 따른 검출 결과

Table 5. Results of discrimination various voiced phonemes using only energy and ZCR.

유성음의 종류	검출된 유성음 음소의 개수	유성음 음소의 개수	%
Front Vowels	579	963	60.12
Middle Vowels	237	450	52.67
Back Vowels	70	72	97.22
Diphthongs	64	123	52.03
Liquids	166	310	53.55
Glides	244	404	60.40
Nasals	150	250	60.00
total	1510	2572	58.71

표 6. 제안된 파라미터와 알고리즘을 이용한 유성음 음소의 종류에 따른 검출 결과

Table 6. Results of discrimination various voiced phonemes using the proposed algorithm.

유성음의 종류	검출된 유성음 음소의 개수	유성음 음소의 개수	%
Front Vowels	694	963	72.07
Middle Vowels	263	450	58.44
Back Vowels	71	72	98.61
Diphthongs	101	123	82.11
Liquids	224	310	72.26
Glides	299	404	74.01
Nasals	180	250	72.00
total	1832	2572	71.23

는 음소의 개수가 1개일 때, 제안된 알고리즘의 성능이 기존의 알고리즘보다 다소 떨어지는 것을 보이는데 이는 제안된 알고리즘에서 사용하는 피치의 변화량은 적어도 2개 이상의 프레임 (40 ms)에 대해서 적용이 가능한데, 실제 유성음 음소 하나의 길이는 평균적으로 40 ms보다 짧기 때문에 이러한 현상이 발생하는 것이라 판단된다.

표 5, 6은 유성음 음소의 종류에 따른 검출결과를 나타낸 것이다. 이중모음과 후설모음에 대해서 높은 검출율을 나타내었으며 에너지와 영교차율을 이용한 것보다 뛰어난 유성음군의 검출결과를 나타내었다. 본 실험에서 사용된 파라미터 값들은 다음과 같다.

$$\underline{\nu}_m = [2\times 96 \quad 102]^T,\ \underline{\nu}_v = [2\times 104 \quad 98]^T,$$
$$\underline{\nu}_u = [2\times 88 \quad 106]^T,\ \underline{\alpha}_v = [0 \quad 0]^T,\ \lambda_v = 10$$

V. 결 론

본 논문에서는 유·무성음 구간을 검출하기 위한 간단한 알고리즘을 제안하였다. 제안된 알고리즘은 음성의 유·무성음의 주기성에 대한 특성을 보완할 수 있는 저대역 에너지와 영교차율, 그리고 주기성의 안정성을 판단하기 위한 피치 변화량을 파라미터로 사용하였다. 유·무성음의 구간검출을 음소단위의 검출이라는 측면에서 접근하여 음소군의 검출율과 음소군 내의 음소의 검출율을 얻었다. 에너지와 영교차율을 이용한 판별알고리즘과 비교하였을 때 프레임별 판별율은 보다 안정적이고 향상된 결과를 나타내었으며, TIMIT 코퍼스를 사용하여 실험했을 때 유성음 음소 검출율이 약 13% 향상되었다. 좀더 정밀한 수준의 음소군 판별과 여러 음소로 이루어진 음소군 내의 음소 판별에 대한 연구가 수행되어야 할 것이다.

감사의 글

이 논문은 한국과학재단이 지원한 목적기초연구로 (과제번호 R01-2000-00259) 얻은 연구 결과의 하나이며, 이에 고마움을 나타냅니다.

참 고 문 헌

1. L. Rabiner, and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, 1978.
2. TIMIT, *Acustic-Phonetic Continuous Speech Corpus CD*, American Helix, 1990.
3. B. S. Atal and L. R. Rabiner, "A pattern recognition approach to voiced-unvoiced-silence classification with applications to speech recognition," *IEEE Trans. Acoust. Speech, Signal Processing*, **ASSP-24**, 201-212, June 1976.
4. C. K. Un and S. C. Yang, "A pitch extraction algorithm based on LPC inverse filtering and AMDF," *IEEE Trans. Acoust., Speech Signal Processing*, **ASSP-25**, 565-572, Dec. 1977.
5. A. Bendiksen and K. Steiglitz, "Neural Networks for Voiced/Unvoiced Speech Classification," *ICASSP-90*, 521-524, 1990.
6. R. P. Cohn, "Robust Voiced/Unvoiced Speech Classification using a Neural Net," *ICASSP-91*, 437-440, 1991.
7. T. G. Crippa and A. E. Jaroudi, "Fast Neural Net Training Algorithm and Its Application to Voiced-Unvoiced-Silence Classification of Speech," *ICASSP-91*, 441-444, 1991.
8. R. Ahn and W. H. Holmes, "Voiced/Unvoiced/Silence Classification of Speech Using 2-Stage Neural Networks With Delayed Decision Input," *ISSPA-96*, 1, 389-390, 1996.
9. TIA/EIA/IS-127, "Enhanced variable rate codec, speech service option 3 for wideband spread spectrum digital systems," 1997.
10. L. J Siegel, "A procedure for using pattern classification techniques to obtain a voiced/unvoiced classifier," *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-28, 398-407, Aug. 1980.
11. D. A. Krubsack, and R. J. Niederjohn, "An Autocorrelation Pitch Detector and Voicing Decision with Confidence Measures Developed for Noise-Corrupted Speech," *IEEE Trans. Acoust., Signal Processing*, 39, 319-329, Feb. 1991.

저자 약력

● 장 규 철 (Gyu-Cheol Jang)

2001년 2월: 한국과학기술원 전자전산학과 졸업 (공학사)
2001년 3월~현재: 한국과학기술원 전자전산학과 석사과정
※ 주관심분야: 신호처리, 음성인식

● 우 수 영 (Soo-Young Woo)

2001년 2월: 한국과학기술원 전자전산학과 졸업 (공학사)
2001년 3월~현재: 한국과학기술원 전자전산학과 석사과정
※ 주관심분야: 음성인식, 음성신호처리

● 박 용 규 (Yong-Kyu Park)

1985년 2월: 한양대학교 전기과 졸업 (공학사)
1987년 8월: 한국과학기술원 전자전산학과 졸업 (공학석사)
1996년 8월: 한국과학기술원 전자전산학과 졸업 (공학박사)
1987년~2000년: 한국전기통신공사 선임연구원
2000년~2002년: 맨투머신(주) 대표이사
2002년 1월~현재: 한국과학기술원 연구교수

● 유 창 동 (Chang D. Yoo)

한국음향학회지 제20권 제3호 참조