제15회 음성통신 및 신호처리 워크샵(KSCSP '98 15권1호)

웨이브렛 변환을 이용한 음성신호의 유성음/무성음/묵음 분류

손 영호, 배 건성
경북대학교 전자·전기 공학부

Voiced/Unvoiced/Silence Classification of Speech Signal Using Wavelet Transform

Young Ho Son, Keun Sung Bae
School of Electronic and Electrical Engineering, Kyungpook National University

요 약

일반적으로 음성신호는 파형의 특성에 따라 파형이 준주기적인 유성음(voiced sound)과 주기성 없이 잡음과 유사한 무성음(unvoiced sound) 그리고 배경잡음에 해당하는 묵음(silence)의 세 종류로 분류된다. 기존의 유성음/무성음/묵음 분류 방법에서는 피치정보, 에너지 및 영교차율 등이 분류를 위한 파라미터로 널리 사용되었다. 본 논문에서는 음성신호를 웨이브렛 변환한 신호에서 스펙트럼상에서의 변화를 파라미터로 하는 유성음/무성음/묵음 분류 알고리즘을 제안하고 제안된 알고리즘으로 검출한 결과와 이에 따른 문제점을 검토하였다.

Ⅰ.서 론

음성신호는 음성이 존재하지 않는 배경잡음에 해당하는 묵음구간과 음성이 존재하는 실제의 음성신호 구간으로 분류된다. 음성신호는 성대(vocal folds)의 자발적인 운동에 의해 발생되는 공기의 흐름이 성도(vocal tract)를 지나면서 변조되어 공기압의 파동형태로 나타나는 것이다. 이러한 음성신호는 성대의 진동 상태에 따라 유성음과 무성음으로 구분할 수 있다. 일반적으로 유성음의 경우 성대의 규칙적인 진동으로 인하여 파형이 준주기적이면서 저주파 대역에 에너지가 집중되어 있는 반면, 무성음의 경우 성대의 진동 없이 생성됨으로써 주기성을 띄지 않게 되고 성도에서의 다소간의 마찰 등으로 인하여 고주파 대역에 에너지가 집중되어 있다.[1] 음성신호처리에서 이와 같은 유성음/무성음/묵음의 정확한 분류는 피치검출, 음성검출, 음성합성, 음성 세그먼테이션 등의 음성분석에서 중요하게 다루어지는 문제이다. 이러한 음성분류 알고리즘에서는 음성신호에서의 피치정보, 에너지, 영교차율 등이 파라미터로 널리 사용되고 있다. 그러나 피치정보의 경우는 짧은 음성구간의 분류시에 어려움이 있으며 에너지나 영교차율의 경우는 다른 파라미터들과 상호 보완적으로만 사용되는 단점이 있다.[2,3] 본 논문에서는 시간 및 주파수 영역에서 동시에 음성신호의 국부적인 특성을 잘 반영하는 웨이브렛 변환을 이용하여 음성신호의 웨이브렛 변환된 신호에서 스펙트럼상의 변화를 검출하기에 적합한 MLR(maximum likelihood ratio)[4,5] 값을 이용하여 음성신호에서 유성음/무성음/묵음을 분류하기 위한 알고리즘을 연구하였다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 웨이브렛 변환에 대하여 소개하며 3장에서는 MLR을 이용한 유성음/무성음/묵음 분류 파라미터와 알고리즘에 대하여 설명한다. 그리고 4장에서는 실험한 결과를 manual하게 검출한 각 구간의 경계와 비교 분석하여 5장에서 결론을 맺는다.

Ⅱ.웨이브렛 변환

웨이브렛 이론은 응용수학에서 처음 소개된 후 최근 컴퓨터비전 분야 등에서 연구되어 온 다중 해상도 표현과 연관성이 있음이 밝혀진 신호해석 방법이다. 일반적으로 웨이브렛 변환은 모 웨이브렛(mother wavelet)의 확장(dilation) 및 축소(contraction) 등의 과정을 통하여 특정 주파수 대역의 신호 성분을 분리하여 해석하는 것을 가능하게 한다. 특히 고주파 대역에서는 축

소된 웨이브렛을 이용하여 시간영역에서의 분해능을 개선시키고, 저주파 대역에서는 확장된 웨이브렛을 이용하여 주파수 영역에서의 분해능을 향상시킬 수 있다. 실제 응용에 이용되는 이산 웨이브렛 변환식은 식 (1), (2)로 주어진다. 한편 식 (2)에서 계수 구현을 용이하게 하기 위하여 a_0를 2로 한 dyadic 웨이브렛 변환(DyWT: Dyadic Wavelet Transform)은 식 (3)과 같이 정의된다.

$$d_{j,k} = \int f(t)\,\varphi^{*}_{j,k}(t)\,dt \qquad (1)$$

$$\varphi_{j,k}(t) = a_0^{\frac{-j}{2}}\,\varphi(a_0^{-j}t - kT) \qquad (2)$$

$$d_{j,k} = \frac{1}{\sqrt{2^j}}\int f(t)\,\varphi^{*}\left(\frac{t}{2^j} - kT\right)dt \qquad (3)$$

DyWT을 이용하여 신호를 분석하는 과정은 그림 1과 같이 트리(tree) 형태의 필터뱅크와 같다. 여기서 H_0는 저역통과 필터이고, H_1은 고역통과 필터이다. 신호처리 관점에서 DyWT는 constant-Q, octave band 특성을 갖는 밴드패스 필터들의 뱅크 출력으로 볼 수 있다.[6] 본 연구에서는 식 (2)에서의 웨이브렛 함수로 quadratic spline[7] 웨이브렛을 사용하여 음성신호를 웨이브렛 변환하였다. 그림 2는 10kHz 샘플링 주파수에서 본 연구에서 사용한 웨이브렛 함수의 스케일 2~5에 해당하는 주파수 응답을 나타낸 것으로 스케일이 증가할수록 주파수 영역에서 중심주파수가 저주파 대역으로 접근함을 확인할 수 있다. 이러한 스케일에 따른 웨이브렛 함수의 주파수 응답 특성은 음성신호 분류시 유성음의 경우는 높은 스케일의 신호에서, 무성음의 경우는 낮은 스케일의 신호에서 잘 검출될 수 있다는 것을 의미한다. 한편 본 연구에서는 웨이브렛 변환시 소요되는 시간을 줄이기 위하여 그림 1과 같이 웨이브렛 변환된 신호를 스케일에 따라 순차적으로 구하는 방법 대신에 각 스케일의 웨이브렛 함수를 직접 구현하여 음성신호를 웨이브렛 변환함으로써 필요한 스케일의 신호만을 직접 얻는 방법을 채택하였다.

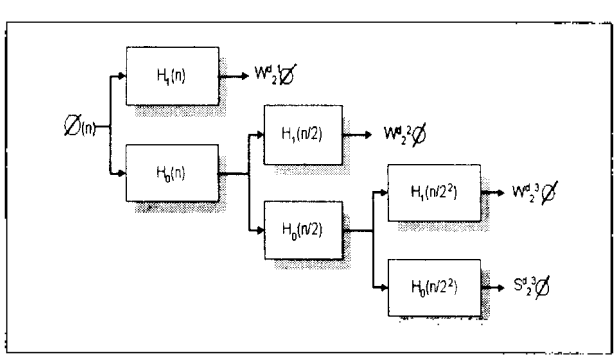


그림 1. 웨이브렛 분해 필터뱅크

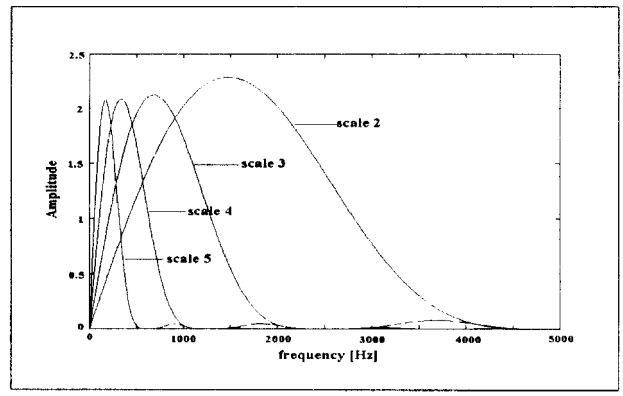


그림 2. 웨이브렛 함수의 주파수 응답

Ⅲ.웨이브렛 변환을 이용한 음성신호의 분류

3.1. 음성분류 파라미터

음성신호를 유성음/무성음/묵음 등으로 분류할 때 문제점은 무성음을 유성음이나 묵음구간과 구분하는 것이다. 특히 진폭이 작은 마찰음(fricative)이나 파열음(plosive)의 경우는 단순히 시간영역의 에너지나 영교차율만으로 정확히 구분하기 어렵다. 본 연구에서는 음성신호의 분류를 위하여 주파수 영역에서 변화를 잘 모델링하는 것으로 알려진 MLR을 웨이브렛 변환한 신호에 적용하였다. 각 스케일에 해당하는 프레임별 MLR에 대한 정의는 식 (4)와 같다. 식 (4)에서 σ^2_k와 σ^2_{noise}는 각각 해당 프레임에 대하여 스케일 k에서 웨이브렛 변환한 신호의 분산과 배경잡음에 해당하는 묵음구간의 분산을 가리키며, N은 MLR을 구할 때 동시에 분석되는 프레임 수로서 1로 설정하였다.

$$MLR_k = \frac{N}{2}\left|\ln\left(\frac{\sigma_k^2}{\sigma^2_{noise}}\right) - \left(\frac{\sigma_k^2}{\sigma^2_{noise}}\right)\right| \qquad (4)$$

그림 3은 음성신호 /fish/에 대한 음성파형과 스케일 1과 4에서의 MLR 값을 프레임 단위로 나타낸 것으로 웨이브렛 함수의 주파수 응답 특성에서 예상했던 것처럼 스케일 1에서는 상대적으로 무성음 /f/, /ʃ/가 스케일 4에서는 유성음 /i/가 잘 나타나고 있음을 볼 수 있다. 본 연구에서는 MLR의 이러한 특성을 이용하여 음성신호를 음성구간과 묵음구간으로 분류한 뒤 검출된 음성구간을 유성음과 무성음으로 분류하였다. 음성구간을 검출하기 위한 파라미터로 식 (5)에 주어진 PS를 정의하였으며, 유성음과 무성음을 구분하기 위한 파라미터로는 식 (6), (7)에 주어진 PU_1, PU_2를 정의하였다.

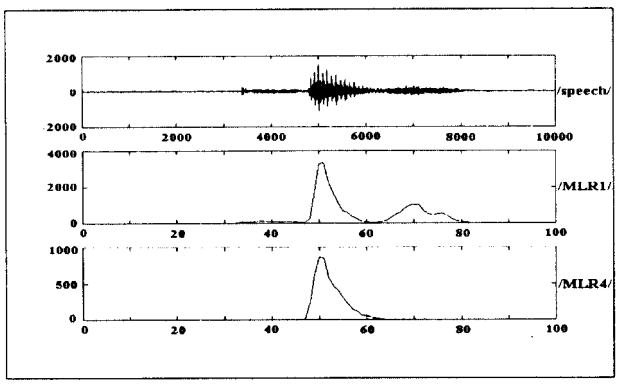


그림 3. 음성신호 /fish/에 대한 MLR test

PS의 경우는 유성음과 무성음을 각각 잘 반영하는 스케일에서의 MLR 값들을 더함으로써 유성음과 무성음의 조합으로 이루어진 음성구간을 검출하는 파라미터이며, PU_1과 PU_2의 경우는 무성음이 유성음에 비해 상대적으로 고주파 대역에 에너지가 집중되어 있는 특성을 이용하여 무성음을 검출할 수 있게 한 파라미터이다. 식 (5), (6), (7)에서 MLR_k는 k번째 스케일에서의 MLR 값을 나타낸다.

$$PS = MLR_1 + MLR_4 \tag{5}$$

$$PU_1 = \frac{MLR_1}{MLR_4} \tag{6}$$

$$PU_2 = 100 \times \frac{MLR_1}{MLR_1 + MLR_2 + MLR_3} \tag{7}$$

그림 4는 단어 /fish/를 대상으로 식 (5), (6), (7)에서 정의한 파라미터 값을 프레임 단위로 추출하여 나타낸 것이다. 이 결과에서 무성음에 해당하는 /f/와 /ʃ/부분에서 PU_1과 PU_2 값이 유성음이나 묵음구간에 비해 상당히 큼을 확인할 수 있다.

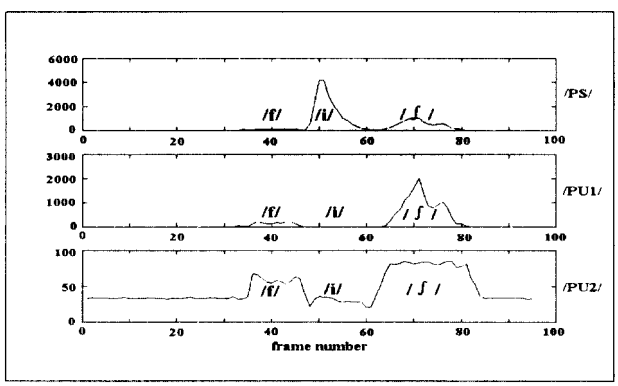


그림 4. 음성신호 /fish/에 대한 파라미터 추출

3.2. 웨이브렛 변환을 이용한 음성 분류 알고리즘

본 논문에서 제안하는 웨이브렛 변환을 이용한 음성신호의 유성음/무성음/묵음 분류 알고리즘은 다음과 같다.

STEP 1: 음성을 프레임 단위로 받아들여 스케일 1~4까지 웨이브렛 변환한다. 이때 각 스케일에서의 필터링으로 인한 delay(각 스케일에서의 웨이브렛 길이, W_L의 1/2)를 고려해주었다. 본 연구에서는 분석 프레임의 길이를 20ms로 하였다.

STEP 2: Step 1의 출력신호를 이용하여 스케일별로 MLR 값을 계산하여 제안한 파라미터들을 계산한다. 음성신호의 앞부분 150ms를 대상으로 해당 음성신호에 대한 문턱값을 설정하는 단계를 거친다.

STEP 3: 음성구간과 묵음구간을 구분하는 단계로 PS 값이 설정된 문턱값 보다 클 경우 음성구간으로 간주하고 그렇치 않을 경우는 묵음구간으로 간주한다.

STEP 4: 음성구간에서 유성음과 무성음을 구분하는 단계로 Step 3에서 음성구간으로 결정된 경우 PU_1, PU_2가 문턱값보다 클 경우 무성음으로 간주하고 그렇치 않을 경우 유성음으로 간주한다.

STEP 5: 한 프레임의 분석이 끝나면 다음 프레임으로 이동하여 Step 1~4를 반복한다. 이때 프레임의 이동 간격은 10ms로 하였다.

각 파라미터에 대한 문턱값의 설정을 위하여 주어진 음성신호의 처음 150ms 정도를 묵음구간이라 간주하고 이 구간을 대상으로 파라미터 PS, PU_1, PU_2에 대한 평균값을 구하여 이 값을 정수배하여 최종적인 문턱값으로 설정하였다. 실험적으로 PS의 경우는 묵음구간 평균값의 5배를, PU_1과 PU_2의 경우는 각각 묵음구간 평균값의 10배와 1.5배를 문턱값으로 설정하였다. 한편 Step 3에서 검출한 음성구간이 100ms 이내인 경우는 nonstationary한 잡음 등으로 인하여 묵음구간이 음성구간으로 잘못 분류된 것으로 간주하여 해당 구간을 묵음구간으로 간주하였다.

Ⅳ. 실험 및 검토

본 연구에서 실험에 사용한 음성 데이터는 연구실 환경에서 여러 화자들로부터 수집되어진 10kHz로 샘플링되고 16비트로 양자화된 음성들을 대상으로 하였다. 그림 5, 6은 단어 /fish/에 대한 유성음/무성음/묵음 분류 결과를 제시한 것이다. 특히 그림 5의 경우는 검출

된 음성구간에서 PU_1과 PU_2를 이용하여 분류한 유성음과 무성음의 경계를 무성음을 중심으로 manual하게 분류한 실제 경계와 비교하여 나타낸 것이다. 그림에서 위쪽 방향의 점선은 manual하게 분류한 각 구간의 실제 경계를, 아래쪽 방향의 실선은 제안한 알고리즘으로 분류한 경계를 나타낸 것이다. 그림 6은 전체 음성에 대한 분류 결과를 나타낸 것으로 위쪽의 실선은 음성구간의 경계를, 아래쪽의 실선은 음성구간에서 유성음과 무성음의 경계를 나타내고 있다. 음소에 따라 다소의 차이는 있으나 분류된 경계의 대부분은 실제의 경계와 비교하여 50 샘플내의 차이를 보였으며, 이 결과는 음성 분류시 20ms의 프레임 단위로 분석한 것을 고려할 때 제안한 알고리즘이 유성음/무성음/묵음 구간을 신뢰성 있게 검출한 것으로 볼 수 있다.

그림 7, 8은 제안한 알고리즘이 실제 문장 내에서도 유성음/무성음/묵음 구간을 잘 분류하는지를 확인하기 위하여 "We saw the ten pink fish."라는 문장을 대상으로 분류 실험한 결과를 나타낸 것으로 문장을 대상으로 했을 경우에도 단어를 대상으로 한 경우와 비슷한 결과를 얻을 수 있었다. 그림 7에서는 문장 내에서 제안한 알고리즘으로 분류한 각 구간의 경계를 그림 5에서와 같이 무성음 /s/, /t/, /f/, /p/, /ʃ/ 등의 경계를 중심으로 manual하게 분류한 각 구간의 실제 경계와 비교하여 나타낸 것이다. 그림 8에서는 전체 음성신호에 대하여 분류한 경계를 나타내고 있다. 위쪽의 실선은 묵음구간과 음성구간을 분류한 경계이며, 아랫쪽의 실선은 음성구간에서 유성음과 무성음 구간을 분류한 경계이다.

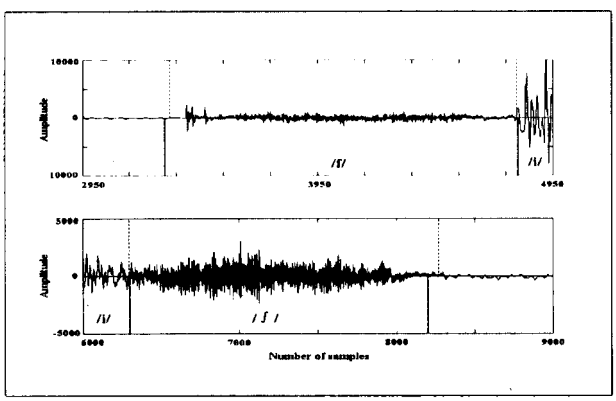


그림 5. 음소의 실제 경계와 검출된 경계와의 비교 /"fish"/

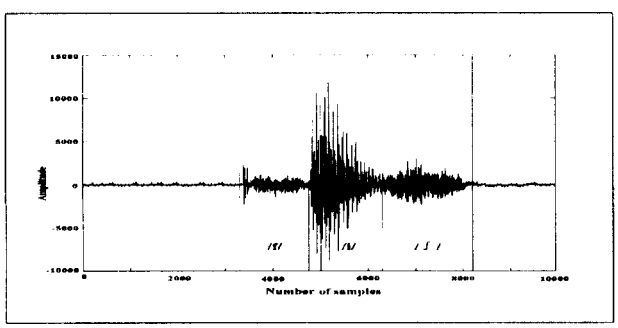


그림 6. 단어에 대한 유성음/무성음/묵음 분류 결과 /"fish"/

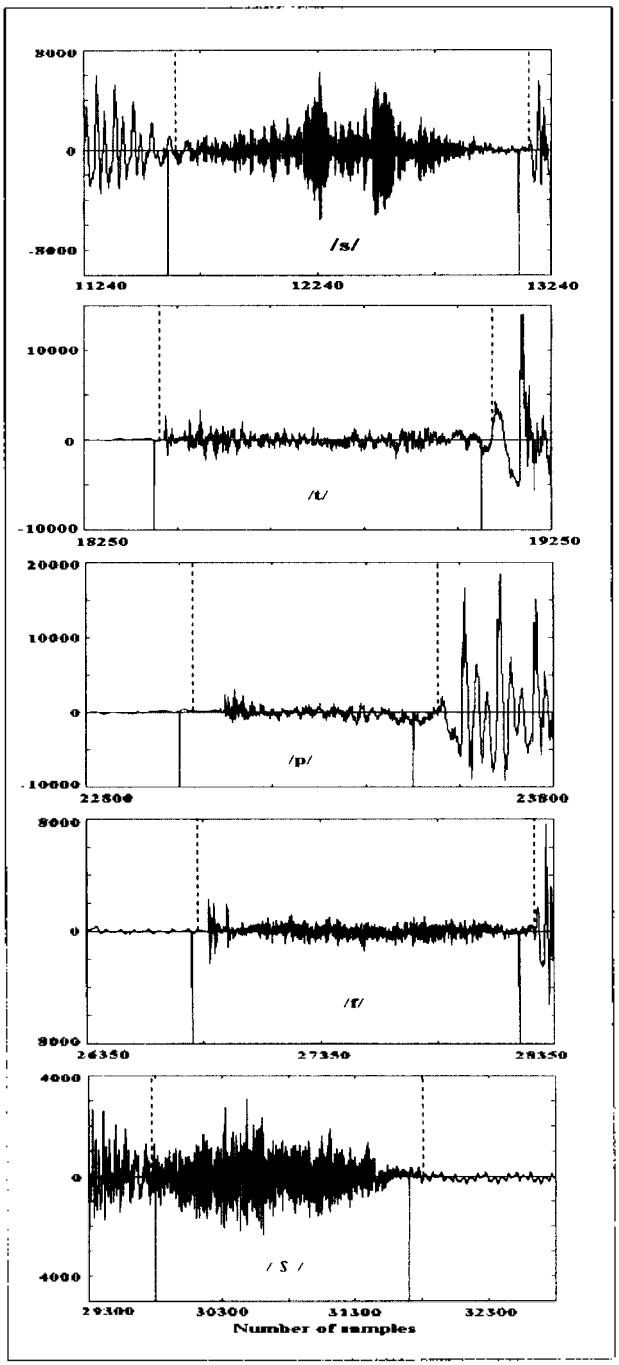


그림 7. 음소의 실제 경계와 검출된 경계와의 비교 /"We saw the ten pink fish."/

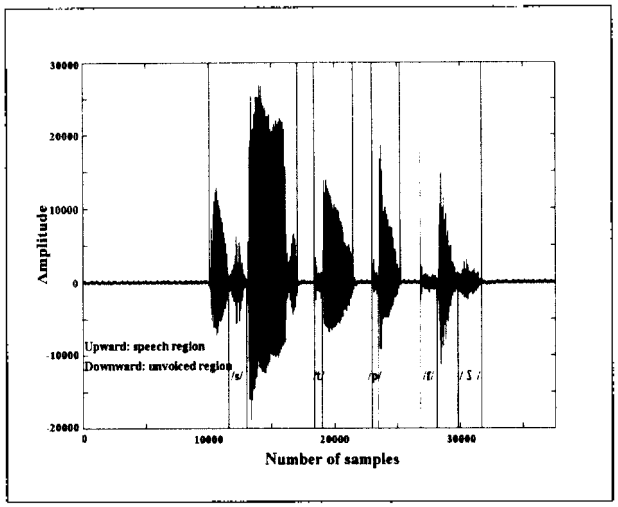
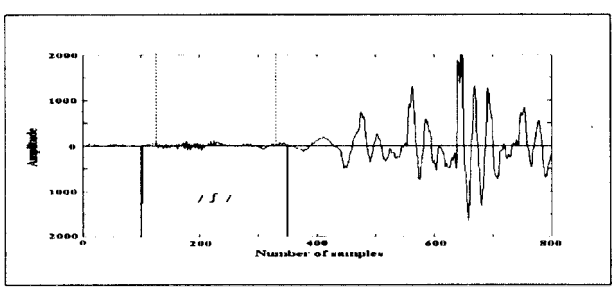
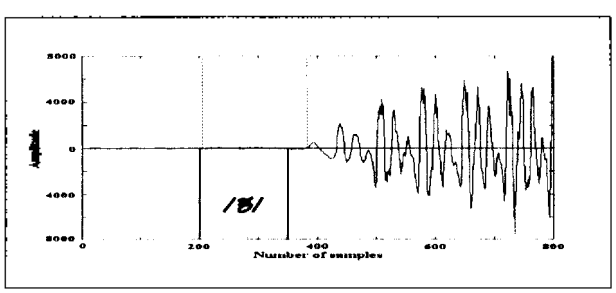


그림 8. 문장에 대한 유성음/무성음/묵음 분류 결과 /"We saw the ten pink fish."/

그림 9는 "Should we chase those cowboys?"라는 문장에서 발음상의 특성으로 묵음구간과 비슷한 진폭을 가지는 마찰음 /ʃ/와 /ð/를 대상으로 제안한 알고리즘이 해당구간을 잘 분류하는지를 조사한 결과로서 그림에서 실선은 검출한 무성음구간의 경계이며 점선은 실제의 경계에 해당한다. 이 결과로 볼 때 제안한 알고리즘이 배경잡음과 비슷한 진폭을 가지는 무성음의 경우도 잘 분류함을 확인할 수 있다.



(a). should의 /ʃ/



(b). those의 /ð/

그림 9. 배경잡음과 진폭이 비슷한 무성음 분류 예

V. 결론

본 논문에서는 음성신호의 웨이브렛 변환된 신호에서 각 스케일별 MLR 값을 이용하여 음성신호의 주파수 영역에서의 특성을 반영하는 파라미터를 정의하고, 이를 이용하여 음성신호를 유성음/무성음/묵음으로 분류하는 알고리즘을 제안하였다. 제안된 알고리즘으로 분류한 음소들의 경계들이 manual하게 분류한 음소들의 실제 경계를 기준으로 대부분 50 샘플내의 차이를 보임을 확인할 수 있었다. 특히 진폭이 작은 마찰음이나 파열음의 경우에 있어서도 주변잡음으로부터 정확하게 분류될 수 있음을 확인할 수 있었다.

본 연구는 한국과학재단의 핵심전문연구비 (과제번호: 971-0917-103-2) 지원으로 수행되었으며, 지원에 감사 드립니다.

참 고 문 헌

1.Douglas O'Shaughnessy "Speech Communication" *Addison Wesley*

2.Bishnu S.ATAL, Lawrence R. Rabiner "A Pattern Recognition Approach to Voiced-Unvoiced-Silence Classification with Applications to Speech Recognition" *IEEE Trans ASSP*, vol. 24, No. 3, Jun. 1976

3.Leah J. Siegel, "A Procedure for Using Pattern Classification Techniques to Obtain a Voiced/Unvoiced Classifier", *IEEE Trans ASSP*, vol. 27, No.1 Feb. 1979

4.B. T Tan, R.Lang, Heiko Schroder, Andrew Spray "Applying wavelet analysis to speech segmentation and classification" *SPIE Wavelet Applications* vol. 2242, 1994

5.V.Ralph Algazi, Kathy L. Brown, Michael J. Ready, David H. Irvine "Transform Representation of the Spectra of Acoustic Speech Segments with Applications-I", *IEEE Trans Speech and Audio Processing*, vol. 1, No.2 , Apr. 1993.

6.Olibier Rioul, Martin Vetterli, "Wavelets and Signal Processing" *IEEE SP Magazine* Oct. 1991.

7.S.Mallat, W. L. Hwang, "Singularity Detection and Processing with Wavelets," *IEEE Trans Information Theory*, vol.38, no.2, Mar. 1992.