제15회 음성통신 및 신호처리 워크샵(KSCSP '98 15권1호)

고립단어 인식시스템에서 음성/비음성 식별에 관한 연구

김 치수, 배 건성
경북대학교 전자·전기 공학부

A Study on The Speech/Nonspeech Identification for Isolated Word Speech Recognition System

Chi Su Kim , Keun Sung Bae
School of Electronic and Electrical Engineering, Kyungpook National University

요 약

음성인식 시스템의 입력인 음성은 실제의 음성부분 외에도 주변잡음을 포함한 기침 소리, 문닫는 소리, 책장 넘기는 소리등과 같은 사용자에 의해서 발생될수 있는 다양한 종류의 비음성을 포함할수 있다. 특히 에너지가 큰 비음성을 포함하는 경우 기존의 끝점검출 알고리듬만으로는 음성부분만의 정확한 검출이 어렵게 되고 이는 음성인식 시스템의 성능을 저하시키는 주요 원인이 된다. 본 논문에서는 음성 발생시 일어날 수 있는 비음성들에 대해서 조사하고 이러한 비음성이 포함될때 음성부분만의 정확한 검출을 가능하게 하는 알고리듬을 제시하였다. 사용된 파라미터로는 자기상관법에 의해 얻어지는 피치정보와 웨이브렛 영역에서의 에너지로써 비교적 낮은 신호대 잡음비(SNR)에서도 음성부 검출을 가능하게 하였다.

1. 서 론

음성인식, 합성 및 분석 등 음성공학의 거의 모든 분야에서 음성신호의 시작점 및 끝점을 주변잡음과 정확하게 분리하여 찾아내는 일은 매우 중요하다. 고립단어 인식시스템에서의 정확한 음성부 검출은 인식률을 향상시킬 뿐만아니라 비음성을 포함한 불필요한 묵음을 사전에 제거시킴으로써 단어 인식에 소요되는 시간을 줄일 수 있다. 음성인식 시스템에 입력되는 음성은 실제 음성부분외에도 주변잡음을 포함한 기침 소리, 문닫는 소리, 책장 넘기는 소리, 전화 끊는 소리와 같은 화자나 녹음환경, 전송매체에 의해서 발생될 수 있는 다양한 종류의 비음성을 포함할 수 있다. 마이크로 음성을 입력받을 경우 문닫는 소리나 의자를 움직일 때 나는 소리와 같이 일시적인 배경잡음들은 잡음제거 마이크를 사용함으로서 어느정도 그 영향을 줄일수 있다[1]. 하지만 전화망을 통하게 될 때 이 방법은 유효하지 않게 된다. 따라서 비음성을 식별할수 있는 알고리듬은 전화망과 연결되어 동작하는 고립단어 인식시스템에서 더욱더 필요하다.

본 연구에서는 비음성의 제거를 위하여 자기상관함수를 구했을 때 나타나는 신호의 주기성 변화를 조사하였다. 또한 잡음 환경하에서의 음성 검출을 위해서는 웨이브렛을 이용한 끝점검출 알고리듬[2]을 수정하여 사용하였다.

2. 음성 검출 방법

2.1 비음성 식별 파라미터

자기상관함수를 이용함으로써 음성신호의 피치를 검출할수 있다는 것은 널리 알려져 있는 사실이다. 그림 1에서와 같이 유성음의 경우 프레임 단위로 자기상관을 구해보면 시작점부터 첫번째 피크 사이의 간격 T_1이

주기가 되는데 첫번째 피크에서 다음 피크 사이의 간격 T_2가 T_1과 거의 동일하게 된다. 하지만 비음성의 경우는 식 (1)에서 정의되어진 피크 사이의 간격 변화가 유성음에 비해서 크게 나타나게 된다. 또한 식 (2)에서 정의된 피치를 구했을 때 비음성의 경우는 음성이 가질 수 있는 피치의 범위를 벗어나는 경우가 많이 생기게 된다. 일반적으로 유성음이 가질수 있는 피치의 범위는 3~15ms이다[3].

표 1은 실제로 발생할 수 있는 비음성들에 대한 자기상관함수 피크 간격의 변화량에 대한 평균값을 측정한 것이다. 수집된 비음성 데이터의 샘플링 주파수는 8kHz이다. 주기성이 뚜렷한 음성의 경우는 0~3의 값의 분포를 가지므로 5를 기준으로 음성과 비음성을 구분하게 된다. 그러나 어느정도 주기성을 갖는 비음성의 경우는 이 값만으로는 완전하게 식별할수 없으므로 피치궤적과 웨이브렛 영역에서 정의된 끝점검출 파라미터 PA[2]의 궤적을 사용하게 된다.

$$PD = |T_1 - T_2| \tag{1}$$

$$Average\ pitch = \frac{T_1 + T_2}{2} \times F_s \tag{2}$$

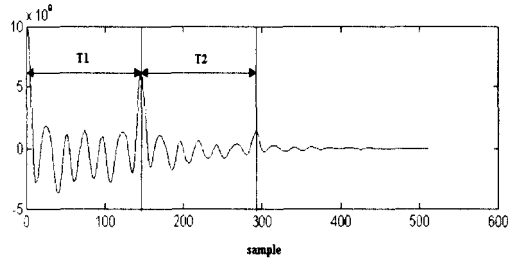

그림 1. /아/ 에대한 자기상관

표1. 비음성 데이터에대한 평균 피치 변화량 [sample]

No.	비음성(Nonspeech)	평균 PD
1	책장 넘기는 소리	10.4
2	책상 두드리는 소리, 문 닫는 소리	9.0
3	입 다시는 소리	9.1
4	기침 소리 및 휘파람 소리	8.1
5	헛기침 /흠/ 하는 소리	5.4
6	전화 끊는 소리	6.9

본 연구에서는 음성의 경우 최소한 세 프레임 이상 연속적인 피치를 가지는 것으로 가정한다. 즉 피치가 존재하는 범위를 만족하고 PD가 5이하인 경우에 이 구간을 음성부분으로 간주하게 된다. 이 과정을 통해서 얻어진 피치궤적을 조사해보면 식 (3)에서 정의된 인접 프레임간의 피치의 변화량이 작다는 것을 알 수 있다. 전체문장에서 식(2)를 이용하여 구한 각각의 고립된 피치궤적에 대해서 피치의 변화량이 5% 이상인 프레임의 수를 PN으로 정의하고 식 (4)에 대입함으로써 scale을 구하게 된다. 이 값을 실험적으로 구해보면 음성의 경우 대체로 0.05 이하의 값을 가지게 되고 주기성을 가지는 비음성의 경우는 이보다 큰 값을 가지게 된다. 식 (5)는 각각의 피치궤적이 음성 또는 비음성으로부터 얻어진 것인지를 최종적으로 판별하기 위한 파라미터로 0.5 이하의 값을 가질 때 비음성으로부터 얻어진 것으로 간주한다.

$$PV = \left| \frac{(T_n - T_{n+1})}{T_n} \right| \times 100\ [\%] \tag{3}$$

T_n , T_{n+1} : 현재 및 인접 프레임의 피치 [sample]

$$scale = \frac{PN}{\text{고립된 피치궤적의 총 프레임수}} \tag{4}$$

$$ratio = \frac{\text{고립된 피치궤적의 총 프레임수}}{\text{고립된 } PA\text{궤적의 총 프레임수}} - scale \tag{5}$$

그림 2는 비음성을 포함하고 있는 음성신호에 대한 피치 검출 결과이다. 그림 (b)는 자기상관 함수에 의한 피치를 나타내고 (c)는 피크 간격의 변화량인 PD를 나타내는데 여기서 수평선은 비음성을 제거하기위한 문턱치이다. (d)는 PD와 ratio를 이용함으로써 최종적으로 얻어지는 피치로서 후보 시작점과 끝점을 얻기 위해서 사용된다. (b)로부터 인접 피치간의 변화량이 음성부분에 비해서 비음성 부분이 더 큰 것을 알수 있다.

2.2 음성 검출 알고리듬

일반적으로 끝점 검출에 많이 사용하는 파라미터는 단구간 에너지와 문턱값을 비교해서 대략적인 끝점을 찾은 뒤에 영교차율로 정확한 끝점을 찾아내는 방법이다[4]. 이러한 검출 방법은 잡음이 없는 음성신호에 대해서는 신뢰할 수 있는 결과를 보이지만 잡음환경에서는 성능이 급속히 떨어지게 된다. 특히 음성의 시작이나 끝부분에 파열음이나 마찰음이 존재할 경우에는 신호의 에너지가 유성음구간에 비해 작기 때문에 잡음환경에서 검출하기가 용이하지 않으며 끝점 검출실패의 주요한 이유중 하나가 된다. 그림 3은 끝점검출을 위한 묵음구간에서의 문턱치 결정과정을 나타내고 있으며 STD1과 STD4는 각각 웨이브렛 영역에서 첫 번째 스케일과 네 번째 스케일의 표준편차를 나타내는데[2] 여

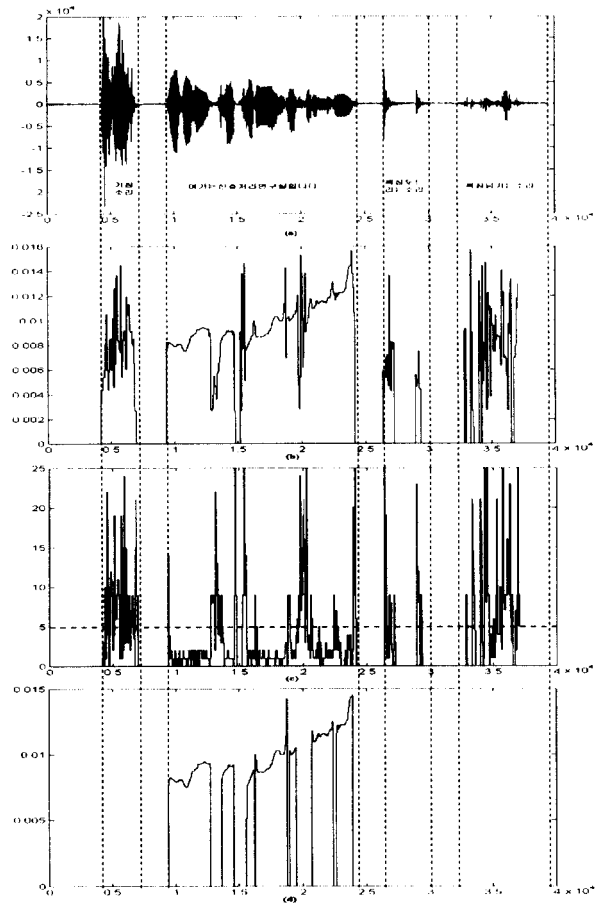

그림 2 (a) 음성신호, (b) 자기상관함수에 의한 피치 궤적, (c) PD, (d) ratio>0.5인 피치궤적

기서는 10프레임에 대한 평균값을 나타낸다. Case 1은 깨끗한 음성, Case 2, 3은 잡음음성에 대한 문턱치를 나타내는데 Case 3은 잡음음성 중에서도 주변잡음이 불규칙한 경우에 적용된다.

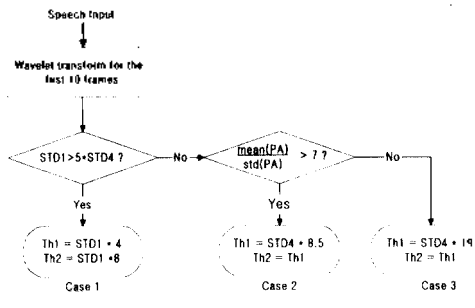

그림 3. 끝점검출을 위한 묵음구간에서의 문턱치 결정

음성/비음성 식별을 위한 알고리듬은 다음과 같다.

STEP 1. 프레임 단위로 웨이브렛 변환을 수행하면서 PA 검출파라미터와 문턱치를 구한다.

STEP 2. PA가 Th2보다 큰 프레임에 대해서 자기상관을 취한후 PD >5이거나 피치의 범위를 넘어서는 경우 이 프레임의 피치를 0으로 둔다.

STEP 3. 3 프레임 이상 연속적인 피치를 가지는 피치궤적에 대해서 scale 및 ratio 값을 구한다.

STEP 4. ratio가 0.5보다 작은 경우 비음성에 해당하는 피치로 간주하고 이 피치값을 0으로 둔다.

STEP 5. STEP 4 를 거친후 구해진 피치궤적으로부터 이 궤적의 시작점과 끝점을 음성의 후보 시작점과 후보 끝점으로 둔다.

STEP 6. 후보 시작점과 끝점에서의 PA값이 Th1보다 크면 Th1이 되는 지점까지 시작점과 끝점을 가져간 후 새로운 후보 시작점과 끝점으로 잡는다.

STEP 7. 후보 시작점과 끝점으로부터 앞뒤 10프레임을 체크해서 Th1이상의 값이 여러 프레임에 대해서 연속적으로 나오는지를 검사한후 최종적인 시작점과 끝점을 찾는다.

3. 실험 및 고찰

본 논문에서 사용한 알고리듬의 타당성을 실험하기 위해 전화음성과 실험실에서 만든 잡음음성을 대상으로 실험을 수행하였다. 실험 데이타는 8kHz 샘플링 되고 16비트 양자화 되었다. 실험실에서 녹음된 데이타는 임의로 백색잡음을 첨가하였고 전화음성은 실제 운용되고 있는 시스템에서 직접 수집한 것으로 주변잡음이 많이 포함된 데이타이다. 아래 그림 4~6은 본 논문에서 제안한 알고리듬을 사용하여 전화음성을 대상으로 실제 음성을 검출한 예를 보여주고 있다. 그림으로부터 화자에 의한 비음성이나 전화끊는 소리, 발신음 등의 전화망에서 발생할수 있는 비음성들을 효과적으로 음성으로부터 분리했음을 볼 수 있다. 발신음과 같은 전자음의 경우는 일정한 피치를 가지는 특성을 이용해서 음성과 쉽게 구분할수 있었다. 하지만 주변잡음이 매우 불규칙한 전화음성에 대해서는 잘못된 문턱치 설정으로 인해 정확한 음성 검출이 실패하는 경우도 있었다. 그림 9는 앞뒤로 비음성을 포함하는 /스포츠/에 대한 검출 결과로써 웨이브렛 변환을 이용하여 약한 마찰음도 잘 찾고 있음을 볼 수 있다. 대체로 15dB까지는 음성부분의 검출이 쉬웠으나 그 이하의 SNR (10dB~5dB)에서는 비음성에 의한 음성 검출실패 확률이 높게 나타났다.

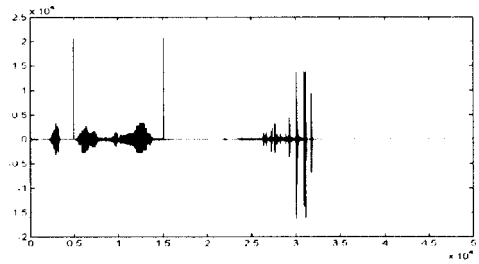

그림 4. /후-호텔신라-수화기놓는소리/

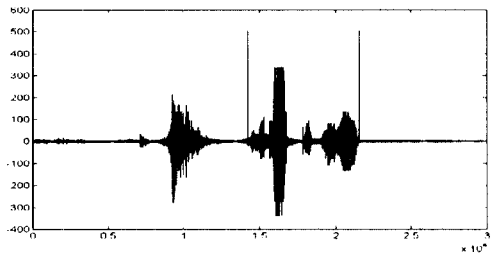

그림 5. /홈-수산중공업/

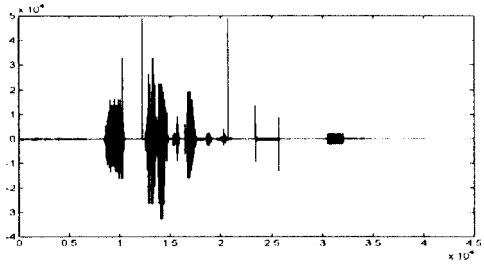

그림 6. /전화끊는소리-종합주가지수-발신음/

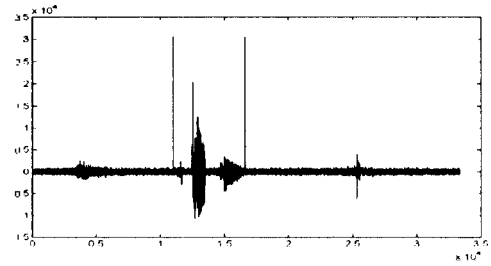

그림 7. /홈-스포츠-책장넘기는소리/, (SNR = 10dB)

4. 결 론

본 연구에서는 피치정보와 웨이브렛 변환을 이용하여 효과적으로 음성과 비음성을 식별함으로써 정확한 음성부 검출을 가능하게 하는 알고리듬을 개발하였다. 특히 에너지가 큰 비음성을 포함하는 경우 기존의 끝점 검출 알고리듬이 큰 오차를 보인 반면 제안된 방법은 비음성과 음성을 분리해냄으로서 끝점 검출의 정확성을 높일수 있었다. 또한 에너지가 약한 마찰음이나 파열음을 효과적으로 검출함으로써 음성 검출의 실패 확률을 줄일수 있었다.

본 연구는 한국통신의 정보통신 기초연구과제 (과제번호 KOSEF : 971-0917-103-2) 지원으로 수행되었으며, 지원에 감사드립니다.

참고문헌

[1] L.F. Lamel et. al., "An improved endpoint detector for isolated word recognition," IEEE Trans. Acoust., Speech, and Signal Processing,Vol.ASSP-29, No.4, pp.777-785, 1981.

[2] 석종원, 배건성 "Wavelet 변환을 이용한 잡음음성의 끝점 검출", 제9회 신호처리합동학술대회, 1996

[3] Ronald W.Schafer, John D.Markel, "Speech Analysis" IEEE PRESS, 1979.

[4] L. R. Rabiner and M. R. Sambur, "An Algorithm for Determining the Endpoints of Isolated Utterances," Bell Syst. Tech. J., Vol. 54, No. 2, pp. 297-315, February 1975.

[5] Kenzo Itoh and Masahide Mizushima, "Environmental Noise Reduction Based On Speech/Non-Speech Identication for Hearing Aids", Vol. I In Proc. IEEE ICASSP-97, pp419-422, 1997

[6] G.S. Ying, C.D. Mitchell, L.H.Jamieson,"Endpoint detection of isolated utterances based on a modified teager energy measurement," In Proc. IEEE ICASSP - 93, pp732-735, 1993.

[7] 안 정모, 계 형철, 구 명완 "전화음성인식을 위한 효율적인 끝점검출," 1994년도 대한전자공학회 추계종합학술대회 논문집, pp.1475-1478, 1994.

고립단어 인식시스템에서 음성/비음성 식별에 관한 연구

김 치수, 배 건성
경북대학교 전자·전기 공학부

A Study on The Speech/Nonspeech Identification for Isolated Word Speech Recognition System

Chi Su Kim , Keun Sung Bae
School of Electronic and Electrical Engineering, Kyungpook National University

요 약

음성인식 시스템의 입력인 음성은 실제의 음성부분 외에도 주변잡음을 포함한 기침 소리, 문닫는 소리, 책상 넘기는 소리등과 같은 사용자에 의해서 발생될수 있는 다양한 종류의 비음성을 포함할수 있다. 특히 에너지가 큰 비음성을 포함하는 경우 기존의 끝점검출 알고리듬만으로는 음성부분만의 정확한 검출이 어렵게 되고 이는 음성인식 시스템의 성능을 저하시키는 주요 원인이 된다. 본 논문에서는 음성 발생시 일어날 수 있는 비음성들에 대해서 조사하고 이러한 비음성이 포함될때 음성부분만의 정확한 검출을 가능하게 하는 알고리듬을 제시하였다. 사용된 파라미터로는 자기상관법에 의해 얻어지는 피치정보와 웨이브렛 영역에서의 에너지로써 비교적 낮은 신호대 잡음비(SNR)에서도 음성부 검출을 가능하게 하였다.

1. 서 론

음성인식, 합성 및 분석 등 음성공학의 거의 모든 분야에서 음성신호의 시작점 및 끝점을 주변잡음과 정확하게 분리하여 찾아내는 일은 매우 중요하다. 고립단어 인식시스템에서의 정확한 음성부 검출은 인식률을 향상시킬 뿐만아니라 비음성을 포함한 불필요한 묵음을 사전에 제거시킴으로써 단어 인식에 소요되는 시간을 줄일 수 있다. 음성인식 시스템에 입력되는 음성은 실제 음성부분외에도 주변잡음을 포함한 기침 소리, 문닫는 소리, 책장 넘기는 소리, 전화 끊는 소리와 같은 화자나 녹음환경, 전송매체에 의해서 발생될 수 있는 다양한 종류의 비음성을 포함할 수 있다. 마이크로 음성을 입력받을 경우 문닫는 소리나 의자를 움직일 때 나는 소리와 같이 일시적인 배경잡음들은 잡음제거 마이크를 사용함으로서 어느정도 그 영향을 줄일수 있다[1]. 하지만 전화망을 통하게 될 때 이 방법은 유효하지 않게 된다. 따라서 비음성을 식별할수 있는 알고리듬은 전화망과 연결되어 동작하는 고립단어 인식시스템에서 더욱더 필요하다.

본 연구에서는 비음성의 제거를 위하여 자기상관함수를 구했을 때 나타나는 신호의 주기성 변화를 조사하였다. 또한 잡음 환경하에서의 음성 검출을 위해서는 웨이브렛을 이용한 끝점검출 알고리듬[2]을 수정하여 사용하였다.

2. 음성 검출 방법

2.1 비음성 식별 파라미터

자기상관함수를 이용함으로써 음성신호의 피치를 검출할수 있다는 것은 널리 알려져 있는 사실이다. 그림 1에서와 같이 유성음의 경우 프레임 단위로 자기상관을 구해보면 시작점부터 첫번째 피크 사이의 간격 T_1이

주기가 되는데 첫번째 피크에서 다음 피크 사이의 간격 T_2가 T_1과 거의 동일하게 된다. 하지만 비음성의 경우는 식 (1)에서 정의되어진 피크 사이의 간격 변화가 유성음에 비해서 크게 나타나게 된다. 또한 식 (2)에서 정의된 피치를 구했을 때 비음성의 경우는 음성이 가질 수 있는 피치의 범위를 벗어나는 경우가 많이 생기게 된다. 일반적으로 유성음이 가질수 있는 피치의 범위는 3~15ms이다[3].

표 1은 실제로 발생할 수 있는 비음성들에 대한 자기상관함수 피크 간격의 변화량에 대한 평균값을 측정한 것이다. 수집된 비음성 데이터의 샘플링 주파수는 8kHz이다. 주기성이 뚜렷한 음성의 경우는 0~3의 값의 분포를 가지므로 5를 기준으로 음성과 비음성을 구분하게 된다. 그러나 어느정도 주기성을 갖는 비음성의 경우는 이 값만으로는 완전하게 식별할수 없으므로 피치궤적과 웨이브렛 영역에서 정의된 끝점검출 파라미터 PA[2]의 궤적을 사용하게 된다.

$$PD = |T_1 - T_2| \tag{1}$$

$$Average\ pitch = \frac{T_1 + T_2}{2} \times F_s \tag{2}$$

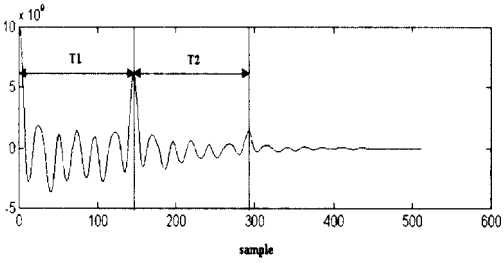

그림 1. /아/ 에대한 자기상관

표1. 비음성 데이터에대한 평균 피치 변화량 [sample]

No.	비음성(Nonspeech)	평균 PD
1	책장 넘기는 소리	10.4
2	책상 두드리는 소리, 문 닫는 소리	9.0
3	입 다시는 소리	9.1
4	기침 소리 및 휘파람 소리	8.1
5	헛기침 /흠/ 하는 소리	5.4
6	전화 끊는 소리	6.9

본 연구에서는 음성의 경우 최소한 세 프레임 이상 연속적인 피치를 가지는 것으로 가정한다. 즉 피치가 존재하는 범위를 만족하고 PD가 5이하인 경우에 이 구간을 음성부분으로 간주하게 된다. 이 과정을 통해서 얻어진 피치궤적을 조사해보면 식 (3)에서 정의된 인접 프레임간의 피치의 변화량이 작다는 것을 알 수 있다. 전체문장에서 식(2)를 이용하여 구한 각각의 고립된 피치궤적에 대해서 피치의 변화량이 5% 이상인 프레임의 수를 PN으로 정의하고 식 (4)에 대입함으로써 scale을 구하게 된다. 이 값을 실험적으로 구해보면 음성의 경우 대체로 0.05 이하의 값을 가지게 되고 주기성을 가지는 비음성의 경우는 이보다 큰 값을 가지게 된다. 식 (5)는 각각의 피치궤적이 음성 또는 비음성으로부터 얻어진 것인지를 최종적으로 판별하기 위한 파라미터로 0.5 이하의 값을 가질 때 비음성으로부터 얻어진 것으로 간주한다.

$$PV = \left| \frac{(T_n - T_{n+1})}{T_n} \right| \times 100\ [\%] \tag{3}$$

T_n , T_{n+1} : 현재 및 인접 프레임의 피치 [sample]

$$scale = \frac{PN}{\text{고립된 피치궤적의 총 프레임수}} \tag{4}$$

$$ratio = \frac{\text{고립된 피치궤적의 총 프레임수}}{\text{고립된 } PA\text{궤적의 총 프레임수}} - scale \tag{5}$$

그림 2는 비음성을 포함하고 있는 음성신호에 대한 피치 검출 결과이다. 그림 (b)는 자기상관 함수에 의한 피치를 나타내고 (c)는 피크 간격의 변화량인 PD를 나타내는데 여기서 수평선은 비음성을 제거하기위한 문턱치이다. (d)는 PD와 ratio를 이용함으로써 최종적으로 얻어지는 피치로서 후보 시작점과 끝점을 얻기 위해서 사용된다. (b)로부터 인접 피치간의 변화량이 음성부분에 비해서 비음성 부분이 더 큰 것을 알수 있다.

2.2 음성 검출 알고리듬

일반적으로 끝점 검출에 많이 사용하는 파라미터는 단구간 에너지와 문턱값을 비교해서 대략적인 끝점을 찾은 뒤에 영교차율로 정확한 끝점을 찾아내는 방법이다[4]. 이러한 검출 방법은 잡음이 없는 음성신호에 대해서는 신뢰할 수 있는 결과를 보이지만 잡음환경에서는 성능이 급속히 떨어지게 된다. 특히 음성의 시작이나 끝부분에 파열음이나 마찰음이 존재할 경우에는 신호의 에너지가 유성음구간에 비해 작기 때문에 잡음환경에서 검출하기가 용이하지 않으며 끝점 검출실패의 주요한 이유중 하나가 된다. 그림 3은 끝점검출을 위한 묵음구간에서의 문턱치 결정과정을 나타내고 있으며 STD1과 STD4는 각각 웨이브렛 영역에서 첫 번째 스케일과 네 번째 스케일의 표준편차를 나타내는데[2] 여

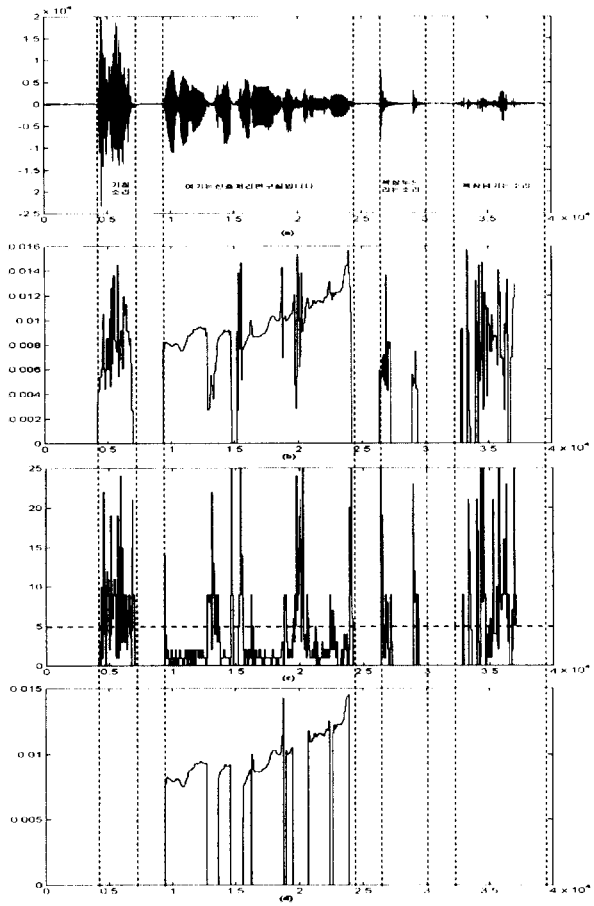

그림 2 (a) 음성신호, (b) 자기상관함수에 의한 피치 궤적, (c) PD, (d) ratio>0.5인 피치궤적

기서는 10프레임에 대한 평균값을 나타낸다. Case 1은 깨끗한 음성, Case 2, 3은 잡음음성에 대한 문턱치를 나타내는데 Case 3은 잡음음성 중에서도 주변잡음이 불규칙한 경우에 적용된다.

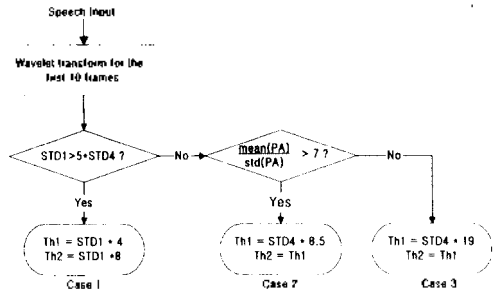

그림 3. 끝점검출을 위한 묵음구간에서의 문턱치 결정

음성/비음성 식별을 위한 알고리듬은 다음과 같다.

STEP 1. 프레임 단위로 웨이브렛 변환을 수행하면서 PA 검출파라미터와 문턱치를 구한다.

STEP 2. PA가 Th2보다 큰 프레임에 대해서 자기상관을 취한후 PD >5이거나 피치의 범위를 넘어서는 경우 이 프레임의 피치를 0으로 둔다.

STEP 3. 3 프레임 이상 연속적인 피치를 가지는 피치궤적에 대해서 scale 및 ratio 값을 구한다.

STEP 4. ratio가 0.5보다 작은 경우 비음성에 해당하는 피치로 간주하고 이 피치값을 0으로 둔다.

STEP 5. STEP 4 를 거친후 구해진 피치궤적으로부터 이 궤적의 시작점과 끝점을 음성의 후보 시작점과 후보 끝점으로 둔다.

STEP 6. 후보 시작점과 끝점에서의 PA값이 Th1보다 크면 Th1이 되는 지점까지 시작점과 끝점을 가져간 후 새로운 후보 시작점과 끝점으로 잡는다.

STEP 7. 후보 시작점과 끝점으로부터 앞뒤 10프레임을 체크해서 Th1이상의 값이 여러 프레임에 대해서 연속적으로 나오는지를 검사한후 최종적인 시작점과 끝점을 찾는다.

3. 실험 및 고찰

본 논문에서 사용한 알고리듬의 타당성을 실험하기 위해 전화음성과 실험실에서 만든 잡음음성을 대상으로 실험을 수행하였다. 실험 데이타는 8kHz 샘플링 되고 16 비트 양자화 되었다. 실험실에서 녹음된 데이타는 임의로 백색잡음을 첨가하였고 전화음성은 실제 운용되고 있는 시스템에서 직접 수집한 것으로 주변잡음이 많이 포함된 데이타이다. 아래 그림 4~6은 본 논문에서 제안한 알고리듬을 사용하여 전화음성을 대상으로 실제 음성을 검출한 예를 보여주고 있다. 그림으로부터 화자에 의한 비음성이나 전화끊는 소리, 발신음 등의 전화망에서 발생할수 있는 비음성들을 효과적으로 음성으로부터 분리했음을 볼 수 있다. 발신음과 같은 전자음의 경우는 일정한 피치를 가지는 특성을 이용해서 음성과 쉽게 구분할수 있었다. 하지만 주변잡음이 매우 불규칙한 전화음성에 대해서는 잘못된 문턱치 설정으로 인해 정확한 음성 검출이 실패하는 경우도 있었다. 그림 9는 앞뒤로 비음성을 포함하는 /스포츠/에 대한 검출 결과로써 웨이브렛 변환을 이용하여 약한 마찰음도 잘 찾고 있음을 볼 수 있다. 대체로 15dB까지는 음성부분의 검출이 쉬웠으나 그 이하의 SNR (10dB~5dB)에서는 비음성에 의한 음성 검출실패 확률이 높게 나타났다.

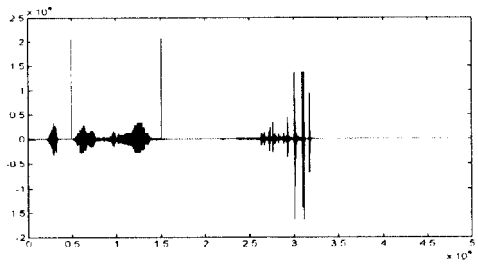

그림 4. /후-호텔신라-수화기놓는소리/

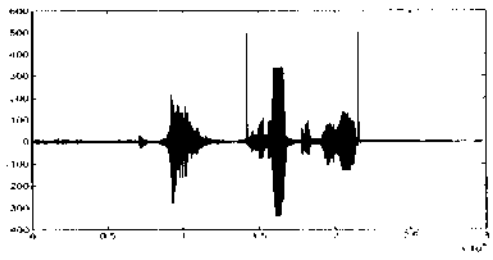

그림 5. /홈-수산중공업/

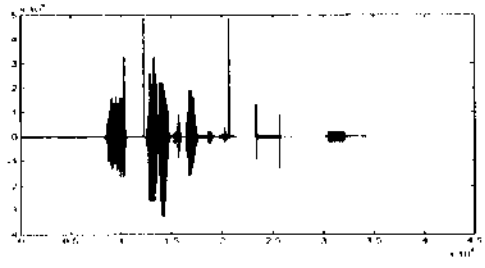

그림 6. /전화끊는소리-종합주가지수-발신음/

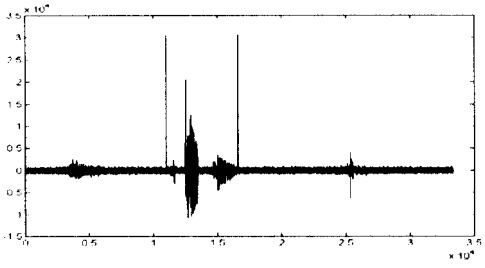

그림 7. /홈-스포츠-책장넘기는소리/, (SNR = 10dB)

4. 결 론

본 연구에서는 피치정보와 웨이브렛 변환을 이용하여 효과적으로 음성과 비음성을 식별함으로써 정확한 음성부 검출을 가능하게 하는 알고리듬을 개발하였다. 특히 에너지가 큰 비음성을 포함하는 경우 기존의 끝점 검출 알고리듬이 큰 오차를 보인 반면 제안된 방법은 비음성과 음성을 분리해냄으로서 끝점 검출의 정확성을 높일수 있었다. 또한 에너지가 약한 마찰음이나 파열음을 효과적으로 검출함으로써 음성 검출의 실패 확률을 줄일수 있었다.

본 연구는 한국통신의 정보통신 기초연구과제 (과제번호 KOSEF : 971-0917-103-2) 지원으로 수행되었으며, 지원에 감사드립니다.

참고문헌

[1] L.F. Lamel et. al., "An improved endpoint detector for isolated word recognition," IEEE Trans. Acoust., Speech, and Signal Processing,Vol.ASSP-29, No.4, pp.777-785, 1981.

[2] 석종원, 배건성 "Wavelet 변환을 이용한 잡음음성의 끝점 검출", 제9회 신호처리합동학술대회, 1996

[3] Ronald W.Schafer, John D.Markel, "Speech Analysis" IEEE PRESS, 1979.

[4] L. R. Rabiner and M. R. Sambur, "An Algorithm for Determining the Endpoints of Isolated Utterances," Bell Syst. Tech. J., Vol. 54, No. 2, pp. 297-315, February 1975.

[5] Kenzo Itoh and Masahide Mizushima, "Environmental Noise Reduction Based On Speech/Non-Speech Identication for Hearing Aids", Vol. I In Proc. IEEE ICASSP-97, pp419-422, 1997

[6] G.S. Ying, C.D. Mitchell, L.H.Jamieson,"Endpoint detection of isolated utterances based on a modified teager energy measurement," In Proc. IEEE ICASSP - 93, pp732-735, 1993.

[7] 안 정모, 계 형철, 구 명완 "전화음성인식을 위한 효율적인 끝점검출," 1994년도 대한전자공학회 추계종합학술대회 논문집, pp.1475-1478, 1994.