

한국어 음성인식을 위한 음소 분할에 관한 연구

송 석일

한 영역

한양 대학교

A study on the phoneme segmentation method for the Korean language recognition

SEOG ILL SONG

Young Yeon Han

Dept. of Electronic Communication Eng. Hanyang Univ.

ABSTRACT

In this paper, two methods of the phoneme segmentation for the Korean language recognition are studied. the phoneme segmentation is accomplished by using the frame power which is filtered by smoothed-second-order-difference-filter after passing band pass filter banks. while the other method comprise of neighbor frames, squaring and summing of difference components and picking the peak values for phoneme segmentation.

1. 서론

음성인식에 있어 음성인식 단위 구분에 의한 인식 단위 분할은 인식 결과와 밀접한 관계가 있다.(1) 우리말에 대한 음성인식을 위해서, 숫자인식 단모음 인식, 음절인식에 대해 연구 되어 왔다.(2),(3),(4) 우리말은 음소단위가 모여 음절, 단어, 문장으로 형성되기 때문에 우리말을 음소 단위로 구분 인식할 수 있다. 본 논문에서는 우리말의 인식을 위해 최소 단위의 음소분할(phoneme segmentation)의 방법으로서는 음성의 각각 프레임의 주파수 성분 비교 방법과, 22-채널 밴드 패스 필터뱅크(BPF banks)의 출력을 평활화 2차분 필터(smoothed-2nd-order-difference-filter)를 통과 시킨 다음 제로크로싱점(zero crossing point)을 갖는 법과 음성의 각 프레임의 절대값 파워를 평활화 2차분 필터를 통과 시켜 제로크로싱점을 찾는 방법을 이용하여 우리말의 음소 분할에 대해서 분석 실험 하였다.

2. 음성자료

아나로그 음성을 LPF(300 - 3400Hz) 통과 시킨 다음 8KHz 샘플링, 12bit로 디지털한 음성신호인 숫자음을 비롯한 20개의 데이터를 한국전자통신 연구소에서 특수, IBM-PC를 이용하여 실험 하였다.

3. 22 채널 밴드 패스 필터뱅크 (BPF)

음성신호의 특징 파라미터를 정확히 산출하기 위해서는 BPF bank의 조합된 주파수 응답은 가능한 진폭특성은 평탄하고, 위상특성은 선형인 특성을 갖는 이상적인 형태로 설계되어야 한다. 그러나 이러한 이상적인 형태를 갖는 필터뱅크의 실제적 구현은 대단히 어렵다.

이점을 고려하고, 사람의 청각특성이 주파수에 따라 비선형적이라는 것과 위상변이에 대해서는 비교적 둔감하다는 사실을 감안하여 비정형 형태로 근사시키는 log 등간격으로 주파수 간격을 정한다.

사람의 청각특성과 Hard Ware 구현시 필터링 시간등을 고려하여 간단한 Biquadratic 형태의 아나로그 BPF를 Bilinear Transformation에 의해 IIR(infinite impulse response) 디지털 필터로 설계한다.

$$H(s) = \frac{\omega_0 S}{S^2 + \frac{\omega_0 Q}{S} + \omega_0^2} \rightarrow \textcircled{1}$$

$$\Rightarrow S = c \frac{1 - z^{-1}}{1 + z^{-1}}$$

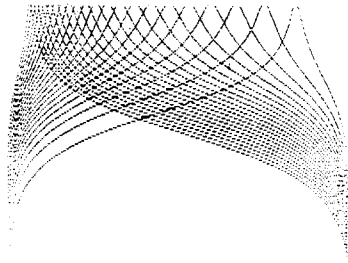
$$H(z) = \frac{B_0 - B_1 z^{-1}}{1 + A_1 z^{-1} + A_2 z^{-2}} \rightarrow \textcircled{2}$$

(ω_0 :공진주파수, Q :quality factor)

이 필터뱅크에서 BPF의 주파수 통과 대역은 300Hz - 3400Hz이며, 이 사이를 log 등간격 (1/2 Octave 간격 $Q=6$)으로 22 채널로 나누어 설계한다.



(그림 1) BPF banks의 각각 주파수 응답곡선



(그림 2) BPF banks의 각각 주파수 응답곡선

4. 평활화 2차분 필터 (9.8, 9.10)

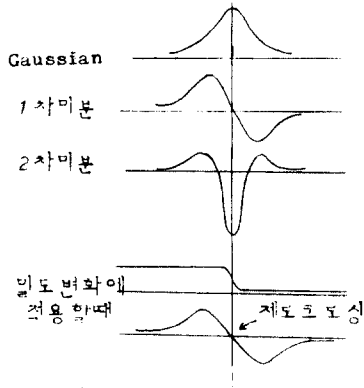


그림 3.

가우시안 곡선(Gaussian Curve) 에 미분을 적용할때와 밀도 변화의 경계점을 찾기 위해서 이차미분을 적용시킨 그림

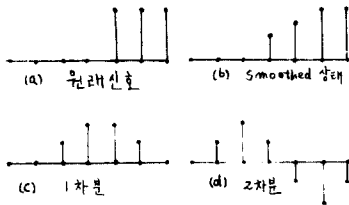


그림 4.

그림(4)에서 7개의 점중에서 4번째와 5번째의 경계점을 찾기 위해서 이루어지는 과정을 설명 하면 신호에서의 잡음의 영향 때문에 일어나는 급격한 변화를 줄이기 위해서 이웃 신호끼리의 평균을 취한다. 이것을 평활화라 하며 식 A_i 와 같다.

$$A_i = \frac{I_{i-1} + I_i + I_{i+1}}{3}, (I_i: i\text{번째 신호의 값})$$

$$A_i \text{의 1차분: } F_{i1} = \frac{(A_{i+1} - A_i) + (A_i - A_{i-1})}{2} = \frac{A_{i+1} - A_{i-1}}{2}$$

$$A_i \text{의 2차분: } S_{i2} = \frac{(F_{i+1} - F_{i1}) + (F_{i1} - F_{i-1})}{2} = \frac{F_{i+1} - F_{i-1}}{2}$$

S_{i2} 의 식에 A_i 을 대입하여 정리하면 다음과 같고 그림으로 표현하면 다음과 같다.

$$S_i = 0.25 A_{i+2} - 0.5 A_i + 0.25 A_{i-2} \rightarrow (16)$$



(그림 5) 평활화 작용을 포함한 2차분 필터

5. 프레임간의 주파수 성분차의 계산 (11.12)

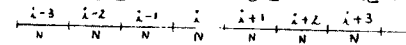


그림 6

그림에서 i 는 각각의 프레임 번호를 나타내며, N 은 M msec 의 샘플수도 $2N$ 개의 주파수 성분으로 FFT 변환한 주파수 성분 개수의 $1/2$ 성분을 의미한다. 예를들면, 8 KHz로 샘플링 할때, $M=16$ msec 라 하면, 그 때의 샘플수는 128개가 되며, 이때 FFT 변환한 주파수 성분도 128개이며, FFT 변환한 주파수 성분 사이의 간격은 62.5 가 된다.

FFT 는 대칭성을 가지므로 128개의 주파수 성분 가운데 64개의 주파수 성분만 필요하다. 프레임 i 번째의 양쪽 이웃의 각각 같은 주파수 성분차 제곱을 구한 다음, 그 값들의 합의 미크 값을 구하는 식은 다음과 같으며, 식에서 B 는 2, 3, 4, 5일때, N 이 64, 128일때 비교 실험한다.

$$TH_i = \sum_{j=1}^B \sum_{l=1}^N (P_{i-1}(l) - P_i(l))^2 \rightarrow (17)$$

6. 실험 및 결과

6-1. 주파수 성분차 미크값 계산

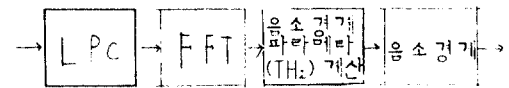
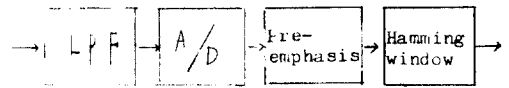


그림 7.

그림에서 300 - 3400 Hz의 LPF를 통과시킨 다음 8 KHz 샘플링 12 bit로 디지털하여 2역을 강조키 위해서 1차 차분필터를 통과시키고, 분석구간을 나눌때 나타나는 영향을 고려하여 Hamming window (16 msec, 32 msec)를 이용하였으며, 음원의 미지 주파수와 폴만트 (Formant) 주파수를 고려하여, 13차 1000 Smoothed Spectrum을 10 msec 단위의 프레임 구간으로 하여 분석 하였다.

6-2. 22 채널 BPF banks와 평활화 2차분 필터
 22 채널 BPF banks와 평활화 2차분 필터를
 이용한 음소 경계검출을 위한 선형 BPF 출력의
 선형대수와 합을 식으로 나타내면 다음과 같다.

$$E = \sum_{i=1}^{22} \log_{10} S_i(x) \quad S_i: \text{프레임간 } 10 \text{ msec의 } i \text{ 번째 BPF 의 출력}$$

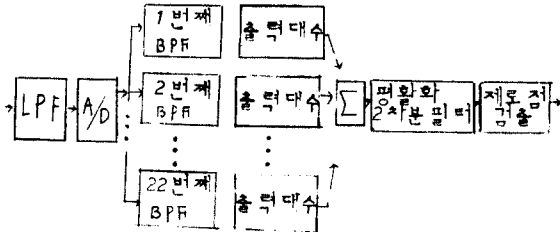


그림 8.

6-3. 각각 프레임 출력과 평활화 2차분 필터

각각의 프레임 길이를 10 msec 한 출력의 절대
 값의 대수를 구하여 2차분 필터를 통과시키는
 것에 대한 식과 그림은 다음과 같다.

$$E = 10 \log \left(\frac{1}{N} \sum_{i=1}^N |S(i)| \right) \rightarrow (9)$$

N: 프레임당의 샘플수
 S(i): i 번째 샘플 출력

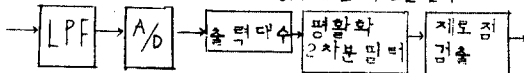


그림 9.

6-4. 실험 결과

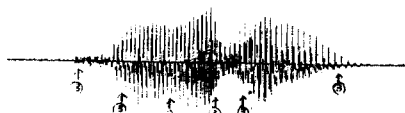


그림 10.a 뒤로



그림 10.b 아니요

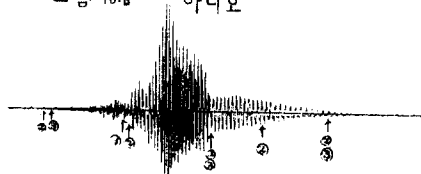


그림 10.c 삼

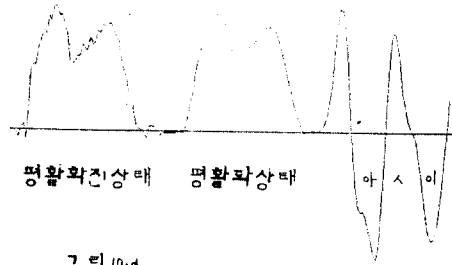
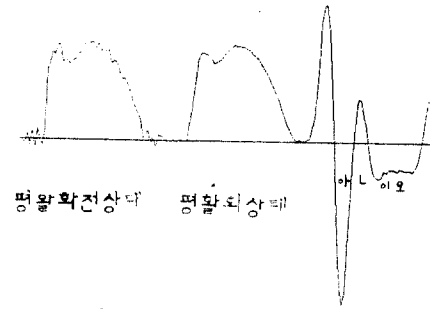


그림 10.d



평활화전상태 평활화상태

그림 10.e 아니요

(그림 10) 실험결과와의 그림

식(7)에 의한 프레임 주위의 주파수 성분 의 피크
 값 검출에서는 B 가 4,5일때는 음소 경계나
 대부분 오류를 나타냈으며, B 가 2일때도
 오류는 4,5보다는 적었으나 많은 오류를 나타
 냈다. B 가 3일때는 음소의 경계가 잘 나타
 났는데, 어두의 음소경계는 오류를 범했으며,
 짧은 어두와 이의의 음소 경계에서 오류를 범했고,
 N 이 64, 128은 거의 같은 결과를 나타냈다.

22-채널 BPF banks 와 2차분 필터에 의한 방법
 에서는 모음구간에서 경계가 나타나지 않았으며,
 프레임 출력과 2차분 필터에 의한 방법에서는 어두
 음소경계가 오류를 범했으며, 모음 구간에서는
 나타나지 않았다.

(그림 10)은 실험결과를 나타낸 그림으로서

㉠, ㉡는 B를 2,3으로 했을때의 개: 의
 피크값을 나타내며 (그림 10) d, e 는 (그림 9)에
 의한 손서에 의한 결과인데 (그림 10) d 는 음성
 '타시'의 결과 (그림 10) e 는 음성 '아니요'의
 결과를 나타낸 그림이다.

7. 결론

본 논문에서 이용한 주파수 성분 비교를 이용한
 방법에서 어두음소의 경계점 추출의 보원과 짧은
 음소 구간 검출의 오차를 보완한다면 좋은 음소
 경계 검출에 이용될 수 있을 것으로 생각되며,
 22채널 BPF banks 와 평활화 2차분 필터,
 프레임 출력과 평활화 2차분 필터를 이용한 방법도
 어두음소와 모음과 모음일때의 경계검출만 보완
 되면 좋은 음소 경계검출에 이용될 수 있을 것
 이라 생각된다.

8. 참고문헌

- 1) M. Jalanco, S. Haltsonen, K. Bry, T. Kohonen, "Application of Novelty Filter to Segmentation of Speech," IEEE Int. Conf. Acoust., Speech Signal Processing, 1978
- 2) 김영일, "한국어 단음의 분석과 인식에 관한 연구", 박사학위논문 연세대학교 대학원, 1985.
- 3) 이길행, "한국어연식을 위한 효과적인 음절 분할에 관한연구", 석사학위논문, 한국과학기술원, 1986.
- 4) 이용주, 차근현, "통계적방법에 의한 단모음의 인식실험" 대한전자공학회, 아계종합학술대회 논문집, Vol 9, NO 1, 1986.6.
- 5) M.E. Van Valkenburg, Analog Filter Design, Holt Saunders, 1982
- 6) N.K. Bose, Digital Filters Theory and Applications, North Holland, 1985
- 7) P.H. Winston, Artificial intelligence 2nd, Addison Wesley, 1984
- 8) E. Charniak, D. McDermott, Introduction to Artificial Intelligence, Addison Wesley, 1983
- 9) 本間 茂, 井出和之, 牧野 城戸, "セグメンテーションによる音素群認識の方法", 音声研究資料, S82-56, 1982-12
- 10) 本間 茂, 井出和之, 川端 豪, 牧野 城戸, "単語音声中の母音子音区間抽出の方法", 電子通信学会, FA81-54, 1982.1
- 11) 安居院 益, 中嶋 正元, コンピュータ音声処理, 産報出版, 1980
- 12) J.D. Markel, A.H. Gray, Linear Prediction of Speech, Springer Verlag, 1976

본 논문 작성에 도움을 주신 한국전자통신
연구소 음향 연구실께 감사드립니다.