Likelihood Ratio에 의한 음소분류에 관한 연구

이 기영 배 철수 최 갑석
명지대학교 전자공학과

A Study on the Phonemic Segmentation by Likelihood Ratio

Ki-young Lee Chul-soo Bae Kap-seok Choi
Dept. of Electronics Eng.,Myong Ji Univ.

Abstract

This paper proposed the phonemic segmentation method that employed two types of Likelihood Ratio that measures the change of spectral structure. By this method,isolated digits and words of VCV form are segmented into phoneme-unit and especially,first-burst part in an aspirated bilabial plosive is divided.

1. 서 론

음성을 기계와의 대화매체로 이용하기 위해 음성인식에 관한 연구가 시작되었으며 궁극적으로 자연음성 인식에 가까운 연속음성 인식에 대한 연구가 계속되고 있다.[1]

고립단어 인식은 패턴 매칭 방법으로 많은 연구가 진행되었고, 반도체 및 신호처리 기술의 발전에 힘입어 수백 단어를 인식하는 시스템이 실용화 되고 있으나,연속음성 인식에서는 단어를 인식단위로 하는경우 인식할 단어의 수가 증가할수록 방대한 기억용량과 많은 계산량이 필요하므로 이를 해결하기 위한 방법으로 입력 음성신호를 언어의 최소 단위인 음소로 분류하고 그것을 식별하므로써 인식하는 방법이 연구되고 있다.[2,3]

Fant는,'음성신호는 그 파형의 구조가 상대적으로 독특하게 구분되는 경계를 가진 음소의 시계열로 구성된다' 고 하였으며 Klatt 와 Stevens 는 Spectrogram 을 관찰함으로써 음소분류를 하였고, Rabiner 는 영교차율과 에너지로써 분류를 시도 하였다. Zue는 LPC 스펙트럼의 대역에너지 및 변화도를 이용하여 분류하였다. 그러나 발성자 또는 문맥에 따라 발성속도와 진폭이 변화하기 때문에 음소단위로 분류하는데에는 많은 어려움이 있다. [4-7]

본 논문에서는 연속적으로 가정된 두개의 음소범위에 있는 두 모델의 Likelihood Ratio을 측정함으로써 음소단위로 분류하는 방법을 제안한다. 이 방법은 음성신호의 음소들이 서로 생성모델과 파형의 구조를 달리 하므로 두 음소의 스펙트럼 구조의 변화도를 Likelihood Ratio로 구하여 음소를 분류한 것이다. 또한 스펙트럼 구조의 변화도를 구하는 것이므로 불특정 화자에게 적용이 가능하며 진폭변화의 영향을 적게 받는다.

2. 음소의 분류 방법

음성은 성대의 진동 여부에 따라 유성음과 무성음으로 나뉘어 진다. 현재 우리말에는 유성음으로 모음과 유성자음(ㄴ,ㄹ,ㅁ,ㅇ)이 있으며, 무성음으로는 자음이 있다.

모음은 성도의 형태가 발성의 시작부터 끝까지 일정하여 자음에 비해 안정된 소리이지만 각 모음의

포만트(formant)주파수 구조가 서로 다른 특징을 가지고 있다. 자음은 성대의 진동을 수반하지 않고 성도의 여러기관과의 마찰을 통해 발성되며 그 방법에 따라 파열음, 마찰음, 비음 및 유음으로 나뉘어 진다.

파열음은 폐쇄, 파열, 휴지기등의 세단계를 거치며, 마찰음은 성도에서 공기가 작은 통로를 지날때 마찰소리이며, 비음은 비강의 공명을 수반하고, 유음은 공기의 흐름이 입안의 어떤 부분을 떨림으로써 나는 소리이다.

이와 같이 조음과정을 달리하는 각 자음과 모음은 서로 다른 스펙트럼 구조를 하고 있으므로, 본 논문에서는 스펙트럼 구조의 차이를 측정하는 Likelihood Ratio(LR) 측정 방법에 의하여 음소를 분류하고자 한다.

그처리과정은 그림.1 과 같이 두 모델로부터 변화를 검출하는 것으로써 다음과 같다.

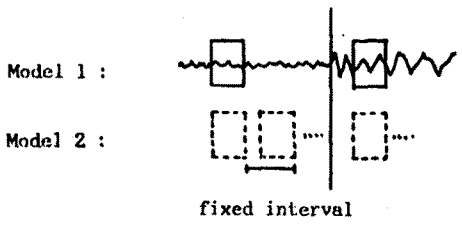

그림 1 두 모델의 변화검출

Fig. 1 Change detection of two models.

(1) 무음부 또는 배경 잡음부에 제 1 모델을 고정한다.

(2) 제 2 모델을 제 1 모델과 일치시켜 LR을 측정한다.

(3) 제 2 모델을 현재의 위치에서 일정한 간격으로 이동시키고 LR을 측정한다.

(4) 측정된 거리와 threshold 와 비교하여 그보다 작으면 (3)의 과정을 반복한다. 만약, 그보다 크면 (5)의 과정으로 간다.

(5) 현재 제 2 모델의 시작점을 분류경계점 후보로 하고 그 위치에 제 1 모델을 고정시키며 (2)의 과정부터 반복한다.

(6) 음성신호의 끝점에 다다르면 (5)에서 정한 분류경계점 후보를 모두 분류경계점으로 하되 두 분류경계점들의 사이시간이 10ms 이면 앞선후보를 선택한다.

3. Likelihood Ratio

음성신호 x(n)이 M개의 과거신호를 선형조합하여 예측한다면 그의 오차신호는 다음과 같다.

$$e(n) = \sum_{i=0}^{M} ai\ x(n-i) \qquad (1)$$

여기서, a0 = 1 일때 자승오차의 합 또는 잔차에너지는 다음과 같다.

$$\alpha = \sum_{i=-\infty}^{\infty} [e(n)] \qquad (2)$$

자기상관 방법에서는 x(n)은 $n < N$ 의 범위에 존재한다고 가정하여 잔차에너지 α가 최소가 될때의 {ai}를 구한다. 이 오차신호 e(n)은 역필터, A(z)의 출력이라고 할 수 있으며, A(z)는 잔차에너지 α를 최소로 한다.

$$A(z) = 1 + \sum_{i=1}^{M} ai\ z^{-i} \qquad (3)$$

1/A(z)는 음성신호 시계열 {x(n)}의 평활스펙트럼으로 나타난다.

만일, {x(n)}이 다른 형태의 필터, 즉 다음 음성신호 시계열 {x'(n)}에 대하여 잔차에너지 α'를 최소로 하는 역필터, A'(z)를 통과한다면 그때 생기는 δ는 α보다 크다.

$$A(z) = \sum_{i=0}^{M} ai'\ z^{-i} \qquad (4)$$

$$\delta = \sum_{n=-\infty}^{\infty} \left[\sum_{i=0}^{M} ai'x(n-i) \right]^2 \geq \alpha \qquad (5)$$

식(5)에서 등식은 A(z) = A'(z)일 때만 가능하다. 역필터 A(z)와 A'(z)를 잔차에너지에 의해서 비교하는 방법은 그림2-1 과 같다.

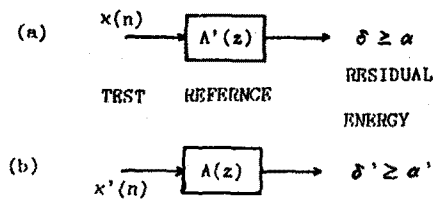

그림2-1 Likelihood Ratio 의 측정 방법

Fig.2-1 Comparison methods of Likelihood Ratio.

그림2-1, (a)의 경우 A(z)을 표준필터라 하고, {x(n)}을 후보신호 시계열이라 하여 식(2), (5)와 같이 α, δ 를 구한다. 또한 그림2-1 의 (b)의 경우는 (a) 와 달리 표준과 후보시계열 및 필터의 위치를 바꾼것이다. 이 두 잔차에너지의 비 δ/α와 δ'/α'는 두 신호의 스펙트럼 구조의 차이로 정의할 수 있으며 Likekihood Ratio 라 한다. 이 두 가지 방법의 Likelihood Ratio 는 그 측정 값이 서로 비선형적이며 그림2-2는 상대적 분포도이다.

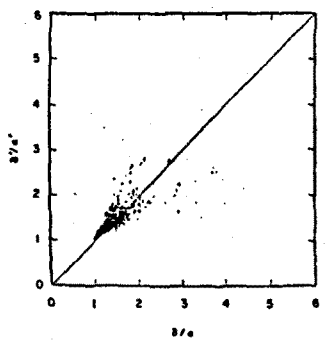

그림2-2 δ/α 와 δ'/α'의 상대적 분포도.

Fig.2-2 Scatter plot of the likelihood ratios $\delta/\alpha, \delta'/\alpha'$.

4. 실험 및 고찰

(1) 실험 데이타.

아나로그 음성신호를 10[kHz]의 sampling rate로 A/D변환하였고 잡음이 있는 실험실에서 마이크로폰을 통하여 입력하였으며 본 실험에서 사용한 모델의 프레임 길이를 10[msec]로 하였다. 그 구성도는 그림3과 같다.

대상음성은 20대 남성 5인이 발성한 단독 숫자음 "공", "일", "이", "삼", "사", "오", "육", "칠", "팔", "구"와 1인의 20대 남자가 발성한 vcv단어를 대상으로 하였다.

(2) 결과 및 고찰

AR모델의 차수는 10차로 하였으며 14차로 할 경우보다 구체적인 부분을 검출할 수 있었다. 그림4 에서는 마찰음, 파찰음, 비음및 유음을 포함하는 "공", "삼", "팔" 등과 vcv단어 "아가"를 Likelihood Ratio에 의해 분류한 결과를 보이고 있다.

이상에서 "공", "팔"의 파열자음을 초성으로 할때와 모음에서 비음의 종성으로 연결되는 "공", "삼"에서 음소분류 경계점의 검출이 용이하였다.. 그러나 "삼" 의 마찰자음을 초성으로 할 때에는 시작점에서 정확히 검출되지 않았다. 또한 vcv단어 "아가"와 같이 파열자음 "ㄱ"이 종성으로 발음 되는 경우에 분류하였다

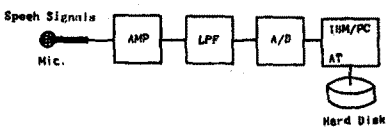

그림3 구성도.

Fig.3 Block diagram.

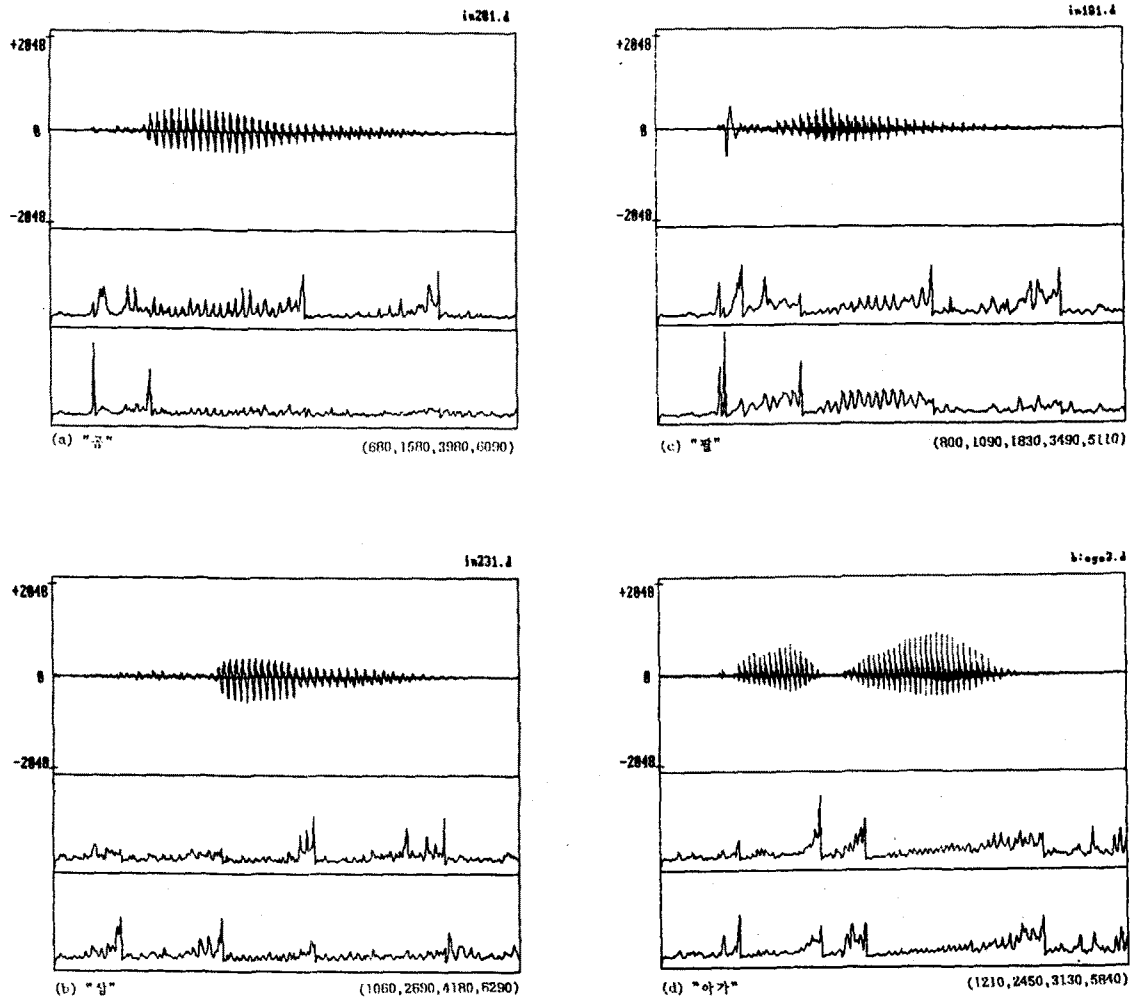

그림4 Likelihood Ratio 에 의한 음소 분류.

Fig.4 Phonemic segmentation by Likelihood Ratio.

5. 결 론

본 연구에서는 음성신호의 가정된 두 개의 음소범위 모델의 Likelihood Ratio를 측정함으로써 음소의 분류를 시도한 결과 다음과 같은 결론을 얻었다.

(1) 두 가지 방법의 Likelihood Ratio를 조합하여 만족 할만한 음소분류를 할 수 있었다.

(2) 초성 기식 양순 파열음(ㅍ ;ph)의 파열부분을 분리할 수 있었다.

(3) 본 방법이 연속음성의 음소분류에 유용하리라 사료된다.

* REFERENCES.

1. L.R.RABINER,S.E.LEVINSON,"ISOLATED AND CONNECTED WORD RECOGNITION-THEORY AND SELECTED APPLICATION",IEEE,TRANS. ON COMM. VOL.COM-29,NO.5,MAY.1981

2.G.BRISTOW,"ELECTRONIC SPEECH RECOGN ITION",COLLINS,1986

3新美康永,"音声認識",共立出版.

4.D.H.KLATT,K.N.STEVENS,"STRATEGIES F-OR RECOGNITION OF SPOKEN SENTENCES FR-OM VISUAL EXAMINATION OF SPECTROGRAM", BBN.INC.,CAMBRIDGE.MASS.,BBN REP.,2154 ,JUNE 1971.

5.L.R.RABINER,M.R.SAMBUR,"SOME PRELIM-IRARY EXPERIMENTS IN THE RECOGNITION OF CONNECTED DIGITS,"IEEE,ASSP-24,NO. 2,APR.1976.

6.V.W.ZUE,et al,"A SYSTEM FOR ACOUST-IC-PHONETIC ANALYSIS OF CONTINNOUS SP EECH",IEEE,ASSP-23,NO.1,FEB.1975.

7.J.R.GLASS AND V.W.ZUE,"MULTI-LEVEL ACOUSTIC SEGMENTATION OF CONTINUOUS SPEECH,"PP.429-432,ICASSP 88.

2.R.SCHWRARTZ,J.MAKHOUL,"WHERE THE P-HONEMES ARE:DEALINGWITH AMBIGUITY IN ACOUSTIC-PHONETIC RECOGNITION",IEEE, ASSP-23,NO.1,FEB.1975

9.A.H.GRAY, et al,"DISTANCE MEASURES FOR SPEECH PROCESSING",IEEE,ASSP-24, NO.5.OCT.1976.

10.M.BASSEVILLE,et al,"SEQUENTIAL DET ECTION OF ABRUPT CHAMGES IN SPECTRAL CHARACTERISTICS OF SIGNALS,"IEEE,IT-29,SEP.,1983

11.허 웅, 국어 음운학, 정음사, 1985.