

초성파찰음의 음소분류에 관한 연구

° 김기운 이기영 배철수 최갑석
 명지대학교 전자공학과

A Study on the Phonemic Segmentation of an Initial Affricate

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

ABSTRACT

In this paper, the starting point of affricate is detected from the first predictor coefficient of a 12-pole linear predictive coding (LPC) analysis and phonemic segmentation is done through measuring short time energy and zero crossing rate. By this segmentation method, the duration of an aspirate can be measured in order to detect an aspirate or not.

1. 서론

고도로 정보화된 현대사회에 있어서, 음성인식은 인간과 기계의 정복교환이라는 점에서 많은 연구가 진행되어 왔으며, 자연음성인식에 가까운 연속음성 인식에 관한 연구가 계속되고 있다.[1]

연속음성 인식은 어휘량의 증가로 인하여, 많은 저장 메모리와 계산량의 증가라는 문제때문에 음소 단위 인식이 활발이 이루어 지고 있으며, 이에 따른 음소단위 분류가 필요 하다.[2]

파찰음의 특징을 추출하는 방법에는, 스펙트럼모델을 이용한 방법과 성도의 단면적 함수를 이용한 방법이 있으며, 적응 Lattice-계수의 거리를 이용하여 파열시점을 검출한 방법이 있다.[3][4] 그러나, 이 방법들은 복잡하고 많은 계산량이 필요하게 된다.

본 논문에서는 한국어 자음 14자 중에서 파찰음(ㄱ, ㅋ, ㆁ)의 특징인 파열시점을 검출하기 위하여

12차 LPC-계수들 중에서 첫번째 계수를 이용 하였으며, 음소분류를 위하여 단시간 에너지와 영교차율을 이용하였다. 그리고, 파찰음의 기음구간을 검출하였다.[5]

2. 파찰음(Affricate)의 특징

한국어 파찰음으로는 "ㄱ", "ㅋ", "ㆁ" 3개가 있으며, 무성음으로서 성대보다 위의 발성기관에 의하여 음을 내게된다. 이들은 초성일경우, 발성시 동일한 조음점(경구개)에서 파열 및 마찰되어 소리가 나게 된다. 즉, 조음점부분을 혀로 막은 다음, 숨을 그쳐 이를 급히 터트리면서 발성을 시작하게 되며, 치아에서 강한 마찰이 생겨난다. 그림1 는 이러한 파열부분을 나타낸 것이다.

이들은 무성 파찰음의 대표로 "ㄱ", 을 비롯하여 기(Breath)가 존재하는 유기음 "ㅋ", 그리고 된소리로 조음점에서의 강한 긴장이 발생하는 "ㆁ"으로 나누어 진다.[6]

이상의 파찰음을 모음과 결합하여 발음할때, 경구개와 접하여 있던 혀의 앞부분이 떨어지면서, 모음의 특징인 성대의 진동이 발생하게 된다. 그리고, 성대의 진동이 발생하기 전에, 과도기적인 음의 변화가 나타나기 시작하는데, 이러한 특징은 기음에 의하여 발생하게 된다. 기음의 시점을 그림2 에 나타내었다.[7]

3. 파열시점 검출 및 음소분류

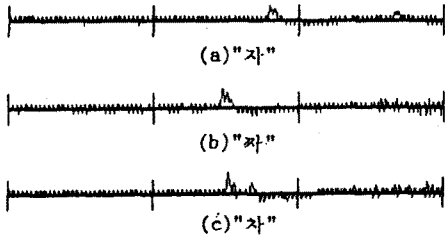


그림1. 파찰음의 파열시점
Fig1. The starting point of affricate

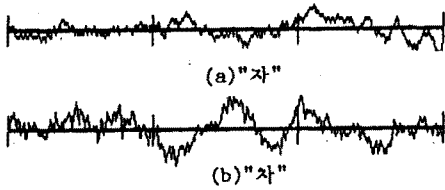


그림2. 기음시점
Fig2. The starting point of breath

(1) 파열시점 검출

본 논문에서는 음성이 파열되는 시점을 구하기 위하여 12차 LPC(Linear Prediction Coding)계수들 중에서 첫번째 계수를 사용 하였다. 무음, 무성음 그리고, 유성음에 대한 3인화자의 첫번째 계수 특징을 그림3 에 나타내었다.

선형예측분석(Linear Prediction-Analysis)이란, 음성신호를 음현에 의하여 방사되는 성도의 출력이라 하고, 성도의 특성을 선형적인 것으로 가정하면, 음성신호를 근사적인 선형임펄스 응답으로 볼수 있으며, 선형미분방정식의 해로부터 음성신호를 직접 구할수 있다. 따라서, 음성신호를 예측하는 것이 가능하며, 성도의 특성을 선형예측계수 $a(i)$ 로 부터 검출할수 있다. 선형예측방정식은 다음과 같이 정의 된다.

$$X(n) = \sum_{k=1}^p a(k)X(n-k) \quad (1)$$

$a(i)$ 를 구하기 위하여, Levinson-Durbin알고리즘을 사용 하였다.[8]

파열시점은 일시적이고, 갑작스런 신호의 변화가 발생하는 부분으로, 첫번째 LPC-계수값은 모음구간보다는 크지만, 무음구간에 비하여 극히 작은 값을 가지고 있다.

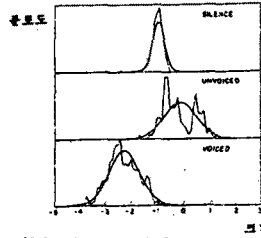


그림3. 1-st 12차 LPC-계수
Fig3. The first coefficient of a 12-pole LPC

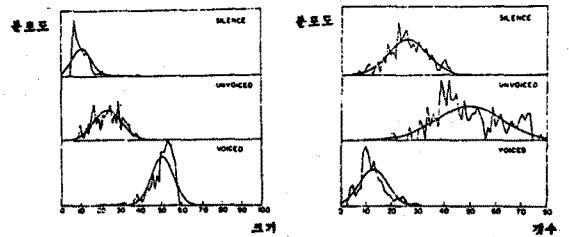
(2) 음소분류

모음의 시작점과 끝점은 단시간 에너지(Short Time Energy)를 사용하여 검출하였으며, 기음구간은 단시간 에너지와 영교차율(Zero Crossing)을 사용하여 검출하였다. 무음, 무성음 그리고 유성음에 대한 두가지 파라미터의 특징을 그림4 에 나타내었다.[9]

i) 단시간 에너지(Short Time Energy)

시간영역에서의 음성신호가 가지는 크기(Amplitude)로, 시간에 대한 진폭의 변동을 잘 나타낸다. 음성신호가 $X(n)$ 이라 하면, 단시간에너지 $E(n)$ 은 $n-N+1$ 에서 n 까지, N 개 데이터의 자승의 합으로 다음 식과 같이 정의된다.

$$E(n) = \sum_{m=n-N+1}^n X(m)^2 \quad (2)$$



(a) 에너지 (b) 영교차율

그림4. 에너지와 영교차율

Fig4. Energy and Zero crossing-rate

ii) 영교차율(Zero-Crossing rate)

음성신호의 부호만을 남겨 1bit로 양자화한 것으로, 스펙트럼의 대체적인 특징을 표현하는데 적합하다. 영교차율 $Z(n)$ 은 다음식 으로 정의 된다.

$$Z(n) = \sum_{m=-\infty}^{\infty} |\text{sgn}(X(m)) - \text{sgn}(X(m-1))| |W(n-m)| \quad (3)$$

식 (3)에서,

$$\text{sgn}(X(n)) = \begin{cases} 1 & ; X(n) \geq 0 \\ -1 & ; X(n) < 0 \end{cases}$$

$$W(n) = \begin{cases} 1/2N & ; 0 \leq n \leq N-1 \\ 0 & ; \text{otherwise} \end{cases}$$
 이 된다.

모음은 무음과 파찰음보다 단시간 에너지값이 큰 구간으로 정하였으며, 기음구간은 파찰음보다 영교차율이 작으며, 모음보다 단시간 에너지가 작은 구간으로 정하였다.

그리고, 음성의 끝점은 파찰음이 파일되는 구간에 비하여, 단시간 에너지가 작은 구간으로 정하였다. 파일시점 검출 및 음소분류를 행한 순서도를 그림5에 나타 내었다.

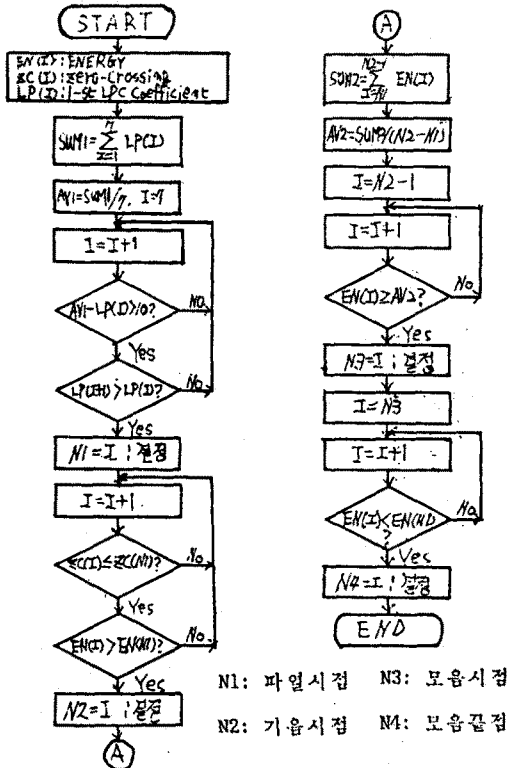


그림5. 파일시점 검출 및 음소분류 순서도
 Fig5. Flow chart of segmentation

4. 실험 및 고찰

(1) 전처리 시스템 및 실험 데이터

본 논문에서 음성을 입력하기 위하여 사용된 전처리 시스템구성도는 그림6 과 같다.



그림6. 전처리 시스템
 Fig6. The preceding system

음성은 마이크로폰을 통하여 입력되고, EQ(Equalizer)AMP를 통하여 -5[V]~+5[V]로 증폭된다. 이 증폭된 신호는 주파수가 100[Hz]~5[KHz]로 제한된 대역 통과 필터를 통과하여, 샘플링 주파수가 10[KHz]이고, 12-Bit의 양자화를 가지는 KSB-8627 A/D CONVERTER 에 의하여 디지털 신호로 변환된다.

대상음성은 20대 남성 3인이 각각 발성한 "자", "자", "자", "저", "저", "조", "조", "조", "주", "주", "주", "지", "지", "지", "즈", "즈", "즈"로서 총 54단어를 사용하였다.

100개의 데이터를 한 프레임(Frame)으로 하여 각 프레임에서의 단시간 에너지와 영교차율 그리고, 12차 LPC-계수를 구하였다.

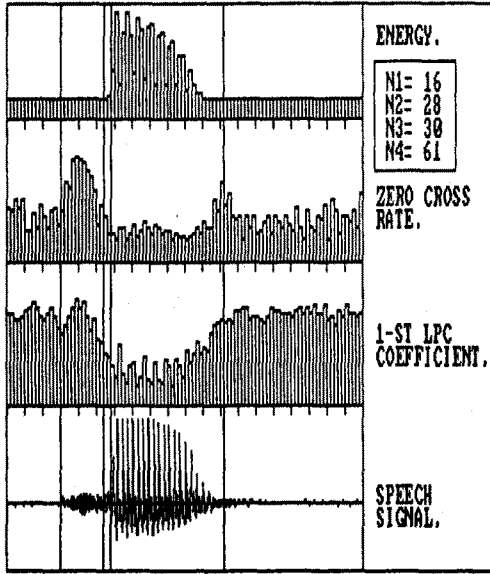
(2) 결과 및 고찰

그림7 에 "자", "자", "자" 에 대한 시점검출 및 음소분류 결과를 나타내었다. 54개의 데이터에 대한 분류결과와 파일 시점에서 아주 작은 값의 LPC-계수값이 나타나 주어 파일시점 검출이 가능하였다. 기음구간은 "ㅅ"이 가장 길며, "ㅈ"은 매우 짧아서, 검출이 되지않았다.

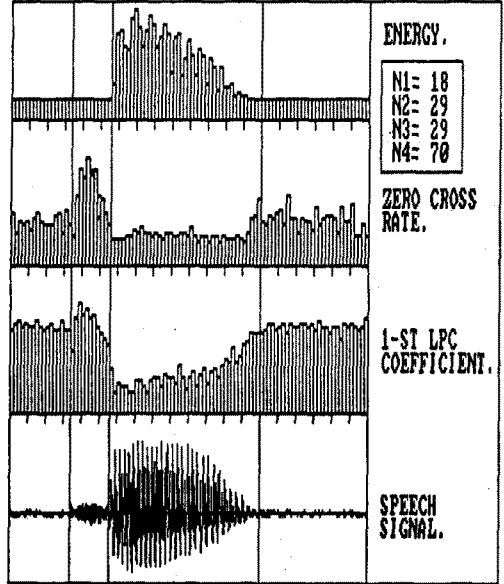
5. 결론

본 논문에서는 12차 LPC-계수들 중에서 첫번째 계수와 단시간 에너지 및 영교차율을 이용하여, 파찰음의 파일시점 검출 및 음소분류를 실험한 결과 다음과 같은 결론을 얻었다.

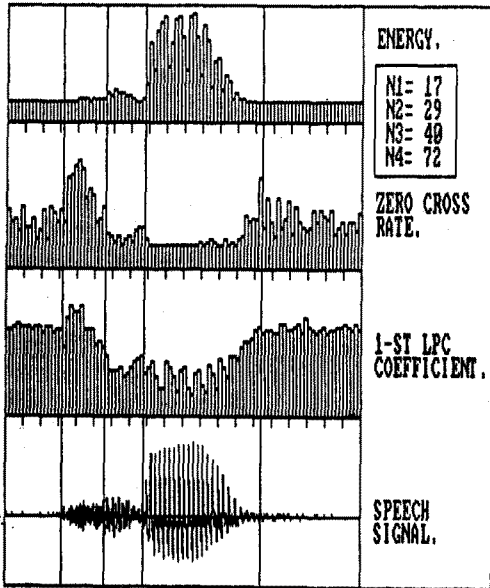
- (1) 첫번째 LPC-계수를 이용하여 3인 화자에 대한 파찰음의 파일부분을 검출할 수 있었다.
- (2) 단시간에너지와 영교차율을 이용하여 음소분류를 행할 수 있었다.
- (3) 유기음과 무기음을 구별할 수 있었으며 기음구간을 검출 할수 있었다.



(a) "자"



(b) "ㅈ"



(c) "ㅊ"

그림 7. 음소분류한 신호

Fig7. Signal of phonemic segmentation

*REFERENCE

1. L.R.Rabiner, S.E. Levinsen, "ISOLATED AND CONNECTED WORD RECOGNITION-THEORY AND SELECTED APPLICATION", IEBE, TRANS. ON COMM. VOL.COM-29, NO.5, MAY. 1981
2. G.Bristow, "ELECTRONIC SPEECH RECOGNITION", COLLINS. 1986

3. 深林計志, 前島正行, "無声破裂音の破裂時点検出の方法", 電子通信学会論文誌, 86/1, Vol. J69-A No.1

4. Hiroyoshi Morikawa, Hiroya Fusisaki . "Adaptive Analysis of Speech Based on a Pole-Zero Representation", IEBE, VOL. ASSP-30, NO.1, FEBRUARY. 1982

5. Bishnu S. Atal, Lawrence R. Rabiner, "A Pattern Recognition Approach to Voiced-Unvoiced-Silence Classification with Applications to Speech Recognition", IEBE, VOL. ASSP-24, NO.3, JUNE. 1976

6. 허웅, "국어 음운학", 정음사. 1985

7. 이현복, "국제 음성문자와 한글 음성 문자"

8. J.D. Markel, A.H. Gray, Jr, "Linear Prediction of Speech"

9. L.R. Rabiner, M.R. Sambur, "An Algorithm for Determining the Endpoints of Isolated Utterances", BSTJ. VOL. 54, NO. 2, FEB. 1979