

# ZCR과 PARCOR 계수를 이용한 숫자음성 인식

○ 김 락 윤      황 영 수      차 일 환  
연세대학교 전자공학과

Spoken digit recognition Using the ZCR and PARCOR Coefficient

Kim Mack-yoon, Whang Young-soo, Cha Il-whan  
Dept. of Electronic Eng. Yonsei University

## 요 약

본 연구는 시간 영역의 parameter를 이용하여 한국어 숫자음(영, 일, 이, 삼, 사, 오, 육, 칠, 팔, 구)을 인식했다. 입력 음성 신호  $X(n)$ 의 Beginning Point와 Ending point를 ZCR(Zero-crossing Rate), Magnitude, Energy, Autocorrelation을 이용 Beginning point와 Ending point를 구하고 자음부의 인식은 위 계수들을 이용하여 행했다. 또, 유성음 부분에서는 PARCOR(Partial Autocorrelation), LPC(Linear Predictive Coding)를 이용 모음부와 유성자음을 인식하여 모음을 6개 부류(ㄱ, ㄴ, ㄷ, ㄹ, ㅁ, ㅂ)로 구분 인식했다. 이 방법에 의하면 입력 음성 신호  $X(n)$ 의 B.P(Beginning Point)와 E.P(Ending Point)를 쉽게 추출 가능하며 또한 각 Parameter를 이용하여 94.4%의 인식율을 얻었다.

## I. 서 론

음성 인식의 기초적 과정은 음성의 물리적 성질, 음향학적 Modeling, 언어학적 해석에 그 근본을 두고 있다. 인간의 음성기관은 호흡기, 후두, 성문상강으로 나눌수 있다. 호흡 행위는 들숨(inspiration)과 날숨(expiration)에 의해 허파에 있는 공기가 배출될때 공기의 진동에 의해 발생된다. 후두는 여린 뼈로 된 통과 같은 것으로 숨통의 상단에 있으며 4개의 연골로 구성되어 있다. 이때 성문의 모양에 따라 유, 무성음이 발생되며, 성대의 진동 속도에 따라 주파수 성분이 달라 짐을 볼 수 있다. 성문상강은 인두, 구강, 비강을 말하는 데 이것에 의해 연구개음, 경구개음, 비음으로 나눈다. 한국어는 우랄 알타이어 족에 속하여 구조상 교착어이므로 모음 조화 현상이 뚜렷하며 자음 14자 모음 10자로 된 음소 문자이다. 이런 음소 문자의 특징을 이용하여 숫자음을 인식하였다. II장에서는 Beginning point와 Ending point를 추출하는 방법을 제시하였고 III장에서는 각 숫자음의 모음 부분을 PARCOR방식으로 인식하는 방법을 서술하였으며 IV장에서는 실험 방법과 결과를 고찰했으며 V장은 인식결과를 비교 검토 했다.

## II. B.P and E.P 검출 방법

### (1) Average Magnitude.

음성 신호  $X(n)$ 의 Magnitude function  $M_n$ 은

$$M_n = \sum_{m=-\infty}^{\infty} |X(n)W(n-m)| \quad \dots\dots(1)$$

이고 Short-time Energy  $E_n$ 은

$$E_n = \sum_{m=-\infty}^{\infty} [X(n)W(n-m)]^2 \quad \dots\dots(2)$$

이며  $W(n)$ 은 Window function이다.

(1)식과 (2)식을 비교하면 Multiplication이 줄으므로 CPU 계산 시간도 단축 되며 신호 Level이 들때 Energy는 Sensitive하므로 이점이 보완된다.

### (2) ZCR (Zero-Crossing Rate)

입력 신호의 주파수가  $f_s$ 의 정현파 일때 ZCR은  $2f_s$ 이다. 이때 평균 ZCR  $Z$ 는

$$Z = 2F_s \cdot \text{Crossings/Sample} \quad \dots\dots(3)$$

어떤 임의점  $n$ 의 ZCR  $Z_n$ 은

$$Z_n = \sum_{m=-\infty}^{\infty} |\text{Sgn}(X(n)) - \text{Sgn}(X(n-1))| \quad \dots\dots(4)$$

이다.

일반적으로 선도 특성에 따라 무성음의 주파수는 유성음보다 크므로 ZCR은 무성음에서 크다.

### (3) Algorithm

입력 음성 신호  $X(n)$ 의 B.P와 E.P를 구하기 위해 다음의 방식을 이용했다. 처음 음성 신호가 어느 Sample까지 Data가 입력되지 않았다고

가정하고 그 구간에서 ZCR과 MAG의 평균 값  $\bar{X}_{ZCR}$ ,  $\bar{X}_{MAG}$ 와 분산 값  $\sigma_{ZCR}$ ,  $\sigma_{MAG}$ 를 구한후 입력 Sample 값을 이동 시키며 MAG'를 구한후 이 값을

$$MAG' > \bar{X}_{MAG} + K\sigma_{MAG} \quad (K: 상수) \quad \dots\dots(5)$$

일때 값을 검출한다. 이 검출된 번지로 부터 역 방향으로 10 Sample 씩 이동하여 ZCR'를 구한 값 이

$$ZCR' < \bar{X}_{ZCR} + K\sigma_{ZCR} \quad \dots\dots(6)$$

인 점을 검출하여 이 점을 Beginning point로 설정 했다. Ending point 검출은 Beginning point를 구한 방법과 유사한 방법으로 했다. 신호의 Peak 점을 찾아 그 점으로 부터 Sample을 택해 이동 하면서

$$MAG' < \bar{X}_{MAG} + K\sigma_{MAG} \quad \dots\dots(5')$$

인 번지를 검출하여 그 번지로 부터 ZCR 값을 10 Sample 씩 순방향으로 이동하여

$$ZCR' < \bar{X}_{ZCR} + K\sigma_{ZCR} \quad \dots\dots(6')$$

이면 그 번지를 택하여 Ending point로 하였다.

### ■. 숫자음 인식 Algorithm

일반적으로 모음은 상대 진동에 의해 생성되는 것으로서 periodic한 신호이다. 이 periodic한 신호를 해석하기 위해 다음과 같은 방식을 도입했다.

#### 1) Auto correlation Function

Discrete-time deterministic 신호의 Autocorrelation Function  $\phi(k)$

$$\phi(k) = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{m=-N}^N X(m) \cdot X(m+k) \quad \dots\dots(7)$$

Autocorrelation 방법은 noise제거가 용이 하며 명백한 pitch를 검출하기 쉽다.

#### (2) PARCOR (Partial Autocorrelation)

Autocorrelation 함수를 정규화한  $P_n$ 은

$$P_n = \frac{(S_i \cdot S_{i-n})}{(S_i^2 \cdot S_{i-n}^2)^{1/2}} \quad \dots\dots(8)$$

PARCOR 계수  $K_n$ 은  $S_i$ 와  $S_{i-n}$ 사이의  $n-1$ 개 Sample 신호 값에서 예측 가능한 부분을 위해 제거한 Autocorrelation 함수로 정의 하므로  $S_i$ 의 예측치  $\hat{S}_i^{(n-1)}$ 는

$$\hat{S}_i^{(n-1)} = - \sum_{j=1}^{n-1} \alpha_j^{(n-1)} \cdot S_{i-j} \quad \dots\dots(9)$$

$$\hat{S}_{i-n}^{(n-1)} = - \sum_{j=1}^{n-1} \beta_j^{(n-1)} \cdot S_{i-j} \quad \dots\dots(10)$$

여기서  $\alpha_j^{(n-1)}$ : 전향 예측 계수

$\beta_j^{(n-1)}$ : 후향 예측 계수

예측값의 실제값의 차를  $\begin{matrix} (n-1) & (n-1) \\ f_i & g_i \\ (n-1) & (n-1) \end{matrix}$ 이라 할때 PARCOR 계수  $K_n$ 은  $f_i^{(n-1)}$ ,  $g_i^{(n-1)}$ 의 정규화 Autocorrelation 함수이므로 다음식과 같다.

$$K_n = \frac{f_i^{(n-1)} \cdot g_i^{(n-1)}}{[(f_i^{(n-1)})^2 \cdot (g_i^{(n-1)})^2]^{1/2}} \quad \dots\dots(11)$$

#### (3) 인식 방법

숫자음의 음소 배열에서 모음부 인식은 PARCOR 계수로서 모음 (ㅏ, ㅑ, ㅓ, ㅕ, ㅗ, ㅛ, ㅜ)을 인식 분리 했다. 자음부는 유성 지음과 무성 자음으로 분리해 PARCOR 계수의 ZCR, MAG로서 시작부의 256 Sample을 택해 인식을 시도 했다. 그 인식 Algorithm flow-chart는 <그림 1>에 제시 하였다.

#### IV. 실험 방법 및 결과

인식 대상은 학습 받지 않은 성인 남성 3명이 숫자음 "영, 일, 이.....구"를 3회 반복 발음하여 녹음 분석 하였다. 이 녹음된 음성 신호는 차단 주파수 6 KHz인 LPF를 거쳐 각 음의 Data를 받았다. 이때 Sampling frequency는 20 KHz이다. 기록된 Data들은 II장에서 제시한 방법으로 Beginning point와 Ending point를 잡은 과정을 <그림 2>에 나타내었다.

또한 숫자음 인식은 II장에 서술한 방법으로 인식 하였다. 이때 Reference pattern은 화자에서 구한 parameter 평균 값에서 잡았다. 인식 방법은 표준 pattern과 입력 pattern사이의 유사도를 결정하기 위해 Euclidean distance를 사용 했다. 표준 pattern을 R, 입력 pattern을 K로 하면 distance D는

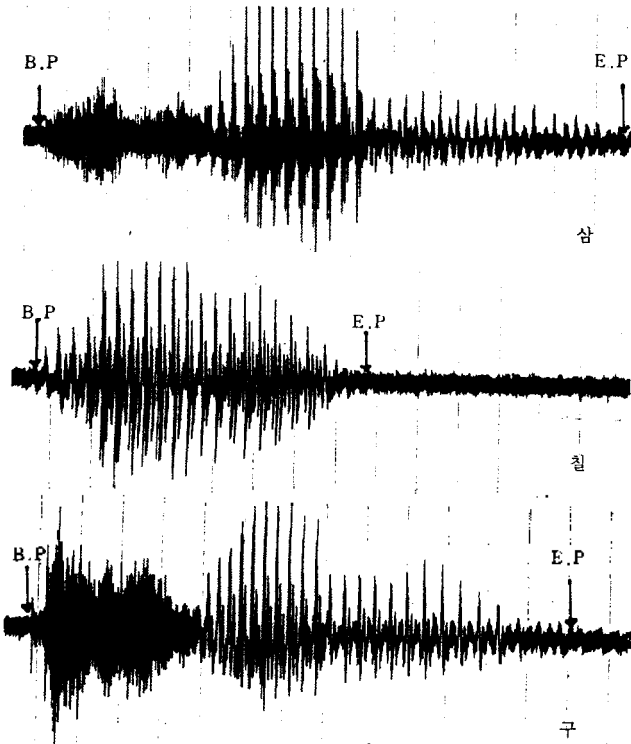
$$D = \sum_{i=1}^P |R_i - K_i| \quad \dots\dots(12)$$

로 되고 이때 D가 최소로 되는 pattern을 입력 음성으로 간주하여 인식 하였다. 이때의 실험 결과는 <표 1>과 같다.

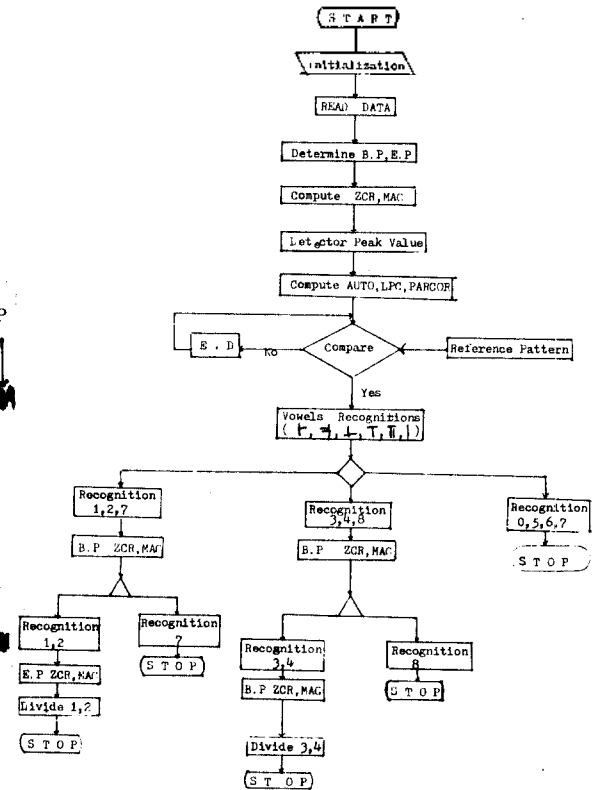
#### V. 결 론

입력 음성 신호 X(n)의 Beginning point와 Ending point 추출 Algorithm을 토대로 시작점과 끝점의 검출은 비교적 쉽게 이루어 졌다. 하지만 각 Threshold level의 차이로 약간의 오차가 생겼으리라 여겨 진다. 숫자음성 인식 Algorithm을 사용하여 각 음들을 인식한 결과 94.4%의 인식율을 얻었다. 이것은 같은 모음 부류 "일, 이, 칠" "삼, 사, 팔"에서 인식 Error가 발생되었는데 이것은 각 음의 끝인 유성자음 인식시 생기는 것으

으로 추측 된다. 음성자음의 인식을 다른 방법으로 인식하면 인식율은 더욱 높아지리라 여겨진다. 앞으로는 숫자음 인식을 분리된 음이 아닌 연속음을 인식하여 실생활에 많은 이용이 되리라 생각한다. 또한 Hardware의 개선으로 인식이 Real Time의 수행이 사료된다.



<그림 2> 숫자음 (삼, 칠, 구)의 B.P와 E.P



<그림 1> 숫자음 인식 흐름도

	영	일	이	삼	사	오	육	칠	팔	구
영	9									
일		8	1							
이			1	8						
삼				8	1				1	
사					8					
오						9				
육							9			
칠								9		
팔									8	
구										9

<표 1> 실험 결과

참 고 문 헌

1. J.M.Baker, "A New Time-domain Analysis and other complex wave forms", Ph.D pillsburgh, 1975.
2. L.R.Rabinar and M.R.Sambur, "An Algorithm for Determining the Endpoints of Isolated Mtterances" Bell syst.Tech.J. Vol154 No.2 February,1975.
3. Nolan, "The phonetic bases of speaker recognition", Cambridge University Press, 1983.
4. Denes, P. & Mathews, M.V. "Spoken digit recognition using tins-freguency pattern Matching", JASA Vol 32, 1960.
5. Davis, K.H.Biddulph. "Automatic recognition of spoken digit", JASA Vol 24, 1980.
6. LR.Rabiner/R.W Schater, "Digital processing of speech signals",prentice-Hall, 1978
7. B.S.Atal, "Effectiveness of linear prediction characteristics of speech wave for automatic speaker identification and veritication JASA, Vol55, 1974.
8. I. Kameny, "Automatic acoustic-phonetic snalysis of vowels and sonorants", In proc IEEE ASSP, Apr. 1976.
9. 차일환, '음향 공학 개론', 한신문화사.