한 국 어 단 모 음 자 동 인 식 에 관 한 연 구

안 동순, 이 창섭, 차 일환
*연세대학교, 전자공학과

A study on the automatic recognition of Korean vowel

○Dong-Soon Ann, Chang-Sub Lee, Il-Whan Cha
*Dept of Electronic Engineering Yonsei Univ

Abstract

In this study, the system is proposed which can be used for recognition of Korean single vowels "ㅏ,ㅓ,ㅗ,ㅜ,ㅡ,ㅣ,ㅐ,ㅔ,ㅚ,", and automatic recognition is processed using μ-computer. 3 men of not-being-studied are participated in this experiment.

Using the period of vowels, one part of the steady state is selected for high speed recognition, and amplitude comparison method, LPC, PARCOR, and Formant are used for parameter of recognition. Formant is obtained by peak picking method using LPC, and then vowels are recognized by amplitude comparison method, LPC, PARCOR, and Formant.

As a result, Recognition rates are 90.1% for amplitude comparison method, 93.1% for LPC, 100% for PARCOR, 88.8% for using formant.

1. 서 론

음성에 대한 연구는 음성 분석, 합성, 화자조회, 화자 식별, 디지탈 전송, 음성 인식 등으로 진행되고 있다. 음성 인식은 인간에서 기계로의 정보 입력을 음성으로 행하는 것을 가능하게 하는 기술이다. [1] Computer 처리 성능의 세속적인 발전에 비하여 정보의 입력 수단은 발전되지 않고 있기 때문에 음성 인식은 man-machine communication으로 매우 중요한 과제이다.

본 연구에서는 정상 상태의 한 주기를 추출하여 amplitude comparison method와 선형 예측계수(LPC), 편자기 상관 함수(PARCOR), Formant를 인식 parameter로 이용하여 한국어 단모음에 대해서 자동 인식을 수행하고 그 결과를 비교하였다.

2. 본 론

음성 특징 parameter의 추출법으로 Fourier변환을 기초로 하는 주파수 영역과 자기 상관 함수를 기초로 하는 시간 영역의 parameter를 크게 나눌 수 있다. 본 연구에서는 4가지 parameter을 이용하였다.

2-1. 주기 추출에 의한 magnitude 비교

모음은 주기적이기 때문에 주기성을 이용하는 것이 처리량을 감소시킴으로 매우 효과적이다.

주기 추출을 위해 처리 시간이 매우 짧은 Envelope Detection method을 사용하였다. 각 파형의 주기 사이에는 진폭차와 시간차가 있음으로 진폭의 정규화와 linear time warping의 preprocessing을 한 후 magnitude을 비교하였다.

2-2. 선형 예측(Linear Prediction) 계수

현재의 음성 신호 S(nT)을 예측하는 경우 과거의 sample의 선형 결합으로 추정할 수 있다.

$$\hat{S}_n(nT) = -(a_1 S_{n-1} + a_2 S_{n-2} + \cdots \quad a_p S_{n-p}) \qquad \cdots\cdots (2\text{-}1)$$

이런 선형 결합에 의해 S_n을 예측하는 방법을 선형 예측법이라 한다.

실제 음성 신호 값 S_n과 예측한 값 $\hat{S}_n$사이의 오차는 다음 식과 같다.

$$e_n = S_n + \sum_{k=1}^{p} a_k S_{n-k} \qquad \cdots\cdots (2\text{-}2)$$

계수 a_i를 구하기 위해서 오차 신호 e_n을 최소로 하면 된다. 즉, 어느 구간의 오차 신호 $e(n)$의 자승의 합 ε^2을 구하면 다음 식과 같다.

$$\varepsilon^2 = \sum_n e_n^2 = \sum_n \left(S_n + \sum_{k=1}^{p} a_k S_{n-k} \right)^2 \qquad \cdots\cdots (2\text{-}3)$$

오차 ε^2을 최소로 하기 위해 ε^2을 a_i에 대해 편미분해서 그 값이 0으로 하는 조건을 이용하면 된다.

$$\frac{\delta \varepsilon^2}{\delta a_i} = 0 \qquad (1 \leqq i \leqq p) \qquad \cdots\cdots\cdots (2\text{-}4)$$

(2-4) 식에서 아래식이 유도된다.

$$\sum_{k=1}^{p} a_k \sum_n S_{n-k} S_{n-i} = -\sum_n S_n S_{n-i} \qquad 1 \leqq i \leqq p$$
$$\cdots\cdots\cdots (2\text{-}5)$$

$\sum_n S_{n-k} S_{n-i}$의 값을 계산하는 여러 가지 방법이 제안되었는데 본 연구에서는 correlation 함수를 이용하였다.

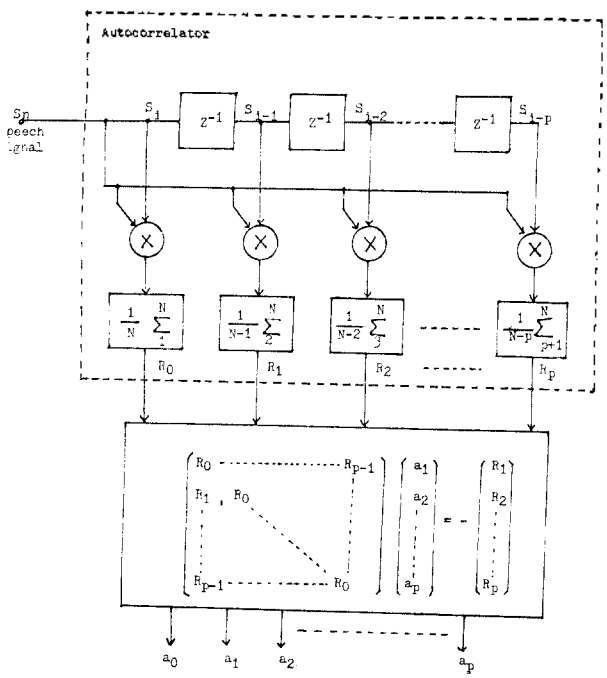
유한개의 data (N개)의 short time auto correlation은 다음 식과 같다.

$$R_i = \frac{1}{N+1-i} \sum_{n=i}^{N} S_n S_{n-i} \qquad \cdots\cdots\cdots (2\text{-}6)$$

(2-6) 식을 이용하여 (2-5) 식을 matrix로 표시하면 다음 식과 같다.

$$\begin{bmatrix} R_0 & R_1 & R_2 & \cdots\cdots & R_{P-1} \\ R_1 & R_0 & R_1 & \cdots\cdots & R_{P-2} \\ \vdots & & & & \\ R_{P-1} & R_{P-2} & R_{P-3} & \cdots\cdots & R_0 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_P \end{bmatrix} = - \begin{bmatrix} R_1 \\ R_2 \\ \\ R_P \end{bmatrix}$$
$$\cdots\cdots\cdots (2\text{-}7)$$

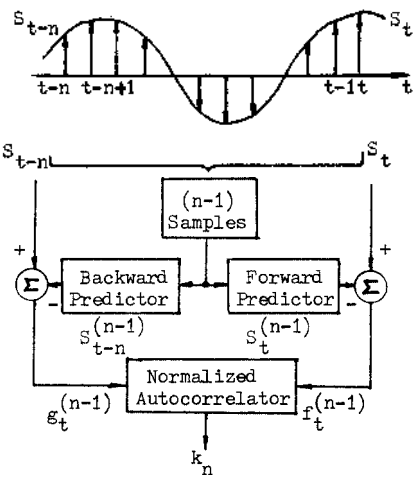
LPC 알고리즘은 그림 1 과 같다.



〈그림 1〉 LPC Algorithm.

2-3. 편자기 상관 (PARCOR) 계수

PARCOR 계수 K_n은 S_t와 S_{t-n} 사이의 (n-1)개의 sample 신호값에서 예측 가능한 부분을 취해서 제거한 것 즉, 오차 신호의 자기 상관 함수로서 정의된다.



〈그림 2〉 PARCOR coefficient.

그림 2에서 S_t와 S_{t-n} 사이의 (n-1)개의 sample에서 선형으로 S_t와 S_{t-n} 값을 예측하면 다음 식과 같다.

$$\hat{S}_t^{(n-1)} = -\sum_{i=1}^{n-1} a_i^{(n-1)} \cdot S_{t-i}$$

$$\hat{S}_{t-n}^{(n-1)} = -\sum_{i=1}^{n-1} B_i^{(n-1)} \cdot S_{t-i} \qquad \cdots\cdots\cdots (2\text{-}8)$$

여기서 $a_i^{(n-1)}$은 전향 예측 계수

$B_i^{(n-1)}$은 후향 예측 계수이다.

예측 값과 실제값 사이의 오차를 $f_t^{(n-1)}$, $g_t^{(n-1)}$로 하면 다음 식과 같다.

$$f_t^{(n-1)} = S_t - \hat{S}_t^{(n-1)}$$
$$g_t^{(n-1)} = S_{t-n} - \hat{S}_{t-n}^{(n-1)} \qquad \cdots\cdots\cdots (2\text{-}9)$$

PARCOR 계수 K_n은 (2-8) 식의 오차 신호의 정규화 자기 상관 함수로서 정의된다.

$$K_n = \frac{\langle f_t^{(n-1)} \cdot g_t^{(n-1)} \rangle}{[\langle |f_t^{(n-1)}|^2 \rangle \langle |g_t^{(n-1)}|^2 \rangle]^{\frac{1}{2}}} \qquad \cdots\cdots (2\text{-}10)$$

윗 식의 분자 $\langle f_t^{(n-1)} \cdot g_t^{(n-1)} \rangle = f_t^{(n-1)} \cdot S_{t-n}$

$$= R(n) \times \sum_{i=1}^{n-1} a_i^{(n-1)} R(n-i) \qquad \cdots\cdots (2\text{-}11)$$

그러므로 PARCOR 계수 K_n은 다음 식과 같다.

$$K_n = \frac{R(n) + \sum_{i=1}^{n-1} a_i^{(n-1)} R(n-1)}{[\langle |f_t^{(n-1)}|^2 \rangle \langle |g_t^{(n-1)}|^2 \rangle]^{\frac{1}{2}}} \qquad \cdots\cdots (2\text{-}12)$$

2-4. 선형 예측법에 의한 Formant

LPC을 이용한 Formant의 추출 방법에는 Root Solving법과 Peak picking법 2가지가 있으나 본

연구에서는 peak picking법을 이용하였다. 예측 계수를 고속 퓨리에 변환(FFT)시켜 이산적 power spectrum 값을 구한 다음, 어느 주파수 간격 f로서 구한 spectrum 값을 순차적으로 비교해서 Local peak의 주파수 mf를 구한다. 이때 주파수(m-1) f, mf, (m+1)에 대한 power spectrum 값을 p(m-1), p(m), p(m+1)로 하면 이 3점에서 2차식 $a\lambda^2+b\lambda+C$로 근사해서 정확한 중심 주파수 Fi와 대역폭 Bi을 구한다. 간단히 하기 위해서 $-f$에서의 spectrum 값을 p(m-1), 0에서 spectrum 값을 p(m), $+f$에서의 값을 p(m+1)로 하면,

$$p(m-1) = af^2-bf+c$$
$$p(m) = C$$
$$p(m+1) = af^2+bf+c \quad \cdots\cdots (2\text{-}13)$$

극대치는 $\frac{d}{d\lambda}(a\lambda^2+b\lambda+c)=0$에서

$$\lambda_p = \frac{-b}{2a} \quad \cdots\cdots (2\text{-}14)$$

로 되고 중심 주파수 Fi는

$$Fi = \lambda_p f+mf$$
$$= \left(\frac{-b}{2a}+m\right) f \text{ 이다.} \quad \cdots\cdots (2\text{-}15)$$

이때 peak 값의 power spectrum P_P는

$$P_P = a\lambda_p^2+b\lambda_p+c$$
$$= \frac{b^2}{4a}-\frac{b^2}{2a}+c$$
$$= -\frac{b^2}{4a}+c \quad \cdots\cdots (2\text{-}16)$$

3. 실험 방법 및 결과

인식 대상어로 학습 받지 않은 연령이 25세, 28세, 32세인 성인 남자 3사람이 한국어 단모음(ㅏ, ㅓ, ㅗ, ㅜ, ㅡ, ㅣ, ㅐ, ㅔ, ㅚ)을 3번씩 반복 발음하여 녹음기에 녹음하였다.

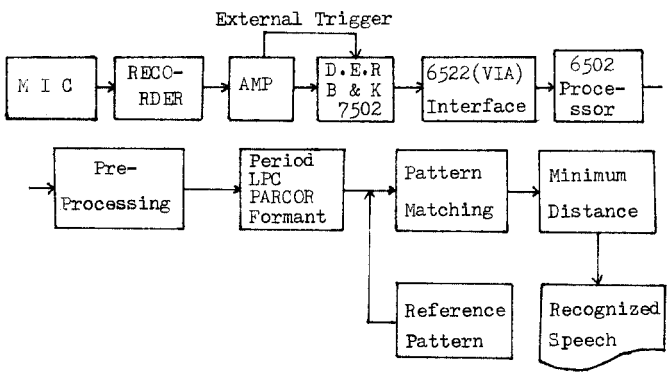


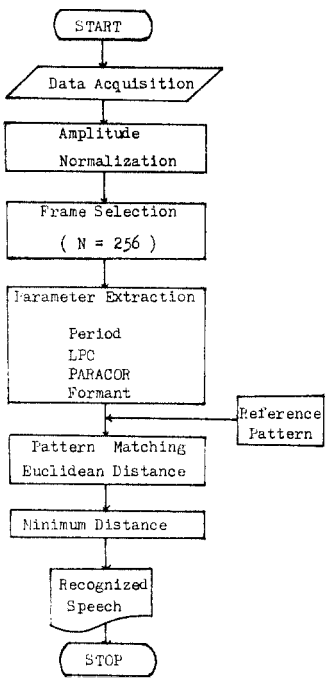
Fig. 3 Experiment Block Diagram for Speech Recognition

인식을 위한 system 블럭도는 〈그림 3〉과 같다.

녹음한 음성은 차단 주파수가 4KHz인 LPF(low-pass filter)을 거쳐 디지탈 이벤트 레코더(D.E.R)안의 A/D 컨버터에 동기를 주기 위해 2(V)의 external trigger을 준 다음, 8bit A/D 컨버터의 sampling 주파수를 10KHz로 하여 sampling하였다. 디지탈 이벤트 레코더와 6502 processor와의 Bus-Interface을 위해 PIO와 CTC 기능을 가진 VIA(6522)로서 Interface을 구성하여 Data을 수집하였다. 모음은 유성음으로 주기성이 있다.

그러므로 과도음 부분을 제거하고 정상음 상태의 한 부분만을 처리하면 대폭적인 수행 시간의 단축을 가져올 수 있다.

인식 처리를 위한 parameter로 정상 부분에서 Envelope detection method로 한 주기를 추출하였고, 1 frame의 길이를 25.6 ms로 취해서 Derbin 알고리즘을 이용하여 15차의 LPC와 PARCOR 계수와 Peak-picking법으로 Formant을 추출하였다.



〈그림 4〉 The flowchart of Recognition.

화자 종속적인 방법으로 음성 인식 실험을 하였음으로 표준 pattern은 각 화자에 대해서 구한 parameter에서 평균 값을 잡았다.

인식 방법으로는 표준 pattern과 입력 pattern사이의 유사도를 결정하기 위해서 Euclidean distance 방법을 이용하였다. 표준 패턴을 R, 입력 패턴을 Ⅱ로 하면 Euclidean distance는,

$$| R - \text{Ⅱ} | = \sum_{i=1}^{p} | R_i - 1_i | \quad \cdots\cdots\cdots\cdots (3\text{-}1)$$

로 되고 이때 거리가 최소로 되는 표준 pattern을 입력 음성으로 간주하여 인식하였다. 인식에 관한 흐름도는 〈그림 4〉와 같다. 각 알고리즘에서 얻은 인식 실험 결과는 표 1, 표 2, 표 3, 표 4와 같다.

〈표 1〉 Experiment Result using Pitch comparison.

입력 \ 출력	아	어	오	우	으	이	애	에	외
아	9								
어		8						1	
오			8						1
우				8		1			
으					8	1			
이						9			
애							9		
에			1	1		1		6	
외								1	8

Correct Rate : 90.1 % (73/81)

〈표 2〉 Experiment Result using LPC

입력 \ 출력	아	어	오	우	으	이	애	에	외
아	9								
어		9							
오			8	1					
우			2	7					
으				1	8				
이						9			
애							9		
에								9	
외								1	8

Correct Rate : 93.8 % (76/81)

〈표 3〉 Experiment Result using PARCOR.

입력 \ 출력	아	어	오	우	으	이	애	에	외
아	9								
어		9							
오			9						
우				9					
으					9				
이						9			
애							9		
에								9	
외									9

Correct Rate : 100 % (81/81)

〈표 4〉 Experiment Result using Formant.

입력 \ 출력	아	어	오	우	으	이	애	에	외
아	9								
어		8	1						
오			7		1			1	
우				9					
으					9				
이		1				8			
애							7	1	1
에								9	
외				1		2			6

Correct Rate : 88.8 % (72/81)

4. 결 론

한국어 단모음 9개에 대해서 3인 화자에 대해 자동 인식을 해본 결과, 인식 수행에 걸린 시간이 추기 추출에 의한 amplitude 비교법이 3분 25초, LPC와 PARCOR 계수를 이용할 때 2분 38초, formant을 이용한 인식 수행 시간은 5분이 걸렸다. 그리고, 인식율은 각각 90.1 %, 93.8 %, 100 %, 88.8 %를 얻었다.

본 연구에서는 한국어 단모음 인식에서 가장 적절한 parameter로 PARCOR 계수가 좋다는 것을 알았다.

앞으로 음성 인식의 연구 방향은 화자에 무관하며, 보다 일반적인 단어나 문장의 인식, 이해와 이를 뒷받침하는 hardware의 개선으로 Real time의 수행이 요구된다.

참 고 문 헌

1. 大泉充郎 外지 "音聲認識合成" JAPAN Industry Engineering Center 昭和 55年.

2. L.R.Rabiner / R.W.Schafer, "Digital Processing of Speech Signals" Prentice-Hall 1978.

3. 安居院猛·中嶋正之 "ユソピュータ音聲處理" 電子科學シリーズ" 昭和 55年.

4. J.D.Markel, A.H.Gray, Jr, "Linear Prediction of Speech" Springer-Verlag 1976.

5. B.S.Atal / S.L.Hanauer, "Speech analysis and synthesis by Linear Prediction of the speech wave" J.Acou. Soc.Am. Vol 50, pp.637-655, 1971.