

추정평면에서 평가한 데이터와 인공신경망에 의한 숫자음 인식

A Numerical Speech Recognition by Parameters Estimated from the Data on the Estimated Plane and a Neural Network

최 일 홍*, 장 승 관**, 차 태 호**, 최 응 세**, 김 창 석***

(Il Hong Choi*, Seung Kwan Jang**, Tae Hoo Cha**, Ung Se Choi**, Chang Seok Kim***)

요 약

본 논문은 추정평면의 데이터로부터 특징파라미터의 평가와 인공신경망에 의한 음성인식방법을 제안한다. 각 프레임에서 평가한 LPC는 매핑함수를 이용하여 추정평면으로 매핑시켰으며, 본 논문에서는 이 추정평면의 데이터로부터 C-LPC, 최대값, 최소값, 3등분할 파워 특징값을 평가하였다. 추정평면에서 평가한 특징 파라미터는 인공신경망에 입력한 음성인식 실험으로부터 원 음성신호의 시간변화에 따른 특징을 포함하고 있음을 확인하였고, 제안한 방법에 의한 인식으로부터 인식율이 약 96.3%이었다.

ABSTRACT

This paper was proposed the recognition method by using parameters which was estimated from the data on the estimated plane and a neural network.

After the LPC estimated in each frame algorithm was mapped to the estimated plane by the optimum feature mapping function, we estimated the C-LPC and the maximum and minimum value and 3 divided power from the mapping data on the estimated plane.

As a result of the experiment of the speech recognition that those parameters were applied to the input of a neural network, it was found that those parameters estimated from the estimated plane have the features of the original speech for a change in the time scale and that the recognition rate by the proposed methods was 96.3 percent.

I. 서 론

음성인식은 인간과 기계와의 통신에서 필수적인 요건으로 되었으며, 최근 컴퓨터의 급속한 보급에 따라 산업이나 사회전반에 걸친 산업의 고도화로 인하여 자동화나 정보화기술이 급속히 발전되면서 가정이나 사회전반에서 인간생활의 편리성을 도모하여 왔다.

최근의 음성인식 기술은 발생시 화자의 발생속도, 감정 및 특징 등이 달라져 시간변동, 주파수변동 및 조음결합 등의 현상으로 동일한 단어를 여러번 발음해도 똑같은 특징을 구할 수 없기 때문에 이러한 문제점을 개선하

고 보완하기 위한 기술의 개발이 과제로서 되어 있다.[1]

일반적으로 음성인식은 특징을 추출하는 전처리과정, 추출된 특징을 압축하는 중간처리과정, 압축된 특징을 이용하여 인식하는 후처리 과정으로 이루어지며 중간처리과정인 압축은 벡터양자화(V.Q)를 많이 사용하고 있다.[2][3] 그러나 벡터양자화 이론은 처리시간이 길고 시간변화에 대한 특징을 분석할 수 없다는 단점이 있다. 이러한 단점을 해결하면서 데이터를 압축 추출할 수 있는 알고리즘이 OFEC(optimum feature extraction & compression) 알고리즘이다.[10]

본 논문에서는 OFEC 알고리즘에 의해서 추출한 C-LPC(Compressed-LPC) 데이터 만으로는 효율적인 음성인식을 수행할 수가 없어서 최적특징 매핑함수에 의해서 매핑된 추정평면의 데이터로부터 최대값 및 최소값의 평가와 3등분할에 의한 LPC파워값을 평가하여 음성신호의 시간변화에 따른 특징분석 및 효율적인 인식을 하고자 한다.

본 논문에서 인식은 인공신경망을 이용하였으며, 인공신경망의 입력은 16차의 C-LPC계수와 추정평면에서 추

*진주전문대학 전자과
Department of Electronic Engineering Chin Ju Technical College

**생산기술연구원 부설 산업기술교육센터
Industrial Technology Training Center, KAITECH

***명지대학교 전자공학과
Department of Electronic Engineering, Myong-Ji University
접수일자:1996년 5월 13일

정된 최대값, 최소값 및 파워값을 정규화시켜서 사용한다. 한국어 숫자음 신호에 대한 실험을 통하여 제안한 방법에 의한 인식률이 제안한 방법을 적용하지 않은 경우보다 인식률이 우수함을 고찰하고자 한다.

II. OFEC 알고리즘

일반적으로 음성인식의 처리과정은 음성특징을 추출하는 전처리과정, 추출된 특징을 압축하는 중간처리과정, 압축된 데이터를 이용하여 인식하는 후처리과정으로 이루어진다. 중간처리과정은 일반적으로 V.Q를 사용하여 압축을 수행하나, 처리시간이 길고 시간변화에 대한 평가를 할 수 없다는 단점이 있다.[4][6]

본 논문에서 사용한 OFEC 알고리즘은 전처리과정에서 추출된 특징파라미터를 압축하면서 시간변화에 대한 특징까지도 평가할 수 있는 알고리즘이다.[8][10]

1. 최적특징매핑함수

일반적으로 성도특성을 구하기 위해 발생기관인 성도를 p 차의 AR모델로 모델링하고 최적의 AR모델 계수들을 구하게 되는데, 이것을 음성신호 $s(n)$ 에 적용시키면,

$$s(n) = \sum_{i=1}^p a_i s(n-i) + w(n) \quad (2.1)$$

이 되고, 여기서 $w(n)$ 은 음성을 생성시키는 신호로서, 백색잡음이나 펄스열로 가정한다.

음성신호 $s(n)$ 을 프레임단위로 분할하여 q 번째 프레임의 음성신호 $s(n)$ 에 식(2.1)을 적용한 후 평가하면

$$\delta_q(n) = \sum_{i=1}^p \hat{a}_{iq} \delta_q(n-i) \quad (2.2)$$

를 얻을 수 있으며, 여기서 $\delta_q(n)$ 은 음성신호 $s(n)$ 의 q 번째 프레임에서 평가한 신호이고, \hat{a}_{iq} 는 p 개로 추정된 계수로서 q 번째 프레임의 음성신호의 특징을 나타내는 LPC계수이다.

따라서 q 번째 프레임의 특징함수를 $F[\delta_q(n)]$ 이라고 정의하면,

$$\begin{aligned} F[\delta_q(n)] &\triangleq \hat{a}_{1q} \delta_q(n-1) + \hat{a}_{2q} \delta_q(n-2) + \dots + \hat{a}_{pq} \delta_q(n-p) \\ &= \sum_{i=1}^p \hat{a}_{iq} \delta_q(n-i) \end{aligned} \quad (2.3)$$

로 표현되고, 여기서 $F[\delta_q(n)]$ 는 추정평면의 함수이다.[10]

그림 2.1은 식(2.3)의 함수를 이용하여 얻어진 추정평면의 데이터이다. 즉, q 번째 프레임의 음성신호가 p 개의 예측계수들로 추정하여 추정평면으로 매핑된 결과로서, 추정평면의 p 개의 이산신호 $\hat{a}_{p1}, \dots, \hat{a}_{1q}$ 는 q 번째 프레임의 음성신호를 대표하는 추정평면의 신호로 볼 수 있으며, 이 추정평면의 이산신호를 추정하면 q 번째 프레임의 원

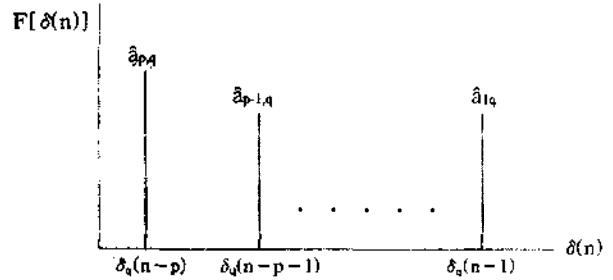


그림 2.1 추정평면의 이산데이터

Fig. 2.1 Discrete data on the estimated plane.

래의 음성신호를 분석할 수 있다.

1번째 프레임부터 m 번째 프레임까지 음성신호에 대하여 AR모델로 평가된 시간과 프레임의 추정순서를 고려하여 추정평면으로 매핑시킬 수 있도록 일반화한 식은 다음과 같다.

$$F[\delta(n)] = \sum_{i=1}^m \sum_{j=1}^p \hat{a}_{ij} \cdot \delta[(n-i) \times i - j] \quad (2.4)$$

식 (2.4)에서 j 는 AR모델의 차수이고, i 는 추정프레임의 순서를 나타내는 것으로 식 (2.4)는 음성신호의 특징을 추정하기 위하여 각 프레임에서 평가한 LPC 계수를 프레임 및 AR모델의 평가순서를 고려하여 매핑평면으로 매핑시키는 최적특징매핑함수이다.[10]

즉 전처리과정에서 얻어진 프레임단위의 LPC계수는 전체 음성신호를 대표하는 특징벡터들로서 이 특징벡터들을 평가순서를 고려하여 추정평면으로 다시 배치하면 화자의 성도특성을 나타내는 또 다른 특성데이터가 된다. 이와같이 LPC계수를 평가된 시간을 고려하여 매핑평면으로 매핑시키는 함수를 최적특징매핑함수라고 하며 이 함수에 의하여 매핑된 데이터를 평가하면 원래 음성신호의 특징을 추정할 수 있다.

2. 특징파라미터 압축

그림 2.2는 음성의 특징파라미터를 압축하기 위한 알고리즘의 흐름도로서, 추정평면 데이터에 자기상관함수를 적용하여 추정평면상의 파워를 분석하면 시계열상의 음성신호의 상대파워를 분석할 수 있고, 추정평면 데이터의 주기관계를 분석하면 프레임단위의 주기형태도 분석할 수 있다. 즉 추정평면 데이터를 분석하면 원래의 음성 고유특성을 추출할 수 있으며, 추정평면 데이터에 q 차의 AR모델을 재설정하여 추정하면, 전체 음성신호의 특징을 나타내는 q 개의 특징파라미터들로 압축된다. 이와 같이 시간영역상에서 p 개의 AR모델의 계수를 추출하여 추출된 LPC 계수를 추정평면으로 매핑시키고, 이 추정평면 데이터에 다시 q 차의 AR모델을 설정하여 계수를 추정함으로써 원래의 음성신호에 근접하는 음성신호의 압

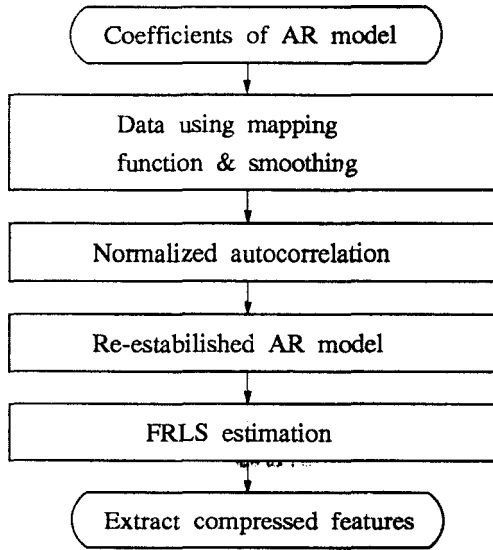


그림 2.2 음성의 특징파라미터 압축 알고리즘 흐름도
Fig 2.2 Flow chart of the compression algorithm of feature parameters of speech.

축된 특징파라미터인 C-LPC(compressed-LPC)를 추정할 수 있다.

특정파라미터 압축의 알고리즘은 다음의 4단계로 처리된다.[10]

1) 첫번째 단계에서는 전처리단계에서 프레임구간별로 추정된 LPC계수를 최적특징매핑함수에 적용하여 추정평면으로 매핑시키는 단계로서, 이렇게 매핑된 프레임구간별 첫번째 데이터는 두번째 이상의 데이터에 비해서 값이 크기 때문에 파라미터 압축시 정확한 특징압축이 이루어질 수 없다.

따라서 첫번째 데이터가 미치는 영향을 줄이기 위하여 식 (2.5)와 같은 평활화함수를 적용하여 프레임의 데이터를 평활화시켰다.

$$F_S[\delta(n)]_{\delta(n)=\text{first_data}} = \frac{F[\delta(n)]}{M_q} \quad (2.5)$$

여기서

$$M_q = \frac{1}{p} \sum_{j=1}^p a_{jq} \quad (2.6)$$

이며, M_q 는 q 번째 프레임에서 선형예측계수의 평균값이고 p 는 AR모델의 차수이다.

2) 두번째 단계는 전처리과정에서 주파수변동을 흡수하기 위하여 적용한 정규화자기상관함수를 추정평면 데이터에 적용하여 추정평면에서 화자의 주파수변동을 흡수하고, 에너지를 균등하게 처리하였다.[11]

3) 세번째 단계는 정규화자기상관함수에 의해 처리된 추정평면의 데이터로부터 p 개의 특징을 평가하기 위하여 전처리과정에서 사용한 p 차의 AR모델을 재설정하였다.

4) 네번째 단계는 C-LPC계수를 추정하는 단계로서 세번째 단계에서 설정된 AR모델의 계수를 추정하기 위하여 선형예측알고리즘인 FRLS를 이용하였다.[7][9]

III. 추정평면에서 최대·최소값 및 파워값 추정

프레임단위로 평가된 LPC계수를 최적특징매핑함수를 이용하여 추정평면으로 매핑시키면 이 추정평면의 데이터들은 시간변화에 대한 원 음성신호의 특징을 포함하고 있어서 이 추정평면의 데이터를 분석함으로써 원 음성신호의 특징을 분석할 수 있고, 분석된 결과를 이용하면 보다 효율적인 인식을 수행할 수 있다.

1. 추정평면의 최대·최소값 및 파워값 추정

음성신호의 분석 및 인식시 원음성신호의 파워는 매우 중요한 요소이나 연산의 어려움 및 연산시간 등의 문제로 실시간처리 시스템에서는 거의 사용을 못하고 있다. 이와같은 문제점을 해결하는 방법으로 본 논문에서는 추정평면의 데이터로부터 파워값을 평가함으로써 원 음성신호의 상대 파워값을 추정하였으며 또한 효율적인 인식을 위하여 추정평면의 데이터의 최대값 및 최소값도 평가하였다.

그림 3.1은 숫자음 "이(2)"와 "삼(3)"의 평활화된 추정평면 데이터들로서, 이 데이터에 균등자기합수를 적용하여 $R(0)$ 를 추정하면 "이(2)"와 "삼(3)"의 음성신호가 갖고 있는 파워값에 비례되는 상대 추정파워값을 구할 수 있으며, 그 값들을 표 3.1에 나타내었다.

표 3.1 추정 파워값(숫자음 "이(2)"와 "삼(3)")
Table 3.1 Estimated power value(numeral speech "I(2)" & "SAM(3)")

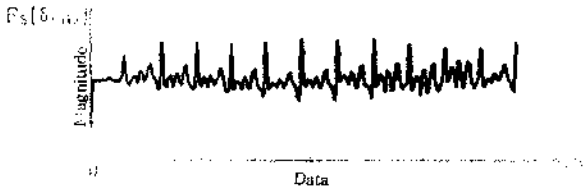
Numeral speech "I(2)"		Numeral speech "SAM(3)"	
Speech power value	Estimated power value	Speech power value	Estimated power value
39452	4570	82239	11497

표 3.1에서 알 수 있는 바와 같이 숫자음 "삼(3)"에 대한 추정평면의 파워값이 숫자음 "이(2)"에 비하여 약 2배 정도 큰 것으로 추정되고, 이 비율은 원래의 숫자음 신호의 파워값에서도 비슷한 것으로 추정되었다.

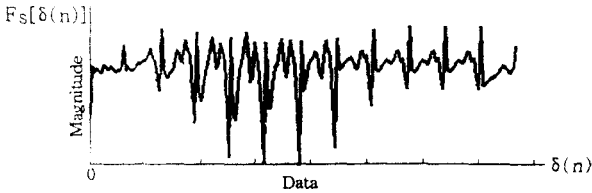
또한, 그림 3.1의 (a)와 (b)에서 알 수 있는 바와 같이 숫자음 "이(2)"의 에너지는 전체의 프레임에 걸쳐 고르게

표 3.2 추정평면의 최대 및 최소값(숫자음 "이(2)"와 "삼(3)")
Table 3.2 Maximum and minimum value on the estimated plane (numeral speech "I(2)" & "SAM(3)")

Numeral speech "I(2)"		Numeral speech "SAM(3)"	
Maximum value	Minimum value	Maximum value	Minimum value
4	-4	14	-1



(a) Data on the estimated plane. (numeral speech "I(2)")



(b) Data on the estimated plane. (numeral speech "SAM(3)")

그림 3.1 추정평면에서의 평활화된 데이터 (숫자음 "이(2)"와 "삼(3)")

Fig 3.1 Smoothing data on the estimated plane. (numeral speech "I(2)" & "SAM(3)").

나타났고, 숫자음 "삼(3)"의 에너지는 가운데 부분에 집중되었음을 알 수 있으며, 이와같은 특징을 평가하기 위하여 본 논문에서는 추정평면의 데이터로부터 최대값과 최소값을 평가하였으며, 그 결과는 표 3.2과 같다.

2. 3등분할에 의한 추정평면의 파워평가 및 정규화

그림 3.1의 추정평면 데이터는 원 음성신호의 시간적인 변화에 대한 특징을 포함하는 신호로서 이 데이터에 FFT를 적용하여 원 신호의 시간적인 변화에 대하여 비례적인 상대 파워값을 평가할 수 있다.[10] 이와같은 특성을 고찰하기 위하여 본 논문에서는 추정평면의 데이터를 3등분할한 후 각각의 분할된 데이터로부터 3등분할된 상대 파워를 평가하였으며 그 결과는 표 3.3과 같다.

즉 숫자음 "이(2)"의 신호에서는 전체적으로 추정평면의 파워변화 즉 시간에 따른 변화가 없었으며 숫자음 "삼(3)"의 경우는 시간에 따른 파워변화가 점진적으로 증가하다가 감소되는 결과를 분석할 수 있었다.

앞 절에서 추정한 p개의 C-LPC계수, 추정평면의 파워값 및 최대 최소값들은 인공신경망을 이용한 효율적인 학습을 위하여 정수값으로 정규화가 요구된다.

표 3.3 추정평면의 3등분 파워값(숫자음 "이(2)"와 "삼(3)")
Table 3.3 Three equal divided power on the estimated plane (numeral speech "I(2)" & "SAM(3)")

Numeral speech "I(2)"			Numeral speech "SAM(3)"		
First power	Second power	Third power	First power	Second power	Third power
1507	1580	1560	3832	5474	2189

추정된 M개의 추정값을 $X_1, X_2, X_3, \dots, X_M$ 이라고 하면, 정규화시키기 위한 함수 $Y_1, Y_2, Y_3, \dots, Y_M$ 은

$$\Delta X_i = \frac{X_{i, \max} - X_{i, \min}}{M} \quad (3.1)$$

$$Y_i = \frac{X_i}{\Delta X_i} \quad (3.2)$$

으로 구할 수 있다.

주식 (3.1)은 추정된 M개 특징값들의 평균을 구하기 위한 함수이고, 식 (3.2)는 정규화 등급의 갯수를 M으로 할 때 M레벨로 정규화시키는 함수로서 본 논문에서 인공신경망의 입력은 위와 같은 방법으로 정규화과정을 수행시켜서 사용하였다.

IV. 실험 및 고찰

제한한 방법의 실효성을 평가하기 위하여 본 논문에서는 5인이 발성한 "영(0)"부터 "구(9)"까지의 한국어 숫자음을 대상으로 인식실험을 수행하였으며, 음성신호의 A/D 변환은 Analog Device사의 7874칩을 이용한 12bit 변환기를 제작하여 사용하였고, 전체 흐름도는 그림 4.1과 같다.

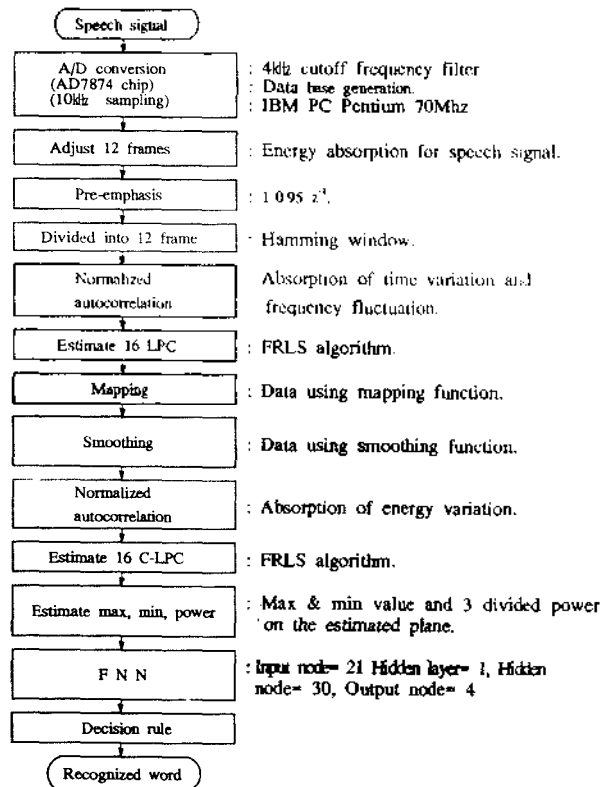


그림 4.1 인식시스템의 전체 흐름도
Fig 4.1 Flow chart of recognition system.

1. OFEC 알고리즘에 의한 C-LPC계수 추출

그림 4.2는 숫자음 "육(6)"의 데이터에 전처리 과정에서 프레임 별로 평가한 LPC계수에 최적특징매핑함수를 적용하여 추정평면으로 매핑시킨 후 OFEC알고리즘을 이용하여 16차의 C-LPC계수를 추정한 것으로서, (a)의 데이터는 추정평면의 데이터에 정규화자기상관함수를 적용하여 처리된 결과이고, 그림 (b)는 (a)의 데이터에 FFT를 적용하여 평가한 파워값이며, 그림 (d)는 (a)의 데이터에 AR모델을 설정하여 추정된 데이터이다. 즉 그림 (d)의 16개 추정데이터는 원 음성신호를 대표하는 C-LPC 계수이다. 그림 (c)의 결과는 (d)의 16개 데이터에 의한 파워값으로, 그림(b)의 파워값과 일치됨을 확인할 수 있었고, 따라서 그림 (d)의 압축추출된 16개 C-LPC계수는 원 음성신호의 특징을 대표하고 있음을 확인할 수 있다.

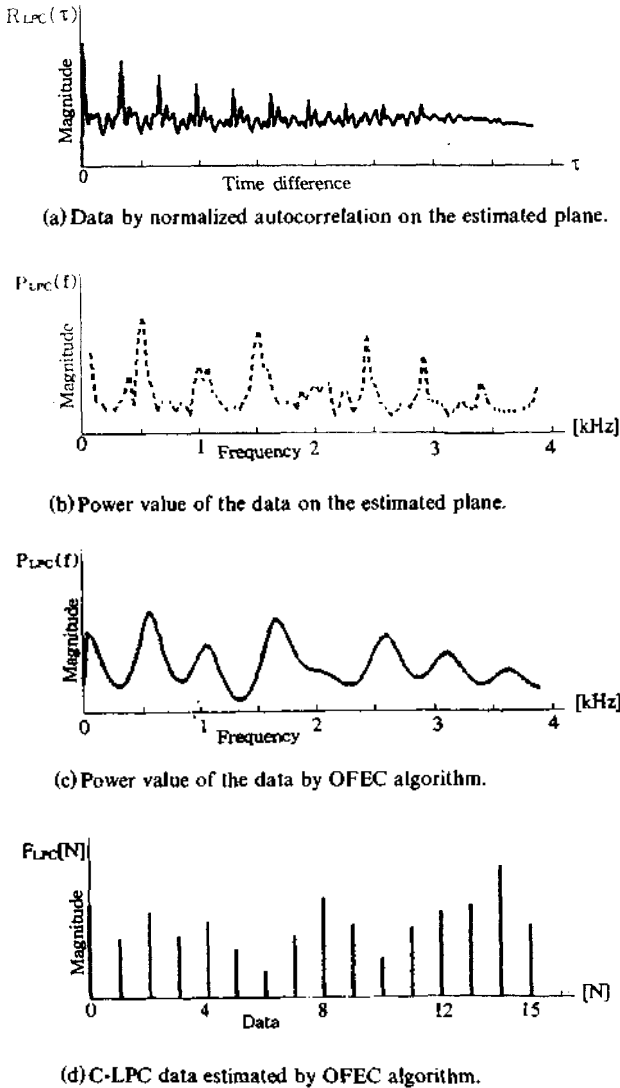


그림 4.2 최종파워스펙트럼 및 C-LPC계수(숫자음 "육(6)")
Fig 4.2 Final power spectrum and C-LPC coefficients (numeral speech "YUK(6)").

표 4.1 OFEC 알고리즘에 의한 추정된 특징파라미터

Table 4.1 Estimated feature parameters by OFEC algorithm.

Order of AR model	Feature parameter									
	YEONG	IR	I	SAM	SA	O	YUK	CHIR	PHAR	KU
0	0.203	0.236	0.234	0.151	0.170	0.228	0.144	0.035	0.161	0.210
1	0.042	0.003	0.023	0.063	0.066	0.035	0.027	0.010	0.070	0.027
2	0.013	0.056	0.070	0.001	0.024	0.007	0.006	0.083	0.021	0.007
3	0.007	0.037	0.005	0.008	0.021	0.001	0.032	0.067	0.010	0.017
4	0.026	0.002	0.020	0.024	0.034	0.021	0.050	0.086	0.030	0.039
5	0.011	0.000	0.034	0.079	0.069	0.018	0.006	0.055	0.068	0.020
6	0.003	0.032	0.062	0.074	0.070	0.005	0.010	0.116	0.053	0.017
7	0.009	0.036	0.031	0.012	0.022	0.008	0.008	0.005	0.019	0.017
8	0.013	0.028	0.002	0.048	0.025	0.014	0.002	0.110	0.009	0.012
9	0.008	0.045	0.020	0.055	0.020	0.006	0.002	0.074	0.027	0.010
10	0.001	0.031	0.050	0.009	0.011	0.004	0.015	0.025	0.003	0.004
11	0.023	0.007	0.034	0.018	0.013	0.032	0.010	0.031	0.016	0.015
12	0.063	0.001	0.007	0.010	0.026	0.058	0.052	0.034	0.028	0.058
13	0.078	0.049	0.014	0.067	0.073	0.071	0.046	0.022	0.081	0.072
14	0.104	0.081	0.069	0.147	0.142	0.090	0.092	0.148	0.145	0.097
15	0.053	0.035	0.040	0.080	0.078	0.045	0.030	0.061	0.072	0.054

2. 3등분할된 추정평면의 파워 및 최대·최소값 평가

전 처리과정에서 "영(0)"부터 "구(9)"까지의 신호로부터 추출한 LPC계수에 최적특징매핑함수를 적용하여 얻어진 추정평면의 데이터로부터 3등분할된 파워값, 최대·최소값을 추정하였으며 그 결과는 표 4.2와 같다.

표 4.2 정규화시킨 최대·최소값 및 3등분 파워값

Table 4.2 Normalized maximum and minimum value and 3 divided power value.

	0	1	2	3	4	5	6	7	8	9
First power	4	1	3	7	11	4	2	9	6	2
Second power	9	1	4	10	21	7	3	7	21	1
Third power	5	3	4	4	15	5	3	5	5	3
Max value	9	1	4	14	21	7	3	7	21	1
Min value	-5	-3	-4	-1	-3	5	-3	-5	-3	-3

표 4.2로부터 알 수 있듯이 숫자음 신호 "일(1), 이(2), 육(6), 구(9)"는 비교적 파워의 변화가 적은 편이며 "삼(3), 사(4), 팔(8)"의 신호는 다른 신호들에 비하여 파워가 크며 특히 중반부에 집중됨을 고찰할 수 있다.

3. 인공신경망에 의한 숫자음인식

OFEC에 의하여 얻어진 C-LPC계수는 1보다 작은 부동소수점값들로서 바로 인공신경망의 입력으로 사용하기에는 인공신경망의 수렴 및 분류기능에 적합하지 않으므로 C-LPC계수에 식 (3.1)과 식 (3.2)를 적용하여 특징을

정규화하였고, 각 숫자음에서 평가된 최대 및 최소값, 3분할에 의한 파워값도 정규화시켰으며 그 결과는 표 4.3과 같다.

표 4.3는 각 숫자음에 대한 C-LPC계수를 정규화시킨 데이터를 나타내었다.

표 4.3 C-LPC계수를 정규화시킨 표본데이터
Table 4.3 Normalized sample data of C-LPC coefficients.

Numeral speech	C-LPC coefficients																				PV	PP	NI	PS	SS	TS
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19						
YEONG(0)	5	-2	0	0	1	0	0	0	0	0	0	0	2	3	4	1	4	9	5	4	9	5				
IR(1)	1	0	1	1	0	0	0	1	1	0	0	0	0	4	2	1	1	3	1	1	3	3				
I(2)	5	2	2	3	3	1	1	2	3	3	2	2	3	3	6	4	2	4	4	3	4	4				
SAM(3)	5	1	5	5	3	0	0	2	4	5	3	0	0	3	8	4	3	14	1	7	10	4				
SA(4)	12	-1	-1	0	-1	-3	-2	0	0	1	0	-1	0	1	3	2	11	21	3	11	21	13				
O(5)	4	-3	-1	0	0	0	0	0	-1	-1	0	2	3	5	2	4	7	5	4	7	5	5				
YUK(6)	2	-1	0	0	0	-1	-2	0	1	0	-1	0	1	2	3	0	2	3	3	2	3	3				
CHIR(7)	6	-1	3	2	0	-1	0	0	1	1	1	-1	0	3	1	9	7	5	9	7	5	5				
PHAR(8)	8	-1	0	0	0	0	0	0	1	1	0	0	0	1	4	2	6	21	5	6	21	5				
KU(9)	2	-1	0	0	1	0	0	0	-1	-1	0	1	2	4	2	2	1	3	2	1	3	3				

1)인공신경망 가중치의 초기화

인공신경망의 초기값들인 가중치 및 오프셋(offset)들은 0.5 + 0.5 사이의 불규칙한 값으로 초기화하였고, OFEC 알고리즘에 의하여 추출된 16개 C-LPC계수, 추정평면의 최대(P.P)값, 최소(N.P)값, 추정파워값들을 정규화하여 인공신경망의 입력으로 하였으며, 효율적인 학습을 위하여 인공신경망에 입력하는 숫자음과 발생자의 순서 및 순번을 불규칙하게 처리하여 학습을 시켰다.

2)인공신경망 학습시 오차의 수렴

본 논문에서는 다음과 같은 단계로 중간처리단계에서 OFEC 알고리즘에 의하여 압축추출된 특징량을 인공신경망의 입력으로 하여 인식실험을 수행하였다.

① OFEC 알고리즘에 의해서 압축추출된 특징량을 인공신경망에 입력.

최종적으로 추출된 16개의 C-LPC계수의 부동소수점 값을 인공신경망의 입력으로 넣은 결과, 인공신경망은 오차값의 범위내에서 수렴하지 못하였다.

② 피크(peak)검출에 의한 특징을 인공신경망에 입력.

16개 C-LPC계수를 식(3.1)과 식(3.2)에 의하여 정규화시킨 후 인공신경망에 입력시킨 결과, 수렴되는 경우가 50%에 미치지 못하였고, 수렴된 경우 인식률도 70%정도였다.

③ ②항의 특징값들에 추정평면의 파워값 및 N.P와 P.P를 추가하여 인공신경망에 입력.

④ ③항에서 사용한 인공신경망의 입력값에 추정평면의

파워값 및 P.P와 N.P를 추가하여 실험한 결과는 수렴되는 경우가 86%정도이고, 인식률도 81%정도로 증가하였다.

④ ③항의 입력에 3등분할 파워값을 추가하여 인공신경망에 입력.

⑤ ④항에서 사용한 인공신경망의 입력값에, 추정평면에서의 시간변화에 대한 파워값 즉, 3등분한 파워값을 포함시켜서 실험하였으며, 그 결과 96.3%의 인식률을 얻었다.

V. 결 론

본 논문에서는 제안한 방법에 의한 음성인식의 실험을 다음의 4 단계로 수행하였으며 그 결과는 다음과 같다.

1 단계: C-LPC계수만에 의한 인식

수렴하지 못하는 경우가 많이 발생하였으며 인식실험을 수행하지 못하였다.

2 단계: C-LPC계수의 정규화에 의한 인식

수렴되는 경우가 약 50%에 근접하였으며 수렴했을 경우의 인식율이 70%이었다.

3 단계: C-LPC계수에 N.P와 P.P를 추가한후 정규화한 값에 의한 인식

2 단계의 실험에 비하여 인식률이 81%, 수렴되는 경우가 약 86%로 개선 되었다.

4 단계: 3 단계의 정규화 값들에 추정평면에서 3등분할 파워값을 정규화값을 포함시킨 인식

수렴이 100%로 되었으며 인식율이 96.3%로 3단계의 인식실험보다 향상된 결과를 얻었다.

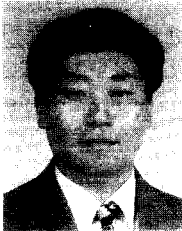
본 논문의 실험에서는 3등분할에 의한 파워평가로부터 얻은 음성신호의 시간변화에 대한 특징을 고찰할 수 있었으며 그 결과를 인식실험에 적용한 인식방법이 인식을 개선에 효율적임을 확인할 수 있었다.


참 고 문 헌


1. J.R. Deller, Jr. et al., *Discrete-Time Processing of Speech Signals*, Macmillan Publishing Company, pp.266~343, 1993.
2. L.R. Rabiner et al., *Digital Processing of Speech Signal* Prentice-Hall, Inc., pp.38~347, 1978.
3. L.R. Rabiner et al., "Speaker-independent recognition of isolated words using clustering techniques", *IEEE Trans. on Acoust., Speech Signal Processing*, vol.ASSP-27, pp.336~349, 1979.
4. 김창석 외 1인, "Neural-HMM을 이용한 고립단어 인식", *한국통신학회논문지*, vol.17 no.11, pp.1199~1205, 1992.
5. N.R. Pal et al., "Generalized Clustering Networks and Kohonen's Self-organizing Scheme", *IEEE Trans. on Neural Networks*, vol.4 no.4, pp.549~557, 1993.
6. T.M. Martinez et al., "[Neural-Gas] Network for Vector Quantization and its Application to Time-Series Prediction", *IEEE Trans. on Neural Networks*, vol.4 no.4, pp.558~569, 1993.

- 7. G. Carayannis et al., "A Fast Sequential Algorithm for Least-Squares Filtering and Prediction", IEEE Trans. on Acoust., Speech Signal Processing, vol.ASSP-31 no.6, pp. 1394~1402, 1983.
- 8. 김창석 외 3인, "음성신호의 최적파형을 적응적으로 추출하는 방법에 관한 연구", 한국통신학회논문지, vol.19 no.2, pp. 373~380, 1994.
- 9. J. Benesty et al., "A Fast Exact Least Mean Square Adaptive Algorithm", IEEE Trans. on Signal Processing, vol.40 no.12, pp.2904~2920, 1992.
- 10. 김창석 외 3인, "최적의 C-LPC계수 추출 알고리즘 제안", 한국통신학회논문지, vol.20 no.1, pp.209~220, 1995.
- 11. 김창석 외 2인, "적용중첩합수에 의한 음성신호의 최적특징추출", 명지대학교 산업기술연구소 논문집 vol.13, pp. 152~159, 1994.

▲최 일 홍(Ii Hong Choi) 1960년 8월 1일생
 1983년 2월: 명지대학교 전자공학과 졸업(공학사)
 1985년 8월: 명지대학교 대학원 전자공학과 졸업(공학석사)
 1992년 9월~현재: 명지대학교 대학원 전자공학과 박사과정
 1988년 3월~1992년 2월: 생산기술연구원 부설 산업기술교육센터 전자기기과
 1992년 3월~현재: 진주전문대학 전자과

▲장 승 관(Seung Kwan Jang) 1960년 1월 8일생

 1984년 2월: 명지대학교 전자공학과 졸업(공학사)
 1986년 2월: 명지대학교 대학원 전자공학과 졸업(공학석사)
 1993년 9월~현재: 명지대학교 대학원 전자공학과 박사과정
 1986년 8월~1992년 8월: 현대전자(주) 산업전자연구소
 1992년 8월~현재: 생산기술연구원 부설 산업기술교육센터 전자기기과

▲차 태 호(Tac Ho Cha) 1960년 10월 20일생

 1985년 2월: 명지대학교 전자공학과 졸업(공학사)
 1987년 2월: 명지대학교 대학원 전자공학과 졸업(공학석사)
 1993년 3월~현재: 명지대학교 대학원 전자공학과 박사과정
 1987년 3월~1988년 8월: 해태전자(주) 기술연구소
 1988년 9월~현재: 생산기술연구원 부설 산업기술교육센터 제측제어과

▲최 응 세(Ung Se Choi) 1955년 12월 11일생

 1981년 2월: 명지대학교 전자공학과 졸업(공학사)
 1983년 2월: 명지대학교 대학원 전자공학과 졸업(공학석사)
 1994년 8월: 명지대학교 대학원 전자공학과 졸업(공학박사)
 1995년 4월~1996년 3월: 일본 국립信州대학 공학부 전기·전자공학과 음성정보신호처리 연구실 Visiting reseacher
 1975년 4월~현재: 생산기술연구원 부설 산업기술교육센터 연구사업3팀

▲김 창 석(Chang Seok Kim) 1938년 9월 18일생

 1968년 2월: 명지대학교 전기공학과 졸업(공학사)
 1972년 2월: 명지대학교 대학원 전기공학과 졸업(공학석사)
 1985년 2월: 명지대학교 대학원 전자공학과 졸업(공학박사)
 1985년 3월~현재: 명지대학교 공과대학 전자공학과 교수