論文2001-38CI-3-6

잡음환경에서 음성-영상 정보의 통합 처리를 사용한 숫자음 인식에 관한 연구

(A Study on Numeral Speech Recognition Using Integration of Speech and Visual Parameters under Noisy Environments)

李 相 栻 *, 朴 仁 政 **

(Sang Won Lee and In Jung Park)

요 약

본 논문에서는 한국어 숫자음 인식을 위해 음성과 영상 정보를 사용하고, 음성에 사용하는 선형예측계수 알고리즘을 영상에 적용하는 방법을 제안한다. 입력으로 얻어지는 음성신호는 0.95 의 매개변수를 통해 고역 신호가 강조되고, 해밍창과 자기상관 분석, Levinson-Durbin 알고리즘에 의해 13 차 선형예측계수를 구한다. 마찬가지로, 그레이 영상신호도, 음성의 자기상관 분석, Levinson-Durbin 알고리즘을 사용하여 13 차의 2 차원 선형예측계수를 구한다. 이러한 음성/영상 신호에 대한 선형예측계수들은 다층 신경회로망에 적용하여 학습이 이루어졌고, 각 레벨의 잡음이 섞인 음성신호를 적용한 결과, 숫자음 '3', '5', '9' 에서 음성만으로 인식한 결과보다 훨씬 좋은 인식결과를 얻을 수 있었다. 결과적으로, 본 연구에서는 영상 신호의 2 차원 선형 예측 계수들이 음성인식에 사용될 경우, 특징 추출에 따른 부가적인 알고리즘이 새로 고안될 필요가 없이, 음성특징 계수를 추출하는 방법을 그대로 사용할 수 있으며, 또한 데이터량과 인식율이 잡음 환경에서 보다 향상되는 효율적인 방법을 제시하고 있음을 알 수 있었다.

Abstract

In this paper, a method that apply LP algorithm to image for speech recognition is suggested, using both speech and image information for recogniton of korean numeral speech. The input speech signal is pre-emphasized with parameter value 0.95, analyzed for 13-th LP coefficients using Hamming window, autocorrelation and Levinson-Durbin algorithm. Also, a gray image signal is analyzed for 2-dimensional LP coefficients using autocorrelation and Levinson-Durbin algorithm like speech. These parameters are used for input parameters of neural network using back-propagation algorithm. The recognition experiment was carried out at each noise level, three numeral speechs, '3','5', and '9' were enhanced. Thus, in case of rocognizing speech with 2-dimensional LP parameters, it results in a high recognition rate, a low parameter size, and a simple algorithm with no additional feature extraction algorithm.

* 學生會員, ** 正會員, 檀國大學校 電子工學科
(Dept. of Electronic Engineering, Dankook Univ.)
接受日字:2000年6月16日, 수정완료일:2001年4月20日

Ⅰ. 서 론

사람과 기계와의 정보전달 체계에서 가장 자연스러운 것은 음성이며, 이 음성신호에는 화자의 나이, 성별, 화술, 감흥 등의 의미 있는 정보들을 제공한다. 이러한

이유로 인해 1950년대부터 음성인식에 대한 연구가 진행되어 왔고[1], 간단한 명령어 인식 정도는 일반 PC 상에서 사용할 수 있을 정도의 결과를 가져왔다. 하지만, 이러한 결실에도 불구하고 잡음이 섞인 환경에서는 아직도 그 결과를 예측하기가 힘들다.

음성인식을 잡음이 적은 실험실 환경이 아닌 주위 생활환경에 적용하기 위해서는 잡음에 강한 음성인식 시스템을 구현해야 할 필요가 있고, 이러한 인식기를 구현연구[2]가 꾸준히 진행되고 있다. 먼저, 음성에 포함될 수 있는 잡음으로는 무엇이 있는가를 살펴보면, 잡음의 원인은 주로 환경 잡음, 잡음을 포함한 음성, 화자의 발성법 및 채널/마이크에 의한 잡음으로 나눌 수 있다. 환경잡음은 팬 소리, 기계의 엔진소리와 같은 연속적인 잡음과 자동차 지나가는 소리, 전화벨 소리 등과 같은 비연속적인 잡음이 있다. 음성 관련 잡음은 음성에 종속적인 잡음으로서 음성의 반사 및 반향이 그 원인이 되고, 화자의 발성에 의한 잡음은 발성의 강약, 주위 잡음과 발성 속도 등에 의한 잡음이며, 채널/마이크에 의한 잡음은 필터의 특성, 주파수 범위 등과 같은 잡음이다[3].

이러한 잡음 요인에 강한 자동 음성인식 연구는 Gong[2]의 연구를 보면, 잡음을 잡음 저항 특성과 비슷한 측정, 음성 강화 및 음성 모델 보상을 통해 특정한 잡음 환경에서의 잡음에 강한 인식결과를 얻었으며, Flanagan 등[4]은 부자연스럽지만, 음성이 입력되지 않는 동안에는 마이크를 끄고, 여러 개의 마이크를 소리가 큰 쪽의 음원에 두는 방법을 제시하였다.

위의 방법들은 그 기본을 음향학적 측면에 두고 음성인식기를 구현하였지만, 최근에는 영상정보를 이용해 잡음에 강인한 음성 인식기 연구가 진행되고 있다.

Petajan[5]은 이미지 프레임에서 4개의 정보를 추출하여 100개의 고립단어 자동음성인식에 적용한 후, 65~78%의 인식률을 보임으로써, 처음으로 음성-영상인식 시스템을 구현하여 음성만의 인식율 보다 영상을 적용한 인식률이 더 증가함을 알 수 있었으며[6], Pentland와 Mase는 얼굴 근육의 움직임을 4개의 영역으로 나눈 구강 이미지의 파라메터를 3 사람이 발음한 5 개의 숫자음 인식에서 75%의 인식률을 얻었다[7]. Yuhas등은 신경회로망을 이용하여 음성에 여러 가지 SNR를 9 개의 모음에 적용하였으며[8], Stork등은 TDNN에 의한 오디오-비디오 인식[9]을, Goldchen은 비디오 정보만을 이용한 경우보다 오디오-비디오에 의한 인식이 25%의 인식률 증가[10]를 보여준다는 것을 증명하였고 또한, 잡음환경을 위한 시스템과 차세대 음성 인식 연구 과제인 잡음에 강한 음성 인식 시스템과 같은 직접적인 시스템 연구에 많은 노력을 기울이고 있다[2]. Bub등은 beam-forming을 사용한 음성 정보와 얼굴의 위치에 의한 영상정보를 활용하여 음성 잡음환경에서 음성정보만을 활용 할 때보다 더 좋은 결과를 얻었다[11].

따라서, 본 논문에서는 영상정보가 음성인식에 좋은 영향을 미친다는 것을 전제로, 시스템 구현이 단순화될 수 있는 방법, 즉 음성처리에서 사용된 자기상관함수 및 Levinson-Durbin 알고리즘을 영상정보를 얻기 위해 사용할 수 있다는 것을 제시한다. 또한, 인위적 잡음 알고리즘인 가우션 잡음 생성기를 이용하여 각 잡음 레벨별로 잡음을 적용한 음성 신호에 대해 2 차원 선형예측계수들의 강인성을 보여준다.

Ⅱ. 선형 예측 및 신경회로망

1. 1 차원 신호에 대한 선형 예측[12]

음성특징 추출은 음성 생성 원리를 이용한 선형 예측방법을 이용한다. 음성신호는 5 에서 100 ms 사이의 충분히 짧은 시간 주기에 대해 조사해 보면, 그 특성이 상당히 정지적(stationary) 이라는 것과, 음성신호는 느리게 변하는 시변신호라는 것을 알 수 있다. 따라서, 음성 신호는 서로 상관적으로 변화하고, 약 20 ms 내에서는 선형적으로 변하는 특성이 있으므로, 선형 시스템으로 가정할 수 있다.

음성신호의 현재 값은 m 개의 과거의 값 s_{n-1}, $s_{n-2}, \cdots, s_{n-m}$ 으로부터 예측된다. 즉,

$$\begin{aligned} \hat{s}_n &= a_1 s_{n-1} + a_2 s_{n-2} + \cdots + a_m s_{n-m} \\ &= \sum_{k=1}^{m} a_k \cdot s_{n-k} \end{aligned} \tag{1}$$

이 때, 예측된 값은 실제 입력되는 값과 차이가 발생하게 되는데, 이것을 선형 예측 오차로 놓고 e_n으로 표시하면,

$$e_n = s_n - \hat{s}_n \tag{2}$$

이 된다.

여기서, e_n 과 s_{n-i} ($i=1,2,\cdots,m$) 에 직교원리(orthogonal principle)를 적용하면,

$$s_n \cdot s_{n-i} - \sum_{k=1}^{m} a_k s_{n-k} \cdot s_{n-i} = 0 \quad (3)$$

$i = 1,2,3,\cdots, m$ 이 되고,

$$R(i) = \sum_{i=1}^{m} s_n \cdot s_{n-i} \quad (4)$$

$$R(k-i) = \sum_{k=1}^{m} s_{n-k} \cdot s_{n-i} \quad (5)$$

이므로, 식 (4) 과 (5)를 식 (3) 에 적용해 보면,

$$\sum_{k=1}^{m} a_k R(k-i) = R(i) \quad i = 1,2,3,\cdots, m \quad (6)$$

이 된다. 이것을 보다 쉽게 하기 위해 행렬식으로 표현하면,

$$\begin{bmatrix} R(0) & R(1) & \cdots & R(m-1) \\ R(1) & R(2) & \cdots & R(m-2) \\ & \cdots & & \\ R(m-1) & & \cdots & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ a_m \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ \cdots \\ R(m) \end{bmatrix} \quad (7)$$

식 (7)을 풀면, 우리가 원하는 선형예측계수들을 구할 수 있다.

2. 2 차원 신호에 대한 선형 예측

음성신호에 사용했던 선형예측 방법을 그레이 레벨 영상에 적용하기 위해서는, 영상의 각 픽셀 주위의 농도값이 급격하게 변하지 않고, 서로 연관되어 변화하는 것처럼 보인다는 가정이 필요하다. 이것은 음성에서의 신호 파형이 갑작스럽게 변화하는 것이 아니라, 서로 상관성을 갖고 변화한다는 것으로 생각할 수 있다. 따라서, 영상에서의 상관성을 식으로 표현하면 다음과 같이 얻을 수 있다. 여기서, $M \times N$ 그레이 레벨 영상을 $G(x,y)$, n 은 차수 증가 인자라 가정하면,

$k = 0$, $n = 0$ 일 때

$$R(n) = \sum_{i=1}^{M} \sum_{j=1}^{N} G(i,j)\, G(i,j) \quad (8)$$

(k,n) 이 각각 $(1,1),(2,4),(3,7),..$ 일 경우

$$R(n) = \sum_{x=1}^{M-k} \sum_{y=1}^{N-k} G(x,y)\, G(x-k,y-k) \quad (9)$$

(k,n) 이 각각 $(1,2),(2,5),(3,8),..$ 일 경우

$$R(n) = \sum_{x=1}^{M} \sum_{y=1}^{N-k} G(x,y)\, G(x,y-k) \quad (10)$$

(k,n) 이 각각 $(1,3),(2,6),(3,9),..$ 일 경우

$$R(n) = \sum_{x=1}^{M-k} \sum_{y=1}^{N} G(x,y)\, G(x-k,y) \quad (11)$$

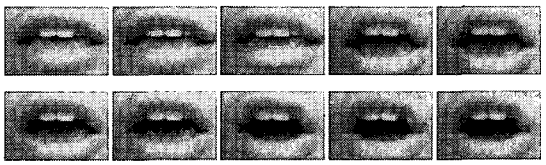
위의 상관계수들을 (7) 에 대입하면, 원하는 예측계수 $a_1, a_2, \cdots, a_n$ 이 구해지며, 다음과 같이 영상에서 어느 위치의 예측 값이 구해진다.

$g(x-1,y-1)$	$g(x,y-1)$	$g(x+1,y-1)$
$g(x-1,y)$	$g(x,y)$	$g(x+1,y)$
$g(x-1,y+1)$	$g(x,y+1)$	$g(x+1,y+1)$

그림 2. 픽셀 $g(x,y)$ 를 중심으로한 각 화소 위치
Fig. 2. Each pixel position centered pixel $g(x,y)$.

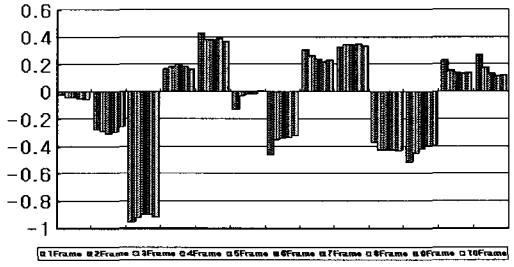
그림 2 의 g 는 영상에서의 각각의 위치에 대응하는 픽셀의 농도를 의미하고, 예측값은 식 (12) 와 같이 구할 수 있다.

$$\hat{g}(x,y) = a_1 \cdot g(x-1,y-1) + a_2 \cdot g(x,y-1) + a_3 \cdot g(x-1,y) \quad (12)$$



(a)

숫자 음 <영>

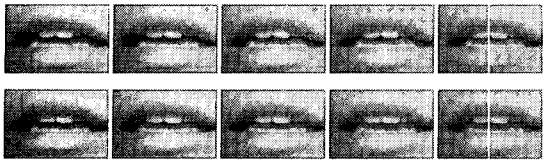


(b)

그림 3. (a) 숫자음 '0' 프레임 영상
(b) 각 프레임별 선형 예측 계수
Fig. 3. (a) Frame image of '0'
(b) LP coefficients per each frame.

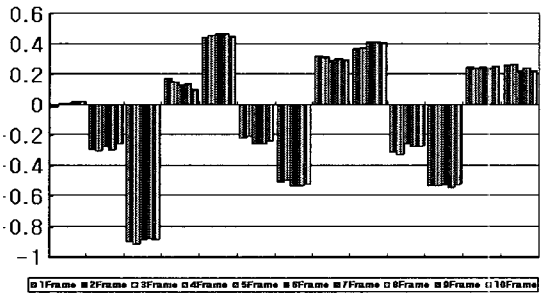
이것은 주위 화소에 의한 것보다도 주사방식에 의해 만들어진 예측 방법이다. Levinson-Durbin[12] 알고리즘

을 영상신호에 적용하였을 경우, 그림 3 의 (b) 그림과 같이 선형예측 계수들이 보이고 있다.



(a)

숫자 음 <일>



(b)

그림 4. (a) 숫자음 '1' 의 프레임 영상
(b) 각 프레임별 선형 예측 계수
Fig. 4. (a) Frame image of '1'
(b) LP coefficients per each frame

3. 신경회로망

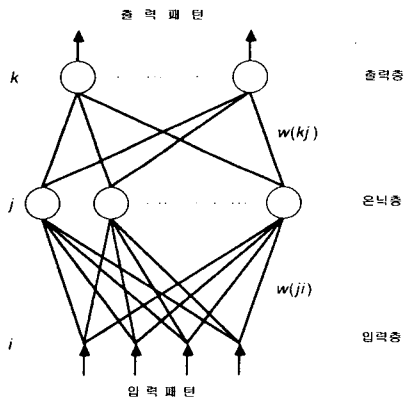


그림 5. 신경회로망 구조
Fig. 5. Neural Network.

본 연구에서 사용한 신경회로망은 역전파 알고리즘을 사용하는 다층 구조 신경망으로써, 학습은 입력단에서의 데이터를 중간노드와 출력노드의 연산을 통해 식 (13) 같이 교사신호와의 오차를 최대한으로 줄이는 방식으로, 각 노드간 가중치를 조정하는 방법을 사용한다.

$$O_{e_k} = o_k(1-o_k)(T_k-o_k) \tag{13}$$

$$H_{e_j} = h_j(1-h_j)\sum_{k=1}^{n} w_{jk}O_{e_k}$$

여기서,

H_{e_j} : j 번째 은닉노드에서의 오차

O_{e_k} : k 번째 출력노드에서의 오차

o_k : k 번째 출력노드의 출력값

h_j : j 번째 은닉노드의 출력값

T_k : k 번째 출력노드에 대한 교사신호

w : 노드간 가중치

Ⅲ. 제안 방법

본 연구에서 제안하는 방법은 음성 인식에 사용되는 자기 상관 분석과 Levinson-Durbin 알고리즘을 이용하여 2차원 선형예측 계수를 구한 후, 음성계수와 영상계수를 인식기에 동시에 적용하여 학습시키는 방법이다. 그림 6의 방법은 음성신호 입력 시 고역 강조와 창 함수, 자기 상관 함수, Levinson-Durbin 알고리즘 적용이 이루어지고, 영상신호 입력 시 그레이 레벨 변환 후, 음성과 동일한 알고리즘이 적용된다는 것을 설명하고 있다. 이러한 방법은 영상에 대해 부가적인 처리 알고리즘이 없어 전체적인 알고리즘이 간단하게 이루어질 수 있다.

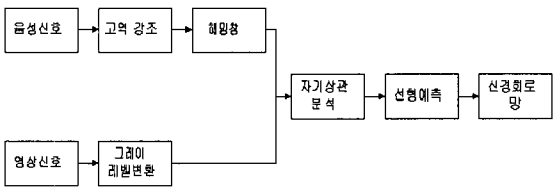


그림 6. 제안된 시스템 구성도
Fig. 6. Suggested system configuration.

Ⅳ. 실험 및 검토

인식에 사용된 음성 데이터는 샘플링 주파수가 8 kHz 이고, Levinson-Durbin 알고리듬을 이용하여, 13 차 선형예측계수를 구하였다. 각 음성 프레임은 음성정

표 1. 음성 인식 강도 출력값(최대값 : 1)
Table 1. Output of only speech recognition strength.

숫자 / dB	영	일	이	삼	사	오	육	칠	팔	구
0	0.9962	0.9960	0.9965	0.9962	0.9958	0.9964	0.9961	0.9963	0.9959	0.9958
10	0.9952	0.9951	0.9954	0.9930	0.9949	0.9900	0.9893	0.9954	0.9954	0.9916
20	0.9940	0.9944	0.9918	0.9835	0.9932	0.9322	0.9653	0.9938	0.9935	0.9564
30	0.9894	0.9927	0.9722	0.8899	0.9893	0.1212	0.8603	0.9895	0.9844	0.5878
40	0.9290	0.9898	0.7301	0.2092	0.9707	0.0006	0.5866	0.9689	0.9295	0.0496

보의 손실을 줄이기 위해 1/2 중첩하여 사용하였으며 각 숫자음별 계수를 구하였다. 또한, 카메라로 부터 얻은 입술정보에 대한 데이터들은 자기 상관 함수를 통한 13차 선형예측계수를 구하였는데, 총 프레임인 10프레임을 기준으로 각 숫자음에 대하여 2 차원 선형예측계수들을 구하였다. 이러한 특징들은 신경회로망의 입력으로 사용되기 위해, 그림 7과 같이 영상 데이터 와 음성 데이터를 서로 합하여 사용되었다. 여기서 사용된 신경회로망은 입력층, 은닉층, 출력층으로 나뉘어 있으며, 역전파 알고리즘을 사용하였다.

원 음성에 잡음을 부가하기 위해 가우션 잡음 생성기를 사용하여 각 잡음 레벨에 따라 각기 다른 선형예측계수를 구하였으며, 영상 데이터를 부가한 실험을 실시하였다. 이때, 입력 데이타의 크기는 한 개의 임의의 단어를 기준으로 할 때, 음성 특징 계수가 260 개, 영상 특징계수가 130 개, 총 390 개의 계수들이 입력으로 사용된다. 이 때, 학습을 위해 사용된 음성 데이터는 잡음 레벨이 0 dB 인 음성 데이터이며, 잡음이 추가된 음성은 시험을 위해 사용되었다.

위 실험은 잡음이 존재한다는 상황을 가정하기 위하여 Box-Muller 방식에 의한 가우션 잡음 발생기[14]를 사용하였으며, 각 잡음 레벨별로 실험을 실시하였다. 각 잡음레벨은 다음과 같이 얻어진다.

$$s_N = M \times G_V, \quad \sigma_N^2 = Var(s_N)$$

$$\text{잡음레벨(dB)} = 10 \cdot \log(\sigma_N^2) \quad (14)$$

M : 증폭상수, G_V : 가우션 잡음

σ_N^2 : 생성된 잡음의 분산

이 때 음성정보에 영상정보를 추가하는 것이 음성적 왜곡이 발생한 숫자음에 대하여 보다 효과적인 결과를 보였다.

표 1 은 음성만으로 인식실험을 하였을 경우에 대한 실험결과를 보이고 있다. 표 1 에서 보면, 숫자음 '3','5','9' 의 출력 결과가 잡음 레벨에 따라 30 dB 이하에서 갑작스럽게 줄어들고 있음을 볼 수 있다. 이러한 숫자음은 외부 잡음에 상당히 영향을 받음을 알 수 있는데, 인식결과를 향상시키기 위해 2 차원 선형 예측계수를 동시 입력하였을 경우의 결과를 표 2에 보였다.

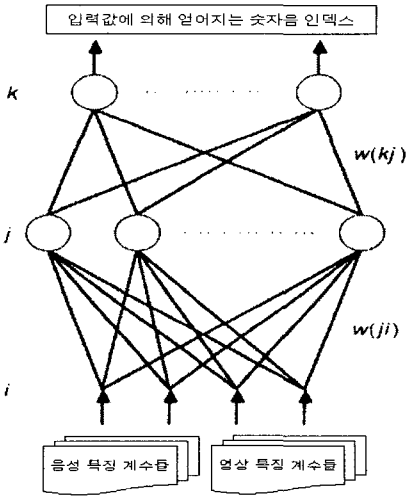


그림 7. 실험에 사용된 신경회로망과 입력계수들
Fig. 7. Neural network and input coefficients used for experiment.

표 2 에서 보면, 음성 그 자체의 정보 보다는 2 차원 선형예측계수를 추가한 것이 좀 더 나은 결과를 보인다는 것을 알 수 있다. 음성만으로 인식할 경우, 잡음 증가시 30 dB 이하부터는 상당한 폭으로 인식 결과가 좋지 않게 떨어짐을 볼 수 있다. 하지만, 영상정보 추가시 인식결과의 저하를 상당히 완화시키고 보다 좋은 결과를 얻을 수 있음을 보이고 있다.

표 2. 음성 및 2 차원 선형예측계수 입력에 대한 신경회로망의 인식강도출력 값
Table 2. Output of recognition strength using speech-only and 2 dimensional LPC.

구분 \ dB	삼		오		구	
	음 성	2D LPC 추가	음 성	2D LPC 추가	음 성	2D LPC 추가
0	0.9962	0.9960	0.9963	0.9964	0.9958	0.9959
10	0.9930	0.9947	0.9900	0.9918	0.9916	0.9950
20	0.9835	0.9919	0.9322	0.9799	0.9563	0.9833
30	0.8898	0.9836	0.1211	0.9049	0.5877	0.8061
40	0.2092	0.9443	0.0006	0.3821	0.0495	0.3523

표 3. 잡음환경에서의 인식률 비교
Table 3. Comparison with recognition rate of noisy environment.

dB	Luettin			Ogihara			제안된 방법		
	영상	음성	통합	영상	음성	통합	영상	음성	통합
0	82.3	94.6	95.1	40.33	98.36	99.02	58.80	71.4	92.0
10		90.4	92.1		65.57	82.6		67.9	85.6
20		81.7	90.0		49.18	70.82		64.3	81.9
30		51.8	79.8		30.49	55.74		57.1	73.6
40		30.7	72.8		15.47	46.89		48.7	68.1

표 3을 보면, Luettin[15]은 HMM을 사용한 고립숫자음에서, Ogihara[13] 는 HMM을 사용한 TV 채널명 인식에서 각각 통합방식의 인식율 증가를 보였으며, 본 논문에서 제안한 방법도 또한 인식율의 증가를 보이고 있다. Ogihara 의 결과와 제안된 방법 비교시, 0 dB에서는 Ogihara 방식이 더 나은 인식유을 보였지만, 잡음 증가에 따른 결과를 보면, 오히려 제안된 방법이 더 좋게 나타남을 볼 수 있다.

V. 결 론

본 논문에서 사용된 음성 인식 방법은 영상정보가 음성인식에 효과적이라는 것과 2 차원 선형예측방법이 전체 알고리즘을 단순화시킬 수 있고 또한 영상에서도 준 주기적인 원리를 적용할 수 있다는 것을 보이고 있다. 또한, 이러한 방법이 잡음이 섞인 한국어 숫자음 인식에 사용될 경우 잡음에 의해 왜곡이 심한 숫자음에 영상정보를 동시 입력하여 보다 좋은 결과를 얻을 수 있다는 것을 보이고 있다. 표 2 에서 보면 음성만의 인식결과와 영상정보를 동시 입력하였을 경우, 잡음의 영향이 커질수록 영상정보의 보상효과가 인식 결과를 향상시킬 수 있다는 것을 볼 수 있다. 또한, 표 3 에서는 HMM을 사용한 Ogihara 의 결과와 제안된 방법과의 결과를 비교하였을 경우, 보다 좋은 결과를 보이고 있다는 것을 나타내고 있다.

따라서, 본 논문에서 제시한 방법은 잡음에 강인한 음성 인식기를 구현하기 위해서 영상정보를 음성정보 사용하는 것이 음성만을 사용한 방법보다 효과적이고, 또한 2 차원 선형 예측 계수 알고리즘을 사용하여 전체 시스템을 단순화시킬 수 있으며, 결과적으로 잡음에 보다 강인한 음성 인식기로 사용될 수 있다는 것을 보여주고 있다.

참 고 문 헌

[1] L. R. Rabiner and B. H. Juang, *Fundamentals of speech recognition*, Prentice-Hall Inc., 1993.

[2] Y. Gong, "Speech recognition in noisy environments: *A survey. Speech Communication*", 16: pp.261-291, 1995.

[3] Alejandro Acero, *Acoustical and Environmental Robustness in Automatic Speech Recognition*, Kluwer Academic Publishers, 1992.

[4] K. Ries, "Hmm and Neural Network based Speech act detection", *ICASSP*, 1999.

[5] E. D. Petajan, "Automatic Lipreading to enhance speech recognition," *IEEE Global Telecom-munications Conference*, pp.265-272, 1984.

[6] E. D. Petajan, B. Bischoff and N. M. Brooke, "An improved automatic lipreading system to enhance speech recognition," *CHI88*, pp. 19-25, 1988.

[7] K. Mase and A. Pentland, "Automatic optically-based recognition of speech," *Pattern Recognition Letters*, vol. 8, no. 3, pp. 159-164, 1988.

[8] B. P. Yahas, M. H. Goldstein and T. J. Sejnowski, "Integration of acostic and visual speech signals using neural networks," *IEEE Communication Magazine*, pp. 65-71, 1989.

[9] D. G. Stork, G. Wolff and E. Levine, "Neural network lipreading system for improved speech recognition," *Int'l. Joint Conf. on Neural Networks*, pp. 285-295, 1992.

[10] A. J. Goldchen, *Continuous Automatic Speech Recognition by Lipreading*, Ph.D. Dissertation, George Washington University, 1993.

[11] J. L. Flanagan, A. C. Surendean and E. E. Jan, "Spatially selective sound capture for speech and audio processing", *Speech Communication*, 13: pp. 207-222, 1993.

[12] Allen Gersho, Robert M.Gray, *Vector Quantization and Signal Compression*, KLUWER ACADEMIC PUBLISHERS, 1992.

[13] A. Ogihara, N. Ishhara, E. Asano and H. Shibata, "Speech Recognition Method by Fusion of Auditory and Visual Information Using Dempster-Shafer's Theorm," *Proc. of ITC-CSCC*, pp. 386-389, Niigata, Aug. 1999.

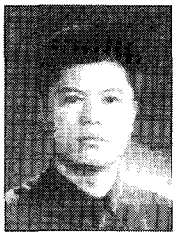
[14] PAUL M. EMBREE, BRUCE KIMBLE, *C Language Algorithms for Digital Signal Processing*, Prentice-Hall International, Inc.

[15] J. Luettin, *Visual Speech And Speaker Recognition*, PhD Thesis, Dept. of Computer Science, University of Sheffield, May, 1997.

저 자 소 개

朴 仁 政(正會員)

1974년 2월 고려대학교 이공대학 전자공학과 공학사, 1980년 2월 고려대학교 컴퓨터공학 공학석사, 1986년 2월 고려대학교 컴퓨터공학 공학박사, 1980년~현재. 단국대학교 전자공학과 교수, 현재: 대한 전자 공학회 이사, xDSL 포럼 의장, 주관심분야 : xDSL, 인터넷방송, 신호처리



李 相 栻(學生會員)

1970년 10월 29일생, 1993년 2월 단국대학교 전자공학과 공학사, 1999년 2월 단국대학교 컴퓨터공학 공학석사, 2000년 8월~현재 단국대학교 전자공학과 박사과정, 주관심분야 : 음성/영상 처리/인식, RTOS, Embedded System