한국음향학회지 제22권 제1호 pp. 28~34 (2003)

잡음 환경에 강인한 이중모드 음성인식 시스템에 관한 연구

A Study on the Robust Bimodal Speech-recognition System in Noisy Environments

이 철 우*, 계 영 철*, 고 인 선*
(Chul-Woo Lee*, Young-Chul Kay*, In-Seon Koh*)

*홍익대학교 전자공학과

(접수일자: 2002년 8월 9일; 채택일자: 2002년 11월 5일)

최근 잡음이 심한 환경에서 음성인식을 신뢰성 있게 하기 위하여 입모양의 움직임 (영상언어)과 음성을 같이 사용하는 방법이 활발히 연구되고 있다. 본 논문에서는 영상언어 인식기의 결과와 음성인식기의 결과에 각각 가중치를 주어 결합하는 방법을 연구하였다: 각각의 인식 결과에 적절한 가중치를 결정하는 방법을 제안하였으며, 특히 음성정보에 들어있는 잡음의 정도와 영상정보의 화질에 따라 자동적으로 가중치를 결정하도록 하였다. 모의 실험 결과 제안된 방법에 의한 결합 인식률이 잡음이 심한 환경에서도 84% 이상의 인식률을 나타내었으며, 영상에 번짐효과가 있는 경우 영상의 번짐 정도를 고려한 결합 방법이 그렇지 않은 경우보다 우수한 인식 성능을 나타내었다.

핵심용어: 음성인식, 영상언어인식, 이중모드 음성인식, 결합 가중치
투고분야: 음성처리 분야 (2,5)

Recent researches have been focusing on jointly using lip motions (i.e. visual speech) and speech for reliable speech recognitions in noisy environments. This paper also deals with the method of combining the result of the visual speech recognizer and that of the conventional speech recognizer through putting weights on each result: the paper proposes the method of determining proper weights for each result and, in particular, the weights are autonomously determined, depending on the amounts of noise in the speech and the image quality. Simulation results show that combining the audio and visual recognizers by the proposed method provides the recognition performance of 84% even in severely noisy environments. It is also shown that in the presence of blur in images, the newly proposed weighting method, which takes the blur into account as well, yields better performance than the other methods.

Keywords: *Speech recognition, Visual speech recognition, Bimodal speech recognition, Combining weights*
ASK subject classification: *Speech signal processing (2,5)*

I. 서 론

최근 컴퓨터 관련기술의 발달로 음성과 영상이 결합된 멀티미디어 형태의 정보체계가 폭넓게 활용되고 있으며, 특히 인간과 기계의 인터페이스 (man-machine interface)를 좀 더 간편하고 정확하게 실현하기 위하여 얼굴 표정이나 방향, 응시 추적, 손동작 그리고 음성 등의 멀티미디어 데이터를 이용한 다중모드 (multimodal) 형태의 인식연구와 이의 상용화가 점차 활발하게 되었다[1]. 음성인식의 경우에 있어서도 이의 상용화가 제대로 이루어지기 위해서는 인식기의 정확도와 주변환경의 영향을 극복할 수 있는 강인성 (robustness)이 무엇보다도 절실히 요구되고 있으나, 보상 (compensation)을 이용하는 기존의 인식기들은 성능면에 있어서 미흡하거나 이의 향상을 위하여 상당히 많은 계산량이 요구되고 있는 실정이다.

보상을 통하여 얻을 수 있는 성능향상의 한계성 때문에 음성정보와 영상언어 (visual speech) 정보를 동시에 결합하여 사용하는 새로운 인식 방법이 활발하게 소개되고

책임저자: 계영철 (yckay@wow.hongik.ac.kr)
121-791 서울시 마포구 상수동 72-1
홍익대학교 전자공학과
(전화: 02-320-1604; 팩스: 02-320-1119)

있다. 하지만 성능의 향상만이 발표되었을 뿐 체계적인 분석이나 최적의 결합방법들에 관하여서는 아직 연구가 되어있지 않은 상태이다[2-5].

따라서 본 논문에서는 영상언어인식 결과와 음성인식 결과를 효과적으로 결합하는 방법을 제안한다. 본 연구실에서 행한 이전의 연구에서는 LPC 분석을 이용하여 음성신호에 존재하는 잡음의 크기를 예측한 후, 이를 이용하여 음성과 영상언어의 인식결과의 결합에 필요한 가중치를 자동적으로 조정하는 방법을 제안하였다[6]. 그러나 이러한 결과는 음성신호에만 잡음이 존재하고 영상신호의 화질은 가장 깨끗한 상태인 상황에서 유도되었다. 그러나 실제로 화자가 발음할 때 얼굴을 움직이거나, 움직임이 많은 발음환경일 경우 입력영상에 blur가 생길 수가 있으므로, 영상과 음성정보의 결합을 위한 가중치의 결정과정에 이러한 상황이 고려되어야 한다.

본 논문에서는 영상신호와 음성신호의 보다 완벽한 결합을 위하여 영상신호의 화질 수준을 추정하는 방법을 제안하였으며, 이를 기반으로 하여 영상언어 인식결과의 신뢰도를 결정하는 방법을 개발하였다. 마지막으로 두 인식결과의 최종가중치를 결정하는 방법을 제안하였다.

II. 본 론

2.1. 음성과 영상언어의 인식값 결합

음성정보와 영상정보를 결합하는 방법으로는 그림 1과 같이 각각의 특징벡터를 결합한 후 인식 알고리즘을 적용하는 방법과 그림 2와 같이 각각 독립적으로 인식을 한 후, 그 인식결과를 결합하는 방법이 있다[4]. 본 논문에서는 두 가지 방법 중 각각을 독립적으로 인식한 후, 그 인식결과를 결합하는 그림 2와 같은 방법을 사용하였다. 특히 그림 1과 2에 표시된 음성정보의 파라메터 추출을 위하여서는 LPC 분석법을 이용하였으며, 영상정보의 파라메터 추출을 위하여서는 입술영역 판별, 입술 경계점 추출 및 입술움직임 모델링과 같은 전처리를 행하였다[6].

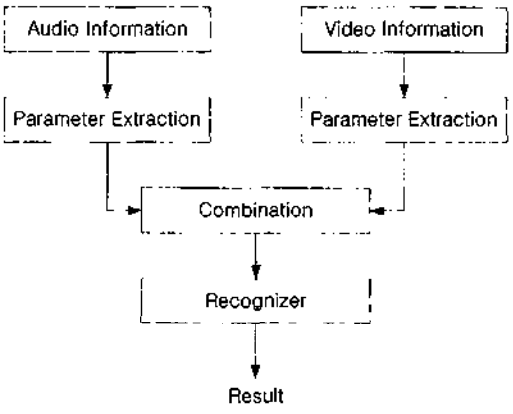

그림 1. 특징벡터의 결합
Fig. 1. Combination of feature vectors.

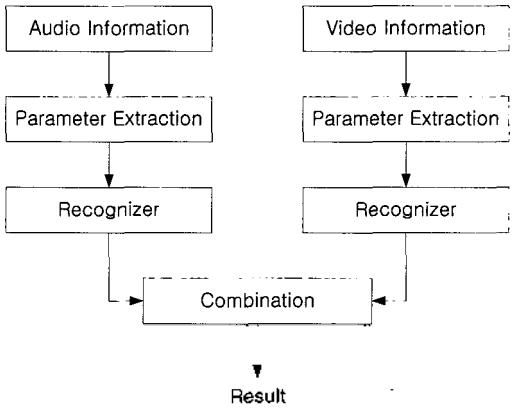

그림 2. 인식 결과의 결합
Fig. 2. Combination of recognition results.

2.2. 전체적인 인식 시스템

화자로부터 영상정보와 음성정보를 입력받아 각각의 특징 파라메터를 추출하여 HMM (Hidden Markov Model)을 사용해 각각의 인식 스코어 (score)를 구한 후, 영상언어 인식 스코어와 음성인식 스코어에 가중치 (weight)를 주어 최종 스코어를 구하여 결과를 얻게 된다[5].

식 (1)은 가중치를 부여하여 최종 스코어를 구하는 방법을 나타낸다.

$$S = W_a S_a + W_v S_v \tag{1}$$
$$W_v = (1 - W_a)$$

S_a : 음성정보에 의한 인식 스코어
S_v : 영상정보에 의한 인식 스코어
W_a : 음성 가중치
W_v : 영상 가중치

음성 가중치 W_a는 0 에서 1 사이의 값을 가지며, 음성신호에 잡음이 적을수록 1에 가까워지고 잡음이 심해질수록 0에 가까운 값을 가지게 된다.

2.3. 제안한 인식 스코어 결합 방법

2.3.1. 음성신호의 잡음 정도 추정

실제신호를 $s(n)$이라 하면, 실제신호와 LPC로 예측된 신호와의 오차 $e(n)$의 MSE (Mean Square Error)는

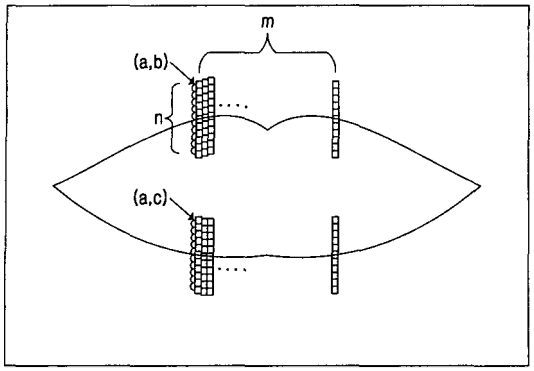

그림 3. 영상의 Blur 정도 측정
Fig. 3. The measurement of the blurred level of an image.

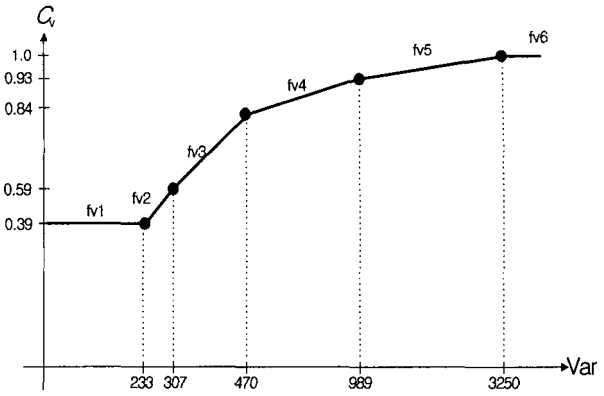

그림 4. 영상 신뢰도 C_v의 결정
Fig. 4. The determination of image confidence C_v.

식 (2)와 같이 표현할 수 있다.

$$E_{MSE} = \frac{1}{M}\sum_{m=1}^{M}\left[s_n(m) - \sum_{k=1}^{n} a_k s_n(m-k)\right]^2 \qquad (2)$$

선형 예측 방법은 잡음에 민감하므로 음성신호에 잡음이 많이 들어갈수록 예측값이 부정확하여 예측 오차가 커지게 된다. 이러한 사실을 이용하여 테스트 음성이 들어오면 일정구간 동안의 E_{MSE}를 측정하여 잡음정도를 예측할 수 있다. E_{MSE}가 클수록 잡음이 많은 것이므로 인식 스코어 결합시 음성 가중치 W_a의 값을 작게 하고 E_{MSE}가 작을수록 W_a의 값을 크게 한다[6,7].

2.3.2. 영상신호의 화질 수준 추정

화자가 발음을 할 때 얼굴을 움직이거나, 움직임이 많은 발음 환경일 경우 입력 영상에 번짐 (blur) 현상이 생길 수 있다. 이러한 경우 영상정보를 이용한 인식률이 저하되므로 인식 스코어 결합시 영상정보 인식 스코어에 부여하는 가중치를 낮게 조정해 주어야 한다. 본 논문에서는 수직 방향으로 번짐 현상이 생긴 경우에 대하여 영상의 화질 수준을 추정하였다.

번짐 정도를 추정하는 방법으로는 먼저 그림 3에서와 같이 입술의 중앙 부분에서 좌우로 m/2 픽셀만큼 떨어진 곳까지 윗입술과 아랫입술의 경계부분을 기준으로 상하 n/2 픽셀 안에 포함되는 모든 픽셀에 대하여 색성분 값의 분산 (variance)[1]을 계산한다.

번짐 정도가 심할 경우 경계부분과 근접 픽셀간의 색성분 값의 차이가 적어서 분산이 작고, 영상이 깨끗할수록 분산이 커지게 된다.

1) 분산을 위한 평균값은 입술 경계부분의 색성분 값으로 정한다.

2.3.3. 가중치 결정 방법

가. 영상언어 인식 스코어의 신뢰도 C_v 결정

입력 영상의 번짐 정도를 측정하여 영상언어 인식 스코어의 신뢰도 C_v를 결정하는 방법은 다음과 같다:

(i) 각각의 번짐 정도 n (n은 영상이 수직 방향으로 n픽셀만큼 번진 것을 나타냄, n=20, 30, ...,clean)에 대한 영상정보만의 인식률을 측정하여 각각의 신뢰도 C_v(n)을 찾아낸다. 여기서 C_v(n)을 결정하는 방법은 식 (3)과 같다.

$$C_v(\mathbf{n}) = \frac{\text{Blur n 영상을 사용한 경우의 인식률}}{\text{clean 영상을 사용한 경우의 인식률}} \qquad (3)$$

(ii) 모든 영상 데이터를 사용하여 분산 (Var)을 측정한 후 (3.2 참고), 번짐 정도가 n인 각각의 경우에 대하여 평균값 $\overline{Var}$(n)을 구한다.

(iii) (i)에서 구한 C_v(n)과 (ii)에서 구한 $\overline{Var}$(n) 정보를 이용하여 Var과 C_v를 대응시키는 함수 $C_v = f_v(Var)$를 찾는다. 여기서 함수 $f_v(Var)$는 모든 구간에 대하여 동일한 것이 아니므로, 그림 4에서와 같이 Var의 값이 속한 구간에 대응하는 함수 $f_{vi}(Var)$[i=1,2,3,...]을 각각 찾아낸다.

(iv) 신뢰도를 자동적으로 결정하는 인식 단계에서는 테스트 영상이 들어오면 Var을 측정한 다음, 그 크기에 따라 어느 구간에 속하는지를 결정하고 그에 대응하는 함수 $f_{vi}(Var)$을 통하여 C_v를 찾아낸다.

나. 최종적인 인식 결과 결합 방법

음성정보의 잡음정도와 영상정보의 신뢰도를 측정하여 최종적으로 가중치를 결정한 후, 각각의 인식 스코어를 결합하는 방법은 다음과 같다:

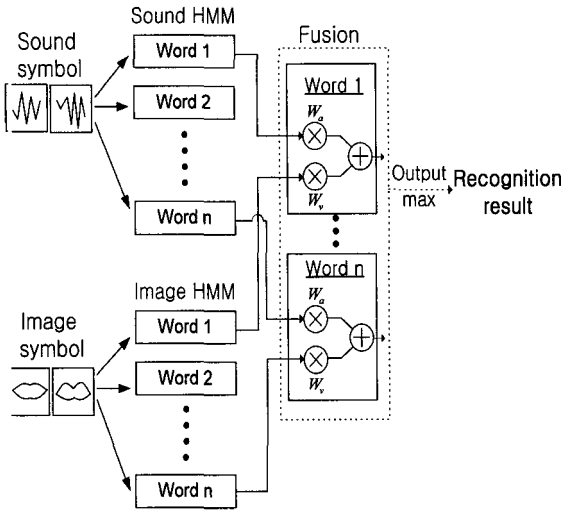

그림 5. 전체적인 인식 시스템 구성도
Fig. 5. Schematic diagram of the entire recognition system.

(i) 잡음 섞인 음성신호로부터 음성정보의 가중치 W_a를 결정한다[6,7]. 영상신호의 가중치는 $1-W_a$가 된다.

$$S = W_a S_a + (1-W_a) S_v \quad (4)$$

(ii) [가]단계에서 찾은 C_v를 이용하여 식 (4)의 영상언어 인식 스코어 가중치 $(1-W_a)$를 $(1-W_a)C_v$로 재조정한다.

(iii) 각 인식결과의 가중치를 다음과 같이 재조정하여 최종적으로 결합한다.

$$S = W_a' S_a + W_v' S_v \quad (5)$$

$$W_a' = \frac{W_a}{W_a+(1-W_a)C_v} : \text{음성 가중치}$$

$$W_v' = \frac{(1-W_a)C_v}{W_a+(1-W_a)C_v} : \text{영상 가중치}$$

그림 5는 전체적인 인식 시스템의 구성도를 나타내는 것으로서, 음성정보와 영상정보가 입력으로 들어오면 각각에 대하여 HMM으로 인식 스코어를 구한 다음, 음성정보의 잡음정도와 영상정보의 신뢰도를 고려한 가중치를 부여하여 최종 스코어를 결정한다. 각각의 모델에 대한 최종 스코어를 구하여 그중 최대가 되는 모델을 찾아 인식 결과를 얻게 된다.

III. 실험 및 결과

3.1. 모의 실험

실험 데이터로 30명의 화자에 대하여 4자리 숫자 10가지를 발음하게 하였으며, 각각 7번씩 발음하여 4개를 학습에 사용하였고 나머지 3개를 이용하여 테스트하였다.

3.1.1. 영상언어 인식

입력 영상은 디지털 캠코더로 촬영한 320×240 픽셀, 30 프레임/sec, 24-bit RGB 컬러 이미지이며, 입술부분의 임계값 Ldown은 1.9를 적용하였다[6]. 16개의 입술 경계점을 추출하여 입술을 포물선으로 모델링하였으며, 포물선 함수를 이용하여 구한 입술의 폭, 높이, 면적과 이들의 프레임간 변화량을 영상 특징 벡터로 사용하였다[6]. 번진 이미지는 깨끗한 이미지를 상용 프로그램인 Photoshop5.0의 Motion Blur Filter기능을 이용하여 만들었다. 인식 알고리즘은 코드북 사이즈 128, 상태 (state) 수 5인 HMM을 이용하였다.

3.1.2. 음성 인식

사용된 음성은 실험실 환경에서 16 bit 양자화, 16 kHz의 샘플링 주파수로 녹음되었다. 잡음 섞인 음성은 이렇게 녹음된 음성에 백색 가우시안 랜덤 잡음을 혼합하여 사용하였다. 12차 LPC-켑스트럼 계수를 추출하여 음성 특징 벡터로 사용하였으며, 인식 알고리즘은 코드북 사이즈 256, 상태수 8인 HMM을 이용하였다.

3.1.3. 인식 결과의 결합

가. 음성인식 스코어 가중치 W_a 결정

학습에 사용되는 음성 데이터를 이용하여 10 ms동안 (160 샘플)의 E_{MSE} 식 (2)가 각각의 SNR에 따라 어느 정도의 값을 가지는가를 훈련단계에서 미리 측정한다. 그리고 각각의 SNR에 대하여 수동으로 가중치를 조절하여 주면서 최적의 인식률을 나타내는 W_a와 $\overline{E_{MSE}}$를 찾

표 1. 음성의 SNR에 따른 W_a 와 $\overline{E_{MSE}}$
Table 1. W_a and $\overline{E_{MSE}}$ determined on various SNRs.

	clean	40 dB	35 dB	30 dB
$\overline{E_{MSE}}$	4244	5885	9558	19984
W_a	1	0.95	0.75	0.55
	25 dB	20 dB	10 dB	0 dB
$\overline{E_{MSE}}$	57327	163829	1586744	16134735
W_a	0.2	0.05	0	0

는다 (표 1).

이 정보를 이용하여 테스트 음성이 들어오면 10 ms동안 (160 샘플)의 E_{MSE}값을 측정한 후 인식결과 결합시 사용할 가중치 W_a를 구한다[6,7]. 함수 $f_{a1}(E_{MSE})$의 형태로 여러 가지 함수를 실험하여 본 결과 선형함수가 가장 간단하고 성능도 우수함을 알 수 있었다[7].

나. 영상언어 인식 스코어 신뢰도 C_v 결정

그림 6부터 그림 9까지는 번짐 정도에 따른 영상을 나타낸다. 'Blur n'에서 숫자 n은 수직방향으로 n 픽셀만큼 번진 것을 의미한다.

학습에 사용되는 영상 데이터를 이용하여 윗입술과 아랫입술의 경계부분의 점 50개를 선택하고 그것을 기준으로 상하 5팍셀에 대하여 색성분 값의 분산 Var이 각각의 번짐 정도에 따라 어느 정도의 값을 가지는가를 훈련단계에서 미리 측정한다. 그리고 각각의 번짐 정도에 대하여 신뢰도 C_v와 $\overline{Var}$을 찾는다 (표 2)

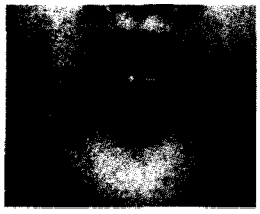

그림 6. 깨끗한 영상
Fig. 6. Clean image.

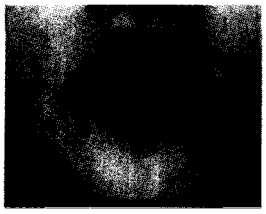

그림 7. blur 20 영상
Fig. 7. Blur 20 image.

그림 8. blur 30 영상
Fig. 8. Blur 30 image.

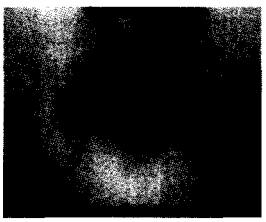

그림 9. blur 40 영상
Fig. 9. Blur 40 image.

표 2. 영상의 번짐 정도에 따른 C_v 와 $\overline{Var}$
Table 2. C_v and $\overline{Var}$ determined on various levels of blur.

	clean	Blur20	Blur30	Blur40	Blur50
$\overline{Var}$	3250.57	989.42	470.14	307.57	233.42
C_v	1	0.93	0.84	0.59	0.39

이 정보를 이용하여 테스트 영상이 들어오면 Var을 측정한 후 그림 4의 함수를 이용하여 C_v를 구한다.

3.2. 실험 결과

3.2.1. 음성과 영상의 독립적인 인식률

표 3과 표 4는 음성정보와 영상정보를 이용하여 각각 독립적으로 인식실험을 하였을 경우의 인식률을 나타낸다. 음성정보의 경우에는 음성신호의 SNR에 따른 인식률을 나타내고, 영상정보의 경우에는 영상의 번짐 정도에 따른 인식률을 나타낸다. 음성정보의 경우에는 잡음에 매우 민감하여 표 3에서 나타나듯이 30 dB 이하에서는 인식 성능이 현저하게 저하되는 것을 볼 수 있다.

3.2.2. 제안된 방법에 의한 인식률

가. 영상의 화질 저하가 없는 경우

화질 저하가 없는 깨끗한 영상정보와 잡음이 섞인 음성정보를 이용한 실험으로, 앞에서 제안한 음성 가중치 W_a를 자동으로 결정하는 방법을 이용하여 각각의 인식 스코어를 결합 (식 (4))하였다. 이러한 방법을 사용했을 때의 인식률을 표 5에 나타내었다[6]. 표 3과 표 5를 비교하여 보면, 제안된 방법에 의하여 음성과 영상을 결합하는 인식기는 음성의 잡음이 심한 경우에도 꽤 좋은 인식률을 보임을 알 수 있다.

나. 영상의 화질 저하가 있는 경우

① C_v를 고려하지 않은 경우

번진 영상에 대해 C_v를 고려하지 않고 W_a만을 결정하여 실험을 한 경우로, 영상 정보의 신뢰도를 반영하지 않은 상태에서 인식 스코어를 결합 (식 (4))하는 방법이다.

표 3. 음성정보의 독립적인 인식률 단위: %
Table 3. Recognition rate for speech information alone.

	clean	40 dB	35 dB	30 dB	25 dB	20 dB	10 dB	0 dB
speech	100	98.7	84	49.7	23	14	10	10

표 4. 영상정보의 독립적인 인식률 단위: %
Table 4. Recognition rate for visual information alone.

	clean	Blur20	Blur30	Blur40	Blur50
image	85	79.7	71.7	50	33.3

표 5. 제안한 방법에 의한 인식률
Table 5. Recognition rate by the proposed method. 단위: %

	clean	40 dB	35 dB	30 dB	25 dB	20 dB	10 dB	0 dB
speech+image	100	98.7	94.7	86.3	85	84.7	85	85

표 6. C_v 를 고려하지 않은 인식률
Table 6. Recognition rate with C_v not included. 단위: %

image \ speech	clean	40 dB	35 dB	30 dB	25 dB	20 dB	10 dB	0 dB
clean_image	100	98.7	94.7	86.3	85	84.7	85	85
blur 20	100	97.3	89	80.3	79.3	77	77.7	77.7
blur 30	99.3	96	87.3	76	72.7	69.3	72.3	71
blur 40	98.7	95.3	75	58.3	52.7	52.3	50	48.3
blur 50	97.7	92.3	72.3	51.7	39	34.3	32.3	34.3

표 7. C_v를 고려한 인식률
Table 7. Recognition rate with C_v included. 단위: %

image \ speech	clean	40 dB	35 dB	30 dB	25 dB	20 dB	10 dB	0 dB
clean_image	100	98.7	94.7	86.3	85	84.7	85	85
blur 20	99.7	97.3	89.3	81	81	79	77.7	77.3
blur 30	99.3	97.7	88.3	76.3	74.3	74	70.7	71
blur 40	100	97.7	78.7	65.3	59.3	51.3	50.3	47
blur 50	100	97	77.3	51.3	39.3	33.7	34.3	33.7

표 6에 인식률을 나타내었다.

② C_v를 고려한 경우

번진 영상인 경우 W_a를 일단 결정한 후, 영상정보의 신뢰도 C_v를 반영하여 영상언어 인식 스코어의 가중치를 재조정하여 인식 스코어를 결합하는 방법이다 (식 (5)). 표 7에 인식률을 나타내었다.

인식률 표에서 알 수 있듯이 음성정보의 SNR이 40 dB 이상의 높은 경우에는 음성정보의 가중치가 매우 커서 영상정보의 화질의 저하가 최종인식률에 크게 영향을 미치지 않지만, SNR이 20 dB~35 dB에서는 C_v를 반영한 경우가 C_v를 반영하지 않은 경우보다 인식 성능이 우수하게 나타났다. 그리고 SNR이 10 dB 이하의 낮은 경우에는 음성정보의 가중치가 매우 작으므로 결합 인식률이 영상정보만의 인식률과 비슷한 성능을 나타내는 것을 볼 수 있다.

IV. 결 론

본 논문에서는 음성정보를 이용한 인식결과와 영상정보를 이용한 인식결과의 결합방법에 대한 새로운 알고리즘을 제안하였다. 기존의 방법들은 이미 알고 있는 음성 SNR에 따라 수동적으로 최적화 된 가중치를 부여하는 방법이었는데 비하여, 본 논문은 테스트음성의 SNR정보를 모르는 상태에서 음성신호의 예측오차를 이용하여 음성에 섞여있는 잡음 정도를 예측한 후 이를 이용하여 자동적으로 가중치를 조절하는 방법을 제안하였다. 또한 입력 영상의 화질 수준을 예측하여 영상언어 인식 스코어에 부여되는 가중치를 재조정한 후, 최종 인식 결과를 얻도록 하였다.

모의 실험 결과 깨끗한 영상과 잡음 섞인 음성을 사용할 경우 전체적으로 84% 이상의 인식률을 나타내었으며 특히 잡음정도가 심해질수록 음성정보만을 이용했을 때의 인식률에 비해 상당한 인식 성능 향상을 볼 수 있었다. 또한 영상의 화질이 떨어질 경우 영상인식의 신뢰도를 고려하여 영상정보에 부여되는 가중치를 재조정하는 방법이 더 우수한 인식 성능을 나타내었다.

감사의 글

이 연구는 정보통신부에서 지원하는 대학기초연구 지원사업으로 수행되었음 (2001-036-2).

참 고 문 헌

1. R. Sharma, V. I. Pavlovic, and T. S. Huang, "Toward multimodal human-computer interface," *Proceedings of the IEEE* 86 (5), May 1998.
2. H. McGurk and J. MacDonald, "Hearing lips and seeing voices," *Nature*, 264, 746-748, 1976
3. J. Luettin, N. A. Thacker, and W. Beet, "Active shape models for visual speech feature extraction," *Technical re-*

port, University of Sheffield, Sheffield, UK, 1995.
4. R. Kover, U. Harz, and J. Schiffers, "Fusion of visual and acoustic signals for command-word recongnition," *ICASSP* 1997.
5. A. Ogihara, and S. Asao, "An isolated word speech recognition based on fusion of visual and auditory information using 30-Frame/s and 24-bit color image," *IEICE Trans. FUNDAMENTALS*, **E80-A** (8), 1997.
6. 이철우, 계영철, 고인선, "강인한 음성인식을 위한 이중모드 센서의 결합방식에 관한연구," 한국음향학회지, **20** (6), 51- 56, 2001.
7. 이동근, 계영철, "음성-영상 인식기 결합을 위한 가중치 결정에 관한 연구," 한국음향학회 추계학술발표대회 논문집, 21 (2), 143-146, 2002.

저자 약력

● 이 철 우 (Chul-Woo Lee)

1998년 2월: 홍익대학교 전자공학과 학사
2002년 2월: 홍익대학교 전자공학과 석사.
2002년 3월~현재: 삼성전자 주식회사 디지털미디어 연구소
※ 주관심분야: 음성인식, 디지탈 신호처리

● 계 영 철 (Young-Chul Kay)

1980년 2월: 서울대학교 전자공학과 학사
1982년 2월: 한국과학기술원 전기 및 전자공학과 석사
1991년 5월: Univ. of Southern California, Electrical Eng, Ph.D.
1991년 9월~현재: 홍익대학교 전자전기공학부 부교수
※ 주관심분야: 디지탈 신호처리, 음성 및 영상인식, 로봇 비젼

● 고 인 선 (In-Seon Koh)

1979년 2월: 서울대학교 전자공학과 졸업 (B.S.)
1987년 5월: Marquette University 졸업 (M.S.)
1991년 5월: Rensselaer Polytechnic Institute, Dept. of ECSE, Ph.D.
1981년~1985년: 대우전자 근무
1991년~1992년: 대우전자 근무
1992년~현재: 홍익대학교 전자전기공학부 부교수
※ 주관심분야: 이산사건 시스템 제어, Computer-Integrated Manufacturing (CIM), Computer Network Analysis, Multimedia, Petri Nets