

동적 환경에서의 립리딩 인식성능저하 요인분석에 대한 연구

A Study on Analysis of Variant Factors of Recognition Performance for Lip-reading at Dynamic Environment

신도성*, 김진영*, 이주현**
(Do Sung Shin*, Jin Young Kim*, Ju Hun Lee**)

*전남대학교 전자공학과, **전남대학교 정보통신공학부 & RRC HECS
(접수일자: 2001년 11월 14일; 채택일자: 2002년 4월 3일)

최근 립리딩에 대한 연구는 음성인식방법에 있어서 부가적인 정보를 제공하여 잡음환경에서 견인한 음성 인식을 하거나 음성정보의 부가적인 특징벡터로 사용하기 위한 방법으로 연구되고 있다. 그러나 립리딩 연구의 대부분은 실험실 환경하의 제한된 결과로서, 실제 다양한 동적환경에서의 견인성에 대해서는 연구된 바가 없다. 현재 우리는 입술정보만을 이용한 자동 22단어 인식기를 만들었으며, 이미지 기반 립리딩의 성능은 53.54%의 성능을 가지고 있다. 본 연구에서는 기구화된 립리딩 시스템을 기반으로 하여, 립리딩 성능이 환경적인 변화에 대해서 얼마나 안정할 수 있는지, 그리고 립리딩의 인식성능 저하를 일으키는 주요 요인이 무엇인지에 대하여 연구하였다. 입술이미지의 동적 변이로서는 이동, 회전, 크기변화와 같은 공간적 변화와 빛에 의한 조명변화를 고려하였다. 실험용 데이터로는 영상변환에 의한 시뮬레이션된 데이터와 동적변화가 심한 자동차 환경에서 수집한 데이터를 사용하였다. 실험결과 입술의 공간 변화가 인식성능 저하의 한가지 요인으로 작용함을 발견하였다. 그러나 실제적으로 공간변화보다 더 심각한 성능저하 원인은 시간 흐름에 따른 조명조건의 변화로써 70% 이상의 왜곡이 발생했다. 따라서 신뢰할 수 있는 립리딩 시스템 구현을 위해서 고려해야 할 가장 큰 요인은 빛의 변화임을 발견할 수 있었다.

핵심용어: 립리딩, 공간적인 변화, 조명의 변화, 음성인식, HMM
투고분야: 음성처리 분야 (2.5)

Recently, lip-reading has been studied actively as an auxiliary method of automatic speech recognition (ASR) in noisy environments. However, almost of research results were obtained based on the database constructed in in-door condition. So, we don't know how developed lip-reading algorithms are robust to dynamic variation of image. Currently we have developed a lip-reading system based on image-transform based algorithm. This system recognize 22 words and this word recognizer achieves word recognition of up to 53.54%. In this paper we present how stable the lip-reading system is in environmental variance and what the main variant factors are about dropping off in word-recognition performance. For studying lip-reading robustness we consider spatial variance (translation, rotation, scaling) and illumination variance. Two kinds of test data are used. One is the simulated lip image database and the other is real dynamic database captured in car environment. As a result of our experiment, we show that the spatial variance is one of degradations factors of lip reading performance. But the most important factor of degradation is not

the spatial variance. The illumination variances make severe reduction of recognition rates as much as 70%. In conclusion, robust lip reading algorithms against illumination variances should be developed for using lip reading as a complementary method of ASR.

Keywords: Lipreading, Spatial variance, Illumination variance, Speech recognition, HMM

ASK subject classification: Speech signal processing (2,5)

1. 서론

현재 영상정보를 이용하여 음성 인식하는 연구는 여러 가지 방식으로 진행되고 있다. 일반적으로 영상정보를 이용한 음성인식은 바이모달 (bi-modal) 형태로 아용되고, 영상정보만을 이용한 독자적인 인식기는 거의 없다. 실제로 다양한 음성구조를 입술의 변화와 모양만을 가지고 인식하기란 어려운 과제이다. 현재까지 여러 연구들을 살펴보면 립리딩 연구는 불과 몇 개의 음절만을 가지고 연구하거나 화자종속적인 피실험자만이 인식할 수 있는 방법들이 대부분이다[1-5]. 더욱이 대부분의 연구는 실험실환경에서 채집한 입술영상에 대하여 립리딩을 수행한 것으로서 립리딩의 강인성에 대해서는 간과되고 있는 실정이다. 그런데 이는 립리딩이 잡음환경에서 음성 인식의 성능을 높여준다는 목적하에 연구가 수행된다는 점을 고려하여 볼 때, 립리딩이 환경변화에 강인하다는 가정은 음성인식과 마찬가지로 옳지 못하다.

본 논문에서는 기개발된 립리딩 시스템을 기반으로 하여 립리딩이 환경변화에 얼마나 강인한지를 연구하였다. 즉, 립리딩 성능이 다양한 공간적 그리고 조명의 변화에 따라서 어느 정도 견인하게 동작하는지에 대한 연구를 수행하였다. 본 논문에서 사용된 립리딩 시스템은 영상변환에 기반한 립리딩 방법으로서 이산코사인변환 (DCT)과 주성분 분석 (PCA)을 사용하여 입술정보를 분석하고 HMM (Hidden Markov Model)을 사용하여 인식하는 알고리즘에 기반하고 있다. 현재 립리딩 시스템을 개발하기 위하여 사용된 데이터 베이스는 70명의 발성한 22단어의 입술영상 코퍼스이다. 이 시스템을 기반으로 하여 립리딩의 강인성에 대하여 컴퓨터 시뮬레이션을 수행하였는데, 고려된 공간변화로는 이동, 회전 그리고 스케일링을 고려하였다. 그리고 조명의 변화로서는 전체적인 조명의 강도변화, 그리고 실제환경에서 채집한 다양한 조명변화를 고려하였다. 따라서 동적환경과 변이를 고려하기 위한 테스트 데이터 베이스는 두 가지를 사용하였는데, 하나는 이미지 변화를 이용한 시뮬레이션된 데이터 베이스

이며, 또 다른 하나는 실제 자동차환경에서 다양한 시간대에 채집한 데이터베이스이다. 컴퓨터 실험결과 조명의 변화가 립리딩 성능을 크게 좌우함을 알 수 있었다. 이 결과로부터 립리딩 실제 환경에서 이용되기 위해서는 조명과 동적변이에 강인한 알고리즘이 개발되어야 함을 알 수 있다.

II. 구현된 립리딩 시스템 구조 및 성능

구현한 립리딩 시스템의 전체적인 구조를 그림 1에 도식하였다. 입력된 프레임은 본 립리딩 시스템에서 파라미터 추출과정은 크게 영상 전처리와 영상변환의 두 과정으로 나누어진다.

영상 전처리 과정은 입력된 입술 영상을 특징 파라미터로 변환시키기 위한 처리과정으로 흑백 영상이나 컬러 영상을 입력으로 받아들여 이진 영상처리를 위해 명암 영상 형태로 변환한다. 2진 영상변환 블록에서는 입력된 영상에서 입술 영역만을 추출하는 작업을 한다. ROI 추출 블록에서는 2진 영상 변환을 통해 찾아낸 입술 안쪽 영역을 기준으로 입술의 중심 높이와 좌우 폭을 찾아 입술만을 따로 분리해 낸다. 분리된 입술 ROI는 다운샘플링

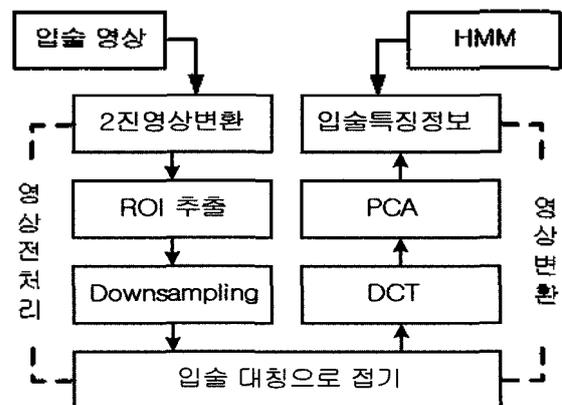


그림 1. 립리딩 알고리즘 블록도
Fig. 1. Block diagram of lip-reading algorithm.

(Downsampling) 과정을 거쳐 입술의 정보손실을 최소화 하는 범위 내에서 데이터 처리량을 줄이기 위한 작업을 수행한다.

입술 대칭 접기 블록에서는 다운샘플링된 영상이 화자의 머리 회전이나 상하좌우 기울어짐이 없다는 전제 하에 인간의 입술형태가 좌우 대칭인 점에 착안하여, 입술 ROI 영상의 절반만을 이용해서 영상 처리한다.

입술 대칭 접기 과정을 마치면 영상 선형 변환 과정을 수행하게 된다. 영상변환의 첫번째 과정은 이산코사인 변환 블록의 수행이다. 여기서는 본래의 ROI 영상 크기를 절반으로 줄여 픽셀 데이터의 처리량을 줄일 수 있게 하고 접어진 8×16 ROI의 픽셀 값들은 대칭되는 픽셀들의 평균값이 되어 영상잡음 요소 및 좌우 측면 조명의 불균형에 대해 강인함을 갖게 된다.

주성분 분석 블록에서는 통계적 알고리즘인 주성분 분석이 갖는 원래 정보를 적절히 선형 변환시켜 정보를 가능한 많이 보존하는 새로운 차원으로 전체 체계의 특성을 요약하는 특징을 이용한다.

매 프레임마다 15개의 파라미터를 추출하고 단어에 대한 음성구간 동안의 파라미터를 생성하고 단어에 대한 파라미터는 HMM의 학습화 과정을 거쳐 각 단어에 대한 평균과 분산과 가중치를 계산하였다. 이렇게 하여 22단어에 대한 각 단어의 특징을 가지고 테스트 데이터의 입력과 비교를 하여 최적의 단어를 짐출한다[6,7].

립리딩 시스템을 이용해서 영상 파라미터를 추출하는

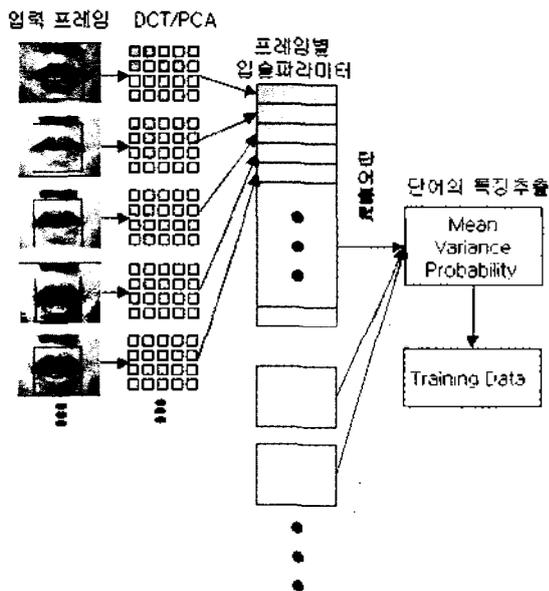


그림 2. 립리딩에서 영상 특징 추출
Fig. 2. Overview of the visual features extraction for lipreading.

표 1. 8×16 접어진 이미지를 PCA 90%와 8×8 DCT를 기반으로 처리한 HMM 인식 결과
Table 1. HMM recognition results (%) on PCA 90%, 8×8 DCT with 8×16 mean half-sized image (52 speakers training / 18 speakers testing).

DCT	St. 3	St. 4	St. 5	St. 6
Mix. 3	39.90	49.24	49.24	53.28
Mix. 4	44.70	46.21	50.00	54.04
Mix. 5	46.72	52.02	49.49	52.78
Mix. 6	43.94	47.47	48.74	53.54

과정을 그림 2에 도식하였다.

화자의 발음 구간은 음성 분석알고리즘을 이용하여 음성의 시작점과 끝점을 영상 정보를 얻기 위하여 동기화를 시킨다. 화자가 발음하는 음성구간동안만 입술 특징 파라미터를 추출한다. 입술 특징 파라미터는 프레임별로 추출하게 되고, 발음하는 구간 전체 프레임에 대해서 입술파라미터를 가지게 된다.

실제 인식 실험에 사용된 영상 데이터는 남성 70명을 대상으로 22개의 단어를 평상시 발음으로 발성한 영상을 30 frame/sec의 속도로 저장하여 데이터 베이스를 구축하였다. HMM 학습화 과정에 52명의 영상데이터가 사용되고 실제 인식을 시험하고자 하는 각 데이터는 서로 다른 18명의 화자가 22 단어를 발성한 입술 움직임 영상에서 얻어졌으며, 표 1과 같이 HMM 알고리즘에서 상태 (state)수와 가지 (mixture)수를 변화시키면서 실험을 수행하였다. 표에서 상태수는 'S'로, 가지 수는 'M'으로 표기하였다. 표1은 실험 결과중 입술의 대칭성을 이용한 알고리즘을 적용한 8×16으로 접어진 입술 ROI 영상에 대해 8×8 이산코사인 변환 영상변환을 하여 주성분 분석을 거쳐 추출된 입술특징 파라미터를 사용하여 HMM 인식 실험을 수행한 결과를 표기한 것이다[8-11].

III. 립리딩 인식성능에 미치는 환경적인 요인

입술의 특징 파라미터 추출은 사용자가 다르거나 발음하는 다양한 습관과 같은 많은 변수가 존재하기 때문에 상당히 어려운 문제이다. 인식성능을 저하시키는 요인은 크게 세가지 형태로 분류할 수 있다. 첫째 사용자 변화에서 발생하는 요인으로, 사용자마다의 다른 피부색을 가지고 있고 화자마다 발음하는 입 모양이 다르다. 그리고 발음하는 모양에서 혀와 치아의 보임 정도에 따라 인식성

능에 영향을 끼치게 된다. 무엇보다 장소별 시간별 조명의 변화가 가장 큰 영향을 준다. 한 예로써 아침과 저녁에 햇빛의 강도가 다르고 같은 시간대에 조사방향에 따라서 영상이 그늘진 왜곡이 발생하고 부분 반사에 의한 밝아지는 부분이 있다. 두번째 인식성능 저하 요인으로 입술의 공간적 변화가 있다. 인식하려는 입술영상이 회전, 입술의 크기 변화, 위치의 이동 등과 같은 공간적 변화로 립리딩의 인식을 저하가 발생한다. 마지막으로 카메라의 입력 효율 등에 따른 영상분석 안정성의 저하를 들 수 있다. 이처럼 립리딩의 성능저하를 일으키는 다양한 요인들이 있다. 본 논문에서는 인식성능에 가장 큰 영향을 주는 공간적인 변화와 조명변화에 대한 요인을 분석하였다.

IV. 공간적인 변화에 의한 인식성능 저하

4.1. 공간적인 영향을 주는 요인 및 성능저하

공간적인 변화 (spatial variant)로 인한 립리딩 인식을 저하 요인은 사용자의 입술이 회전을 하거나 이동하거나 크기 변화에 대한 것들이 있다. 이러한 변화의 립리딩에 미치는 영향을 알기 위해서 물리적인 변화를 각 단계별로 적용을 하여 공간변화에 견인한 알고리즘을 가지고 인식성능을 비교하여 보았다.

그림 3은 공간적인 변화의 한 샘플을 보여주고 있다.

4.2. 공간 변화에 보상 단어인식 실험

공간변화에 대한 단어인식 실험은 공간적인 변화에 어느 정도의 견인성을 가지고 있는지 대한 실험을 위해 테스트 데이터를 가변적인 공간변화를 시켜서 실행하였다. 테스트 데이터는 표 2처럼 회전, 이동, 크기 변화를 하였으며, 각각의 공간변화는 단계별로 변화를 주었다. 인식 실험의 결과는 연속적인 HMM (Hidden Markov Model) 방법을 이용하여 얻어냈다. 음성언어의 특징은 각 상태



그림 3. 공간변화의 예제
Fig. 3. Example images of spatial variance.

표 2. 공간변화에 대한 실험을 위한 변화정도, 테스트크기는 18명이 22단어에 대한 발음한 프레임수를 나타낸다.

Table 2. The degree of variance divided into three or four levels for spatial invariance experiments. Test set sizes are shown as number of words multiplied by speakers (18 × 22).

공간 변환	정도	테스트 Set
회전 (rotation)	5°	18 × 22 × nfw (nfw: 단어에 대한 프레임 수)
	10°	
	15°	
이동 (translation)	3%	
	6%	
	9%	
크기 (scale)	× 0.8	
	× 0.9	
	× 1.1	
	× 1.2	

(state)의 시간적인 흐름으로 분류되어 있으며 여러 개의 가우시안 밀도함수의 분포를 가지고 있다. 이 실험의 안정성을 위해서 상태는 다중 상태 (multi states)를 가지며 각각의 상태에 가우시안 가지 (mixture)를 가지고 있다. 이 실험에서는 4~8 상태를 가지며 각 상태별로 4~6가지를 가지고 인식실험을 하였다.

각각의 립리딩에서 사용한 학습화 과정에 참여한 집단은 70명이 22단어에 대한 연속적인 이미지를 이용하였다 (1144 MJPEG). 테스트 과정에 참여한 집단은 18명이 22단어에 대한 연속적인 이미지를 사용하였다 (396 MJPEG).

이러한 공간적인 변화가 립리딩의 인식성능에 어느 정도 영향을 미치는지에 대한 실험을 하였다. 그 결과는 그림 4, 그림 5, 그림 6에 각각의 공간적인 변화의 정도에 따라 인식성능의 저하 정도를 분석하였다. 각 인식성능 저하 실험에서 공간적인 변화를 각 단계별로 적용을 하였다. 그 기준은 4.2절의 표 2에서 테스트를 위한 단계별 공간적인 변화에서 상세히 나와 있다. 그림 4는 회전 변화에 대한 인식성능 변화에 대한 22단어 인식실험을 한 결과이다. 입술을 약 15° 기울였을 경우에 약 20%의 인식성능이 저하됨을 알 수 있다.

그림 5는 이동 변화에 대한 인식성능저하 정도를 보여주고 있다. 이동 변화의 정도는 입술의 크기를 기준으로 전체길이에서 대각선으로 이동을 시켜서 인식성능을 비교해 보았다. 이미지가 12% 이동되었을 경우 인식률이 4/10으로 낮아졌다.

마지막으로 그림 6에서는 입술 이미지를 확대·축소하여 크기 변화가 발생했을 때의 인식성능 저하가 발생함을 보여주고 있다. 결과를 보면 테스트 크기가 0.8로 줄어들었을 때 3/10으로 인식 성능저하가 발생했음을 볼 수 있다.

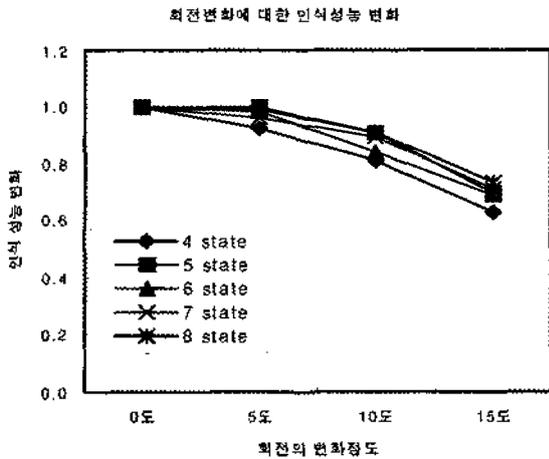


그림 4. 회전변화에 대한 인식성능 저하 결과
Fig. 4. The results of word recognition rate by rotation variance.

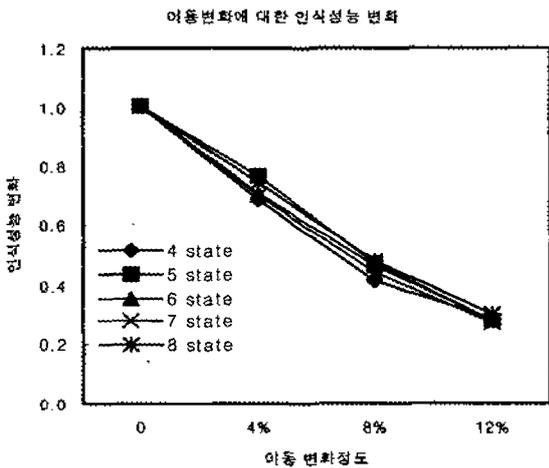


그림 5. 이동변화에 대한 인식성능 저하 결과
Fig. 5. The results of word recognition rate by translation variance.

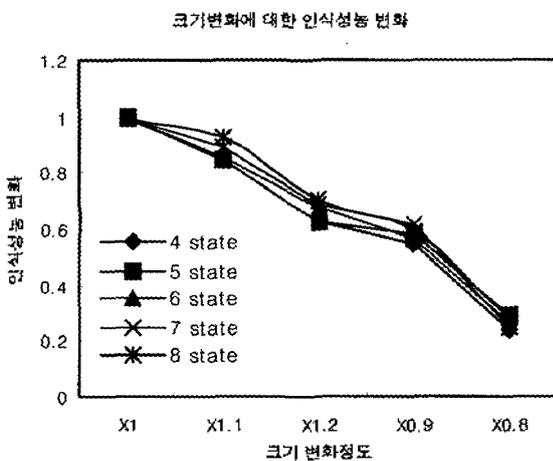


그림 6. 크기변화에 대한 인식성능 저하 결과
Fig. 6. The results of word recognition rate by scale variance.

표 3. 자동차내에서 데이터베이스 정보

Table 3. The information of the our database at vehicle environment.

정보	회자의 수	5명 (남성 20대)
	음성 데이터	22개의 독립된 단어
	반복횟수	12회
	데이터 조건	날씨
시간		3 (9시경, 12시경, 6시경)
조사 방향		4 (정면, 후면, 좌측면, 우측면)
Video	이미지 포맷	320×240 pexels (jpeg)
	Frame Rate	30Hz (30 frames/sec)
	Pixel resolution	8bits gray level
	총 이미지 수	43,384

V. 조명 변화에 의한 인식성능 저하

5.1. 동적환경에서 조명에 대한 데이터 수집

실제 자동차 내의 환경은 명암이나 조사 방향과 같은 다양한 조명의 변화로 인하여 입술 이미지 명암에 큰 영향을 미친다.

예를 들어 한 사람이 동일한 단어를 같은 위치에서 발음할 경우에도 아침과 오후의 명암 정도에 따라 다른 입술 이미지가 생성된다.

또한 햇빛의 조사방향에 대해서도 명암변화가 일어난다. 얼굴부위에서 입술은 돌출부이므로 정면에서나 후면에서 빛이 조사한 경우에는 전체 이미지의 명암이 어두워지거나 밝아진다. 하지만 측면에서의 빛이 조사할 경우는 조사하는 방향을 받는 면은 밝아지지만, 반대면은 상대적으로 어두워진다.

이런 관계로 자동차 내에서의 조명에 대해서 조사하였는데 저녁처럼 극단적으로 어두운 환경에서는 조사하지 않았다. 그리고 그늘진 곳도 피하였다.

논문에서는 자동차 내에서 다양한 빛의 방향에 따른 이미지를 얻기 위해 아침부터 저녁까지 시간대 별로 입술의 모션 이미지를 저장하였다. 시간은 각각 오전 9시, 12시, 18시로 설정하였으며 햇빛의 조사 방향은 정면, 후면, 좌우 측면 내 가지로 분할하였다.

5.2. 조명변화에 대한 요인

입술 안에서의 불규칙적인 명암 (Intensity)을 발생시키는 경우는 대부분 입을 벌리거나 이와 혀가 보이는 경우이다. 그리고 입술 외부에서는 화차마다 다른 피부색과 뺨의 주름이나 수염, 기미와 같은 요소 때문에 산란한

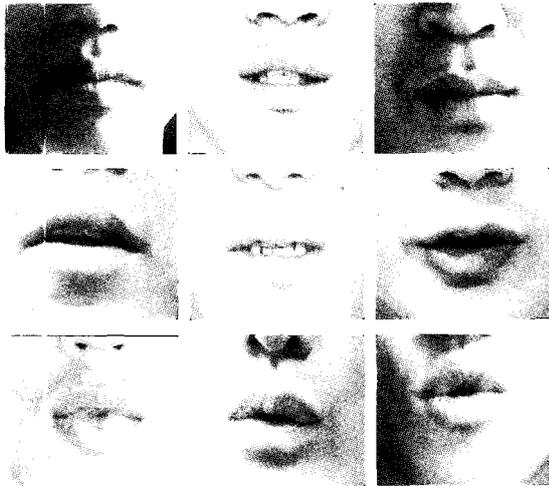


그림 7. 데이터베이스에서 샘플 이미지
Fig. 7. Example images from the our database.

명암잡음이 발생한다. 하지만 이런 요인들보다는 조명환경에서의 변화가 가장 큰 명암변화를 일으키는데 장소마다의 조명 강도가 다르고 자연 조명의 경우는 시간대별로 큰 차이를 가지고 있다.

실험을 위한 환경 설정 정보를 표 3에서 보여주고 있으며, 그림 7에서는 자동차 환경에서 구축한 데이터 베이스의 샘플 이미지를 보여 주고 있다.

조명환경이 립리딩에 미치는 영향을 실험적으로 증명한 결과를 아래에 기술하였다.

그림 8에서 보여주는 것처럼, 평범한 실내 환경 조건에서 자동차 내부 조건으로 빛의 조건의 변화함에 따라 인식률이 매우 급격하게 떨어지는 것을 알 수 있다. 결과는 실내 환경에서 립리딩의 인식성능을 100%로 잡고 자동차

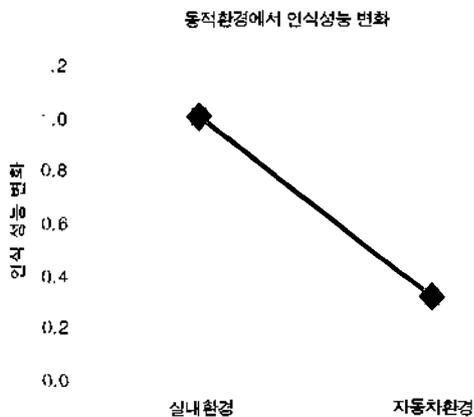


그림 8. 자동차 환경에서의 인식성능 저하 결과
Fig. 8. The results of word recognition rate in vehicle environment.

내에서 시간대 별로 빛의 조사 방향별로 변화에 대한 인식 성능과 비교한 것이다. 실험 결과에서 보여주듯이 인식성능이 70% 이상의 왜곡을 발생함을 알 수 있다.

상당히 좋은 환경에서 트레이닝된 립리딩 인식기는 자동차 환경에서 저장한 이미지를 이용하여 테스트를 했을 때 인식성능을 유지하는데 실패했다.

이는 립리딩 시스템이 자동차 환경에서와 같은 실외 조건에서 동작시켰을 때, 신뢰할 수 있는 성능을 보여준다고 장담할 수 없다는 것을 의미한다. 이것은 그림 9의 결과처럼 실험실 내에서 구축한 데이터에 명암을 임의적으로 가산을 가여 선행적인 명암이 인식성능에 미치는 것을 살펴봄으로써 확인할 수 있었다.

VI. 결론

본 논문에서는 공간적 변화와 빛에 의한 조명 변화와 같은 동적 환경에서 립리딩 인식성능 저하의 다양한 요인들과 립리딩 인식기의 인식률 사이의 관계를 연구하였다.

실험을 위해 우리는 자체 개발한 HMM기반의 립리딩 시스템을 사용하여 안정적인 실내 환경에서 데이터 베이스를 획득하고 그 후 이 데이터베이스에 이동, 회전, 크기의 변화와 같은 왜곡을 주어 공간적 변화에 의해 발생하는 인식률의 변화에 대해 테스트를 하였다. 그리고 조명의 변화가 발생했을 경우의 인식률 변화를 알아 보기 위해 아침에서 오후 늦은 시간까지 각 세 번에 걸쳐서 야외에 노출된 자동차의 내부에서 생성한 데이터베이스를 사용하여 립리딩 인식기를 테스트를 했다.

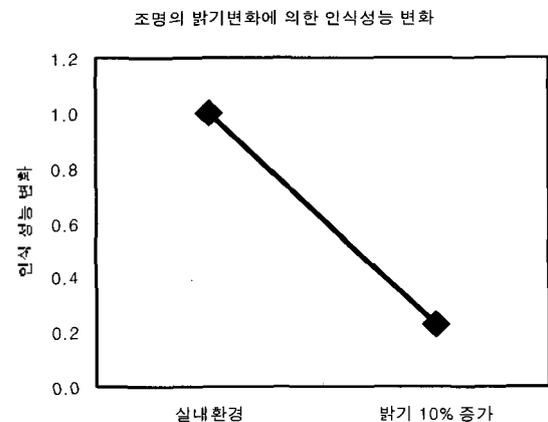


그림 9. 조명의 밝기 변화에 의한 인식성능 결과
Fig. 9. The results of word recognition rate in different illumination environment.

실험결과부터 우리는 화자의 입술 모양의 공간적인 변화가 인식 능력을 떨어뜨리는 큰 한 원인이 되었음을 발견할 수 있었다. 입술이 약 15° 정도 기울었을 때 약 20% 정도의 인식성능의 저하가 발생함을 알 수 있었다. 그러나 실제적으로 공간적 변화 요인보다 더 심각한 성능 저하의 원인 요소는 시간의 흐름에 따른 조명 조건의 변화가 발생하는 것이다. 단지 10%의 조명 밀도가 상승하여도 전체적인 시스템은 신뢰할 수 없는 상태가 되었다. 빛의 변화나 주위 환경의 변화가 거의 없는 실내에서의 인식률을 100%로 기준을 설정하고 야외에 노출된 자동차 내에서 실험을 했을 때와 비교하면 약 70% 이상의 왜곡이 발생됨을 알 수 있었다. 이 실험을 통해 신뢰할 수 있는 립리딩 시스템 구현을 위해서는 빛의 밀도나 조사방향을 포함한 다양한 변화 요인에 강인한 시스템의 개발이 요구됨을 알 수 있었다.

감사의 글

이 논문은 전남대학교 지역연구센터 RRC HECS에 의해 부분적으로 지원받았음.

참고 문헌

1. G. Potamianos, H. P. Graf, and E. Cosatto, "An image transform approach for HMM based automatic lipreading," *Processing Of the Int. Conf. On Image Processing*, 173-177, 1998.
2. C. Bregler and Y. Konig, "Eigenlips' for robust speech recognition," *Proc. IEEE Int. Conf. On Acoustics, Speech and Signal Processing*, 669-672, 1994.
3. T. Chen, H. P. Graf, and K. Wang, "Lip-synchronization using speech-assisted video processing," *IEEE Signal Processing Lett.*, 2, 57-59, 1995.

4. M. Liévin and F. Luthon "Lip features automatic extraction," *Proc. Of the 5th IEEE Int. Conf. On Image Processing*, Chicago, Illinois, 1998.
5. G. Potamianos, G. H. P. Graf, and E. Cosatto, "An image transform approach for HMM based automatic lipreading," *Image Processing, ICIP98, Proceedings, International Conference*, 3, 173-1777, 1998.
6. D.-S. Min, J.-Y. Kim, "Robustness of Lipreading against the Variations of Rotation, Translation and Scaling," *ITC-CSCC*, 1, 15-18, 2000.
7. D.-S. Min, J.-Y. Kim, "Robust Lip Extraction and Tracking of the Mouth Region," *ITC-CSCC*, 2, 927-931, 2000.
8. 박병구, 김진영, 임재열, "입술 파라미터 선정에 따른 바이모달 음성인식 성능 비교 및 검증," *한국음향학회지* 8 (3), 68-72, 1999.
9. 박병구, 김진영, 임재열, "입술 파라미터 선정에 따른 바이모달 음성인식 성능 비교 및 검증," *한국음향학회지*, 18 (3), 68-72, 1999.
10. 박병구, 김진영, 최승호, "바이모달 음성인식의 음성정보와 입술 정보 결합방법 비교," *한국음향학회지*, 18 (4), 31-37, 1999.

저자 약력

● **신 도 성 (Do Sung Shin)**



1993년 2월: 동신대학교 정보통신공학과 졸업
 1998년 2월: 전남대학교 대학원 전자공학과 (석사)
 1999년 3월 ~ 현재: 동대학원 (박사수료)

● **김 진 영 (Jin Young Kim)**

1986년: 서울대학교 전자공학과 졸업
 1988년: 서울대학교 전자공학과 석사
 1994년: 서울대학교 전자공학과 박사
 1994 ~ 1995년: 한국통신 소프트웨어 연구소
 1995 ~ 현재: 전남대학교 전자공학과 부교수

● **이 주 현 (Ju Hun Lee)**

1995년: 서울대학교 전자공학과 박사
 1997년 ~ 현재: 동아방송대학교 인터넷방송학과 교수
 2000년 3월 ~ 2002년 2월: 일본 와세다 대학 방문 연구
 * 주관심분야: 음성인식 및 화자인증, 멀티모달 다중 생체인식, 얼굴인식