

## 얼굴 특징점을 이용한 한국어 8모음 독화 시스템 구축

김선옥\*, 이경호\*

\*한라대학교 정보통신공학부

e-mail:sokim@halla.ac.kr

khlee@halla.ac.kr

### Design & Implementation of Speechreading System using the Face Feature on the Korean 8 Vowels

Sun Ok Kim\*, Kyong Ho Lee\*

\*School of Information & Communication Broadcasting Engineering, Halla University

#### 요 약

본 논문은 한국어 8 단모음을 인식하는 자동 독화 신경망 시스템을 구축한 것이다. 얼굴의 특징들은 휘도와 채도 성분으로 인하여 다양한 색 공간에서 다양한 표현 값을 갖는다. 이를 이용하여 각 표현 값들을 증폭하거나 축소, 대비시킴으로서 얼굴 특징들을 추출되게 하였다. 눈과 코, 안쪽 입의 외곽선, 이의 외곽선을 찾았고, 그 후 한국어 8모음 발화시 구분되게 변화하는 값들을 파라미터로 설정하였다. 한국어 8모음을 발화하는 2400개의 자료를 모아 분석하고 이 분석을 바탕으로 신경망 시스템을 구축하여 실험하였다. 이 실험에 정상인 5명이 동원되었고, 사람들 사이에 있는 관찰 오차를 정규화를 통하여 수정하였다. 5명으로 분석하였고, 5명으로 인식 실험하여 좋은 결과를 얻었다.

키워드 : 자동독화(Automatic Speechreading), 립리딩(Lipreading), 신경망(Neuralnet)

#### I. 서론

독화란 발화자의 시각적 정보를 분석하여 발화를 읽는 기술이다. 오래전부터 청각 장애인들을 위한 구화 교육에서 제한된 범위나마 입의 모양으로 발화를 이해할 수 있는 교육들을 실시하고 있다[1]. 청각 장애인뿐 아니라 정상인도 화자의 얼굴에서 얻는 시각 정보가 음성 인식에 도움을 준다는 것은 널리 알려진 사실이다. Sumby 등은 잡음 환경에서 수행하는 대화에서 시각적 정보가 인식에 이득이 있음을 알아냈고[2], McGurk 등은 대화의 인식에 시각적 정보와 음향적 정보가 융해됨을 알아내고 'McGuck 현상'이라 명명하였다[3]. Summerfield도 독화가 발화자의 국소화 기능, 발화의 범위 설정 및 조음 장소 정보를 제공하는 중요한 요소임을 보였다[4]. 그 밖의 많은 연구에서 독화는 음성인식에 매우 큰 기여를 할 수 있음을 명백히 하고 있다.

#### II. 관련 연구

초기 연구자인 Petajan은 얼굴을 찾은 것을 전제한 영상에서 2진화하여 입의 높이와 폭, 넓이, 입의 원주를 계산하고 보기로 준비된 템플릿과 연결시키는 방법으로 자동 독화를 수행하였다[5,6]. Finn과 Montgomery는 표식자를 사용하여 입술 사이, 입술과 코 사이, 입술과 턱 사이의 거리를 수작업으로 추적하고 가중치 유클리디안 거리 측정 방법 알고리즘으로 독화를 수행했다[7]. Stork 등은 입주위와 코와 턱 상의 마커를 이용하여 다섯 개의 거리를 측정하여 이를 이용하여 독화하였다[8]. Yuhus 등은 입주위 20x25 그레이 스케일 이미지를 별다른 처리 없이 신경망 시스템에 직접 제공하여 독화하였다[9]. Mase 등은 연속된 두 화상에서 입 주위 상하 좌우 이미지들의 광학적 흐름을 계산하여, 입을 중심으로 한 사각형에서 여덟 개 요소의 평균값을 계산하고, 입의 열리고

단히는 평균을 기초하여 분류되는 벡터 값으로 독화하였다 [10]. Silbee도 입 주위 80x80 이미지를 벡터 양자화 하여 독화하였다[11]. Bregler는 입 주위 13x13 영역을 Fourier 변환을 통해 파라미터로 사용하였으며, 나중에는 active contour model 화상에서 80개의 파라미터로 입모양을 추출하여, 맵핑하는 방법으로 독화하였다[12]. Chiou는 칼라 이미지로부터 특징점을 추출하는 주성분분석과 active contour model 두 개의 기술을 결합하여 매 프레임에서 두 개의 특징 벡터를 추출하였는데 하나는 주성분분석을 통하여 하나는 active contour model을 통하여 계산하여 독화하였다[13]. Adjoudani 등은 머리에 장착시킨 카메라를 이용하여 입력되는 영상 전체를 이용하여 독화하였다[14]. Meier 등은 통계적 피부색 모델을 이용하여 얼굴 영역을 추적한 다음 입술 영역을 추출하고 입술의 크기를 24x18로 정규화를 하고 그 픽셀 그대로 인식기의 데이터 벡터로 사용하거나 주성분 분석이나 선형 판별 분석하여 구한 16개의 개수를 파라미터로 사용하여 독화하였다[15]. 김진범 등은 카메라 입력을 입주위로 제한하여 이미 입을 찾은 것으로 하고, 이진화를 통한 입 찾기와 이미지기반의 이산 코사인 변환과 이산 웨이블릿 변환을 이용한 파라미터 추출과 주성분 분석을 통한 독화를 수행하였다 [16,17]. 이지은 등은 코에서 턱까지만 입력하여 그레이레벨에서 선형 명암 마스크를 이용하여 영상 보정을 하고 크기를 줄여 주성분 분석과 은닉 마르코브 모델을 이용하여 자동 독화를 수행하였다[18,19]. 민덕수는 입술로 제한된 범위 내에서 입술 움직임의 가지고 단어를 인식하기 위한 연구를 하였다[20]. 민소희 등은 얼굴에 마커를 붙이고, 거울을 이용하여 정면과 측면을 관측하는 노력을 하였다. 입술의 폭과 높이, 윗 입술에서 턱까지의 거리, 코로부터 윗입술까지 거리, 코부터 턱까지의 거리 등의 데이터를 이용하여 자동 독화를 수행하였다[21,22]. 백준준 등은 센서를 부착한 카메라로 입술의 높이, 폭, 안과 바깥의 입술 경계, 윤곽의 파라미터로 입술 정보를 추출하여 음성인식 보조 자료로 이용하였다[23]. 신도성 등은 입술 영역으로 제한된 입력으로 은닉 마르코브 모델을 이용하여 인식을 하였다[24-27]. 서재영은 다양한 각도로 입력되는 영상으로부터 자동 독화를 하려고 하였다[28].

### III. 본론

#### 3.1 자동 독화 과정

자동 독화 전체 과정은 '입력 영상에서 얼굴 검출 작업, 특징 점을 위한 얼굴 요소 검출 작업, 정규화 및 파라미터 추출 작업, 발화 인식'으로 구성되어 있다. '입력 영상에서 얼굴 검출 및 특징 검출 작업'은 다양한 세부 작업으로 나누어진다. 이 과정은 먼저 사전 분석 작업을 선행한다. 180명의 증명사진을 통해 두 눈 사이를 기준으로 두 눈 중점으로부터 코끝과 윗입술 턱 끝까지 분포를 조사하여 항 후 영상 처리 과정에서

통계 자료를 바탕으로 접근하여 정확성을 배가하고 계산량을 줄였다. 또한 눈, 코, 입의 다양한 색공간 통합 및 단일 성분에서 표현 상태를 분석하여 동일 요소가 색 공간에 따라 구별되는 것을 파악하여 증폭 또는 축소, 대비로 특징 점 추출이 용이하도록 하였다. 이후 과정은 컴퓨터를 이용한 처리 과정으로 얼굴 추출 과정은 피부색 추출이 용이하도록 휘도 성분에 의존적인 보정을 하여 어두운 곳은 좀 밝게 조정하고 너무 밝은 곳은 좀 어둡게 되도록 조명 보정하였다. 피부색 추출은 Rein-Lien Hsu의 피부색 추출 방법을 이용하여 입력 영상에서 피부색이 추출되도록 하였으며[29]. 모폴로지 연산을 통한 잡음의 제거 및 돌출 픽셀의 정리 전처리 작업을 하여 외곽을 정리하였으며, Jankowski의 8-연결 Two Pass CCL 함수 이용하여 피부색 블록을 구성되게 하고, Graham scan algorithm을 이용하여 불록다각형 형성하여 얼굴 후보 마스크를 구성하였다. 이 과정까지 수행하면 얼굴 후보 마스크의 범위 내에는 눈 코 입이 포함된다. 따라서 이후 얼굴 구성 요소를 추출하는 모든 작업은 얼굴 후보 마스크에서 수행한다. 눈은 YCbCr 색공간의 Cb상에서 높은 값을 갖는 점과 Cr상에서 낮은 값을 갖는 점을 이용하여 두 단일 영상을 대비시킨 채도상의 결과와 휘도상 Y에서 눈동자가 낮은 값을 갖는 점과 흰자위가 높은 점을 갖는 점을 이용하여 한 번은 밝은 색이 번지도록 한 모폴로지 결과와 어두운 색이 번지도록 한 결과를 대비시킨 휘도 상의 결과를 결합하여 눈을 찾는다. 눈을 찾은 후에는 앞에서 분석한 통계 정보를 이용하여 더 축소된 범위에서 작업한다. 입은 채도 성분이 약한 입안과 보통의 피부와 구별되는 입술의 채도 성분을 이용하여 대략적 영역인 입 마스크를 형성하고, 형성된 입마스크 안에서 Canny 외곽선 추출 후 모폴로지 연산을 통한 후처리로 입 안쪽 경계선, 이의 외곽선을 추출한다. 마지막으로 눈의 위치와 입의 위치를 참고하여 통계 정보 범위 내에서 코의 대략적 범위를 선정하며 Canny 외곽선 추출을 이용한 외선 길이 분포 정보를 통한 코 위치 추출한다. 이 모든 작업은 이경호 등의 "색상 정보를 이용한 자동 독화 특징 추출"의 방법으로 구성하였다[29].

표 1. 얼굴 특징점 거리 분석

구분	코끝	윗 입술	턱끝
최소	65.6%	86.3%	169.2%
최대	99.2%	126.2%	226.1%
평균	80.0%	104.2%	195.8%
표준편차	6.4%	9.2%	11.6%
범위	33.6%	40.0%	56.9%



원영상

조명보정후 영상

피부색 추출



피부색 추출 후처리 TwoPassCCL 이용한 연산 결과 화면(8-연결) 얼굴마스크구성



마스크적용영상 Cb<sup>2</sup> 영상 (255-Cr)<sup>2</sup> 영상



EyeMapChroma 결과 영상 : Cb, Cr 성분이 비슷한 곳은 검게 출력되었다. 스케일 스페이스 팽창 연산 후 영상 스케일 스페이스 침식 연산 후 영상



EyeMapChroma와 EyeMapLuma의 'AND' 연산 결과 밝기 순 상위 10% 이진화 영상 눈의 대칭성과 위치를 이용한 눈찾기 결과 영상



형성된 입 마스크 피부색 안으로 제한된다. 입 찾기 입의 범위 내에서 이의 외곽선 찾기 결과



코찾기 턱 찾기 특징 점 추출

그림 2. 특징 점 추출 과정

### 3.2 파라미터 추출

앞의 과정을 통하여 추출할 수 있는 파라미터는 입의 면적, 입의 폭과 높이, 이 보임, 코와 입술 상하단과의 거리 등이다. 한국어 '아/에/애/이/어/우/오' 8모음 인식을 위하여 추출된 영상은 정상인 5명으로부터 각각 8모음 발화이미지 60 세트 2400개의 영상이다. 영상 취득은 컴퓨터에 연결된 카메라로 하였고, 카메라 거리는 두 눈간 거리가 대략 90~100 픽셀 정도이도록 하였다. 영상 취득시 정확한 발음을 위하여 발화 방법에 대한 설명하고 발화 연습을 한 후 3초 이상 발화하게 하며 발화 영상을 취득하였다. 2400개의 영상에서 눈 찾기 실패가 118장, 입 영역 찾기 실패가 144장, 입 외곽선 추출 실패가 12장 있었다. 입영역 찾기 실패는 주로 '우' 와 같이 발화 시 입의 굴곡이 심할 때의 영상들이었다.

입 외곽선 추출 실패는 혀가 관찰되는 영상에서 혀와 입술의 영역을 구분하지 못하여 발생하였다. 파라미터 추출은 총 11.4%의 실패가 발생하였다. 취득한 영상에 대한 두 눈 간 거리의 기계적 특징 점 추출에 의한 통계는 평균 94.8픽셀, 표준편차 4.2픽셀 최고 105.3픽셀 최저 86.4픽셀 이었다. 또한 평균으로부터 좌우 10% 범위 내에 94.8%의 데이터가 있었다. 1 표준 편차 거리 내에 84.6%의 자료가 있었고, 2 표준 편차의 거리 내에 98.7%의 데이터가 있었다. 파라미터 정규화는 두 눈 사이의 평균 거리를 기준으로 비례식을 이용하여 수행하였고, 추출한 얼굴 특징 점으로부터 구성된 파라미터는 입 면적, 입 면적 대비 이 보임 비, 입 높이, 입 너비, 코끝과 입술 상단 간의 거리, 입술 하단 간의 거리이다. 동 자료들 간의 의미는 개인 간의 약간의 차이가 있었으나 평균의 차이가 모두 10% 이내이었고 분포도 역시 유사하였다. 표 2와 그림 3은 발화자 #1과 #2에 대한 통계를 표와 그래프로 표현한 것이다.

표2. 발화자 #1에 대한 파라미터 별 통계

입 면적 통계								
	아	에	애	이	어	으	우	오
Ave	821.9	464.7	754.9	476.3	479.9	392.1	138.4	81.8
Stdev	48.5	52.2	66.1	57.0	26.3	13.0	25.0	10.0
Max	889.0	546.0	841.0	516.0	517.0	418.0	183.0	96.0
Min	759.0	392.0	660.0	318.0	443.0	377.0	100.0	70.0
입면적 대비 이 외곽선 길이 비								
	아	에	애	이	어	으	우	오
Ave	0.032	0.010	0.011	0.104	0.009	0.067	0.006	0.028
Stdev	0.023	0.008	0.005	0.011	0.009	0.025	0.012	0.036
Max	0.054	0.023	0.020	0.122	0.023	0.106	0.031	0.094
Min	0.000	0.000	0.000	0.081	0.000	0.034	0.000	0.000

\*. 분포도 계급은 100배 한 것임

입 높이								
	아	에	애	이	어	으	우	오
Ave	16.7	9.1	14.4	10.2	12.7	7.1	6.8	3.50
Stdev	1.7	1.1	1.1	0.8	0.7	0.8	2.3	0.53
Max	18.0	11.0	16.0	11.0	14.0	8.0	12.0	4.00
Min	14.0	8.0	13.0	9.0	12.0	6.0	4.0	3.00

입 너비								
	아	에	애	이	어	으	우	오
Ave	63.5	68.0	66.2	72.8	50.0	73.9	26.3	18.9
Stdev	2.4	2.7	0.8	1.3	2.7	0.9	2.1	1.5
Max	67.0	72.0	67.0	75.0	53.0	75.0	29.0	21.0
Min	61.0	64.0	65.0	71.0	45.0	73.0	23.0	17.0

코와 입술 상단 거리								
	아	에	애	이	어	으	우	오
Ave	38.9	37.9	36.9	34.0	39.5	37.8	32.2	42.3
Stdev	3.7	3.4	4.7	2.5	1.7	2.1	0.6	5.9
Max	43.0	42.0	42.0	37.0	41.1	40.0	33.0	49.0
Min	30.0	34.0	31.0	31.0	35.1	33.0	31.0	30.1

	아	에	애	이	어	으	우	오
Ave	55.6	47.0	51.4	44.2	52.2	44.9	38.1	47.4
Stdev	3.2	3.6	4.7	2.2	1.8	2.3	0.6	4.0
Max	59.0	52.0	57.0	47.0	54.1	46.0	39.0	52.0
Min	48.0	42.0	45.0	41.0	48.0	40.0	37.0	40.0

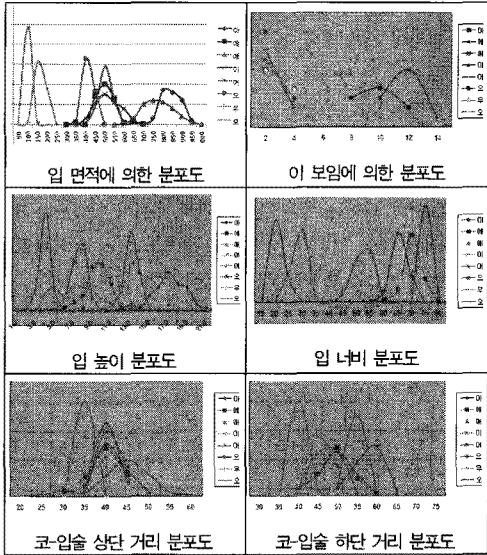


그림 3. 발화자 #2에 대한 파라미터 별 분포도

두 개의 정보들을 살펴보면 유사함을 알 수 있다. 그림 3의 '입 면적 분포'를 보면 '오/우'는 면적이 작아 왼편에 위치하며, '이/에'는 오른편에 위치함을 관찰 할 수 있다. '입 면적 대비 이 보임 비율'을 보면 '이/애'가 다른 집단과 다르게 오른쪽에 따로 분포되어 있음을 알 수 있다. '입 높이'와 '입 너비'도 비교적 잘 구분된다. 그러나 코끝과 입술 상단 하단과의 거리는 예상보다 구분성이 떨어지고 있다. 이는 입술의 바깥 경계선이 아닌 입의 안쪽 경계선으로부터 거리를 측정하였기 때문으로 사료되며, 특징 점 코끝의 좌표가 일관된 위치를 갖지 못함에 기인한 것으로 보인다. 그러나 다른 자료와 동시에 살펴보면 개연성이 반영될 수 있음을 알 수 있다.

### 3.3 인식 시스템 구축

신경망은 사람의 인지 과정을 흉내 내어 패턴 인식 문제를 해결해 보려는 시스템이며, 환경의 변이에 적응하는 능력이 있으며, 병렬 제약조건을 만족시키는 문제를 해결할 적합한 구조를 가지고 있어 인식 시스템은 신경망으로 구축하였다.

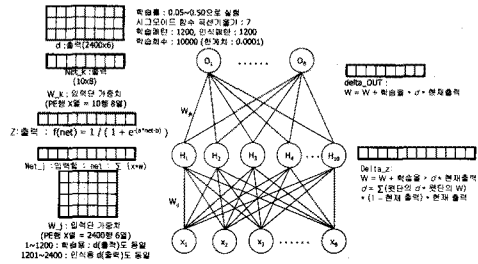


그림 4. 구축된 신경망

구축된 신경망은 입력층 6, 은닉 10, 출력 8층 24 유닛으로 구성하였으며, 학습률은 0.05~0.50, 시그모이드 함수 곡선 기울기에 반영되어 수렴을 조절할 모멘텀은 7, 학습은 20000회 또는 에러 값 경계치 0.0001이하까지로 구성하였다.

## IV. 성능평가 및 분석

인식 실험은 1인 10세트로 훈련하고 5인 각 30세트로 인식하기(실험 #1), 각각 20세트로 훈련하고 20세트로 인식하기(실험 #2), 5인 각각 10세트로 훈련하고 5인 각각 30세트로 인식(실험 #3)하는 3가지 실험을 하였다. 실험 #1의 의미는 파라미터가 화자 독립적인가를 확인하기 위한 것이며, 실험 #2의 의미는 화자 종속일 경우 인식률을 보고자 하였으며, 실험 #3은 여러 사람의 자료로 훈련하였을 경우 인식률을 보고자 하였다. 실험 #1의 훈련과 인식 평균 제곱 오차 변화도는 그림 5와 같으며, 훈련 오차는 0.093651였으며, 인식 세트에 대한 평균 제곱 오차는 0.222551, 최고 인식률은 90%, 평균 인식률은 87.4%를 보였다. 실험 #2의 훈련과 인식 평균 제곱 오차 변화도는 그림 6과 같으며, 인식률은 최고 92.1%, 평균 90.2%의 결과를 얻었다. 실험 #3의 훈련과 인식 평균 제곱 오차 변화도는 그림 7과 같으며, 89.3%의 결과를 얻었다. 표 3에서와 같이 모든 실험의 결과에서 대부분 '아/에', '이/으' 간에 인식 혼란을 일으켰다.

표 3. 실험1의 인식 결과

	아	에	애	이	어	으	우	오
아	25		5					
에		30						
애	4		26					
이				17		13		
어					30			
으				1		29		
우							29	1
오								30

는 본 실험에서 효율적인 파라미터를 추출하였음을 보이고 있는 증거라 하겠다.

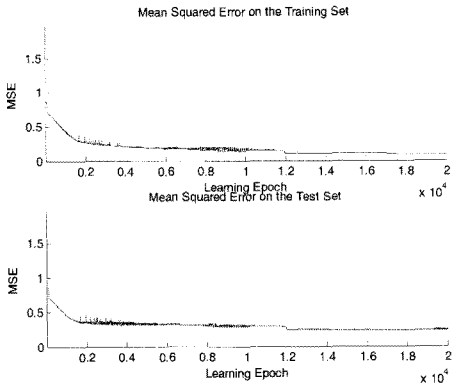


그림 5 그림 5 실험 #1의 훈련 셀과 실험 셀의 MSE

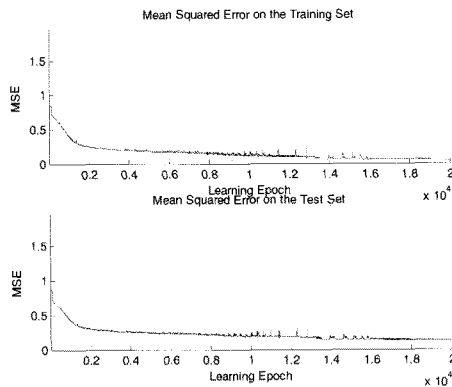


그림 6 실험 #2의 훈련 셀과 실험 셀의 MSE

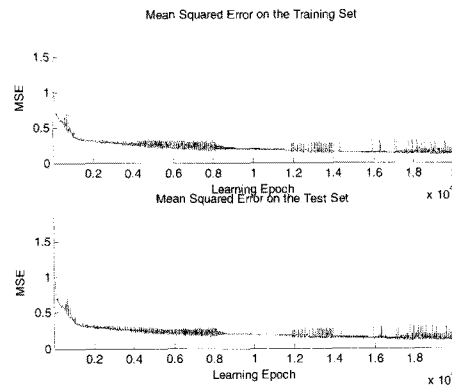


그림 7. 실험 #3의 훈련 셀과 실험 셀의 MSE

동 실험과 유사한 실험으로는 서재영[28]과 이경호[31]의 '아/에/이/오/우' 5모음 인식 실험이 있다. 단순히 입의 폭과 높이만을 이용하여 인식한 서재영의 70% 인식보다는 매우 높은 인식률을 보이고 있으며, 입의 폭과 너비, 코와 입술 상단, 코와 입술 하단, 코와 턱의 거리를 파라미터로 한 이경호의 91.1% 보다 3개나 더 많은 8개 모음을 인식하는 데도 과히 떨어지지 않는 89.3% 인식률을 보이고 있다. 이

표 4. 독화 결과 비교

연구	인식대상	파라미터	인식률	비교
[28]	한국어 5모음 '아/에/이/오/우'	입의 폭, 입의 너비(2개)	70%	입주위만 영상인력
[31]	한국어 5모음 '아/에/이/오/우'	입의 폭, 입의 너비, 코끝과 입 상단 거리, 코끝과 입 하단 거리, 코끝과 턱끝 거리(5개)	91.1	다양배경 다양조명
This	한국어 8모음 '아/에/애/이/어/ 오/우'	입의 폭, 입의 너비, 입 면적, 이 보임 정도, 코 끝과 입 상단 거리(6개)	89.3	다양배경 다양조명

### V. 결론

본 논문에서는 한국어 8 단모음의 기계적 인식을 위한 자동 독화 시스템 구축을 수행하였다. 컴퓨터에 의한 영상 처리 과정을 통해 두 눈의 위치와 안쪽 입의 외곽선, 이(齒)의 경계선과 코의 위치를 추출하고 이것으로 입의 면적, 입의 폭과 높이, 이 보임 정도, 코와 안쪽 입술 상하단과의 거리를 파라미터로 사용할 수 있는지를 분석하고 이를 바탕으로 자동 독화 신경망 시스템을 구축하여 화자 중속에서는 화자 독립 인식실험을 하였다. 모든 실험에서 90% 내외의 인식률을 얻어 파라미터와 시스템의 효율성을 증명하였다.

본 연구가 음성인식 시스템의 인식률을 높이기 위한 방편으로 수행한 연구이므로 정지 영상에서 정보처리 뿐 아니라 동 영상에서 정보 취득 연구가 필요하다. 아울러 본 연구는 파라미터를 적절히 이용하면 휴먼컴퓨터 인터페이스와 컴퓨터 보안 등에 응용할 수 있다. 현재 비교적 좋은 결과를 보이고 있으나 다양한 환경에서 사람의 눈과 같이 작동할 수 있는 좀 더 정밀한 특징점 추출 기술이 필요하며 음성 신호 처리와 결합하는 연구가 필요하다.

### 참고문헌

- [1] 최병문, 구화교육, 한국구화학교, 1970
- [2] Bailly, G., Vatikiotis-Bateson, E., & Perrier, P., Issues in Visual and Audio-Visual Speech Processing, MIT Press, 2004.
- [3] MCGurk, H., & MacDonald, J., "Hearing lips and seeing voices," Nature, 264:746-748, 1976.
- [4] Summerfield, A. Q., "Some preliminaries to a comprehensive account of audio-visual speech

- perception," *Hearing by Eye: The Psychology of Lip-Reading*. London, United Kingdom: Lawrence Erlbaum Associates, 3-51, 1987.
- [5] Petajan, E. D., "Automatic lipreading to enhance speech recognition," Ph.D. Dissertation, University of Illinois at Urbana-Champaign, 1984.
- [6] Petajan, E. D., "Automatic lipreading to enhance speech recognition," *Proceedings of the IEEE Communication Society Global Telecommunications Conference*, November 26-29, Atlanta, Georgia, 1984.
- [7] Finn, E. K. & Montgomery A.A. "Automatic optically based recognition of speech," *Pattern Recognition Letters*, 8(3):159-164, 1988.
- [8] Stork, D. G., & Hennecke, M.E., *Speechreading by Humans and Machines*, Berlin, Germany: Springer, 1996.
- [9] Yuhas, B. P., Goldstein, M.H. & Sejnowski, T.J., "Integration of acoustic and visual speech signals using neural networks," *IEEE Communications Magazine*, 27:65-71, 1989
- [10] Mase, K. & Pentland, A., "Automatic lipreading by computer," *Trans. Inst. Elec. Info. and Com. Eng.*, J73-D-II(6):796-803, 1990.
- [11] Silsbee, P. L., "Computer Lipreading for Improved Accuracy in Automatic Speech Recognition," Ph.D. dissertation, The University of Texas at Austin, 1993.
- [12] Bregler, C., Omohundro, S. M. & Konig, Y., "A Hybrid Approach to Bimodal Speech Recognition," in *28th Annual Asilomar Conference on Signals, Systems, and Computers*, 1:556-560, 1994.
- [13] Chiou, G. I. & Hwang, J. N. "A neural network based stochastic active contour model (NNS-SNAKE) for contour finding of distinct features," *IEEE Trans. on Image Processing*, 4(19):1192-1195, 1995.
- [14] Adjoudani, A. et al., "A multimedia platform for audio-visual speech processing," *Proc. European Conference on Speech Communication and Technology*, Rhodes, Greece, 1671-1674, 1997.
- [15] Meier, U., Stiefelhagen, R., Yang, J., Waibel, A., "Towards Unrestricted Lipreading," *International Journal of Pattern Recognition and Artificial Intelligence*, 14(5):571-785, 2000.
- [16] 김진범, 김진영, "이미지 변환과 HMM에 기반한 자동 립리딩", *대한전자공학회 추계학술대회 논문집*, 22(2):585-588, 1999.
- [17] 김진범, 김진영, "입술 대칭성에 기반한 효율적인 립리딩 방법", *전자공학회논문지*, 37(5):455-464, 2000.
- [18] 이지은, 김진영, 이주현, "시간영역 이미지 필터링에 의한 립리딩 성능 향상", *한국음향학회 학술발표대회논문집*, 20(2):45-48, 2001.
- [19] 이지은, 시간영역 이미지 필터링에 의한 립리딩 성능 향상, *전남대학교대학원 석사학위논문*, 2002.
- [20] 민덕수, "동적 환경에서 립리딩 성능저하 요인 분석 및 인식성능 향상에 관한 연구", *전남대학교대학원 석사학위논문*, 2002.
- [21] 민소희, 김진영, 최승호, "입술 정보를 이용한 음성 특징 파라미터 추정 및 음성 인식 성능 향상", *대한음성학회지*, 44:83-92, 2002.
- [22] 김진영, 민소희, 최승호, "음성인식에서 입술 파라미터 열화에 따른 견인성 연구", *음성과학*, 10(2):27-33, 2003.
- [23] 백성준, 김진영, "입술정보 및 SFM을 이용한 음성의 음절향상알고리즘", *음성과학*, 10(2):77-84, 2003.
- [24] 신도성, "입술영상접기와 프레임간 필터링을 이용한 립리딩 성능 개선", *전남대학교대학원 박사학위논문*, 2004.
- [25] 김진영, 신도성, "상태공유 HMM을 이용한 서브워드 단위 기반 립리딩", *음성과학*, 8(3):123-131, 2001.
- [26] 신도성, 김진영, 최승호, "시간영역 필터를 이용한 립리딩 성능향상에 관한 연구", *한국음향학회*, 22(5):375-382, 2003.
- [27] 신도성, 김진영, 이주현, "동적 환경에서의 립리딩 인식 성능저하 요인분석에 대한 연구", *한국음향학회*, 21(5):471-477, 2002.
- [28] 서재영, "단순 특징값과 활용 각도에 따른 한국어 모음의 오디오 비주얼 인식에 관한 연구", *성신여자대학교대학원 석사학위논문*, 2004.
- [29] Hsu, R. L, Abdel-Mottaleb, M., Anil K. J., "Face Detection in Color Images," *IEEE Trans. on Pattern Analysis*, 24(5):696-706, 2002.
- [30] 이경호, 양룡, 이상범, "색상 정보를 이용한 자동 독화 특징 추출", *한국컴퓨터정보학회 논문지*, 13(6):107-116, 2008
- [31] 이경호, 금종주, 이상범, "한국어 5모음의 조음적 제어 분석을 이용한 자동 독화에 관한 연구", *컴퓨터산업교육학회*, 8(4):281-288, 2007