韓國컴퓨터情報學會 論文誌
第14卷 第3號, 2009. 3.

2009-14-3-4-4

한국어 8모음 자동 독화에 관한 연구

이 경 호*, 양 룡**, 김 선 옥*

A Study on Speechreading about the Korean 8 Vowels

Kyong Ho Lee *, Ryong Yang **, Sun Ok Kim*

요 약

본 논문은 한국어 8단모음을 인식하기 위한 효율적인 파라미터의 추출과 자동 독화 시스템의 구축에 관하여 연구한 것이다. 얼굴의 특징들은 다양한 칼라 공간에서 다양한 값으로 표현되는 것을 이용하여 각 표현 값들을 증폭하거나 또는 축소, 대비시켜 얼굴 요소들이 추출되도록 하였다. 눈과 코의 위치, 안쪽 입의 외곽선, 윗입술의 상단, 이의 외곽선을 특징 점으로 찾았으며, 이를 분석하여 안쪽 입의 면적, 안쪽 입의 높이와 폭, 이의 보임 비율, 코와 윗입술 상단과의 거리를 파라미터로 사용하였다. 2400개의 영상으로 분석하였고 이 분석을 바탕으로 신경망 시스템을 구축한 후 인식 실험을 하였다. 정상인 5명이 동원되었고, 사람들 사이에 있는 관찰 오차를 정규화를 통하여 수정하였으며, 실험하여 파라미터의 유용성 관점에서 좋은 결과를 얻었다.

Abstract

In this paper, we studied about the extraction of the parameter and implementation of speechreading system to recognize the Korean 8 vowel. Face features are detected by amplifying, reducing the image value and making a comparison between the image value which is represented for various value in various color space. The eyes position, the nose position, the inner boundary of lip, the outer boundary of upper lip and the outer line of the tooth is found to the feature and using the analysis the area of inner lip, the hight and width of inner lip, the outer line length of the tooth rate about a inner mouth area and the distance between the nose and outer boundary of upper lip are used for the parameter. 2400 data are gathered and analyzed. Based on this analysis, the neural net is constructed and the recognition experiments are performed. In the experiment, 5 normal persons were sampled. The observational error between samples was corrected using normalization method. The experiment show very encouraging result about the usefulness of the parameter.

▸ Keyword : 자동독화(Automatic Speechreading), 립리딩(Lipreading), 신경망(NeuralNet)

• 제1저자 : 이경호 교신저자 : 김선옥
• 투고일 : 2009. 1. 28, 심사일 : 2009. 2. 12, 게재확정일 : 2009. 3. 16.
* 한라대학교 정보통신방송공학부 교수 ** 인하공업전문대학 컴퓨터정보과 교수
※ 이 논문은 2009년 한국컴퓨터정보학회 제39차 동계학술대회에 발표한 "얼굴 특징점을 이용한 한국어 8모음 독화 시스템"을 확장한 것임.

I. 서 론

독화란 발화자로부터 시각적 정보를 분석하여 발화를 읽는 기술이다. 오래전부터 청각 장애인들을 위한 구화 교육에서 제한된 범위나마 입의 모양만으로 발화를 이해할 수 있는 교육들을 실시하고 있다[1]. 청각 장애인뿐만 아니라 정상인이라도 화자의 얼굴에서 얻는 시각 정보가 음성을 지각하는 동안에 음성 정보와 통합되어 인식에 도움을 준다는 것을 밝혔다. Sumby 등이 잡음 환경에서 수행하는 대화에서 시각적 정보가 발화의 인식에 이득이 있음을 알아냈다[2]. McGurk 등은 대화에서 시각적 정보와 음향적 정보가 융해됨을 알아내고 융해 현상을 'McGuck 현상'이라고 명명하였다[3]. Summerfield도 독화가 발화자의 국소화 기능, 발화 구간 설정 기능, 조음 장소를 알아내는 중요한 요소임을 보였다[4]. 그 밖의 많은 연구에서 독화는 음성인식에 매우 큰 기여를 할 수 있음을 명백히 하고 있다.

본 논문은 한국어 단모음을 인식하기 위한 효율적인 파라미터의 추출로 합당한 인식 시스템의 구축을 실험한 연구이다. 인식 대상은 한국어 '아/에/애/이/어/으/우/오' 8 단모음으로 하였으며, 구축된 시스템의 대략적 처리 과정은 컴퓨터에 연결되어 있는 카메라에서 컬러 영상을 취득하고, 얼굴의 각 요소들이 다양한 색 공간에서 다양한 값으로 표현되는 점을 이용하여 각 값들을 수학적으로 증폭 또는 축소, 대비 시켜 얼굴 요소들이 추출되도록 하고, 발화시 구분되게 변화하는 값들을 파라미터로 하여 한국어 8모음을 인식하는 신경망 시스템을 구축하였고 인식 실험하여 파라미터와 시스템의 유용성을 보였다.

II. 기존 연구

자동 독화 초기 연구자인 Petajan은 얼굴을 찾은 것을 전제한 단일 휘도 영상을 2진 이미지로 변화시켜 입안은 검게 만들어 입의 높이와 폭, 넓이, 입의 원주를 계산하는 파라미터를 추출하여 보기로 준비된 템플릿과 연결시키는 방법으로 자동 독화를 수행하였다[5,6]. Finn과 Montgomery는 특별한 표식자를 사용하여 입술 사이, 입술과 코 사이, 입술과 턱 사이의 다양한 거리를 수작업으로 추적하고 가중치 유클리디안 거리 측정 방법 알고리즘으로 최근접 이웃을 찾기 알고리즘으로 독화를 수행했다[7]. Stork 등도 입주위와 코와 턱 상의 마커를 이용하여 변화하는 다섯 개의 거리를 측정하였고 이를 파라미타로 독화하였다[8]. Yuhas 등은 입주위 20x25 그레이 스케일 이미지를 별다른 처리 없이 신경망 시스템에 파라미타로 직접 제공하여 독화하였다[9]. Mase 와 Pentland는 연속된 두 화상에서 입 주위 상하좌우 이미지들의 광학적 흐름을 계산하였다. 각 장면마다 입을 중심으로 한 사각형에서 x, y의 여덟 개의 요소의 평균값을 계산하였고 수평과 수직으로의 확장뿐만 아니라 입의 열리고 닫히는 평균을 기초로 하여 분류에 사용되는 벡터 값으로 파라미터로 사용하여 독화하였다[10]. Silbee는 조명의 차이를 해결하기 위해 전처리와 정규화를 하고, 입 주위의 이미지를 80x80의 벡터 양자화 하여 파라미터로 사용하여 독화하였다[11]. Bregler는 입 주위 13x13 영역을 Fourier 변환을 통해 파라미터로 사용한 바 있으며, 나중에는 active contour model 화상에서 80개의 파라미터로 묘사되는 입모양을 추출하여, 입 모양을 맵핑하는 방법으로 독화하였다[12]. Chiou는 칼라 이미지로부터 특징 점을 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 발화자의 입 주위 영역을 견고하게 추출하기 위하여 머리에 장착시킨 카메라를 이용하여 자동 독화를 수행하는 멀티미디어 플랫폼에서 발화자를 찾는 과정을 생략하여, 입력되는 영상 전체가 특징 파라미터를 추출하는 이미지로 직접 이용하여 독화하였다[14]. Meier 등은 통계적 피부색 모델을 이용하여 얼굴 영역을 추적한 다음 입술 영역을 추출하고 입술의 크기를 일정한 형태로 변환하는 방법으로 입술 이미지의 크기는 24x18로 정규화를 하고 그 픽셀 그대로 인식기의 데이터 벡터로 사용하거나 주성분 분석이나 선형 판별 분석을 적용하여 구한 16개의 개수를 파라미타로 사용하여 독화였다[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 입력 되게 촬영한 이미지로 그레이레벨에서 선형 명암 마스크를 이용하여 영상 보정을 하고 다운 샘플링하여 크기를 줄여 주성분 분석과 은닉 마르코브 모델을 이용하여 자동 독화를 수행하였다[18,19]. 민덕수는 입술로 제한된 범위 내에서 이미지에 기반을 둔 방법으로 입술 움직임을 가지고 단어를 인식하기 위한 연구를 하였다[20]. 민소희와 김진영은 1인의 얼굴에 마커를 붙이고, 정면 얼굴과 옆얼굴을 동시 확보하는 방편으로 거울을 이용한 측면 영상을 함께 확보하는 방법으로 정면과 측면을 관

측하는 노력을 하였다. 시각적 정보 취득은 추적 도구를 사용하였다고 하고 있으며, 입술의 폭과 높이, 윗입술에서 턱까지의 거리, 코로부터 윗입술까지 거리, 코부터 턱까지의 거리 등의 데이터를 이용하여 자동 독화를 수행하였다[21,22]. 백성준 등은 특별 센서를 부착한 카메라로 입술의 높이, 폭, 안과 바깥의 입술 경계, 윤곽의 파라미터로 입술 정보를 추출하여 음성인식의 보조 자료로 이용하였다[23]. 신도성 등은 입술 영역으로 제한된 입력으로 이미지를 기반으로 하는 방식의 파라미터를 추출하였고 은닉 마르코브 모델을 이용하여 인식을 하였다. 동적 환경에서 인식을 목표로 연구하였으며, 파라미터를 추출하기 위한 영상 처리 부분의 기술이 미약하고 얼굴 검출에 대한 기술이 없으며 입술 검출도 다중 선형 회기분석으로 임계값을 설정하는 2진화 임계값과 같이 기초적인 기술을 이용하였다[24-27]. 서재영은 다양한 각도로 입력되는 영상으로부터 자동 독화를 하려고 하였다[28].

자동 독화 성능은 발화자를 잘 찾아 시각적으로 보이는 발화자의 조음기관 변화를 정밀히 인식하여 그 변화를 파라미터로 추출하고 인식에 반영하는 것이다. 입력된 영상에서 발화자를 찾는 것은 고도의 영상 처리를 이용한 얼굴 검출 기술이 필요하고 또 얼굴에서 파라미터 추출을 위한 관심 점의 설정과 그 관심 점의 변화 검출 기술이 필요하다. 그러나 영상 처리의 어려움으로 인하여 많은 연구가 이미 얼굴을 검색한 것을 전제로 하거나 입 주위 영상을 검색한 것을 전제로 하고 있다. 또 입술만 촬영하여 학습 및 인식실험을 하거나, 입술에 특별한 색칠을 한 후 얼굴을 촬영하여 입술 영역을 검출하거나, 얼굴에 마커를 붙여 영역을 확인하기도 하고, 헤드 셋 카메라를 이용하여 고정된 영상을 취득하기도 하고 수작업으로 관심 점들의 변화를 추출하기도 하여 진정한 자동 독화라고 보기 어렵다.

III. 특징점 추출 및 파라미터 설정 과정

3.1 특징점 추출 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. 정상인 180명의 증명사진을 통해 두 눈 사이 거리를 기준으로 하여 두 눈 중점으로부터 코끝과 윗입술 턱 끝까지 분포를 조사하고 통계정보를 추출하여 향 후 영상 처리 과정에서 특징 점 추출 시 통계 자료를 바탕으로 접근하도록 하여 정확성을 배가하고 계산량을 줄이도록 하였다. 또한 컬러 얼굴 영상을 다양한 색공간 배치하고 눈, 코, 입의 요소가 통합 및 단일 성분에서 표현 상태를 분석하여 동일 요소가 색 공간에 따라 다른 특징 점과 구별되게 높은 값 또는 낮은 값을 갖는 것을 파악하여 향 후 증폭 또는 축소, 대비로 특징 점 추출이 용이하게 되도록 분석하였다. 이후 과정은 컴퓨터를 이용한 처리 과정으로 얼굴 추출 과정은 피부색 추출이 용이하도록 휘도 성분에 의존적인 보정을 하여 어두운 곳은 좀 밝게 되게 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 볼록다각형 형성하여 얼굴 후보 마스크를 구성하였다. 이 과정까지 수행하면 얼굴 후보 마스크의 범위 내에는 눈 코 입이 포함된다. 따라서 이 후 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 Y상에서 눈동자가 낮은 값을 갖는 점과 흰자위가 높은 점을 갖는 점을 이용하여 한 번은 밝은 성분이 번지도록 한 모폴로지 결과 영상과 어두운 성분이 번지도록 한 결과영상을 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 입안쪽 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 방법과 "색상 정보를 이용한 자동 독화 특징 추출(입술 상단 검출)"으로 구성하였다[30, 31]. 이 방법에 의한 통계 정보가 표 1에 있으며 처리되는 과정은 그림 1로 표현하였다.

표 1. 얼굴 특징점 거리 분석
Table 1. Facial Feature Distance analysis

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%

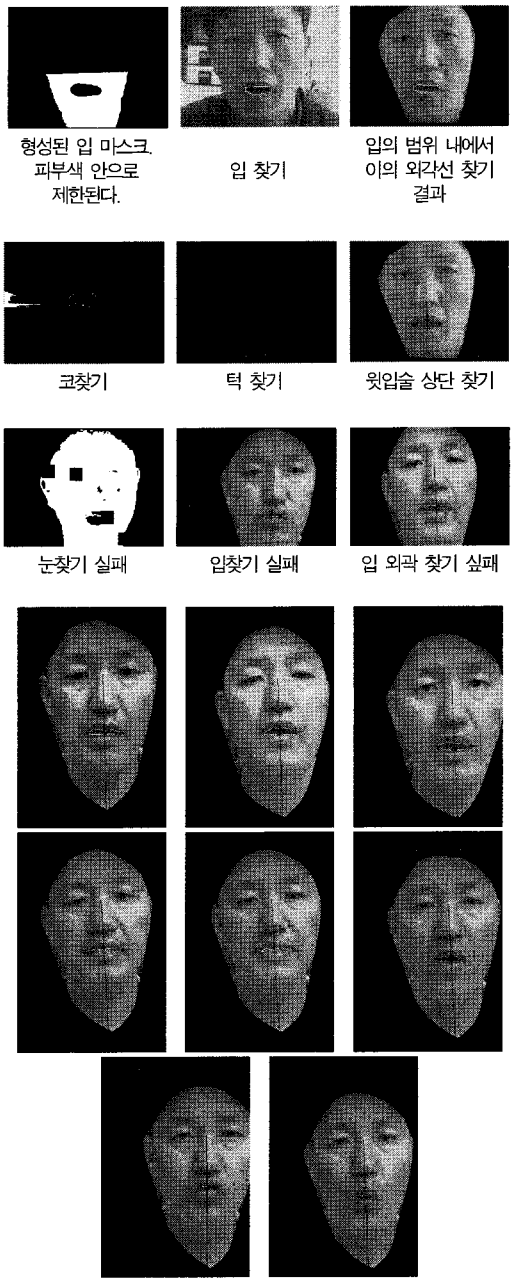


그림 2. 특징점 추출 과정과 결과
Fig 2. Feature Detection Process and Result

3.2. 파라미터 설정

앞에서 기술한 과정을 통하여 추출한 파라미터는 입술 안쪽의 면적, 입술 안쪽의 폭과 높이, 이 보임 비율, 코와 윗 입술 상하단과의 거리 등이다. 한국어 '아/에/애/이/어/으/우/

오' 8모음 인식을 위하여 추출된 영상은 정상인 5명으로부터 각각 8모음 발화이미지 60 세트 2400개의 영상이다. 영상 취득은 컴퓨터에 연결된 카메라로 하였고, 카메라의 위치는 두 눈간 거리가 대략 90~100 픽셀 정도가 하도록 하였다. 영상 취득시 정확한 발음을 하게 하기 위하여 발화 방법에 대한 설명 후 수 차례 발화 연습을 한 후 3초 이상 발화하게 하며 발화 영상을 취득하였다. 2400개의 정면 영상에서 특징점 추출에 실패한 것은 눈 찾기 실패가 118장, 입 영역 찾기 실패가 144장, 입 외곽선 추출 실패가 12장, 입술 상단 찾기 실패는 120 이었다. 입영역 찾기 실패는 주로 '우' 발화시 발생하였으며, 발화 시 입의 굴곡이 심할 때의 영상들이었다. 입 외곽선 추출 실패는 혀가 관찰되는 영상에서 혀와 입술의 영역을 정확히 구분 하지 못하여 주로 발생하였다. 윗입술 상단 추출에 실패한 이유는 Canny 외곽선 추출 시 일괄적으로 적용한 문턱 값으로 추정된다. 파라미터 추출은 총 16.4%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.

파라미터의 정규화는 두 눈 사이의 평균 거리를 기준으로 비례식을 이용하여 수행하였고, 추출한 얼굴 특징 점으로부터 구성한 파라미터는 발화 입 면적, 발화 입 면적 대비 이 보임 비율, 입 높이, 입 너비, 코끝과 입술 상단 간의 거리, 코끝과 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 파라미터들에 대한 통계표
Table 2. The statistical table about parameters of speaker #1

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

입 면적 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

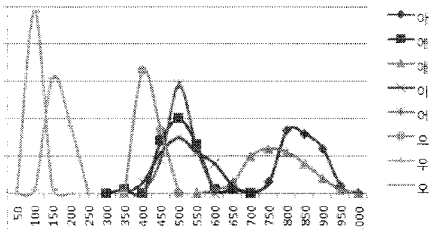
입 높이								
	아	에	애	이	어	으	우	오
Ave	16.7	9.1	14.4	10.2	12.7	7.1	6.8	3.50
Stdev	1.7	1.1	1.1	0.8	0.7	0.8	2.3	0.53
Max	18.0	11.0	16.0	11.0	14.0	8.0	12.0	4.00
Min	14.0	8.0	13.0	9.0	12.0	6.0	4.0	3.00

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

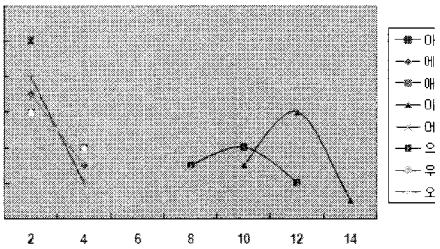
코와 윗입술 하단 거리								
	아	에	애	이	어	으	우	오
Ave	38.9	37.9	36.9	34.0	39.5	37.8	32.2	42.3
Stdev	3.7	3.4	4.7	2.5	1.7	2.1	0.6	5.9
Max	43.0	42.0	42.0	37.0	41.1	40.0	33.0	49.0
Min	30.0	34.0	31.0	31.0	35.1	33.0	31.0	30.1

코와 윗입술 상단 거리								
	아	에	애	이	어	으	우	오
Ave	24.0	24.0	18.0	22.0	27.0	26.0	19.0	24.0
Stdev	31.0	31.0	28.0	26.1	30.1	29.0	21.0	31.0
Max	28.4	27.8	23.3	24.5	28.2	27.7	20.1	27.9
Min	2.3	3.0	4.6	1.7	0.9	1.1	0.8	2.5

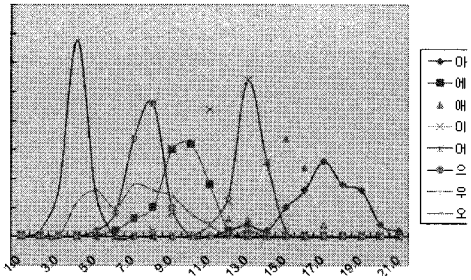
코와 입술 하단 거리								
	아	에	애	이	어	으	우	오
Ave	55.6	47.0	51.4	44.2	52.2	44.9	38.1	47.4
Stdev	3.2	3.6	4.7	2.2	1.8	2.3	0.6	4.0
Max	59.0	52.0	57.0	47.0	54.1	46.0	39.0	52.0
Min	48.0	42.0	45.0	41.0	48.0	40.0	37.0	40.0



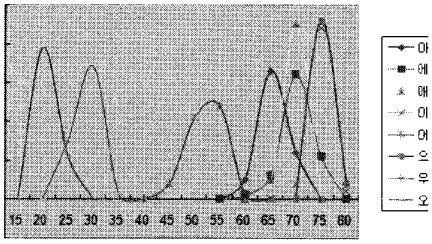
입 면적 분포도



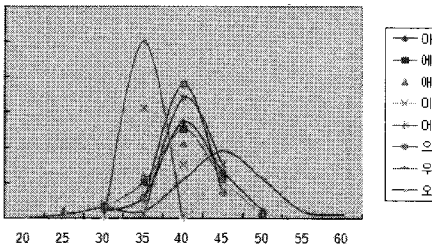
입 면적 대비 이 외곽선 길이 분포도



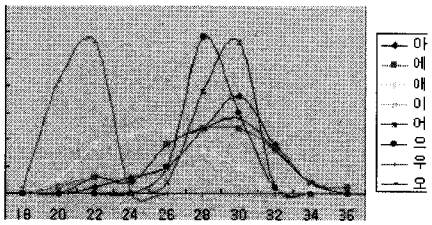
입 높이 분포도



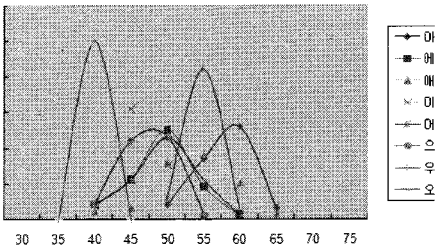
입 너비 분포도



코-윗입술 하단 거리 분포도



코-윗입술 상단 거리 분포도



코-입술 하단 거리 분포도

그림 2 파라미터별 분포도

Fig 2. a distribution chart of each parameter

위 표와 분포도들을 살펴보면 서로 다른 사람의 자료지만 정규화를 통해 보정을 하여 서로 유사한 분포를 갖고 있음을 볼 수 있다.

또한 그림 2의 '입 면적 분포'를 보면 '오'/'우'는 면적이 작아 왼편에 위치하며, '이'/'애'는 오른편에 위치함을 관찰 할 수 있다. '입 면적 대비 이 보임 비율'을 보면 이의 보임 비율이 높은 '이'/'애'가 다른 집단과 다르게 오른쪽에 따로 분포되어 있음을 알 수 있다. '입 높이'와 '입 너비'도 발화별로 비교적 잘 구분될 수 있음을 알 수 있다. 코끝과 윗입술 하단과의 거리는 구분성이 떨어지나 코와 윗입술 상단 거리 분포에서는 '우'가 명확히 구분됨을 알 수 있다. 코끝과 입술 하단과의 거리도 적절한 구분성을 가지고 있으며, 아울러 각각의 자료를 다른 자료와 동시에 살펴보면 8모음이 구분이 될 개연성이 충분히 있음을 알 수 있다.

이번 연구에서는 표와 그래프를 시각적인 관점에서 평가를 하여 입 면적, 이 보임, 입 높이, 이 너비, 코끝 윗입술 상단 거리를 파라미터로 추출하였다. (실험에 사용된 파라미터를 심사에 참고하도록 첨부 1로 추가하였다.)

IV. 자동 독화 실험

4.1 인식 시스템 구축

발화 인식을 위하여 구성한 시스템은 신경망이다. 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.

구축된 신경망은 입력층, 은닉층, 출력층으로 구성하였으며 입력층은 5유닛, 은닉층 10유닛, 출력층 5(5 모음 인식) 유닛 또는 8(8모음 인식)유닛으로 구성되었으며, 학습률은

0.05, 시그모이드 함수 곡선 기울기에 반영되어 수렴을 조절할 모멘텀은 7, 학습은 20000회 또는 에러 값 경계치 0.0001이하로 설정하여 인식 실험하였다.

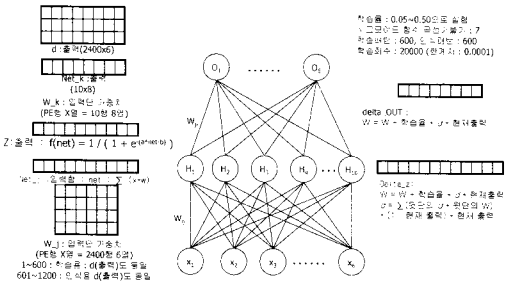


그림 3. 구축된 신경망

Fig 3. Implemented Neuralnet System

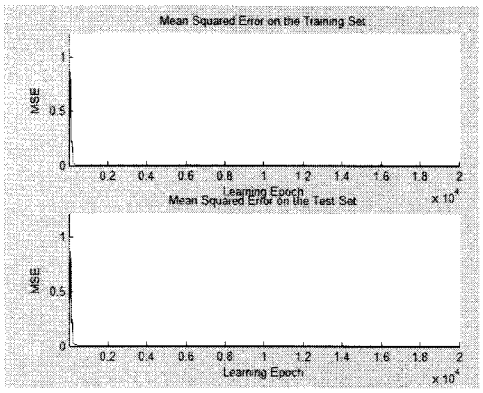
4.2 인식 실험 및 결과

인식 실험은 동일 관점에서 연구된 과거 논문과 비교를 위하여 '아/에/이/오/우' 5 모음 인식 실험과 한국어 단모음 관점에서 '아/에/애/이/어/으/우/오' 8모음 인식 실험을 하되, 각각 화자 독립적 관점과 화자 종속의 관점에서 실험을 하였다. 실험에 이용된 자료는 영상 처리 과정에서 특징점 추출에 실패한 것을 제외한 자료 세트 중에서 임의로 1인 1모음당 40개씩 추추하여 총 5인 x 8모음 x 40 = 1600개를 실험 데이터로 추출하여 실험하였다.

4.2.1 5모음 인식 실험

5모음 인식에 이용된 파라미터는 정규화된 입면적, 입의 높이, 입의 너비, 입 면적 대비 이 외곽선 길이 비, 코와 윗입술 외곽 상단간의 거리이다. 실험은 화자 종속 관점에서 5인 각각 훈련 10세트, 평가 30세트, 화자 독립 관점에서 '훈련 1인 10세트, 평가 5인 150세트'와 '훈련 5인 50세트, 평가 5인 100세트'로 하였다.

동 파라미터를 이용한 5모음 인식은 위에 열거한 모든 실험에서 인식률이 거의 100%로, 영상 처리 상에서 인식 할 수 있다면 동 파라미터로 100%인식할 수 있는 것으로 판단된다.



*.학습 - 5인 50셋트, 평가 - 5인 100셋트

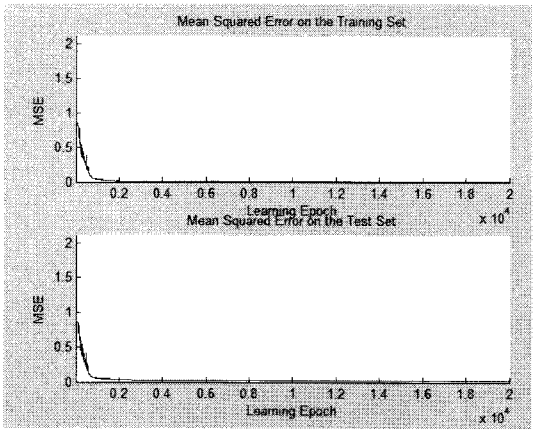
그림 4. 5모음 인식 실험

Fig 4. 5 Vowels Recognition Experiment

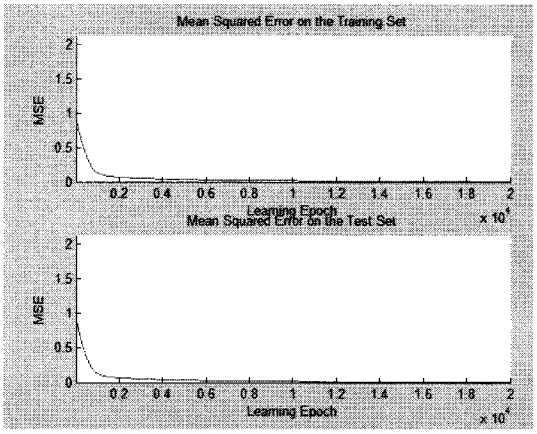
그림 4는 '훈련 5인 50세트, 평가 5인 100세트'로 실험한 것으로 20,000번 학습 후 인식하였을 때 학습 세트의 Mean Squared Error(MSE)는 0.00038, 평가 세트의 MSE는 0.000155 였다.

4.2.2 8모음 인식 실험

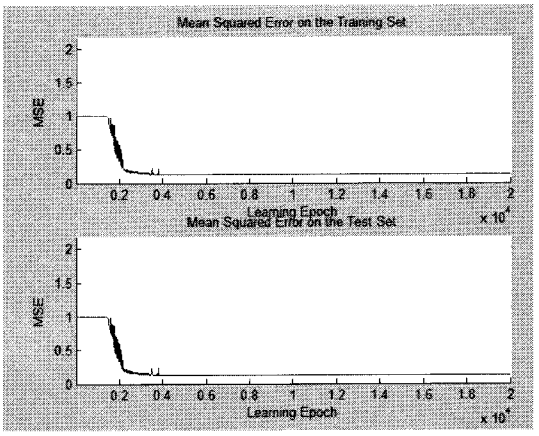
8모음 인식에 사용된 파라미터도 5모음 인식에 사용된 것과 동일한 정규화된 입면적, 입의 높이, 입의 너비, 입 면적 대비 이 외곽선 길이 비, 코와 윗입술 외곽 상단간의 거리이다. 이 역시 화자 종속과 화자 독립의 관점에서 실험하였으며 화자 종속 실험은 5인 각각 '학습 10세트, 평가 30세트'와 '학습 20세트 평가 20세트'를 수행하였으며, 화자 독립 실험은 5인 통합 '훈련 50세트 평가 150세트', '훈련 100 세트 평가 100세트'를 수행하였다.



*. 훈련 10 세트 평가 30 - 최고의 결과

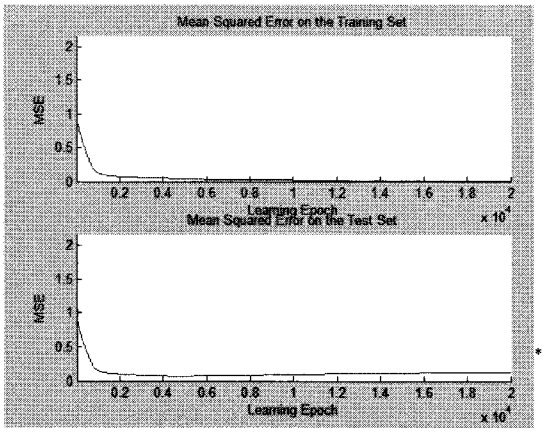


*. 훈련 20 세트 평가 20 - 최고의 결과
그림 5. 8모음 화자 종속 실험
Fig 5. 8 Vowels Speaker Dependence Recognition Experiment



*. 훈련 100세트 평가 100세트 - 최고의 결과
그림 6. 화자 독립 실험
Fig 6. 8 Vowels Speaker Independence Recognition Experiment

위 실험에서 화자 종속 실험의 경우 매우 높은 인식률을 보이는 반면, 화자 독립 실험의 경우 다소 인식률이 떨어짐을 보였다. 그림 5는 8모음의 화자 종속 실험 중 최고의 결과를 보인 것으로 훈련 세트가 많을수록 학습 속도가 느림이 관찰되었다. 최고 인식률을 보인 경우는 100%의 인식을 보이고 있으며, 나머지도 비교적 높은 인식률을 보이고 있었다. 화자 독립 '학습 50세트 평가 150세트' 실험 시 훈련 세트를 바꾸어 가며 실험하였고, 그림 6의 상단에서 보는 바와 같이 학습 세트는 MSE가 줄어듬에도 평가 세트는 오히려 MSE가 늘어 인식률이 떨어지는 현상도 볼 수 있었다.



*
. 훈련 50세트 평가 150세트 - 최고의 결과

표 3 독화 결과 비교
Table 3. Speechreading Result Comparisons

연구	인식대상	파라미터	인식률	비고
[28]	한국어 5모음 '아/에/이/오/우'	2 개 입의 폭, 입의 너비	70%	입주위만 영상입력
[32]	한국어 5모음 '아/에/이/오/우'	5 개 코끝과 입 상단 거리, 코끝과 입 하단 거리, 코끝과 턱끝 거리, 입의 폭, 입의 너비	91.1	다양배경 다양조명
This	한국어 5모음 '아/에/이//오/우'	5 개 입 면적, 이 보임 정도, 입의 폭, 입의 너비, 코끝과 윗입술 상단까지 거리	99.9	다양배경 다양조명
[33]	한국어 8모음 '아/에/애/이/어/으/오/우'	5 개 입 면적, 이 보임 정도, 입의 폭, 입의 너비, 코끝과 윗입술 하단까지 거리	89.3	다양배경 다양조명 화자종속 화자독립 평균
This	한국어 8모음 '아/에/애/이/어/으/오/우'	5 개 입 면적, 이 보임 정도, 입의 폭, 입의 너비, 코끝과 윗입술 상단까지 거리	93.7	다양배경 다양조명 화자종속 화자독립 평균

V. 결론

본 논문에서는 한국어 8 단모음의 기계적 인식을 위한 연구를 하였다. 컴퓨터에 의한 영상 처리 과정을 통해 두 눈의 위치와 안쪽 입의 외곽선, 이(齒)의 외곽선, 코의 위치, 윗입술 외곽의 상단 위치를 추출하고 이것으로 안 쪽 입의 면적, 안 쪽 입의 폭과 높이, 입 면적 대비 이 외곽선 길이 비, 코와 안쪽 입술 상하단 간의 거리 및 윗입술 외곽 상단과의 거리를 파라미터로 사용할 수 있는지를 분석하고, 안 쪽 입의 면적, 안 쪽 입의 폭과 높이, 입 면적 대비 이 외곽선 길이 비, 코와 윗입술 외곽 상단과의 거리를 파라미터로 하여 자동 독화 신경망 시스템을 구축하고, 화자 종속 및 화자 독립 인식실험을 하였다. 그리고 입의 면적, 이의 보임 비, 입의 폭과 너비, 코 끝에서 윗입술 상단까지의 거리를 파라미터로 한 실험에서 90% 이상의 인식률을 얻어 파라미터와 시스템의 효율성을 확인하였다.

본 연구는 음성인식의 인식률을 높이기 위한 방편으로 수행한 연구이므로 정지 영상에서 정보처리 뿐 아니라 동 영상에서 정보 처리 연구가 필요하며, 발전된 영상 처리를 통해 좀 더 정밀하고 정확한 특징점의 추출과 효율적인 파라미터의 추출을 위한 연구와 음성 신호 처리와 결합하는 연구가 필요하다. 아울러 본 연구 파라미터를 적절히 이용하면 휴먼컴퓨터 인터페이스와 컴퓨터 보안 등에 응용할 수 있다.

참고문헌

[1] 최병문, "*구화교육*," 한국구화학교, 1970년 11월

[2] Sumby, W.H. and Pollack, I., "Visual Contribution to Speech Intelligibility in Noise," Journal of the Acoustical Society of America, Vol. 26, No. 2, pp. 212~215. Mar. 1954

[3] MCGurk, H., & MacDonald, J., "Hearing Lips and Seeing Voices," Nature, No. 264, pp. 746-748, Des. 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, pp. 3-51, 1987.

[5] Petajan, E. D., "*Automatic lipreading to Enhance Speech Recognition*," Ph.D. Dissertation, University of Illinois at Urbana-Champaign, Feb. 1984.

[6] Petajan, E. D., "Automatic Lipreading to Enhance Speech Recognition," Proceedings of the IEEE Communication Society Global Telecommunications Conference, Atlanta, Georgia, USA, pp. 26-29, Nov. 1984.

[7] Finn, E. K. & Montgomery A.A. "Automatic Optically Based Recognition of Speech," Pattern Recognition Letters, Vol. 8, No. 3, pp. 159-164, Oct. 1988.

[8] Stork, D. G., & Hennecke, M.E., "*Speechreading by Humans and Machines*," ISBN 3-540-61264-5, Springer, Dec. 1996.

[9] Yuhas, B. P., Goldstein, M.H. & Sejnowski, T.J., "Integration of Acoustic and Visual Speech Signals Using Neural Networks," IEEE Communications Magazine, Vol. 27, pp. 65-71, Nov. 1989

[10] Mase, K. & Pentland, A., "Automatic Lipreading By Computer," Trans. Inst. Elec. Info. and Com. Eng., Vol. J73-D-II, No. 6, pp. 796-803, Nov. 1990.

[11] Silsbee, P. L., "*Computer Lipreading for Improved Accuracy in Automatic Speech Recognition*," Ph.D. dissertation, The University of Texas at Austin, Sep. 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, no. 1, pp. 556-560, Nov. 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, Vol. 4, No. 19, pp. 1192-1195, Oct. 1995.

[14] Adjoudani, A. et al., "A Multimedia Platform for Audio-Visual Speech Processing," Proc. European Conference on Speech Communication and Technology, Rhodes, Greece, pp. 1671-1674, Sept. 1997

[15] Meier, U., Stiefelhagen, R., Yang, J., Waibel, A., "Towards Unrestricted Lipreading," International

Journal of Pattern Recognition and Artificial Intelligence, vol. 14, no. 5, pp. 571-785, Jun. 1999.
〔16〕 김진범, 김진영, "이미지 변환과 HMM에 기반한 자동 립리딩," 대한전자공학회 추계학술대회 논문집, 제22권, 2호, 585-588쪽, 1999년 11월
〔17〕 김진범, 김진영, "입술 대칭성에 기반한 효율적인 립리딩 방법," 전자공학회논문지, 제37권, 5호, 55-464쪽, 2000년 9월
〔18〕 이지은, 김진영, 이주헌, "시간영역 이미지 필터링에 의한 립리딩 성능 향상," 한국음향학회 학술발표대회논문집, 제20권, 2호, 45-48쪽, 2001년 11월
〔19〕 이지은, *"시간영역 이미지 필터링에 의한 립리딩 성능 향상,"* 전남대학교대학원 석사학위논문, 2002년 2월
〔20〕 민덕수, "동적 환경에서 립리딩 성능저하 요인 분석 및 인식성능 향상에 관한 연구," 전남대학교대학원 석사학위논문, 2002년 2월
〔21〕 민소희, 김진영, 최승호, "입술 정보를 이용한 음성 특징 파라미터 추정 및 음성 인식 성능 향상," 대한음성학회지, 44호, 83-92쪽, 2002년 12월
〔22〕 김진영, 민소희, 최승호, "음성인식에서 입술 파라미터 열화에 따른 견인성 연구," 음성과학, 제10권, 2호, 27-33쪽, 2003년 6월
〔23〕 백성준, 김진영, "입술정보 및 SFM을 이용한 음성의 음질향상알고리듬," 음성과학, 제10권, 2호, 77-84쪽, 2003년 6월
〔24〕 신도성, *"입술영상접기와 프레임간 필터링을 이용한 립리딩 성능 개선,"* 전남대학교대학원 박사학위논문, 2004년 2월
〔25〕 김진영, 신도성, "상태공유 HMM을 이용한 서브워드 단위 기반 립리딩," 음성과학, 제8권, 3호, 123-131쪽, 2001년 9월
〔26〕 신도성, 김진영, 최승호, "시간영역 필터를 이용한 립리딩 성능향상에 관한 연구," 한국음향학회, 제22권, 5호, 375-382쪽, 2003년 7월
〔27〕 신도성, 김진영, 이주헌, "동적 환경에서의 립리딩 인식 성능저하 요인분석에 대한 연구," 한국음향학회, 제21권, 5호, 471-477쪽, 2002년 7월
〔28〕 서재영, *"단순 특징값과 촬영 각도에 따른 한국어 모음의 오디오 비주얼 인식에 관한 연구,"* 성신여자대학교대학원 석사학위논문, 2005년 2월
〔29〕 Hsu, R. L, Abdel-Mottaleb, M., Anil K. J., "Face Detection in Color Images," IEEE Trans. on Pattern Analysis, vol. 24, no. 5, pp. 696-706, May. 2002.
〔30〕 이경호, 양룡, 이상범, "색상 정보를 이용한 자동 독화 특징 추출," 한국컴퓨터정보학회 논문지, 제13권, 6호, 107-116쪽, 2008년 11월
〔31〕 이경호, 김선옥, " 색상 정보를 이용한 자동 독화 특징 추출(입술 상단 검출)," 한라대학교 논문집, 11집, 125-135쪽, 2009년 2월
〔32〕 이경호, 금종주, 이상범, " 한국어 5모음의 조음적 제어 분석을 이용한 자동 독화에 관한 연구," 컴퓨터산업교육학회, 제8권, 4호, 281-288쪽, 2007년 10월
〔33〕 김선옥, 이경호, " 얼굴 특징점을 이용한 한국어 8모음 독화시스템 구축," 한국컴퓨터정보학회, 제16권, 2호, 135-140쪽, 2008년 12월

저 자 소 개

이 경 호

1987: 인하공업전문대학.
1991: 한국방송통신대학교.
1994: 한국과학기술원 공학석사.
2008: 단국대학교 공학박사.
1996 - 현재: 한라대학교 정보통신공학부 교수
관심분야: 패턴인식, HCI, 디지털 신호처리, 컴퓨터 기술 응용

양 룡

1972: 한국항공대학교 공학사.
1980: 동아대학교 공학석사.
1990: 단국대학교 공학박사
1979 - 현재: 인하공업전문대학 컴퓨터정보과 교수
관심분야: 병렬처리컴퓨터, HCI, 디지털콘텐츠, 컴퓨터윤리,

김 선 옥

1991 서강대학교 이학석사
1998 서강대학교 이학박사
2005-현재 한라대학교 정보통신방송공학부 전임강사
〈관심분야〉 추천시스템, 정보검색, 음성인식