論文
8-4-9

<< 한국어 5모음의 조음적 제어 분석을 이용한 자동 독화에 관한 연구 >>

Design & Implementation of Lipreading System using the Articulatory Controls Analysis of the Korean 5 Vowels

이경호(Kyong Ho Lee)[1] 금종주(Jong Ju Kum)[2] 이상범(Sang Bum Rhee)[3]

요약

In this paper, we set 6 interesting points around lips. Analyzed and characterized is the distance change of these 6 interesting points when people pronounces 5 vowels of Korean language. 450 data are gathered and analyzed. Based on this analysis, the system is constructed and the recognition experiments are performed. In this system, we used the camera connected to computer to measure the distance vector between 6 interesting points. In the experiment, 80 normal persons were sampled. The observational error between samples was corrected using normalization method. We analyzed with 30 persons and experimented with 50 persons. We constructed three recognition systems and of those the neural net system gave the best recognition result of 87.44 %.

이 논문에서 우리는 입주위에 6개의 관찰 점을 설정하고, 한국어 중 '아/에/이/오/우' 5 모음을 발음할 때 생기는 관찰 점간의 거리 변화를 계수화 하였다. 약 450개의 자료를 모아 분석하고 이 분석을 바탕으로 시스템을 구축하여 실험하였다. 우리 시스템에서는 컴퓨터에 연결된 카메라를 사용하였으며, 6개의 영역 간의 변화를 계수로 하였다. 이 실험에 정상인 80명이 동원되었고, 사람들 사이에 있는 관찰 오차를 정규화를 통하여 수정하였다. 30명으로 분석하였고, 50명으로 인식 실험을 하였다. 3개의 시스템을 구축하였는데 신경망이 가장 좋은 결과를 보였다. 신경망의 인식 결과는 87.44%였다.

key word : lipraeding

논문접수 : 2007. 10. 4.
심사완료 : 2007. 10.28.

1) 정회원 : 한라대학교
2) 정회원 : 국방품질관리원
3)종신회원 : 단국대학교 대학원 컴퓨터응용 교수

1. 서 론

최근 음성 인식 분야에서는 심한 잡음 환경에서도 높은 인식률을 갖게 하기위한 연구가 활발히 진행되고 있다. 실험실과 같이 거의 잡음이 없는 환경에서는 높은 인식률을 보이나 소음이 많은 환경에서는 인식 성능이 많이 저하되고 있기 때문이다.

자동독화(Lip-reading)는 음성인식 분야 중 잡음 환경에서 현저하게 떨어지는 인식률을 높이기 위한 보상 방법의 하나로서, 시각적 관점에서 화상을 이용하여 조음 현상을 반영하는 입술 주위로부터 얻은 정보를 추가하여 인식률을 높이고자 하는 연구이다. [1-6]

발화라는 것이 무한의 자유도를 갖지만 한 개의 언어에서 사용되는 모음의 종류가 유한이고, 이산적인 언어 정보를 전달하고 있음은 모음 생성에 관여하는 제어도 비교적 소수의 이산적이고 단계적이라는 것이 기존 연구결과이다. [12]에서 한국어 모음이 조음 차원에서 제어의 이산성을 가짐을 명확히 하였고, 발화를 위한 턱의 열림각의 제어가 3단계, 혀의 제어가 3단계, 입술의 오므림의 유무를 2단계로 분석하였다. 즉 조음 영상은 언어정보의 이산성을 반영한다는 것을 확실히하였다.

우리말의 시각적 관점에서의 인식 연구는 미미한 편이나, 입술 영역 검출 및 입술 형태 표현은 많은 연구가 되고 있는 편이다. 자동 독화에 관한 연구로는 [14],[15],[16] 등이 있으며, [14],[15]는 입술 특징 파라메터를 음성 파라미터로 변환하여 84%의 인식률을 얻은 결과를 보이고 있다. [16]은 입술 영상 한 프레임을 주성분 분석을 통한 특징 계수를 구하여 립리딩으로만 40%의 인식률을 얻고 있다. 기존 연구 중 [12]는 납 조각을 관심 영역에 붙여 X선을 이용한 장치로 조음 제어를 조사한 것이므로, 자동 독화 장치로는 적합하지 않으며, [14], [15]에서는 1명을 대상으로 인식 실험하였고, [16]너무 낮은 인식 결과를 보이고 있다.

본 연구에서는 컴퓨에 연결한 카메라를 이용하여 영상을 획득하였으며, 획득한 영상에서 조음 시 제어되는 턱과 입술의 오무림을 관찰 할 수 있는 관심 영역의 위치를 추정하여, 거리벡터의 변화를 계수로 하여 인식하는 실험을 하였다. 관심 영역은 턱의 열림 각과 입술의 오무림을 추정해 볼 수 있는 얼굴 주위의 6곳으로 입술 상하좌우와 코끝 턱끝이다. 물론 관심 영역의 추정의 어려움 때문에 준비된 실험실 환경에서 자료를 획득하였다. 인식 대상은 한국어 5개 단모음 '아/에/이/오/우'로 하였고, 30인을 이용하여 발화를 하고, 얻은 이미지로부터 계수화/정규화 하여 추출된 계수들을 통계적으로 분석하고, 분석 결과를 반영한 3가지 시스템을 구축하고 인식 실험을 하였다.

관측을 통한 분석에서 코와 입술 상위의 거리 벡터의 변화는 크게 3그룹(우, 아/에/이, 오)으로 분류됨이 관측되었고 아/에/이는 유사 분포 곡선을 가지고 있어 거의 서로 구분되지 않았다. 입술의 상하 벌어짐 정도는 크게 3그룹(우/오/이, 에, 아)이나 뭉쳐있는 한 그룹은 각각이 특징적인 분포 곡선을 가지고 있었다. 아래 입술과 턱과의 거리 벡터는 크게 2 그룹(우, 아/에/이/오)이나 입술 상하에서와 마찬가지로 뭉쳐있는 그룹은 각각이 특징적인 분포곡선을 가지고 있었다. 코끝에서 턱까지의 거리는 앞의 세 관심 거리 벡터의 통합된 결과이나 '우', '오/이', '아/에'로 분류된 3그룹이었고, 뭉쳐진 그룹은 구별되는 분포 곡선을 가지고 있었다. 마지막으로 입술 좌우 양끝의 벌어짐은 우가 매우 길게 분포한 3그룹(오, 아/에/우, 이)으로 관측되었고, 각각은 구별되는 분포 곡선을 가지고 있었다.

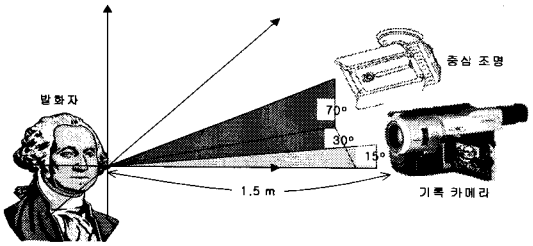
이렇게 관측된 데이터를 바탕으로 통계적 관점하에서 1가지 방법과 확신도를 이용한 전문가 시스템 방법으로 1가지, 신경망을 통한 1가지 방법 총 3가지 방법으로 시스템을 구축하여 인식 수행하였다. 이 실험에서 다양한 사람의 거리 벡터도 정규화를 한다면 자동 독화에 이용할 수 있음을 확인하였고, 각각의 시스템이 특징을 가지고 있지만 시스템1은 '아'와 '에'의 혼

동이 매우 심하고, '에'와 '이'도 혼동이 심한 상태에서 61.8%의 인식율을 보였고, 시스템2가 '아'와 '에', '에'와 '이'의 혼동이 2/3로 줄어든 상태에서 80.9%, 시스템3이 '아'와 '에', '에'와 '이'의 혼동이 1/2로 줄어든 상태에서 87.44%의 인식률을 기록하였다. 또한 음향만을 이용한 음성 인식기와 상호 보완성이 강하여, 두 시스템을 통합하면 우리말에서도 매우 좋은 결과를 얻을 수 있다는 개연성을 확인하였다.

2. 시스템 구축을 위한 자료 수집 및 분석

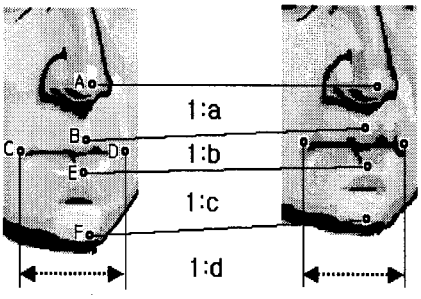
음성 인식을 위한 음향 관점에서 특징 파라메터를 잡는 방법은 모음과 유성 자음에만 존재하여 성도의 공진 특성을 반영하는 포먼트와 에너지, 시간 영역에서의 피치, LPC(Linear prediction coefficient), PARCOR(Partial auto corelation), 켑스트럼 등이 있으며, 시각적 관점에서 입술 정보를 이용하는 방법은 입술 형태를 모델링 하는 방법과 입술 주위의 관심 영역을 설정하여 이들의 변화를 계수화하여 이용하는 방법이 있다.

본 논문에서는 턱의 열림 각의 변화의 반영과 입술의 오무림을 반영할 수 있는 관심 영역으로 입술의 상하좌우와 코끝 턱끝을 설정하였고, 이 관심 영역간의 거리가 발화시 발생하는 변위를 계수화하였다.



(그림 67) 이미지 획득 환경

관심 영역의 인식은 YCbCr의 기반의 영상에서 얼굴 후보 영역을 검출하고, 후보 영역에서 색차와 휘도 영역에서의 특징 추출을 통하여 eye-map을 통하여 눈을 추출하고, 얼굴 후보 마스크를 이용한 필터링을 통하여 mouth-map을 통하여 입을 추출하여 이들의 존재를 바탕으로 얼굴 영역을 확정하며, 찾은 눈과 입의 좌표를 통하여 눈과 눈의 중점과 입 사이의 휘도 성분을 이용하여 입의 범위와 코와 턱의 위치를 추정하여 거리 벡터를 계수화 하였다.[8], [10], [11].



(그림 68) 개인차 정규화

2.1 이미지 획득 환경

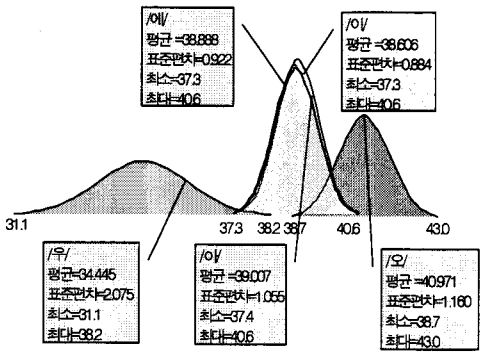
시각적 관점에서 관심 영역의 추출은 조명과 같은 다양한 환경에서 상당한 오차를 발생하기 때문에 그림과 같은 구성된 환경에서 이미지를 획득하였다. 이미지 획득을 위하여 구성한 실험 환경은 발화자의 앞쪽 위 약 70도 상에 조명이 있었으며, 약 1.5미터 앞에 발화자보다 약간 높게 카메라를 장착하였고, 획득된 화상은 코끝에서 턱끝까지가 최소한 150픽셀 정도가 되도록 충분히 크게 하였다.

또한 관심 영역들 사이 거리가 사람들 간의 차이가 있어, 오차를 보정하기 위해 이를 정규화할 필요가 있었다. 정규화는 발화자 중 무작위 1인의 관심 영역 간의 거리벡터를 기준으로 하였으며, 다른 발화자는 무음 시 관심 영역 간의 거리 벡터를 기준과 비교 비율을 계산하여,

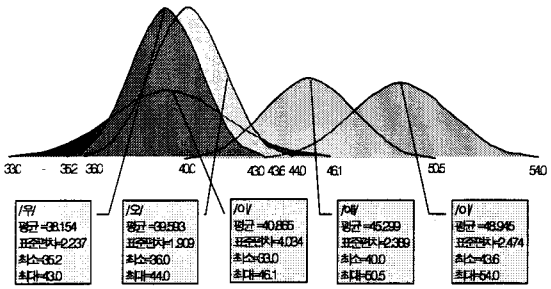
이 비율로 발화시의 측정한 거리 벡터에 적용하는 방법으로 정규화를 하였다.

2.2 습득 자료의 분석

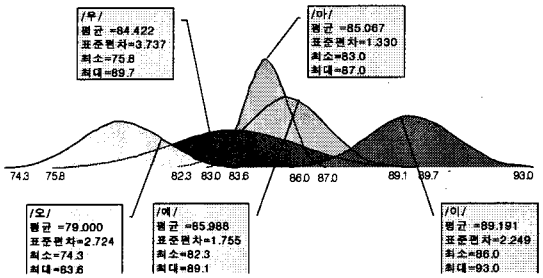
분석을 위하여 준비된 30명의 발화자는 정상인으로 대부분 20대의 남녀이며, 약간의 3~40대가 포함되어 있었다. 발화자 30명은 각각 5 모음을 반복하여 30회씩 발화하게 하였다. 각 음절과 음절 사이 무음의 지속은 최소한 1초 이상이 되게 하였으며, 발화 시간도 1초 이상 지속되게 하여 화상에서 발음이 가장 정상적으로 진행되는 곳을 얻을 수 있도록 하였다.
다음 분포도와 표들은 30인을 통합한 평균과 최빈값, 표준편차, 범위, 최소값, 최대값에 대한 분석표이다. 이 분포는 개인별로도 정규분포하였으며, 정규화를통한 통합 자료에서도 정규분포 하였다.



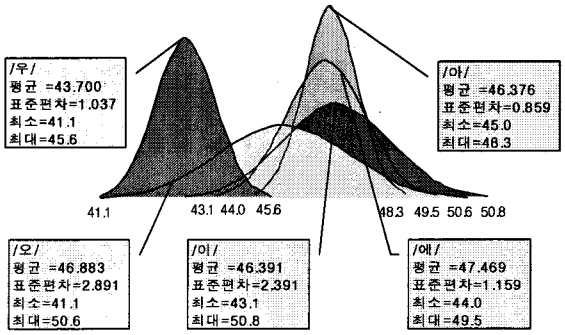
(그림 69)30인의A-B간 거리 변위 통계량



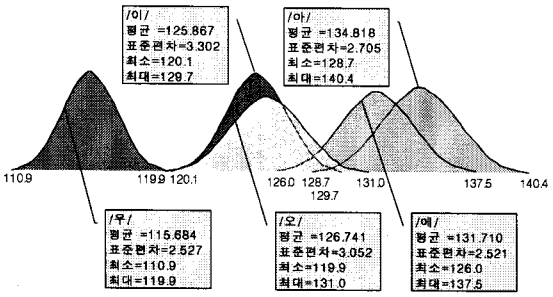
(그림 70) 30인의 B-E간 거리 변위 통계량



(그림 71) 30인의 C-D간 거리 변위 통계량



(그림 72) 30인의 E-F간 거리 변위 통계량



(그림 73) 30인의 A-F간 거리 변위 통계량

발화	위치	평균	최빈값	표준편차	범위	최소값	최대값
아	AB	39.007	39.1	1.055	3.2	37.4	40.6
	BE	48.945	48.3	2.474	10.4	43.6	54.0
	EF	46.376	45.6	.859	3.3	45.0	48.3
	AF	134.818	136.0	2.705	11.7	128.7	140.4
	CD	85.067	84.1	1.330	4.0	83.0	87.0
에	AB	38.888	39.0	.922	3.3	37.3	40.6
	BE	45.299	43.0	2.389	10.5	40.0	50.5
	EF	47.469	46.8	1.159	5.5	44.0	49.5
	AF	131.710	131.0	2.521	11.4	126.0	137.5
	CD	85.988	87.1	1.755	6.8	82.3	89.1
이	AB	38.606	38.7	.884	**3.3**	37.3	40.6
	BE	40.865	41.9	4.034	13.0	33.0	46.1
	EF	46.391	43.1	2.319	7.8	43.1	50.8
	AF	125.876	128.2	3.302	9.6	120.1	129.7
	CD	89.191	86.9	2.249	7.0	86.0	93.0
오	AB	40.971	40.9	1.160	4.3	38.7	43.0
	BE	39.593	39.1	1.909	8.0	36.0	44.0
	EF	46.883	46.0	2.819	9.5	41.1	50.6
	AF	126.741	123.8	3.052	11.1	119.9	131.0
	CD	79.000	81.4	2.724	9.3	74.3	83.6
우	AB	34.445	37.3	2.075	7.1	31.1	38.2
	BE	38.154	36.8	2.327	7.8	35.2	43.0
	EF	43.700	42.3	1.037	4.5	41.1	45.6
	AF	115.684	111.9	2.527	9.0	110.9	119.9
	CD	84.422	88.8	3.737	13.9	75.8	89.7

(표 39) 화자 30명의 모음별 표식자간 통계량

코끝과 윗 입술간의 거리 벡터는 크게 3(우, 아/에/이, 오)그룹으로 분류됨이 관측되었다. '우'는 가장 짧은 거리로 가장 넓게 분포되어 있었다. 또한 분포 범위 상에서 '오'와는 겹침이 없었고, '아/에/이' 세 모음과는 매우 적게 겹쳐져 있었다. 발화 시 조음을 위하여 입술의 내 뻗음과 오무림에 의하여 형성된 결과로 분포가 넓음은 '우'의 발화가 넓은 조음 범위를 가지고 있는 것 같다. '아/에/이'는 첨도가 매우 높은 편으로 좁은 분포 범위에서 발견되었다. 이는 이 세 모음이 조음 시 입술의 오무림이 거의 없음을 반영하는 것이다. 그러나 거리 벡터의 관점에서 범위의 반 정도가 '오'와 겹치고 있었다. '오'는 세 그룹 중 중간 정도의 첨도를 가지고 있으며, '아/에/이'와는 2/5정도로 범위가 겹쳐지고 있었으나 분포 면적으로는 10% 정도 중복되어 있었다.

입술의 상하 벌어짐 정도에서는 크게 3그룹(우/오/이, 에, 아)으로 분류됨이 관측되었다. 심하지는 않으나 전체적으로 겹쳐짐 정도가 있었다. 첫 번째 그룹으로 '우/오/이' 가 있으나 평균은 '우/이/오'순이며, 양 끝에 있는 '우'와 '오'는 높은 첨도를 가지고 있고, '이'는 낮은 첨도로 넓게 분포 되어 있다. 거리 벡터 상으로는 매우 심하게 중복되어 있으나, 분포 면적상으로는 70%내외의 중복을 가지고 있었다. 두 번째 그룹인 '에'와는 거리 벡터의 큰 값 쪽에서 '우/오/이'순으로 1/3, 2/5, 3/5 정도로 겹쳐 있었으나, 분포 면적으로는 10%내외였다. '에'는 중간 정도의 첨도를 가지고 있으며, 각 모음과 거리 벡터 상으로는 많게는 70% 적게는 30% 정도 중복되어 있다. '아'는 첫 번째 '우/오/이' 그룹과는 미미하게 범위가 중복되어 있으나, '에'와는 거리 벡터 상으로 2/3, 면적상으로는 1/4 정도로 중복 되어 있었다.

입술과 턱과의 거리 벡터는 대략 두 그룹 정도로 분류된다. 표면적으로 '우'와 '아/에/이'가 큰 두 그룹을 대표하고 '오'는 전역에 걸쳐 분포하고 있다. '아/에/이'는 평균점은 유사하나 첨도가 현저히 다른 상태로 관측되었다.

코끝에서 턱까지의 거리는 앞의 거리 벡터의 통합이나 1/2/2로 분류된 3그룹(우, 오/이, 아/에)이며, 각 뭉쳐진 그룹은 구별되는 분포 곡

선을 가지고 있었다. '이/오'는 유사한 분포와 유사한 평균을 가지고 있었으나, '아/에'는 유사한 첨도와 범위 평균의 위치가 달라 분포가 의미를 부여하고 있었다. 특히 [12]의 연구에서 밝힌 턱의 열림각의 제어가 3단계가 일치하고 있다.
입술 좌우 양끝의 벌어짐은 우가 매우 길게 분포한 3그룹(오, 아/에/우, 이)으로 관측되었고, 각각은 구별되는 분포 곡선을 가지고 있었다.

3. 인식을 위한 시스템 구축

인식을 위한 시스템은 통계적 관점 하에서 1가지, 확신도를 이용한 전문가 시스템 방법으로 1가지, 신경망을 통한 1가지 총 3개를 구성하였다.

시스템 #1

가장 기본적인 시스템이며 통계적 분석을 바탕으로 모든 분포에서 자신의 범주 내에 있는 것만을 인식하도록 하였다. i를 5모음 j를 5개의 거리 벡터 분포, V_j를 추출된 분포 j에 적용될 값, $GR(V_j)$는 값 V_j를 적용한 인식함수 라고 할 때 인식시스템은 다음과 같이 표현될 수 있다.

$$f(V_j)=\Pi GR_i(V_j)=GR_1(V_1)\oplus GR_2(V_2)\oplus GR_3(V_3)\ \oplus GR_4(V_4)\oplus GR_5(V_5)$$

시스템 #2 :

확신도는 불확실성을 표현하는 한 방법으로 MYCIN을 개발하는 과정에서 Shortliffe와 Buchman이 고안한 방법이다. 본 시스템에서는 추정 오차를 위해 모든 분포의 범위를 하한과 상한으로 5% 확대하고, 겹치지 않은 공간 값에 대하여 최고의 확신도 값을 부여하고, 중복된 공간의 값에 대하여 정규 분포함수를 바탕으로 배분하며 5 분포 공간 중 1표준편차 범위 내의 값의 수와 범위 밖의 수에 따라 확신의 정도와 불확신의 정도를 자승하여 가감하였다. 인식 시스템은 다음과 같이 표현될 수 있다.

$$f(V_j) = Max(\ \Sigma G_jCF(H_i, E_j)\ )$$

$G_jCF(H_i, E_j) = MB(H_i, E_j) - MD(H_i, E_j)$
G_jCF : 증거 E_j가 주어졌을 때 가설 H_i에 대한 확신도.
MB : E_j로 인한 H_i에 대한 증가된 확신의 정도.
MD : E_j로 인한 H_i에 대한 불확신의 정도.
H_i : 아/에/이/오/우

$$MB(H_i,E_j)=\begin{cases}1 & P(H)=1인\ 경우\\ \dfrac{\max[\ P(H_i|E_j),\ P(H_i)\ ]-P(H_i)}{1-P(H_i)} & 그\ 외의\ 경우\end{cases}$$

$$MD(H_i,\ E_j)=\begin{cases}1 & P(H)=0인\ 경우\\ \dfrac{P(H_i)-\min[\ P(H_i|E_j),\ P(H_i)\ ]}{P(H_i)} & 그\ 외의\ 경우\end{cases}$$

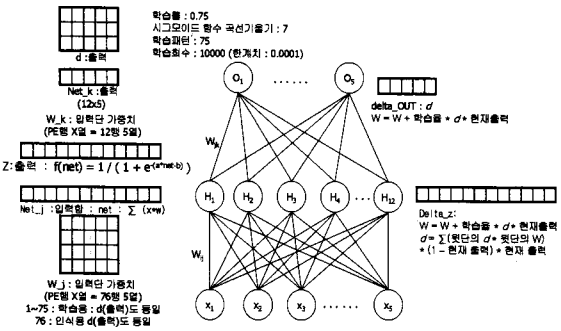
$$P(H_i)=\begin{cases}1 & V_j\in(\ GR_j^{i}\ )\ \&\&\ V_j\notin\{\ GR_j^{\neq i}\ \}경우\\ \dfrac{P'(H_i)}{\Sigma P'(H_i)} & 그\ 외의\ 경우\end{cases}$$

$$P'(H_i)=\frac{1}{\sqrt{2\Pi}*\sigma}\ \exp\{-\frac{(x-m)^2}{2\sigma}$$

시스템 #3 :

분석 자료를 바탕으로 학습시킨 신경망으로 구성하였다. 신경망은 사람의 인지 과정을 흉내내어 패턴 인식 문제를 해결해 보려는 시스템이며, 신경망은 환경의 변이에 적응하는 능력이 있으며, 병렬 제약조건을 만족시키는 문제를 해결할 적합한 구조를 가지고 있다. 본 실험에서 사용된 Multi Layer Perceptron은 입력, 히든, 출력 3층으로 구성하였다. 입력은 5, 히든은 12, 출력은 5 총 22유닛으로 구성되어 있다. learning rate는 0.75, 시그모이드 함수 곡선 기울기에 반영되어 수렴을 조절할 momentum은 7, 학습패턴은 분석을 위하여 준

비한 5모음 데이터 중 무작위로 각각 75개를 추출하였으며, 학습은 10000회 또는 에러 값 경계치 0.0001이하까지로 하였다. 학습은 최적치라고 할 수 없으며, 데이터 수를 5개씩늘려 나가며, 비교적 좋은 값에서 멈춘 것이다.



4. 인식 실험 결과

인식을 위한 데이터는 시스템 구축을 위해 분석 시 사용한 것이 아닌 다른 데이터로 50명으로부터 얻은 5모음 각각 1000개씩 총 5000개를 사용하였다.

4.1 인식 실험에 사용된 데이터 분석

구축한 시스템의 인식 실험에 사용된 데이터는 시스템 구축을 위해 분석한 데이터의 통계적 범주를 넘는 데이터가 28.97%나 되었다. 이는 이미지의 관점에서 벡터 거리에 초점을 맞춘 발화라는 것이 무한의 자유도를 갖지만 한 개의 언어에서 사용되는 모음의 종류가 유한이고, 이산적인 언어 정보를 전달하고 있음은 모음 생성에 관여하는 제어도 비교적 소수의 이산적이라는 것을 바탕으로 충분히 이해될 수 있다.

4.2 인식 결과

본 실험에서 구축한 각 시스템별 인식 결과는 다음과 같다.

	인식	Sys#1	Sys#2	Sys#3	비고
전체	인식	61.8	80.9	87.44	
아	인식	70.0	48.6	81.2	정상 데이터 인식
	미인식	30.0	51.4	18.8	정상 데이터 미인식
	오인식	6.9	0.8	2.8	오 데이터 인식
에	인식	82.1	80.3	82.6	
	미인식	17.9	19.7	17.4	
	오인식	13.5	11.3	6.3	
이	인식	52.6	83.6	81.7	
	미인식	47.4	16.4	18.3	
	오인식	1.54	3.4	1.4	
오	인식	67.5	94.9	95.4	
	미인식	32.5	5.1	4.6	
	오인식	0.2	1.6	0.3	
우	인식	65.1	97.1	96.3	
	미인식	34.9	2.9	3.7	
	오인식	0.0	2.1	1.8	

5. 연구 평가 및 향후 추가 연구

본 연구는 한국어 모음이 조음 차원에서 제어의 이산성을 가짐을 밝혀낸 기존 연구들을 바탕으로 시각적으로 관찰 할 수 있는 입술 주위의 6개 관심 점으로부터 조음 시 변화를 조사 분석하고, 이를 바탕으로 시스템을 구축하고 인식 실험을 하였다. 특히 본 연구의 카메라를 이용한 시각적 관점에서의 분석이 X선을 이용한 [12]의 분석과 턱의 열림에 대하여는 일치하였지만, 입술의 오무림에 대하여는 다른 결과를 보이고 있다. 서로 평가한 모음이 다르다 하더라도 본 연구에서는 입술의 상하가 3단계 좌우가 3단계로 벌어짐을 관찰하였다. 이는 마커 부착을 통한 조사와 화상을 통한 조사의 차이로 보이나 자세한 결과는 면밀한 분석이 필요한 것으로 사료된다. 또한 [13], [14]에서 1인을 대상으로 실험하였으나, 본 실험에서는 총 80명이 실험에 이용되었고, 이들의 오차를 정규화를 통하여 줄였으며, 30인을 대상으로 분석하고, 50명이 발화한 것으로 인식 실험을 하여 견고성을 확인하였다. 또한 음성 인식에서는 '오/우'에서 오류가 많음[13]비하여 화상을 이용한 인식에서는 '오/우'에서 거의 오류가 없음을 확인하였고, '아/에/이' 또한 음향과 화상이 상호 보완될 수 있음을 확인하였다.

화상 관점에서 시스템의 인식률의 향상을 위하여 입안의 혀와 치아의 보임을 계수화하여 정보를 추가한다면 더 높은 인식률을 얻을 수 있을 것으로 사료되며, 음향적 관점의 음석인식과 결합하여 각각의 인식 강점에서 확신도를 높인 시스템은 인식률이 매우 향상될 것으로 보인다.

참고문헌

[1] Rajeev Sharma, Vladimir I. Pavlovic, Thomas S, Huang, "Toward Mult-imodal Human-Computer Interface", Proceeding of the IEEE Vol. 86. No 5. May 1998.

[2] Gerasimos Potamianosm, Hans peter Graf, Eric Cosatto, "An Image Transform Approach for HMM based Automatic LipReading", Proceeding of the Int. Conf. On Image Processing. pp. 1731177, 1998.

[3] C. Bregler and Yochai Konig, "Eigenlips' for Robust Speech Recognition", Proc. IEEE Int. Conf. On Acoustics, Speech and Signal Processing, pp. 669-672, 1994

[4] T. Chen, H. P. Graf, and K. Wang, "Lip-synchronization using speech-assisted video processing", IEEE Signal Processing Lett., vol 2, pp. 57-59, 1995

[5] Devi Chandramohan, Peter L. Silsbee, "A Multiple Deformable Template Approach for Visual Speech Recognition", Proc. ICSLP, Voll, pp. 434-437, 1996

[6] Iain Matthnews, Timothy F.Cootes, J. Andrew Banghan, Stephen Cox and Richard Marvey, "Extraction of Visual Features for Lipreading," IEEE Tans. on Pattern Recognition and Machine Analysis, Vol 24, No. 2, pp. 198-213, Feb 2002.

[7] Prasad, K.V., Stork D.G., Wolff G. Preprocessing video images for neural learning of lipreading. Ricoh California Research Center, Technical Report CRC-TR-93-26. (1993)

[8] Watanabe, T. & Kohda, M. "Lip-reading of Japanese vowels using neural networks." 1990

[9] 민소희, 김진영, 최승호, "입술 정보를 이용한 음성 특징 파라메터 추정 및 음성인식 성능 향상", 대한 음성 학회지 : 말소리, 1226-1173, 제44호, pp 83-92, 2002

[10] Rein-Lien Hsu, Mohamend Abdel-Mottaleb, Anil K. Jain, "Face Detection in Color Images", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.24, no.5, May 2002, pp.696-706.

[11] Prasad, K.V., Stork D.G., Wolff G. (1993) Preprocessing video images for neural learning of lipreading. Ricoh California Research Center, Technical Report CRC-TR-93-26. >

[12] 김부일, 양룡, 이태원, "한국어 모음의 조음적 제어에 관한 연구", 한국정보과학회 논문지 '87.8, Vol.14. No.3, pp.194~202

[13] 조용덕, 김기철, 맹승렬, 조정완, "Multi-Layer Perceptron을 이용한 백색 잡음이 섞인 모음의 인식", 한국정보과학회 가을 학술발표논문집 1989, Vol.16 No.2, pp.629~632

[14] 민소희, 김진영, 최승호, 대한음성학회지: 말소리 , 1226-1173 , 제44호 , pp.83-92 , 2002

[15] 김진영, 민소희, 최승호, 음성과학 , 1226-5276 , 제10권2호 , pp.27-33 , 2003

[16] 이은숙, 이호근, 이지근, 김봉완, 이상설, 이용주, 정성태, "견고한 입술 영역 추출을 이용한 립리딘 시스템 설계 및 구현", 한국멀티미디어학회, 춘계학술논문발표집, pp.524~527, 2003