

한국어 파열음 인식을 위한

피쳐 셋 입력 인공 신경망 모델에 관한 연구

김 기석, 김 인범, 황 희음

서울대학교 컴퓨터공학과

A STUDY ON THE IMPLEMENTATION OF ARTIFICIAL NEURAL NET MODELS WITH FEATURE SET INPUT FOR RECOGNITION OF KOREAN PLOSIVE CONSONANTS

Kiseok Kim, Inbum Kim and Heeyung Hwang
Dept. of Computer Engineering, Seoul National University

ABSTRACT

The main problem in speech recognition is the enormous variability in acoustic signals due to complex but predictable contextual effects. Especially in plosive consonants it is very difficult to find invariant cue due to various contextual effects, but humans use these contextual effects as helpful information in plosive consonant recognition. In this paper we experimented on three artificial neural net models for the recognition of plosive consonants. Neural Net Model I used "Multi-layer Perceptron". Model II used a variation of the "Self-organizing Feature Map Model". And Model III used "Interactive and Competitive Model" to experiment contextual effects.

The recognition experiment was performed on 9 Korean plosive consonants. We used VCV speech chains for the experiment on contextual effects. The speech chain consists of Korean plosive consonants /g, d, b, K, T, P, k, t, p/ (/ㄱ, ㄷ, ㅂ, ㅋ, ㅌ, ㅍ, ㄲ, ㄸ, ㅍ/) and eight Korean monothongs.

The inputs to Neural Net Models were several temporal cues - duration of the silence, transition and vot -, and the extent of the VC formant transitions to the presence of voicing energy during closure, burst intensity, presence of aspiration, amount of low frequency energy present at voicing onset, and CV formant transition extent from the acoustic signals. Model I showed about 55 - 67 %, Model II showed about 60%, and Model III showed about 67% recognition rate.

INTRODUCTION

음성인식은 크게 두 단계를 거친다. 첫째, 음성 신호로부터 음소나 음절 등 음성의 최소 단위로 인식하는 과정이요, 둘째는 음성의 최소 단위로부터, 발음된 음성의 문장 및 의미불 해석해내는 과정이다. 첫째 단계에서는 FFT 연산, 스펙트럼 변환 등의 음성 신호 처리 과정을 거쳐, 음향 음성학, 형태학, 음운론적 지식 등을 이용하여 음소로 해석된다. 둘째로는, 인공 지능 기법에 의한 어휘 지식, 문법 및 의미 지식 등을 통하여 주어진 음소 스트림으로부터 문법 및 의미에 맞는 문장을 추론 해내는 과정으로 음성 이해 시스템으로 일컬어지기도 한다. 과거 수십년 동안의 음성 인식 연구 결과는, 정확한 음소인식, 음성인식 과정에 있어서 첫번째 단계일 뿐만이 아니라 가장 중요한 단계임을 보여준다.

그러한 음성 인식의 과정이 어려운 이유는 다음과 같다. 즉 입력된 음성 신호를 처리하여 그 특성을 추출하고 이를 인식하는 단계에 있어서 음성 자체 내에 내재하고 있는 모호성, 불확실성, 오류성 등은 특징 추출의 과정에 많은 어려움과 정보 손실을 야기시킨다. 즉 연속 발음 시 단어가 단어, 구와 구, 문장과 문장의 구별이 불확실하며 단어들이 연속 발음될 때 서로의 발음에 미치는 영향 등의 인식에는 많은 한계가 있다. 또한 발

음의 속도 차이, 문장 내의 단어의 위치에 따른 발음의 차이, 방언의 차이, 개인마다 다른 성대의 차이, 잡음 등 많은 문제들이 제기된다.

이와 같은 다양한 환경 하에서의 음성 신호로부터, 신뢰할 수 있는 음소의 출력을 얻기 위하여, 음성 피쳐의 추출 및 이를 퍼지 개념으로 표현하여 인식하는 실험[4] 등 많은 연구가 이루어져 왔다. 기존의 많은 피쳐 추출을 통한 규칙 기반 시스템은 각종 피쳐를 추출하는 과정 및 이를 결합하는 모든 과정에서 프로그래밍과 통계적 분석 등의 많은 노력이 들어간다. 본 연구에서는 이들 피쳐의 추출 및 결합 과정을 인간의 인지 모델에 근접한 인공 신경망 모델을 통하여 이루어지도록 한다. 본 논문은 한국어 음소 인식을 위한 인공 신경망 모델의 도입을 위하여, 한국어 파열음을 대상으로 실험하였다. 파열음은 한국어 음소 중에서 가장 인식하기가 어려운 음소이다. 즉 그 특징 추출이 매우 어렵고, 특히 특징이 시간적으로 분포되어 있어 이들을 결합하여 인식하기가 매우 어렵다. 또한 주위의 음소에 따라 많은 영향을 받는다. 따라서 그 특징의 결합에 있어서 주변 모음을 고려하여야 한다. 그것을 위하여 본 논문은 세 가지의 인공 신경망 모델을 통하여 파열음의 인식을 위한 피쳐의 선택 및 결합을 위한 인공 신경망 모델의 구성 방법을 실험하였다.

본 논문은 먼저 한국어 파열음의 특징을 살펴보고, 본 논문에서 사용한 음성 데이터의 수집 및 처리 과정을 설명하였다. 그리고 본 논문에서 사용한 세 모델의 원리, 실험 방법 및 결과를 기록하고, 기타 문맥에 의한 실험 및 토의를 행하였다.

한국어 파열음과 그 특징

파열음은 혀에서 올라오는 공기를 혀와 이외의 어떤 지점에서 일단 완전히 막아 공기의 압력이 높아진 다음에 갑자기 파열시켜 얻는 소리이다. 한국어 파열음은 일반적으로 6가지 종류를 갖는 외국의 파열음과는 달리 모두 9가지 종류의 파열음이 있다. [1] 9개의 파열음은 조음의 방식에 따라 크게 3가지의 파열음으로, (a) slightly aspirated and lax, (b) unaspirated and tense, and (c) heavily aspirated and tense in initial position로 구분이 된다. (a)에 해당하는 파열음은 /ㄱ, ㄷ, ㅂ/ 이 있으며 이는 평음 파열음이라 칭하고 (b)에 해당하는 파열음은 /ㄲ, ㄸ, ㅍ/ 이 있으며 이는 무기 정음 파열음이라 칭하고 (c)에 해당하는 파열음은 /ㅋ, ㅌ, ㅍ/ 이 있으며 이는 유기 정음 파열음이라 칭한다.

조음 위치에 따라 한국어 파열음은 양순 파열음, 치경 파열음, 연구개 파열음으로 구분이 된다. 양순 파열음에는 /ㅁ, ㅂ, ㅍ/ 이 있으며 치경파열음은 /ㄷ, ㄸ, ㅌ/ 연구개 파열음은 /ㄱ, ㅋ, ㄲ/ 이 있다. 따라서 조음 방식과 조음 위치에 따른 각기 3가지의 구분은 9가지 유형의 폐쇄음(stops)을 구성하게 된다.

파열음을 컴퓨터로 인식하기 위한 음향 특성(acoustic cues)들은 포르만트의 추이(formant transitions), 파열(bursts)부의 스펙트럼의 특성 및 시상(timing)에 관한 특성 등으로 나누어 볼 수 있다[2]. 본 연구에서는 기존의 음성학적으로 연구된 파열음 구분을 위한 여러 변별적 요소(distinctive features)를 정리하고

그 중 프로그램과 할 수 있는 11 가지의 특징을 선정하여 구현 해 보았다. 본 연구에서 사용된 특징변수에 대한 리스트가 표 1에 나타나있다.

그림 1은 조음 방식 및 조음점에 대한 파열음의 특성을 스펙트로그램을 통하여 살펴본 것이다. 먼저 앞의 두 그림에서 (a)의 연구개 파열음 /g/와 (b)의 양순 파열음 /b/의 두 파열음에 있어서 조음점이 다른 경우의 포르مان트의 추이 형상이 다른 것을 관찰할 수 있다. 즉 (a)에서는 제 3 포르مان트가 앞으로에서 자음부로 갈때 아래로 향하지만, 두번째 그림에서는 위로 향하는 것을 볼 수 있다. 그리고 (b)의 평음 파열음과 한국어에만 존재하는 (c)의 무기 경음 파열음의 경우에는 SILDU 즉 파열음 형성시 폐쇄 기간의 차이가 뚜렷이 나타남을 볼 수 있다.

데이터 수집과 전처리

실험은 모음+자음+모음 연쇄 음성(VCV speech chain)에서 행하였다. VCV speech chain은 연속 음성상에서의 문맥효과(context effects)를 실험하기 위해서 사용된다[3]. 실험 데이터는 8개의 한국어 단모음인, V = [아,어,오,우,이,에,애]와 9 개의 파열음인 C = [ㄱ, ㄷ, ㅂ, ㄱ, ㅌ, ㅍ, ㄷ, ㅌ, ㅍ]이다. 학습 데이터는 72개의 파열음이고, 테스트는 576개의 파열음으로 실시하였다. 한 명의 남성 화자가 조음한 방에서, 감음 방지 마이크로폰으로 두번째 발음한 음성을 테이프에 녹음하였다. 이 녹음된 음성신호는 적절한 크기로 증폭이 된 후, 4.95 KHz의 저대역 필터를 거쳐 필터링을 한다. 그 후에 12 bit A/D converter로 10KHz로 샘플링을 하고, 256 point Hamming Window를 사용하여 매 6.4 msec마다 계산된 256 point FFT로 LPC와 포르مان트를 얻는다. 그리고 각 프레임마다 영교차율, 에너지, 레지듀얼 에러 등을 구하였다.

그림 2은 /어가/ 발음의 대수에너지, 영교차율, 레지듀얼 에러의 추출 결과를 그래프로 그린 것이다. 그림 3은 /어가/ 발음의 포르مان트를 그린 것이다.

기본 모델을 이용한 실험(MODEL I)

본 실험에서는 기본이 될 수 있는 모델을 다층 퍼셉트론 모델(MLP)을 이용하여 실험하였다. MLP(Multi-layer Perceptron)모델[5]은 입력층, 출력층, 히든층으로 구성되어 있다. 각 층은 인접한 층과 완전연결이 되어 있고, 각 연결은 연결 강도(connection weight)를 가지고 있다. MLP에서의 학습은 연결 강도를 조절하는 것이다. MLP는 Error Back Propagation Learning Rule라고 잘 알려져 있는 Generalized Delta Rule을 사용한다.

먼저 표 1에서의 피쳐 중에서 7개 그리고 11개의 피쳐를 골라 이를 구한 값을 각각의 최대 최소 값을 구하였다. 그리고

이들 최대 최소 값을 통하여 피쳐 각각을 10등분하여 10개의 벡터를 형성하였다. 이를 Multi-layer Perceptron의 입력으로 사용하였다. 7개의 피쳐를 사용한 경우에는 입력 노드가 70개, 출력 노드가 파열음 9개에 해당하는 9개, 그리고 히든 층의 노드가 41개이다. 약 1500번의 반복 학습 뒤에 얻은 인식율은 50.5% 이었다. 그러나 피쳐를 4개 더 추가하여 입력 노드를 110개, 출력 노드 9개, 히든 노드를 41개로 하였을 때에는 185번의 반복 학습 뒤에 약 55%의 인식율의 향상을 보였다. 이를 통하여 피쳐의 추가가 인식율의 향상에 기여함을 볼 수 있다. 표 2에는 기본 모델의 인식 결과가 나타나 있다.

표 1. 파열음 인식을 위한 음향 피쳐
Table 1. Acoustic Features of Plosive Recognition

Acoustic Features	Description
FOSLF2, FOSLF3	FOLLOWING SLOPE FORMANT 2,3
PRSLF2, PRSLF3	PRECEDING SLOPE FORMANT 2,3
FOPSF2, FOPSF3	FOLLOWING PSEUDOLOCI FORMANT 2,3
PRPSF2, PRPSF3	PRECEDING PSEUDOLOCI FORMANT 2,3
SILDU	SILENCE DURATION
FTRDU	FORMANT TRANSITION DURATION
VOT	VOICE ONSET TIME
VOICE	VOICING
RATIOEN	EN(897 - 1697)/EN(1697 - 5000)
ICOFG	EN(1014 - 4010)
IBRENG	EN(897-1697) - EN(1697 - 5000)
POWER	LOG ENERGY
ZCR	ZERO CROSSING RATES
ALPHA(RER)	TOTAL LPC ERROR RESIDUAL ERROR
ABRUPT	CHANGE IN RMS POWER OVER TIME STOPS/VOWEL TRANSITIONS
TENSE/LAX	THE DEGREE OF ENERGY DISPLAYED BY THE PUFF

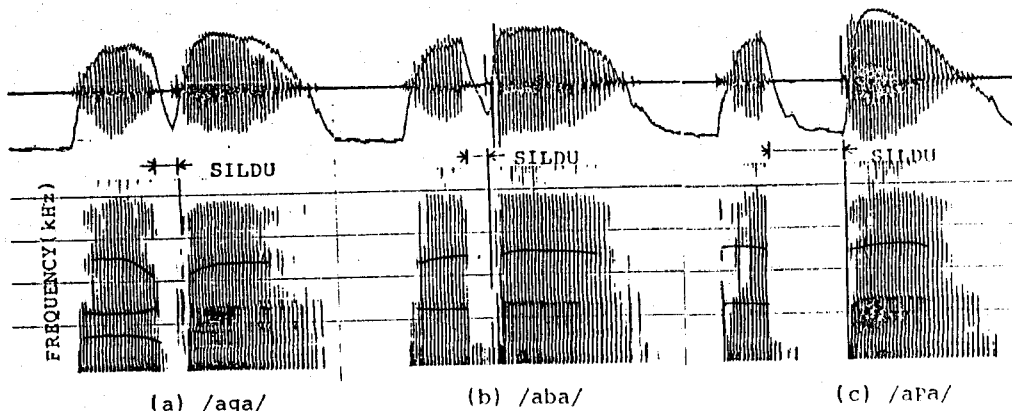


그림 1. 조음 방식 및 조음점에 따른 한국어 파열음의 SILDU와 포르مان트 추이형상의 차이
Fig. 1. The differences of SILDU and the pattern of formant transition in Korean plosive consonants by the manner and place of articulation

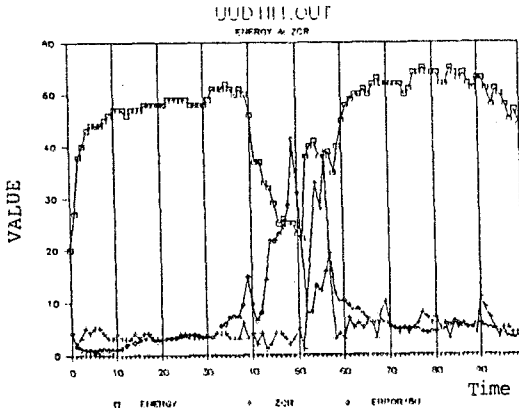


그림 2. 대수 에너지, 영교차율, 레지듀얼 에러(/우디)
Fig. 2. Log Energy, zero crossing rate and residual error of /udi(/우디)

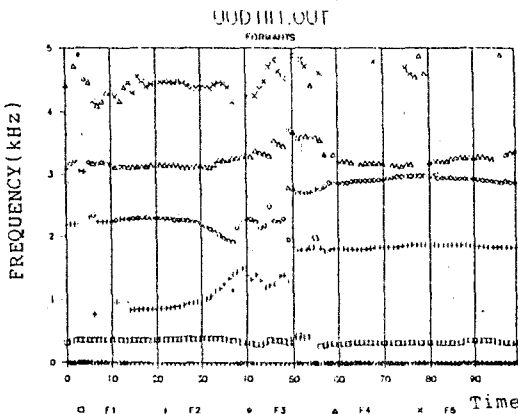


그림 3. /우디/ 음성체의 포르만트 값
Fig. 3. Formant Values of Speech /udi(/우디/)

SELF-ORGANIZING FEATURE MAPS(MODEL II)을 사용한 실험

Kohonen은 두꺼에서 일어나는 인지 현상과 유사한 Self-Organizing Feature Maps란 알고리즘을 발표하였다[5,6]. Kohonen의 알고리즘은 입력 노드에서 2차원 그리드(grid)에 배열되는 출력 노드 사이의 weight를 조절하는 벡터 quantizer를 생산하는 것이다.

이 모델의 특징은 EBP모델등과는 달리 관리자가 없는 학습 방법을 사용한다는 것이다. 즉, 원하는 출력 값의 지정 없이 입력이 들어가고, 프로세스가 진행됨에 따라 weight는 자동적으로 조절이 되는 것이다. 어떤 입력값과 가장 유사한 출력 노드 j의 근처에 있는 출력 노드의 집합을 neighbor라고 부르는데, 이것들과 연결된 weight를 조정하게 된다. 각 iteration마다 노드 j로부터 정해진 거리 d 내의 노드들은, 그 때의 neighbor가 되고, 반복될수록 d는 줄어들게 된다. 각 iteration마다 weight는 현재의 weight와 입력의 차에 비례해서 보정이 된다.

우리가 실험한 모델의 구조는 입력 데이터가 들어갈 수 있는 input node가 있고, 그것들 각각과 full connect된 25 x 25, 2차원 feature map의 output node, 그리고 그들 사이의 weight로 구성되어 있다.

표 2. 기본 모델(MODEL I)의 인식율
Table 2. Recognition Result of the baseline model(MODEL I)

	g	d	b	K	T	P	k	t	p	
g	64	27	6	7	8	1	0	7	4	4
d	64	3	41	2	4	2	6	3	2	1
b	64	3	12	31	7	2	4	0	0	5
K	64	3	7	4	29	0	7	2	8	4
T	64	0	3	0	4	32	25	0	0	0
P	64	0	2	0	4	6	51	0	0	1
k	64	4	2	0	9	0	1	39	4	5
t	64	2	7	0	4	1	0	5	37	8
p	64	6	4	4	4	0	2	7	6	31
Total	576	correct			318	rates			55.2083 %	

표 3. 모델 II의 인식율
Table 3. Recognition Result of MODEL II

	g	d	b	K	T	P	k	t	p	Total	Rates
Model I	27	41	31	29	32	51	39	37	31	318/576	55.2%
Exp.2-1	34	28	16	51	14	12	65	17	16	253/576	43.9%
Exp.2-2	41	28	16	57	59	36	65	34	12	348/576	60.4%

특성 데이터를 두가지로 정의해서, 두가지 실험을 실시했다. 첫번째는 자음 중심점으로부터 5개의 frame씩 formant1, formant2, formant3값을 읽어, 그 평균값들과의 차이를 15개의 입력 노드에 입력시켰다. 인접 formant값의 편이 상태가 positive의 특징을 잘 반영한다는 결과가 음향, 음성학적으로 밝혀진 상태이기 때문에, 이것을 적용한 실험이다.

두번째 실험은 한국어 plosive를 구별할 수 있는 feature중 formant2와 formant3를 가지고 얻은 7가지(fops2, fops3, foslf2, foslf3, sildu, ftrdu, vot)를 선택해서, 그것을 입력으로 정의해서 실험하였다.

두 실험 모두 V + C + V chain의 72개의 데이터를 학습 데이터로 하였고, weight를 임의의 0과 1사이의 값으로 초기화 하여 학습을 시작하였다. Kohonen의 SOFM의 알고리즘에 따라 weight를 조정한다. test는 V + C + V chain의 576개의 data를 가지고 실시했는데, 각 실험의 결과는 표 3에 나타나 있다.

IAC MODEL(MODEL III)을 이용한 실험

이 실험은 피쳐(feature), 음소(phoneme), 워드(word)등 3가지의 계층으로 구성된 IAC model[7,8]을 실험하였다. 피쳐 계층은 입력 신호의 음향학적 피쳐를 가지고 있는 बैं크(bank)들로 구성되어 있다. 음소 계층에서는 피쳐 계층에서 추출한 음향학적 피쳐를 기반으로 음소를 구별하게 된다. 워드 계층에서는 음소 계층에서 추출한 음소를 가지고, 실제 워드를 찾는다.

이 모델의 특징은 interactive와 competition 프로세스를 하는 것인데, 여기서 interactive란 음소를 찾기 위해서 피쳐에서, 또 워드를 찾기 위해서 음소에서의 bottom up 프로세스와 하위 계층을 위한 상위 계층에서의 정보 전달인 top down 프로세스가 동시에 존재하는 것을 의미한다. Competition란 같은 계층에 있는 여러 가설(hypotheses)들은 서로 억제하며 경쟁을 하게 되고 그 자신을 지원하는 인접 계층의 가설들은 강조하는 것을 말한다.

각 layer안에서는 full로 연결이 있고, layer간의 node는 역시 full로 연결이 되어있다. 입력 도메인은 VCV chain에서 모음과 파열음들. 각 모음과 파열음들의 feature node 들은 80개의 입력 벡터로 구성되어있다. 그러나 모음과 파열음의 입력 벡터들은 다르게 만들어진다. 즉 모음인 경우는 모음 중심점에서의 0-4000Hz 사이의 formant 1,2,3,4,5 값을 50으로 나눈 값을 80개의 vector로 mapping시켜 그 값을 feature layer의 입력으로 정의한다. 피쳐로는 파열음을 분류하기에 유용한 7가지 특징(fopsf2, fopsf3, fosl2, fosl3, sildu, firdu, int)을 구해서 각각을 10개의 vector로 mapping시키고 마지막 10개는 모음과의 통일을 위해, null로 첨가한다. 이것을 역시 feature layer의 입력으로 정의한다.

phonem layer에는 225개의 phonem node로 구성되어 있는데, 앞모음 V에는 8개, C에 9개, 뒷모음 V에 8개, 모두 25 종류의 음소들에 대해서 각각 9 가지의 형태로 구성되어있다. word layer에서는 모두 576개의 word로 구성이 되어있는데, 모두 576 개이다. 인식 실험 결과는 표 4에 나타나 있다.

문맥 효과에 관한 실험

파열음의 피쳐는 문맥 즉 전후 모음 또는 자음의 영향을 받아 변화한다.[2,3,4,9]. 본 실험에서는 문맥을 고려하여 인공 신경망 모델에 입력하였을 때 어떻게 인식이 향상되는지를 보았다. 표 5에는 기본 모델 I과 두가지의 실험 결과가 나타나 있다. 여기에서 실험 4-1은 파열음의 후속 모음 8가지를 조음 위치에 따라서 전설, 중설, 후설 모음으로 구분하여 이를 다층 퍼셉트론 모델(모델 I과 동일)에 입력하였을 경우이다. 즉 입력 노드는 모델 I의 110개와, 전설,중설, 후설 모음을 표시하는 각각 3개씩의 9개의 노드로 입력은 모두 119개이다.

그리고 실험 4-2는 파열음의 후속 모음 8개를 그대로 나타 내었다. 즉 입력 노드는 모델 I의 110개와 후속 모음 8개를 표시하는 각각 2개씩의 19개 노드로 입력 노드는 모두 126개이다. 이와 같이 실험하였을 때 실험 결과는 표 5에 나타나 있다. 즉 후속 모음을 정확히 분류할 수록 파열음의 인식이 높아짐을 볼 수 있다.

결론

본 논문에서는 한국어 파열음의 인식을 위하여 3개의 인공 신경망 모델을 설계하여 학습 및 인식 실험을 하였으며, 인공 신경망 모델의 입력으로 파열음의 인식을 위한 특징 변수를 조사하여 그중 일부를 사용하였다. 또한 추출된 특징 변수에 의하여 후속 모음의 여러 환경에서 파열음의 인식 실험을 행하였다. 실험 결과 퍼셉트론 모델에서는 약 55-67%의 인식을 보였으며, Self-Organizing Feature Map에서는 60%, 전후 모음의 인식을 같이 수행하는 IAC 모델의 경우에는 67%의 인식을 보았다. 파열음의 구분을 위한 특징 변수들은 파열음의 인식에서도 복합적으로 작용한다. 특히 문맥적 정보에 대하여 IAC MODEL이 어느 정도의 지식을 결합하는 데 성공적으로 작용 하였음을 볼 수 있었다.

표 4. IAC 모델(모델 III)의 인식율
Table 4. Recognition Result of IAC MODEL(MODEL III)

	g	d	b	K	T	P	k	t	p	Total	Rates
Model I	27	41	31	29	32	51	39	37	31	318/576	55.2%
Exp. 3	44	39	39	41	51	50	48	37	37	386/576	67.0%

표 5. 문맥 효과에 대한 실험 결과
Table 5. The Comparison of Recognition Result of Experiments on Contextual Effect

	g	d	b	K	T	P	k	t	p	Total	Rates
Model I	27	41	31	29	32	51	39	37	31	318/576	55.2%
Exp. 4-1	37	32	44	48	40	34	55	29	30	349/576	60.6%
Exp. 4-2	42	45	43	44	41	46	44	44	39	388/576	67.4%

파열음의 분류에 있어서 후속 모음을 좀 더 정확히 구분하고 난 후 이에 의하여 파열음을 구분했을 경우에 인식이 많이 증가되는 것을 볼 수 있었다. 이를 통하여 파열음의 구분은 파열음 전후 문맥에 대한 사전 정보를 갖는다면 좀더 정확해진다는 것을 알 수 있다.

특징 변수의 정확하고 빠른 추출을 위한 알고리즘과 더불어 더 효율적인 새로운 특징 변수를 발견하려는 노력도 역시 병행되어야 하겠다. 이러한 실험이 계속 된다면 경제성을 고려한 효율적인 특징 변수들의 집합을 결정하게 되며, 이에 의하여 빠르고 정확하게 파열음의 분류가 가능하다고 본다. 특히 앞으로는 음소의 피쳐를 추출할 수 있는 인공 신경망 모델이 개발되어야 한다. 본 논문에서는 특징 변수의 추출을 알고리즘적으로 하였으나 앞으로는 이들 특징 변수들도 인공 신경망을 통하여 추출할 수 있도록 연구하고 있다. 특히 시간적 특징 변수의 추출 및 결합법은 파열음의 인식을 위하여 가장 절정적인 문제라고 볼 수 있다.

이 연구 결과는 다른 음소의 분류 작업에도 사용될 수 있다. 각 음소의 음향 특성을 계속 조사하고 이를 알고리즘으로 구현하는 작업이 병행하면 좋은 음소 인식 시스템을 구현할 수 있을 것이다. 이 연구는 1인의 남성의 음성만을 사용하였으나 여러 명의 남성 그리고 여성 화자에도 확장 연구가 필요하다. 이 경우 특징 변수의 설정 등등이 새로이 연구될 필요가 있다. 또한 이 연구에서는 VCV Speech chain만을 고려하였으나 VC, CV, VCCV Speech chain도 고려하여야만 한다. 그리고 연속 단어 상에서의 파열음의 추출 실험도 앞으로 연구되어야 할 과제이다.

REFERENCES

- [1]. Yang-soo Moon, "Some Aspects of Korean Stop Consonants", 언어학 제 3호, pp 81-88.
- [2]. M.S.Han and R.S.Weitsman, "Acoustic Features of Korean /P,T,K/, /p,t,k/ and /p^h,t^h,k^h/", *Phonetica* 22, 1970, pp 112-128.
- [3]. Hiroya Fujisaki, Masahiko Tominaga, "Automatic Recognition of Voiced Stop Consonants In CV and VCV Utterances", *IEEE ICASSP*, 1982, pp 1996-1999.
- [4]. Piero Demichelis, Renato De Mori, Pietro Laface, Mary O'Kane, "Computer Recognition of Plosive Sounds Using Contextual Information", *IEEE TRANS.ON ASSP*, VOL. ASSP-31, NO. 2, APRIL 1983, pp 359-377.
- [5]. R.Lippman, "An Introduction to computing with neural nets", *IEEE ASSP*, pp. 4-22, April 1989
- [6]. Teuvo Kohonen "The Neural Phonetic Typewriter", *Computer*, vol 21, No.3, March 1988, pp.11-22.
- [7]. Jeffrey L. Elman and James L. McClelland, "Exploiting Lawful Variability In the Speech Wave, In Invariance and Variability In Speech Processes, PP 360 - 385.
- [8]. J.L McClelland and J.L. Elman, "Interactive Processes in Speech Perception: The Trace Model", *Parallel Distributed Processing* vol.2 MIT press, 1986, pp. 59-121.
- [9]. Hong C.Leung and Victor W.Zuc, "Some Phonetic Recognition Experiments Using Artificial Neural Nets", *ICASSP* 88, pp.422-425.