창 립
40주년학술대회
논문 87-I-20-5

한국어 파열음의 자동 인식에 대한 연구 : 한국어 치경 파열음의 자동 분류에 관한 연구

최 윤석°, 김 기석, 황 희웅
서울대학교 공과대학 전자계산기공학과

A Study On The Automatic Discrimination Of The Korean Alveolar Stops.

Yunseok Choi, Kiseok Kim, Heeyeung Hwang
SEOUL NATIONAL UNIVERSITY

(ABSTRACTS)

This paper is the study on the automatic discrimination of the Korean alveolar stops. In Korean, it is necessary to discriminate the asperate/tense plosive for the automatic speech recognition system because we, Korean, distinguish asperate/tense plosive allphones from tense and lax plosive.In order to detect acoustic cues for automatic recognition of the [ㄱ,ㅋ,ㄲ],we have experimented the discrimination of [ㄷ,ㅌ,ㄸ].We used temporal cues like VOT and Silence Duration,etc., and energy cues like ratio of high frequency energy and low frequency energy as the acoustic parameters.The VCV speech data where V is the 8 Simple Vowels and C is the 3 alevolar stops, are used for experiments. The 192 speech data are experimented on and the recognition rate is resulted in about 82% - 95 %.

1.서론

음성 인식 시스템의 구현에 있어 크게 두 가지 방향의 연구가 진행되고 있다.그 하나는 상위 레벨의 언어학적 지식 즉 음성의 음운론적 규칙(Phonological Rule), 문법 및 의미론적 지식 등을 표현하고 이를 적절한 제어 전략을 이용하여 음성의 정확한 인식에 이용하려는 연구이다. 또 하나는 하위 레벨의 음향 음성학적 지식을 발견하고 이를 표현하여 음성 신호의 일차적 해석의 정확도를 높이려는 연구이다.

이러한 두 방향의 연구는 상호 보완적인 관계에 있으며 대규모의 빠르고 정확한 음성 인식 시스템의 구현에 있어 모두 필요한 연구라 볼 수 있다. 본 연구에서는 상위 레벨의 지식의 구현에 앞서 필요한 하위 레벨의 일차적 해석을 위한 음성 신호의 특징을 기술하는 음운론적 발견에 주안점을 두고 있다. 즉 음성 인식 시스템의 구현시 최소 인식 단위로 가장 많이 쓰이고 있는 음소(Phonemes)와 음성 신호에서 추출한 특징(features)과의 관계를 기술하는 음운론적 규칙 (phonological rule)을 발견하는 것을 말하며 이는 음성 인식 시스템의 효율을 결정짓는 중요한 작업이라 하겠다. 음소란 한 언어에서 서로의 사이에 의미적인 구분을 이루지 않고 같은 음으로 받아들여지는 이음(allophone)들의 집합을 말한다.

음향 신호의 정확한 해석을 위한 음운론적 규칙을 음성 인식에 이용하는 절차는 먼저 음성 신호에서 음소 간의 구별을 주는 변별적 요소(distinctive features)를 발견하는 것이며 , 또 이를 근거로 하여 프로그램에 의한 자동 추출을 위해 특징 변수의 추출 알고리즘을 구현하는 것이고 , 또한 이것이 음소간 구별에 어느 정도 기여하는가의 효용성을 조사하는 것을 들 수 있다.

본 연구에서는 한국어 음성 이해 시스템의 구현의 전단계로 한국어 음소의 인식 시스템을 구성한다는 목표 아래 한국어 음소 중 치경 파열음에 해당하는 /ㄷ,ㅌ,ㄸ/을 전후 문맥에 의하여 그 특징들을 발견한 후 발견된 특징들에 의하여 음소간의 분류를 행하는 실험을 수행하였다. 한국어 파열음은 모두 9가지 종류로 외국에서의 6가지 종류의 파열음보다 3가지가 더 많으며 이는 외국의 음성에서는 구별되지 않는 유기 경음 파열음이 한국어에서는 다른 음소로 인식되기 때문이다. 이러한 유기 경음 파열음을 다른 파열음과 구분시키는 특징 변수를 발견하는 것은 매우 중요한 의미가 있으며 본 논문에서는 이를 위하여 /ㄷ,ㅌ,ㄸ/중 /ㅌ/을 구별하는 유용한 특징 변수가 무엇인지 조사하기 위해 /ㄷ,ㅌ,ㄸ/의 분류 실험을 행하였다.

2.파열음의 음성학적 특징

파열음은 허파에서 올라오는 공기를 허파 이외의 어떤 지점에서 일단 완전히 막아 압력이 높아진 다음에 갑자기 파열시켜 얻는 소리를 말하며 음향 분석에 있어서 어느 일정 기간 동안의 목음과 폐쇄 해제에 따르는 진폭의 급격한 상승으로 특징지워진다.그 폐쇄 해제시에는 약간의 마찰 잡음(frication noise)과 파열 잡음(burst noise)이 동반된다.

이러한 특성을 가진 파열음의 분류를 위한 변별적 요소(distinctive features)로서는 파열 시작점과 후속 모음의 정상 발성 지점과의 시간 간격인 VOT(voicing onset time)이 있으며 이는 무성 파열음과 유성 파열음을 구분하는 중요한 특징이 된다.또한 제 1 포르만트의 추이 및 fundamental frequency contour 등이 파열음의 유무성을 파악하는 주요 단서이다.

파열음의 조음점의 위치를 파악하기 위한 파라미터로서는 파열시의 주파수 특성과 인접 모음의 포르만트의 추이 등을 들 수 있다.또한 VOT도 조음점에 따라 규칙적으로 변화하는 것이 발견되어 있다.

본 연구에서는 한국어 파열음의 자동 분류를 위해, 그 동안의 실험 음성학에서 밝혀진 여러 변별적 요소들을 구현하는 특징 변수 추출 알고리즘을 실험하였으며, 그 중 파열음의 조음방식을 추론하기 위해 유용한 시상 변수 및 에너지의 변화 정도를 조사하여 이에 따르는 여섯 가지의 특징 변수를 선정,이를 가지고 /ㄷ,ㅍ,ㅌ/에 대하여 인식하는 실험을 행하였다.

3.음향 분석

(1).실험 시료

본 연구에서 쓰인 음성 시료로는 한국어 연속 음성의 모든 문맥의 모음 조합인 VCV 음성 데이타를 사용하였다. V=[아,어,오,우,으,이,에,애],C=[ㄷ,ㅍ,ㅌ]을 사용하였으며 모두 8*3*8=192개의 데이타를 갖고 실험하였다. 조용한 방에서 메탈 테이프에 녹음한 뒤 4.95KHZ로 low-pass filtering하였다. 10KHZ로 표본화한 후 음성부만 골라 diskette에 저장하였다.화자는 1인의 남성 화자로 하였다.

(2).음향 분석

음향 분석을 위해 쓰인 분석 파라미터는 대수 에너지, 영교차율,Residual Error, 포르만트,대역별 에너지 등이며 이들을 각각의 음성 데이타에 관하여 그 결과를 출력하여 각 자음당 특성을 분석하는데 사용하였다. 다음 그림 3-1은 /어타/음성으로부터 계산된 각각의 음향 파라미터의 계산된 결과를 LOTUS_123에 의하여 그래프로 출력한 결과이다.

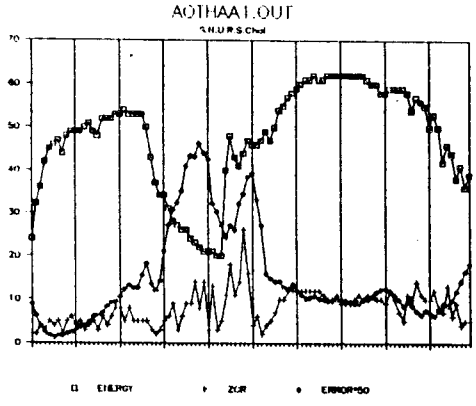

그림 3-1.대수 에너지,영교차율,Residual Error (/어타/)

4.특징 변수의 정의 및 추출 방법

각각의 음성 데이타 192개에 대한 음향 분석 그래프를 모두 얻어 이를 분석하여 파열음의 분류를 위한 특징 변수를 정의하고 이를 분류한다.치경 파열음의 자동 분류를 위해 유용한 특징 변수라고 생각되는 특징 변수로서는 SILDU, FTRDU, VOT, RATIOEN, ICOFG, IBRENG 등을 들 수 있다. SILDU란 파열 시작 이전에 존재하는 폭음 즉 폐쇄 기간을 말하며 이는 파열음의 tense/lax 정도를 분류하는 주요 단서가 된다. FTRDU는 파열음과 후속 모음간의 포르만트의 추이 시간 간격을 말하며 VOT는 폐쇄 시작점으로부터 후속 모음 시작점 사이의 시간 간격을 말한다. 기타의 특징 변수들은 에너지에 관한 변수들로 RATIOEN은 자음부의 800-1600 대역의 에너지와 1600-5000 대역의 에너지의 비를 말하며 ICOFG는 1000-4000 대역의 에너지의 합을 말하고 IBRENG는 800-1600 대역의 에너지와 1600-5000 대역의 에너지의 차이를 말한다.다음에는 이들 특징 변수들을 구하는 알고리즘이 나타나있다.

(step 1)
후속모음 중앙부(IMDFV)를 추출한다.
파열 시작점(ISTBURST)를 추출한다.
후속모음 안정 시작점(IENDTR)을 추출한다.
전위모음 안정 종료점(ISTTR)을 추출한다.
전위모음 후의 폐쇄 시작점(ISTSIL)을 추출한다.

(step 2)

SILDU = ISTBURST - ISTSIL
FTRDU = IENDTR - ISTBURST
VOT = ISTVOT - ISTBURST

(step 3)

VPOINT = (ISTVOT-ISTBURST)/2 + ISTVOT

RATIOEN = EN(800-1600) / EN(1600-1500) at VPOINT
IBRENG = EN(800-1600) - EN(1600-1500) at VPOINT
ICDFG = EN(1000-4000) at VPOINT

여기에서 특징 변수의 추출에 중요한 작용을 하는 것으로 각 기준점들의 추출을 들 수 있다. 기준점들로는 IMDPV, ISTTR, ISTSIL, ISTBURST, ISTVOT, IENDTR, IMDFV 등을 들 수 있다. IMDPV,IMDFV는 각 파열음 인접 모음의 중심점을 말하며, ISTSIL은 파열음의 폐쇄 개시점, ISRBURST 는 폐쇄 해제점, ISTVOT는 후속 모음의 성대 진동 개시점, IENDTR은 포르만트의 변이가 끝나는 점이다. 본 연구에서는 특징 변수의 정확한 의미와 변별적 효용성을 조사하기 위하여 이들 기준점들을 자동 추출하지 않고 눈으로 조사하여 입력시켰다. 다음 그림 4-1에는 음성 신호상에서 본 이들 기준점들의 위치이다.

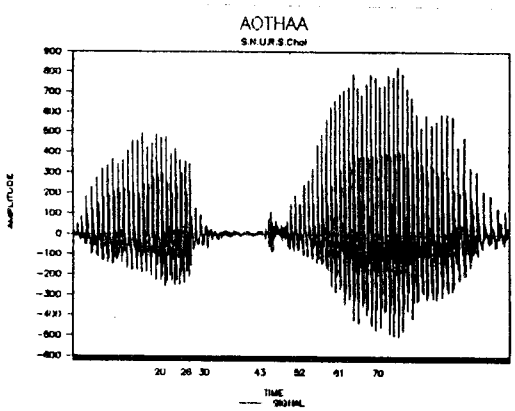

그림 4-1. 기준점들의 위치

5.특징 변수 추출 결과의 분석

위의 알고리즘에 의하여 192개의 음성 데이타에 대하여 특징 변수를 모두 구한 뒤 각각을 후속 모음의 종류에 따라 분석한 뒤 이들 분석 결과에 따라 표준 패턴 및 유효성을 출력한다.

(1).표준 패턴(TEMPLATE)의 작성

여섯 개의 특징 변수 Pi 에 대해 다음과 같은 방법으로 표준 패턴을 얻어낸다. 먼저 n개의 training data dj 로부터 Pi를 추출한 후 Pi의 표준 편차가 최소가 되는 70%의 Pi를 취해 그것의 평균을 표준 패턴의 값으로 갖는다.

SILDU	VOT	FTRDU	RATIOEN	ICOFG	IBRENG
9.04	4.58	22.93	0.15	214.26	233.79
20.78	3.58	18.44	0.16	304.48	275.99
14.29	8.82	26.04	0.16	143.52	190.28

표 5-1.Template table(후속 모음을 구별하지 않을 때)

SILDU	VOT	FTRDU	RATIOEN	ICOFG	IBRENG
9.35	4.94	23.29	0.12	202.27	273.89
21.24	3.71	18.29	0.12	285.38	307.32
14.59	9.47	26.76	0.11	116.77	231.48

표 5-2.Template table(후속 모음이 전설모음일 때)

SILDU	VOT	FTRDU	RATIOEN	ICOFG	IBRENG
8.00	3.00	22.67	0.11	270.03	276.20
20.33	4.17	20.33	0.11	254.18	229.33
13.80	9.00	26.40	0.12	136.22	211.14

표 5-3.Template table(후속 모음이 /아/일 때)

(2).각 특징 변수의 집중도

각 자음 Cj 에 대한 특징 변수 Pi의 유효성을 알아보기 위해 n개의 training data dx 에 대해 다음과 같은 분류를 시도하여 Cj의 Pi에 의한 집중도 Kij를 얻어낸다.

$$Kij = \frac{(dx = Cj)\text{인 횟수}}{Pi\text{가 } dx\text{를 } Cj\text{라고 인식한 횟수}}$$

(%)

SILDU	VOT	FTRDU	RATIOEN	ICOFG	IBRENG
96.9	51.4	55.9	33.0	44.2	36.8
93.3	60.2	81.5	31.6	66.7	48.2
85.3	84.1	74.6	0.0	64.9	53.5

표 5.4 특징 변수의 집중도 (후속 모음을 구분하지 않을 때)

(%)

SILDU	VOT	FTRDU	RATIOEN	ICOFG	IBRENG
100.0	55.6	60.7	28.6	33.3	40.0
92.0	80.0	90.9	0.0	64.0	53.6
95.7	91.7	77.3	35.1	69.2	44.8

표 5.5 특징 변수의 집중도 (후속 모음이 전설모음일 때)

(%)

SILDU	VOT	FTRDU	RATIOEN	ICOFG	IBRENG
100.0	66.7	71.4	41.7	50.0	54.5
100.0	62.5	70.0	0.0	28.6	16.7
88.9	100.0	85.7	33.3	77.8	42.9

표 5.6 특징 변수의 집중도(후속 모음이 /아/일때)

(3).각 특징 변수의 유효성

각 특징 변수가 자음의 특성을 나타내는 정도를 측정하기 위해 12가지의 모든 경우에 대해서 다음과 같은 분석을 행하였다.

(%)

SILDU	VOT	FTRDU	RATIOEN	ICOFG	IBRENG
96	28	59	48	35	21
87	82	82	48	68	62
90	90	68	0	75	59

표 5-7.각 특징 변수의 /ㄷ,ㅃ,ㅌ/ 인식율(후속 모음을 구별하지 않을 때)

(%)

SILDU	VOT	FTRDU	RATIOEN	ICOFG	IBRENG
100	41	70	41	29	25
95	75	83	0	66	62
91	91	70	54	75	54

표 5-8.각 특징 변수의 /ㄷ,ㅃ,ㅌ/ 인식율(후속 모음이 전설모음일 때)

(%)

SILDU	VOT	FTRDU	RATIOEN	ICOFG	IBRENG
100	75	62	62	50	75
87	62	87	0	25	12
100	87	75	50	87	37

표 5-9.각 특징 변수의 /ㄷ,ㅃ,ㅌ/ 인식율(후속 모음이 /아/일 때)

조사된 결과로는, 후속 모음의 분류가 전혀 선행되지 않거나 혹은 전설,중설, 후설 정도의 분류만이 선행될 때, SILDU는 /ㄷ,ㅃ,ㅌ/ 모두의, VOT는 /ㅌ/만의, FTRDU는 /ㅃ,ㅌ/의 특징을 잘 나타냄을 발견하였다. 이 경우 에너지 변수 RATIOEN, ICOFG, IBRENG 는 특징 변수로서의 유용성이 없었다. 그러나, 후속 모음을 분명히 구별할 경우, 본 실험에서 사용한 6 개의 변수 모두가 어떤 자음에 대해서는 유용한 특징으로서의 타당성이 있음을 알 수 있었다.

그러나 어느 한 특징 변수가 /ㄷ,ㅃ,ㅌ/을 각각 명확히 구분하기를 기대할 수 없으므로 어떤 자음의 특징을 잘 나타내는지를 파악,이를 지식으로 이용하여 전체적인 분류를 시도하였다.

표 5-2에서 5-5는 특징 변수의 집중도 및 각 특징 변수별 인식율을 나타낸다.

6.인식 모듈의 구현과 실험 결과

위에서 구한 특징 변수 값을 근거로 하여 그 중 유효한 6가지의 특징 변수를 사용하여 인식 모듈을 설계하였다. 인식 모듈은 크게 세 가지 방법으로 구성하여 실험하였다. 첫째, 후속 모음을 인식한 후 그 인식된 모음에 해당하는 표준 패턴과 비교하는 법, 둘째로는 후속 모음을 음소 단위로 인식하지 않고 단지 전설,중설,후설의 여부만을 인식하여 그 결과에 따라 표준 패턴과 비교하는 법,셋째로 후속 모음을 고려하지 않고 구해진 표준 패턴과 비교하는 법 등으로 인식 실험을 해 보았다.

(1).인식 결정 방법

테스트 데이타 dx 에 대해서 특징 변수 Pi가 자음 Cj라고 인식하면 Cj에 대해 Pi가 유용하다고 판단된 경우 Cj의 Pi에 의한 집중도 Kij의 합이 최대가 되는 자음을 인식된 결과로 결정한다.

(2).실험 결과

***** TEST1 *****

:	D :	DD :	T :	RATIO
D :	36 :	16 :	12 :	56 (%)
DD :	1 :	62 :	1 :	96 (%)
T :	2 :	1 :	61 :	95 (%)
				82 (%)

표 6-1.문맥을 고려하지 않은 자음 인식 결과

***** TEST2 *****

:	D :	DD :	T :	RATIO
D :	45 :	9 :	10 :	70 (%)
DD :	2 :	61 :	1 :	95 (%)
T :	0 :	1 :	63 :	98 (%)
				88 (%)

표 6-2.전설,중설,후설 인식 후 자음 인식 결과

***** TEST3 *****

:	D :	DD :	T :	RATIO
D :	59 :	4 :	1 :	92 (%)
DD :	2 :	62 :	0 :	96 (%)
T :	1 :	1 :	62 :	96 (%)
				95 (%)

표 6-3.후속 모음 인식 후 자음 인식 결과

7.결론

본 실험에서는 6 가지의 특징 변수를 이용하여 /ㄷ,ㅍ,ㅌ/의 분류를 시도하였다. 특징 변수 Pi의 유효성을 검토한 결과 /ㅍ/과 /ㅌ/에 대해서는 많은 수의 Pi가 발견되나 /ㄷ/의 분류에 유효한 Pi는 비교적 많지 않았다. 그 결과 /ㄷ/의 인식율이 저조하였다. 또한 어떤 한 Pi가 /ㄷ,ㅍ,ㅌ/에 대하여 공통적으로 유효하기를 기대하기는 힘들므로 적절한 패턴 매칭 알고리즘의 설계가 뒷받침될 때 인식율의 향상을 기대할 수 있다.

본 실험 결과 후속 모음의 정보 여하에 따라 파열음의 인식율이 좋게 나오는 것은 파열음의 분류에 앞서 후속 모음의 분류가 필요함을 예증한다. 또한 전위 모음의 구별 역시 후속 파열음의 분류에 영향을 미치리라고 예측한다.

8.참고 문헌

[1].오 영환,국어 음성의 모음.자음.모음 연쇄의 음향 분석,한국 전자 통신 연구소 연구논문집,1987.

[2].오 영환,숫자음 인식을 위한 패턴 매칭 알고리즘 개발 연구,한국 전자 통신 연구소 연구논문집,1986.

[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].Renato De Mori,Pietro Laface, and Yu Mong,"parallel Algorithm for Syllable Recognition in Continuous Speech ",IEEE TRANS. ON PAMI. VOL.PAMI-7, NO,1.JANUARY,1985,pp 56-69.

[6].V.W.ZUE, ACOUSTIC CHARACTERISTICS OF STOP CONSONANTS: A CONTROLLED STUDY, 1980.

[7].J.D.Markel and A.H.Gray, Linear Prediction of Speech,1976.

% 본 논문은 문교부 자유 학술 진흥 연구비에 의한 연구 결과임.