

한국어 마찰음 및 파찰음의 분석과 인식

Analysis and Recognition of Korean Fricatives and Affricates

정 석 재* 정 현 열** 이 무 영**

(Seok-Jae Jeong, Hyun-Yeol Chung, Mu-Young Lee)

요 약

음소를 인식의 기본 단위로 하는 소규모 음성 인식 시스템을 구현하기 위한 기초 연구로서 마찰음(/s(s), ʃ(s'), ㅎ(h)/)과 파찰음(/t(c), ʈ(c'), ʧ(cʰ)/)에 대하여 지속시간, 평균패턴, 분산비를 이용하여 각 음소의 특징을 분석하고 각 음소군 내에서의 식별에 유효한 parameter들을 추출하여 인식 실험을 실시하였다.

분석과 인식에 있어서는, 각 음소를 조음방법별(마찰음군 /s(c), ʃ(c'), ㅎ(cʰ)/)에 대하여 지속시간, 평균패턴, 분산비를 이용하여 각 음소의 특징을 분석하고 각 음소군 내에서의 식별에 유효한 parameter들을 추출하여 인식 실험을 실시하였다.

지속시간의 분포, 평균패턴의 분포, 분산비의 분포를 이용하여 분석한 결과 6차원(c0~c5) 정도의 cepstrum 계수만으로 마찰음 및 파찰음의 식별이 가능하고, 시간방향의 정보는 음성의 始端으로부터 14 frame(140ms) 정도의 특징을 인식 파라미터로 할 경우가 최적임을 알 수 있었다.

이론 이용한 인식실험 결과에서는 조음방법별로 분류된 음소군내의 각 음소에대한 인식실험의 인식률(85% 이하) 보다는 발음방법별 인식실험시의 인식률(91% 이상)이 높게 나타나 동일 음소군 내에서의 각 음소에 대한 식별이 더 어려울 수 있었고, 특징 파라미터의 길이를 음성의 시단으로부터 14 frame 정도로 했을 때 조음방법별 인식률은 평균 81.1%, 발음방법별 인식률은 평균 97.9%로 최고의 인식률을 나타내었다. 특징 파라미터의 길이를 14 frame 이상으로 증가시켜도 인식률은 큰 변화가 없어 분석결과를 잘 설명하고 있음을 알 수 있었다.

ABSTRACT

As a basic research for constructing a small size Korean speech recognition system, analysis and recognition for Korean fricatives and affricates were carried out with 21 order of cepstral coefficients as a feature parameter.

From our analysis using the distributions of durations, average patterns and fisher ratios, we found that c0 to c5 of cepstral coefficients of each frame for 140ms of duration were enough to discriminate phonemes in each of 5 groups (i. e., fricatives /s, s', h/, affricates /c, c', cʰ/ that divided by the manner of articulation and laxies /s, c/, unaspirated tenses / s', c' /, aspirated tenses / h, cʰ / by the manner of pronunciation).

As the recognition experiments, multi-speaker dependent recognition with S-Mahalanobis distance were carried out for the fricatives and affricates. The duration was varied from 50 ms to 200 ms in the experiment to confirm our analysed results.

*금성정밀 제1연구소
**영남대학교 전자공학과

For the groups divided by the manner of pronunciation, the highest average recognition rate of 91.1% correct was obtained by use of c0 to c5 of cepstral coefficients of each 10 ms in 140 ms of duration, while the rate was 81% for the groups divided by the manner of articulation, which shows a good agreement with our analyzed results.

I. 序 論

반도체와 통신기술의 발전과 컴퓨터의 처리속도 향상으로 Man-machine interface의 문제가 현실적으로 부각되고, 이를 해결하기 위한 수단으로 음성인식과 합성에 대한 연구가 활발하게 진행되고 있다.

최근 각 나라별 음성 인식 연구에 관한 현황을 살펴보면, 일본에서는 "음성언어에 의한 man-machine interface에 관한 종합적 연구"의 일환으로 대어휘(3000단어), 연속음성인식 system의 완성을 목표로 하고 있으며, 미국에서는 "DARPA Spoken Language System Program"으로 10,000단어 어휘로 구성되는 연속음성을 95% 이상, 실시간인식, 이해율 목표로 연구를 진행하는 등 선진국에서는 국가적 차원의 음성정보 처리 연구를 활발하게 추진하고 있다.

최근 우리 나라에서도 숫자음성, 단어음성, 모음 및 자음음소, 대용량 단어인식에 관한 연구가 활발하게 진행되고 있다.¹⁾

한국어는 단모음 10개 자음 19개의 음소로 구성되며, 이들의 조합에 의해 실제 약 1000개의 단음절이 사용되고 있다.²⁾ 또 한국어의 어휘는 명사의 약 90%, 용어간의 약 90%, 용어어간의 약 86%가 1음절 및 2음절로 구성된다.³⁾

따라서, 인식의 기본 단위로서는 음소 혹은 그 이하가 유리할 것으로 생각되며, 이 경우 각 음소의 성질을 정확하게 이해하고 있지 않으면 인식시스템 설계시 발생하는 오인식의 규명이 곤란하게 된다.

그리고, 자음중 경음 / ㅃ, ㅌ, ㄱ, ㅋ, ㆁ / 은 외국어에서는 평음 / ㅂ, ㄷ, ㄱ, ㅋ, ㄴ /, 격음 / ㅍ, ㅌ, ㅋ, ㆁ / 과 구별하고 있지 않으나, 우리말에서는 그 구별이 분명하며 일상회화에서 빈번히 사용되고 있는 한국어 특유의 음소이다.

그러나 음성의 자동인식의 입장에서부터 이들 음소의 성질이 다른 음소들과 어떻게 상이하며 외국의 그것과는 어떤 차이가 있는지도 아직 밝혀지지

않고 있다.

음성인식을 전제로 한 자음의 분석과 인식에 대한 예는 파열음 9개에 대한 1989년 정현열 등의 연구⁴⁾에서 찾아볼 수 있으나, 파찰음, 마찰음, 비음, 유음등에 대한 연구는 아직 불충분한 상태에 있다.

따라서 본 논문에서는 음소를 인식의 기본 단위로 하는 소규모 음성 인식 시스템을 구현하기 위한 기초 연구로서 아직까지 그 성질이 명확히 밝혀지지 않고 있는 마찰음 (/ ㅃ, ㅌ, (ㅎ) /)과 파찰음 (/ ㅈ, ㅉ, ㅊ /)의 특징분석을 행하여 이들 음소의 성질과 상호관계를 밝히고자 한다. 또 각 음소군내에서의 식별에 유효한 parameter들을 추출하여 이를 이용하여 인식실험을 실시한다.

II. 한국어 자음의 분류 및 음성 자료

2.1 한국어 자음의 분류

한국어 자음은 조음의 방법(Manner of Articulation) 별로 분류하면 파열음, 마찰음, 파찰음, 비음, 유음으로 분류되어지며 그림 1 과 같은 組織을 가진다.⁵⁾

2.2 음성 자료

단음절 자료로서는 한국어역순사전중의 단어를 출현 빈도순으로 나열하여 누적빈도 90% 이내에 들어가는 501개의 단음절과 음소의 수가 적은 경음 / ㅃ, ㅌ / 에 대해서는 누적빈도 99.9%까지에 들어가는 단음절로부터 뽑은 48개를 추가한 549개로 구성한다.

마찰음 및 파찰음의 단음절자료는 이중에서 선택된 160개를 방음실에서 한국인 성인 남성 3인이 3회 발성한 마찰음 720개, 파찰음 720개로 한다.

각 음소별 자료의 수는 표 1과 같다.

표 1. 마찰음 및 파찰음의 음성자료
Table 1. Phoneme data for fricatives and affricates.

phoneme	number of phonemes	phoneme	number of phonemes
ㅅ (s)	3 4 2	ㅆ (c)	3 4 2
ㅆ (s')	1 6 2	ㅅㅅ (c')	1 6 2
ㅎ (h)	2 1 6	ㅅㅅ (c ^h)	2 1 6

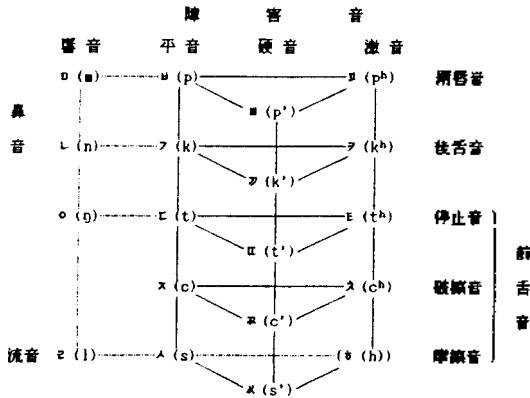


그림 1. 국어 자음의 조직
Fig. 1. Structure of Korean consonants.

Ⅲ. 마찰음 및 파찰음의 분석

음성자료는 그림 2에 보인 바와 같이 4.5KHz Low-Pass Filter를 통과한 후, 10KHz 12bit A/D converter를 통해 이산데이터로 변환되고, 29 Channel Band-Pass Filter(Q=6 단공진, 1/6 octave 간격, 중심 주파수 250Hz-6300Hz)를 통과시켜 분석된다. (frame 길이: 10ms, 분석 window 길이: 20ms). 각 음소에는 視察에 의해 음소의 개시점이라고 생각되는 시점을 시단 frame, 그 음소의 특징을 가장 잘 표현하고 있는 부분을 중심 frame, 음소의 끝 부분을 종단 frame 이라고하는 시간적 label을 부여한다.

이 Label을 참고하여 각 frame별 21차원 LPC cepstrum 계수를 추출하여 음소별 데이터 베이스를 구성한다.

분석에 있어서는 각 음소의 지속시간 및 그 음소의 LPC cepstrum 계수의 평균편편, 각 음소군내의 음소상호간의 분산비(Fisher Ratio)를 이용하여 각 파라미터의 인식에 대한 공헌도, 인식에 필요한 시간

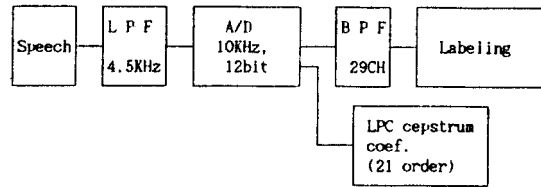


그림 2. 분석의 흐름도
Fig. 2. Flow of analysis

정보의 길이 등을 조사한다.

인식에 있어서는 각 음소를 조음방법별 2group (마찰음군 / ㅅ(s), ㅆ(s'), ㅎ(h) / 과 파찰음군 / ㅆ(c), ㅅㅅ(c'), ㅅㅅ(c^h) /), 발음방법별 3 group (평음군 / ㅅ(s), ㅆ(c) /, 경음군 / ㅆ(s'), ㅅㅅ(c') /, 격음군 / ㅎ(h), ㅅㅅ(c^h) /)으로 분류하여 식별실험을 실시한다. 거리척도로서는 공분산 공통의 Mahalanobis 범거리를 이용한다.

3.1 지속시간의 분포

표 2에 단음절내에서의 각 음소환경이 자음+모음(C+V)인 경우, 자음+모음+자음(C+V+C)인 경우에 대한 음소의 지속시간의 평균(M)과 표준편차(SD)를, 그림 4에 각 음소군별 지속시간의 분포를 나타낸다.

그림 3a)로부터 마찰음의 경우 경음 / ㅆ(s') / 은 평음 / ㅅ(s) / 과 다소 중첩은 되지만 분리가능함을 알 수 있고, 격음 / ㅎ(h) / 은 경음 / ㅆ(s') / 평음 / ㅅ(s) / 과 대부분 중첩되므로 지속시간 정보만으로는 분리하기 힘들 것으로 생각된다.

또, 그림 3b)로부터는 파찰음의 지속시간 분포로 격음 / ㅅㅅ(c^h) / 은 경음 / ㅅㅅ(c') / 과 거의 완전히 분리할 수 있으며 평음 / ㅆ(c) / 과도 다소 중첩되지만 분리가능함을 알 수 있다.

따라서, 이 지속시간정보는 마찰음에서는 평음

표 2. 지속시간의 평균과 표준편차(단위 : ms)
Table 2. Mean and standard deviation of duration (unit : ms)

phoneme	Phoneme Environment				phoneme	Phoneme Environment			
	C+V		C+V+C			C+V		C+V+C	
	M	SD	M	SD		M	SD	M	SD
ㅅ (s)	180	36	170	33	ㅈ (c)	82	17	83	18
ㅆ (s')	174	42	157	53	ㅉ (c')	62	15	58	12
ㅎ (h)	122	34	109	30	ㅊ (c ^h)	146	24	127	21

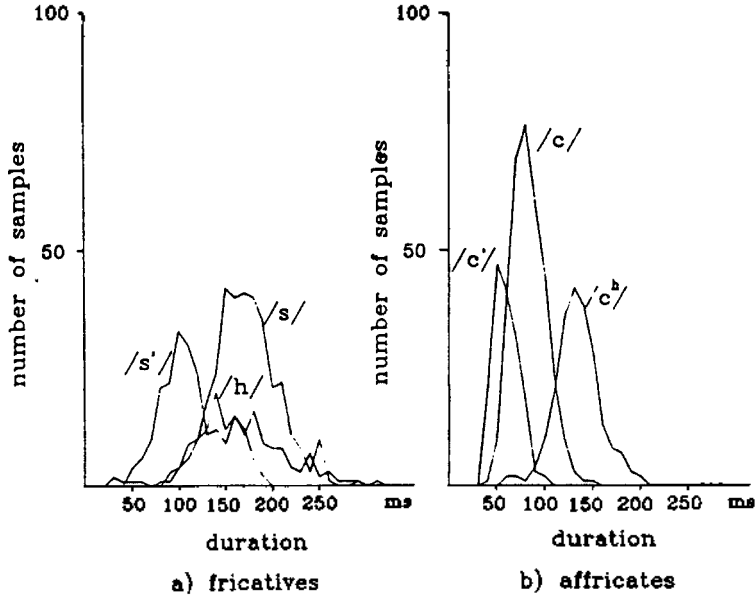


그림 3. 마찰음, 파찰음의 지속시간분포
Fig. 3. Histogram of duration fricatives and affricates.

/ㅅ(s)/과 경음 /ㅆ(s')/의 분리에, 파찰음에서는 경음 /ㅉ(c')/, 평음 /ㅈ(c)/으로 부터 격음 /ㅊ(c^h)/의 분리에 유효하다고 생각된다.

3.2 cepstrum 계수 및 POWER의 평균패턴 분포

각 frame별 LPC cepstrum 계수의 평균패턴을 조사함으로써 각 음소의 특징을 알 수 있다.

그림 4에 제1frame(음성시단으로부터 첫 10ms 구간)과 제2 frame(시단 10ms~20ms 구간)에서의 평균패턴분포중 후속모음이 /ㄱ(o)/인 경우의 일례를 보인다.

이로부터 마찰음군과 파찰음군의 분리가 가능하며, 마찰음군 내부에서는 /ㅅ(s), ㅆ(s')/와 /ㅎ(h)/를 분리할 수 있음을 알 수 있다.

그림 5에 power의 평균패턴 분포(후속모음이 /ㄱ(o)인 경우)의 일례를 보인다.

마찰음이 파찰음보다 시단 30-50ms구간의 평균 power가 적으며 그 기울기도 다르게 나타남을 알 수 있다. 따라서 power성분의 평균패턴을 조사함으로써 마찰음 및 파찰음의 분리가 가능하리라 생각된다.

3.3 분산비의 분포

특징량이 scalar양 (단일특징량)인 경우에 있어서 class의 식별에 이용하는 특징량으로서는 class간의 분산이 큰 동시에 class내의 분산이 작은 것이 요구된다.

이와 같은 특징의 평가량으로서 양 분산의 비가

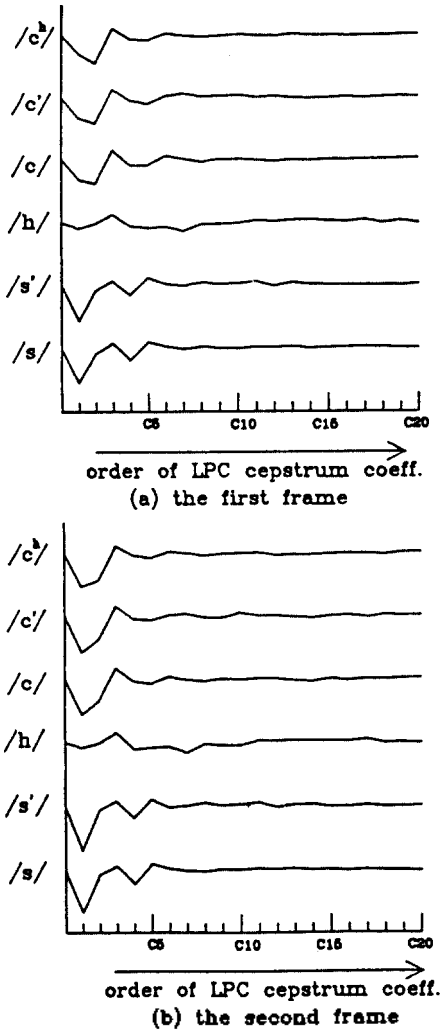


그림 4. 커스트럼 평균패턴의 예
(후속모음이 /ㄱ/인 경우)
Fig. 4. Examples of average cepstrum pattern (followed by /o/).

이용되며, 그 분산비의 정의식은

$$F = \frac{1 / (g-1) \cdot \sum_{k=1}^g (\mu_k - \mu)^2}{1 / \{g(n-1)\} \cdot \sum_{k=1}^g \sum_{i=1}^n (X_{ik} - \mu_k)^2} \quad (1)$$

여기서, X_{ik} : class k 특징량의 i번째의 관측치
 ($i=1,2, \dots, n; k=1,2, \dots, g$)
 μ_k : class k 특징량의 평균치
 μ : 전체 특징량의 평균치
 와 같이 나타난다.

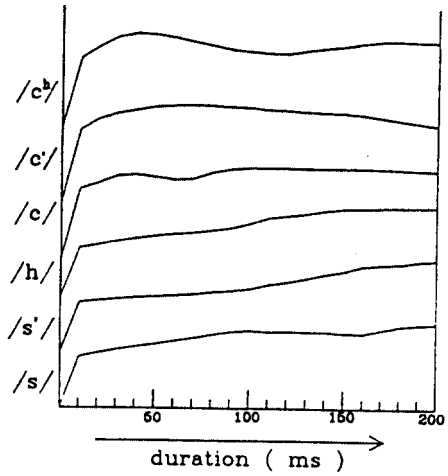


그림 5. 시단으로부터 200ms 구간에 대한 파워패턴의 일예
(후속모음이 /ㄱ/인 경우)
Fig. 5. An example of average power pattern for the first 200 ms(followed by /o/)

본 실험에서는 group별로 분류된 음소군내에서의 각 음소를 후속모음 별로 분류한후 해당음소의 LPC cepstrum 계수의 시단에서부터 150ms 구간(15frame)의 시간변화에 대하여 식(1)에 의해 분산비의 분포를 계산한다. 그림 6에 후속모음이 /ㄱ/인 경우의 일예를 보인다.

즉, 그림 6a)는 마찰음군, 그림 6b)는 파찰음군, 그림 6c)는 평음군, 그림 6d)는 경음군, 그림 6e)는 격음군내의 각 음소 상호간의 분산비를 나타낸 것이다.

그림 6의 분산비의 분포로부터 다음 사항을 확인할 수 있다.

1) 마찰음군 /ㄱ(s), ㅅ(s'), ㅎ(h)/ 내의 각 음소상호간의 분산비 (그림 6a)에서는 cepstrum 계수 c1의 분산비가 가장 크고 그 다음이 c5로서 c1, c5가 식별에 대한 공헌도가 큼을 알 수 있고, 파찰음군 /ㅈ(c), ㅉ(c'), ㅊ(c^h)/ 내에서의 분산비의 분포(그림 6b)에서는 c0, c1, c3가 식별에 대한 공헌도가 큼을 알 수 있다.

2) 평음군 /ㅏ(s), ㅑ(c)/ 상호간의 분산비 (그림 6c)에서는 c0, c1, c2, c5가, 경음군 /ㅓ(s'), ㅕ(c')/ 상호간의 분산비 (그림 6d)에서는 c0, c1이, 격음군 /ㅗ(h), ㅛ(c^h)/ 상호간의 분산비 (그림 6e)에서는 c0, c2, c3의 분산비가 크게 나타나 全體의으로

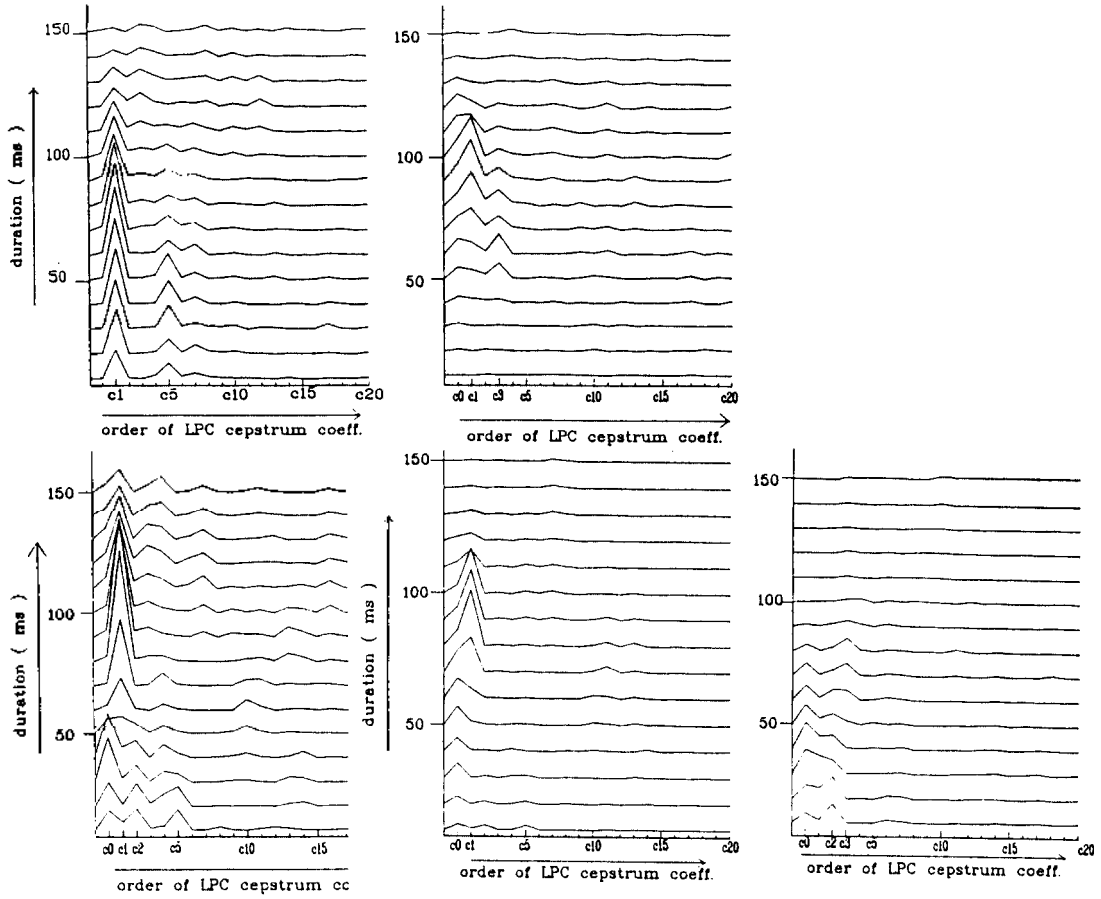


그림 6. 시간으로부터 150ms 구간에 대한 제스트루계수의 분산비의 분포(후속모음이 /ɪ/인 경우)
 Fig. 6. F Ratio of each of cepstrum coefficients for the first 150ms in each phoneme group (followed by /ɪ/).

c0~c5의 cepstrum 계수로 마찰음 및 파찰음의 식별이 가능함을 보여주고 있다.

3) 마찰음 /s(s), ʃ(s'), ㅎ(h)/ 상호간에서는 음성의 시간으로부터 제 14 frame(140ms) 까지의 분산비가 크게 나타나고, 파찰음 /ʒ(c), ㅈ(c'), ㅊ(ch)/ 상호간에서는 제4frame까지는 분산비가 적어 식별에 대한 공헌도가 적게 나타나지만 제5 frame 부터 제11 frame까지의 분산비는 크게 나타나서 이 부분의 Cepstrum 계수(특히 c0, c1, c3)가 식별에 중요함을 알 수 있다.

4) 평음 /ɪ(s), ɹ(c)/ 상호간에서는 제14 frame 까지, 경음 /ʃ(s'), ㅈ(c')/ 상호간에서는 제11 frame까지, 격음 /ㅎ(h), ㅊ(ch)/ 상호간에서

는 제8 frame 까지의 분산비가 크게 나타나 여기에서도 시간방향으로 14 frame 정도의 길이에 대한 Cepstrum 계수를 인식파라미터로 할 경우 식별가능함을 알 수 있다.

이상의 결과를 종합하면, 인식실험에 있어서는 14frame 정도의 시간 방향정보에 대하여 각 frame 별 c0~c5의 cepstrum 계수를 인식 파라미터로 사용하여 인식할 경우 마찰음 및 파찰음의 식별이 가능함을 알 수 있다.

IV. 마찰음 및 파찰음의 인식

이상과 같이 각 음소에 대한 지속시간의 분포,

LPC Cepstrum 계수의 평균패턴의 분포와 분산비의 분포를 이용하여 분석한 결과, 14frame 정도의 시간 방향정보에 대하여 각 frame 별 6차원(c0~c5) 정도의 cepstrum계수를 인식 파라미터로 사용할 경우 마찰음 및 파찰음의 식별이 가능함을 알 수 있었다.

여기서는 이러한 정보를 이용하여 각 음소군별 인식실험을 실시하기로 한다. 거리척도로서는 공분산 공통의 Mahalanobis 범거리를 이용한다.

4.1 공분산 공통의 Mahalanobis 범거리

특징 파라미터의 평가를 위한 거리척도로서는 다차원정규분포를 가정한 통계적인 거리척도가 자주 이용된다. 이 중 Bayes 판정에 기초한 거리척도는 일반적으로 높은 식별률이 얻어지지만, 특징 파라미터의 차원수가 증가하는 데 비례하여 자료수도 증가시키지 않으면 안되며 또 계산량도 상당히 증가하게 된다.

따라서, 본 논문에서는 계산량은 적은 반면 소량의 데이터로서도 비교적 높은 식별률을 얻을 수 있는 분산, 공분산행렬을 공통으로하는 Mahalanobis 범거리(S-Mahalanobis' Generalized Distance)⁽¹⁰⁻¹²⁾를 이용한다. 이하 이에 대해 설명한다.

입력 파라미터인 LPC cepstrum 계수를 $C = \{(c_1, c_2, c_3, \dots, c_d) : d : \text{차원수}\}$, 음소 i에 대한 표준패턴의 평균치를 $\mu_i = (\mu_{i1}, \mu_{i2}, \dots, \mu_{id})$, 공분산 행렬을 Σ_i 라고 할 때 음소 i에 대한 입력 파라미터의 확률 밀도는

$$P_i = \frac{1}{(2\pi)^{d/2} |\Sigma_i|^{1/2}} \exp\{-1/2 (C - \mu_i)^t \Sigma_i^{-1} (C - \mu_i)\} \quad (2)$$

로 나타난다.

각 음소에 대해 식(2)를 계산하고, 그 값이 최대가 되는 음소를 식별 음소로 하는 방법이 Bayes판정에 의한 음소 식별법이다.

여기서 exp 항만을 취한 것이 일반적인 Mahalanobis 거리어며, 이것을 M_i 로 하면,

$$M_i = (C - \mu_i)^t \Sigma_i^{-1} (C - \mu_i) \quad (3)$$

가 되고, 식(3)에 대해서 $\Sigma_i = \Sigma$, 즉, 전체음소에 대한 분산, 공분산 행렬을 같은 값으로 하여 이것을 MS_i 로 하면 식(4)와 같이 나타낼 수 있으며 이를 '공분산공통의 Mahalanobis 범거리'라 부른다.

$$MS_i = C^t \Sigma^{-1} C - 2C^t \Sigma^{-1} \mu_i + \mu_i^t \Sigma^{-1} \mu_i \quad (4)$$

식(4)를 전개하면,

$$MS_i = C^t \Sigma^{-1} C - 2C \Sigma^{-1} \mu_i + \mu_i^t \Sigma^{-1} \mu_i \quad (5)$$

로 되며, 위 식의 제1항은 음소 i에 관계가 없으므로 값의 크기에 따라 음소를 판별하는 경우에는 고려하지 않아도 좋다.

제 3항은 각 음소만의 관계이고, 입력 파라미터에는 무관한 정수이기 때문에 이것을 T_i 로 놓으면 식(5)는

$$MS_i = T_i - 2C \Sigma^{-1} \mu_i = T_i - \sum_{j=1}^d a_{jk} \cdot c_j \quad (6)$$

여기서, $T_i = \mu_i^t \Sigma^{-1} \mu_i$

$a_{jk} = 2 \sum_{l=1}^d \sigma_{jkl} a_{lk} : \sigma_{jkl}$ 는 Σ 의 역행렬의 요소와 같이 나타난다.

식(6)의 계산은 matrix의 계산이 불필요하고, 한 음소당 차원수 d회의 적화(積和)와 1회의 합으로 된다.

이것은 일반적 Mahalanobis 범거리에 비해서 적화(積和)연산량이 2/(d+1)이 된다.¹⁰⁾

4.2. 조음방법별, 발음방법별의 인식 실험

여기서는 각 특징파라미터들이 실제로 6음소의 식별에 어느정도 공헌하는가를 조사하기 위해 식별 실험을 행한다. 인식에 있어서는, 3.의 분석에 있어서와 마찬가지로 각 음소를 조음방법별 (마찰음군 / s(s), ㅅ(s'), ㅎ(h) /, 파찰음군 / ㅈ(c), ㅉ(c'), ㅊ(ch) /), 발음방법별 (평음군 / ㅏ(s), ㅑ(c) /, 경음군 / ㅓ(s'), ㅕ(c') / 격음군 / ㅗ(h), ㅛ(ch) /)의 5group으로 나누어 인식실험을 실시하여 각 특징 파라미터의 인식률이 어느 정도인가를 조사한다.

인식실험에 있어서는 3인 화자의 2회 발성을 표준 패턴으로 하고, 나머지 1회 발성을 입력으로 해서 인식하며, 이를 3회 반복한 평균을 인식률로 한다.

이때 자료취득의 제한으로 인해 후속모음의 영향은 고려하지 않는 것으로하며 거리척도는 위에서 기술한 공분산행렬 공통의 Mahalanobis 범거리척도를 이용한다.

그림 7에 LPC cepstrum 계수를 6차로 했을 때 지속시간의 변화에 따른 인식률의 변화를 나타내었다.

그림 7로부터 다음을 알 수 있다.

1) 지속시간에 따라 각 음소를 조음방법별 2group으로 나누어 인식실험을 실시한 경우 (그림 7a),

인식 파라미터의 길이를 14frame으로 했을 때 인식률이 최고로 나타남을 알 수 있고 마찰음군 /s(s'), ʃ(ʃ'), h(h)/의 경우는 제2frame 이후부터는 시간정보가 인식률에 커다란 변동이 없으며, 파찰음군 /c(c'), ʈ(ʈ'), ʈ(cʰ)/의 경우도 제1frame에서 제4frame까지는 인식률이 낮게 나타나 앞 절의 분산비의 분석결과 (그림 6a), b)와 일치함을 알 수 있다.

2) 각 음소를 발음방법별 3 group으로 나누어 인식실험을 실시한 경우 (그림 7b))에서도 약 14 frame 부근에서 인식률이 높게 나타나 역시 분산비의 분석결과(그림 6c), d), e))와 거의 일치함을 확인할 수 있다.

3) 조음방법별 인식실험의 인식률 (85% 이하) 보다는 발음방법별 인식실험시의 인식률 (91% 이상)이 높게 나타나, 동일 음소군 내에서의 각 음소에 대한 식별이 어려움을 알 수 있다.

4) 특징 파라미터의 길이를 14frame 정도로 했을 때 조음방법별 인식률은 평균 81.1%, 발음방법별 인식률은 평균 97.9%로 최고 인식률을 나타내었고, 특징 파라미터의 길이를 14frame 이상으로 증가시키도 인식률에는 큰 변화가 없었다.

이상의 결과로부터 특징 파라미터의 길이는 140ms 정도가 적당하며 이때 최고 인식률을 나타냄을 알 수 있다.

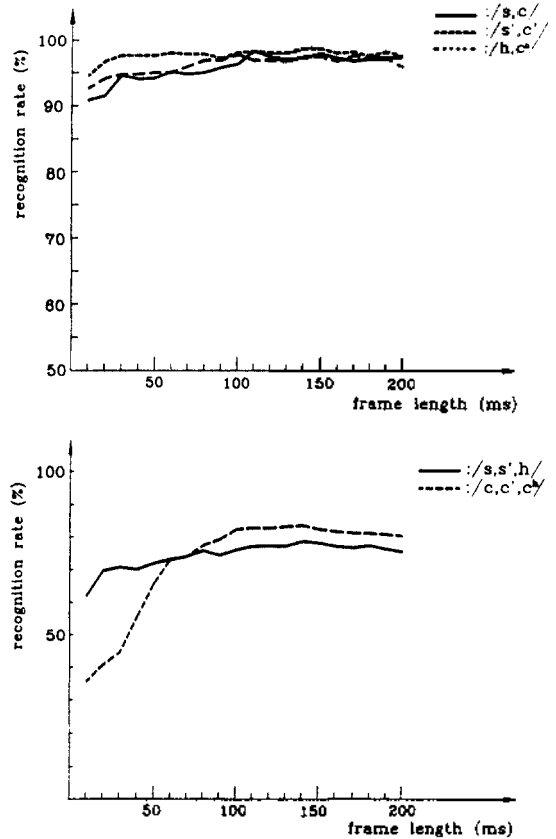


그림 7. 프레임길이에 따른 인식률 a) 발음방법에 따라 분류한 음소그룹내 b) 조음방법에 따라 분류한 음소그룹내
Fig. 7. Recognition rates according to the frame length in each group of phonemes divided by a) the manner of pronunciation, b) the manner of articulation.

V. 결 론

음소를 인식의 기본단위로 하는 소규모 음성인식 시스템을 구현하기 위한 기초연구로서 마찰음과 파찰음에 대하여 지속시간, 평균패턴, 분산비를 이용하여 특징분석을 행하여 각 음소군 내에서의 식별에 유효한 parameter들을 추출하고, 그 결과를 이용하여 인식실험을 실시하였다.

분석결과 다음과 같은 사실을 확인할 수 있었다.

1) 지속시간의 분포로부터는 지속시간 정보가 마찰음에서는 평음과 경음의 분리에, 파찰음에서는 평음, 경음으로부터 격음의 분리에 유효하고, cepstrum 계수의 평균패턴 분포와 power의 평균패턴 분포로부터는 마찰음 및 파찰음의 분리가 가능함을

알 수 있었다.

2) 분산비의 분포로부터 14frame 정도의 시간방향 정보에 대하여 각 frame 별 c0~c5의 cepstrum 계수를 인식 파라미터로 사용하여 인식할 경우 식별이 가능함을 알 수 있었다.

3) 분산비와 평균패턴의 조사결과로부터는 6차원 정도의 cepstrum 계수만으로도 마찰음, 파찰음의 식별이 가능하다고 판단되었다.

인식실험 결과 다음과 같은 사실을 알 수 있었다.

1) 조음방법별 인식실험의 인식률(85% 이하) 보다는 발음방법별 인식률(91% 이상)이 높게 나타나 같은 음소군내에서의 각 음소에 대한 식별이 더 어려움을 알 수 있었다.

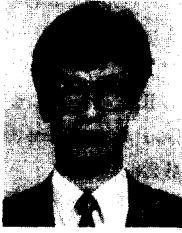
2) 시간방향의 특징 파라미터의 길이를 14frame 정도로 했을 때 조음 방법별 인식률은 평균 81.1%, 발음 방법별 인식률은 평균 97.9%로 최고의 인식률을 나타내었으며, 특징의 길이를 14frame 이상으로 증가시켜도 인식률은 큰 변화가 없었다. 따라서, 시간방향의 특징파라미터의 길이는 140ms 정도가 적당하며 이때 최고 인식률을 나타냄을 알 수 있었다.

參 考 文 獻

1. B.I.Kim and H. Fujisaki: "Analysis and Recognition of Korean Vowels" Annual report of the engineering Reserch institute, University of Tokyo Vol. 32.

2. Y.H.OH and T. Agui: "A Spoken Korean-Digits Recognition System Based on Linear Prediction Spectra" K.I.E.E Vol. 17 No.3(1980.6)
- 3) K.T.Kim, J.Miwa and K. Kido: "Recognition of Isolated Korean Digits Using Band-pass Filters Based on FFT" J.Acoust. Soc. Japan(E) 4,4(1983).
4. H.Y.Chung, S.Makino, and K.Kido: "Analysis and Recognition of Korean isolated Vowels using formant frequency" J.Acoust. Soc. Jpn.(E) 9.5 225-232(1988).
5. 은종관 외: "음소모델링을 이용한 한국어 대용량단어 인식에 관한 연구" 音聲通信 및 信號處理 WORKSHOP, pp.19-24.(Aug. 1989)
6. 許 能: "국어 음운학" 정음사, pp190-213,(1985)
7. 梅田博之: "韓國語의 音聲學的 研究" 형설출판사, pp. 35-37 (1983).
8. 鄭鉉烈, 牧野正三, 城戶健一: "韓國語破裂子音의 分析" 日本音聲學會講演論文集 1-3-3(Oct. 1988)
9. 鄭鉉烈, 牧野正三, 城戶健一: "韓國語語頭破裂子音의 認識" 日本音聲學會講演論文集 1-2-21 (Mar. 1989)
10. 二夫田勝行: "計算屋 が少ない 統計的距離尺度の評價" 松下技研, 技術報告書 (昭和 58.5)
11. 芳賀敏郎 외: "多變量解析法", 日科技連出版社, pp.4 25-429(1985).
12. Douglas O'shaughnessy: "Speech communication" ADDISON-WESLEY PUBLISHING COMPANY, (1990)
13. L.R.Rabiner, P.W.Schafer: "Digital Processing of Speech Signals" Prentice Hall,(1982)
14. J.D.Markel, A.H.Gray: "Linear Prediction of Speech" Springer-Verlag.(1976)
15. 新美康永: "音聲認識".

▲정 석 재(정회원) 1968년 2월 3일생



1989년 2월 : 영남대학교 전자공학과 졸업(공학사)
1991년 2월 : 영남대학교 대학원 전자공학과 졸업(공학석사)
1991년 ~ 현재 : 금성정밀(주) 제1연구소

▲정 현 열(정회원) 1951년 11월 26일생



1975년 2월 : 영남대학교 전자공학과 졸업(공학사)
1981년 2월 : 영남대학교 대학원 전자공학과 졸업(공학석사)
1985년 5월 ~ 1986년 3월 : 일본 東北대학 응용정보학 연구센터 연구생
1989년 4월 : 일본 東北대학 대학원 정보공학과(공학박사)
1989년 3월 ~ 현재 : 영남대학교 전자공학과 조교수
※관심분야 : 음성신호처리 및 그 응용

▲이 무 영(정회원) 1932년 11월 29일생



1959년 : 한양대학교 전기공학과 졸업
1972년 : Canada BC 주립대학원
1975년 2월 : 영남대 대학원(공학박사)
1966년 2월 ~ 현재 : 영남대 전자공학과 교수