Journal of the Industrial Technology Research Institute.
Korea Maritime University. Vol. 17. 1999.

# 화자 종속 모음 인식에 대한 연구

하 동 경* · 신 옥 근**

# A Study of Speaker-Dependent Vowel Recognition

Dong-Kyung Ha, Ok-Keun Shin

## 요 약

화자 종속 음성 인식이나 화자 검증은 불특정 화자들의 음성에서 동일한 특징을 찾는 화자 독립 음성 인식과는 다르게 음성에서 각 화자를 구분할 수 있는 특징을 찾아야 한다. 화자에 따라 각기 다른 특성을 가진 발성 기관을 지니고 있으므로 이 발성기관을 통해 생성되는 모음에는 화자를 구분할 수 있는 특성들이 포함되어 있다. 그러나 동일 화자가 같은 음소를 발화하더라도 전후 문맥에 따라 음향학적인 특성이 다르게 발화될 수 있다. 본 논문에서는 단일 화자가 발성한 모음에 대한 특징 벡터들 중 포만트(formant)와 MFCC (Mel-Frequency Cepstral Coefficient)의 변화량을 조사해 각 특징 벡터들의 범위를 정하였다. 모음의 특징은 화자의 상태나 문장 내의 모음의 위치에 따라 변하는데 본 논문에서는 화자의 심리적 상태는 고려하지 않았다. 본 논문에서 제안한 포만트와 MFCC의 특징들의 범위를 각 모음의 대표값으로 하여 화자 종속 음성 인식에 적용한 결과, 포만트와 MFCC를 같이 적용하고 포만트는 F1에서 F3까지를 적용한 경우의 대표값을 적용한 경우가 가장 좋은 결과를 보였다. 본 논문에서 제안한 모음의 대표값은 화자 종속 음성 인식, 화자 검증, 컴퓨터를 이용한 발음 학습 시스템등에 적용할 수 있을 것이다.

## 1. 서 론

음성 인식은 인식할 대상으로 삼는 화자에 따라 화자 종속 음성 인식과 화자 독립 음성 인식으로 분류할 수 있다. 화자 독립 음성 인식은 전화번호 안내나 기차표 예약 시스템과 같이 불특정 화자의 음성을 인식의 대상으로 삼기 때문에 각기 다른 화자에 의해 발성된 특정 단어나 문장에 대하여 동일한 특징을 추출해 내는 것이 중요하며 현재 사용되는 음성 인식에 관한 시스템들은 화자 독립적인 시스템이 많다. 화자종속 시스템은 특정 화자의 음성을 인식하기 위한 시스템으로, 현재 휴대폰에 탑재되어 사용되는 음성 다이얼링(voice dialing) 시스템이 대표적인 예이다. 화자종속 시스템에서는 일반적으로 시스템의 사용전에, 사용자의 음성을 저장, 등록시키고, 실제 인식을

* 한국해양대학교 대학원 컴퓨터공학과
** 한국해양대학교 컴퓨터공학과 조교수

수행할 때는 입력된 음성의 패턴(pattern)과 저장된 음성의 패턴을 비교하는 패턴 정합(pattern matching) 기법이 사용된다. 화자 종속적 음성 인식은 음성의 일반적인 특징뿐 아니라 화자의 고유한 특징도 같이 추출해 내어야 한다. 현재 인식률은 화자종속 음성인식이 95~98%, 화자독립 음성인식이 90~97% 정도이다[10].

본 논문에서는 단일 화자가 발성한 음성 중에서 화자의 음성적 특성을 구별할 수 있는 모음을 분석하였다. 단일 화자의 모음 특성을 분석하기 위하여 먼저 음성 데이터를 음소 단위로 분할하였고 음성 데이터 전체의 특징 벡터들을 추출하였다. 분할된 음소들 중에서 모음 부분을 찾아내었다. 찾아내 모음 부분에 대하여 다시 안정된 모음 구간과 불안정한 모음 구간으로 구분을 하였는데 모음 부분이라 할지라도 양쪽 가장자리에는 인접 음소에 영향을 받는 불안정 구간이 포함되어 있기 때문에 이 구간을 제외한 안정된 구간만의 특징 벡터들을 적용한 것과 모음 부분 전체에 대한 특징 벡터를 적용시킨 것을 비교하기 위해서이다[12].

본 논문에서는 2장에서 음성신호에서 특징 벡터인 포만트와 MFCC에 관하여 설명하고 3장에서는 모음구간을 찾는 방법과 평균 특징 벡터 추출에 대하여 설명하고 4장에서 실험 및 결과를 보인다.

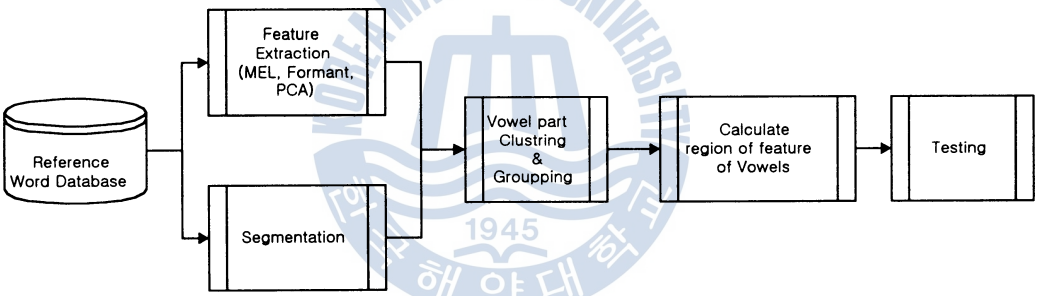


그림 1. 모음의 특징 벡터들의 범위 설정을 위한 처리 흐름도

## 2. 음성 신호에 대한 특징 추출

그림 1은 음성 데이터에서 모음을 찾아 특징 벡터를 추출하는 과정을 보여준다. 먼저 수작업으로 음성 데이터를 가지고 음성 신호의 파형과 음성 신호에 대한 주파수 스펙트럼을 이용하여 음소 단위로 분할한다. 음소 단위로 분할하기 위해 TIMIT에서 사용한 61개의 음소를 사용하였다[14]. 계산량을 줄이기 위해 일반적으로 음성 데이터를 대표할 수 있는 특징만을 추출하여 사용한다. 음성에 대한 특징을 추출하기 위해서는 다음과 같은 전처리 과정을 수행한다.

### 2.1 전처리

음성을 생성하는 조음기관의 특성상 하나의 음소는 최소한 10~30msec 시간 동안은 발성되기 때문에 단어나 문장에 대한 전체 음성 신호를 10~30msec 보다 적은 간격으로 나누어 처리를 하면 전체 음성을 한번에 처리하는 것에 비하여 계산량을 감소시켜서 음성 특성을 얻을 수 있다[1].

본 연구에서는 실험용 음성 데이터로 영어를 모국어로 하는 원어민 여자 1명이 읽은 단어들의 음성 신호를 22050Hz로 샘플링하여 8bit로 양자화하여 저장하였다. 이 음성 데이터를 한 프레임당 5msec 크기로 블록킹을 하여 프리엠파시스(preemphasis)와 해밍 윈도우(hamming window)를 이용해 전처리를 하였다[1]. 프리엠파시스는 음성 신호의 특정 부분을 강조하기 위한 것으로 식 (1)을 사용하였다. 음성을 프레임 단위로 나누어 주파수 변환을 하면 프레임의 가장자리의 신호는 비선형적이 되어 매우 높은 주파수를 띄게 되는데 이를 보상하기 위해 일반적으로 해밍 윈도우를 사용한다. 식 (2)에 해밍 윈도우 함수를 나타내었다.

$$\tilde{s}(n) = s(n) - \tilde{a}s(n-1) \quad \cdots\cdots (1)$$

여기서 $s(n)$은 음성 신호이고, $\tilde{a}$ = 0.95를 사용하였다.

$$w(n) = 0.54 - 0.46\cos(\frac{2\pi n}{N-1}), \quad 0 \le n \le N-1 \quad \cdots\cdots (2)$$

## 2.2 음성 신호의 특징 벡터 추출

음성 인식을 위해 흔히 이용되는 특징 벡터에는 Pitch[3], LPC 계수[4], MFCC, PLP 계수[5], 포만트(formant) 등이 있다. 본 논문에서는 사람의 청각 기능을 모델링한 MFCC[6]와 모음의 특징을 잘 나타낼 수 있는 포만트[7]를 사용하였다.

### 2.2.1 MFCC 추출

MFCC(Mel-Frequency Cepstral Coefficient)는 잡음환경에 강인한 것으로 알려져 있으며 필터뱅크를 통한 스펙트럼 분석(filterbank spectral analysis)을 이용하여 얻어진다[6][8].

인간의 인지특성을 이용한 MFCC는 전체 주파수 대역을 임계 대역(critical band) 단위로 나눈 다음 비균일 필터뱅크를 적용하는데 각 필터들의 간격은 1000Hz 미만의 주파수에서는 선형적이고 그 이상의 주파수에서는 대역폭이 주파수 f의 로그 단위로 이루어져 있다. MFCC를 추출하기 위해 먼저 FFT(Fast Fourier Transform)[9]를 통해 음성신호를 주파수 영역으로 변환한 다음 MEL 단위의 비균일 필터뱅크를 이용한다. 그림 2에 MFCC를 추출하는 과정을 보인다.

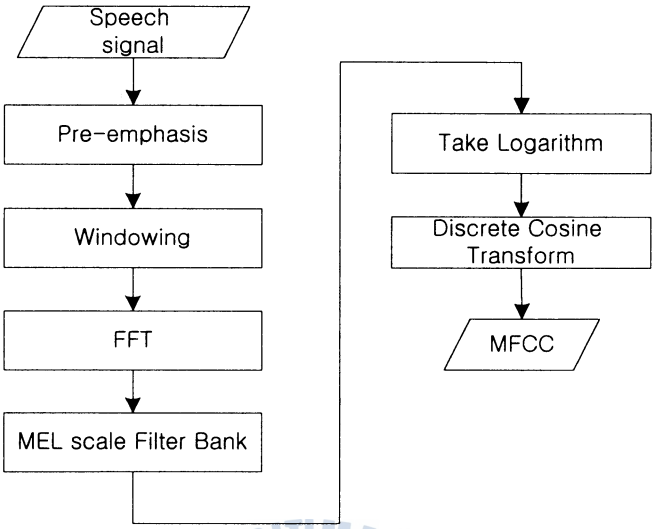


그림 2. MEL 단위 비균일 필터뱅크를 이용한 MFCC 추출

$\widetilde{S_k}$를 신호 $s(n)$의 파워 스펙트럼이라하고 L을 MFCC의 차원, K를 필터뱅크의 길이라 할 때, MFCC $\widetilde{c_n}$은 다음 식 (3) 과 같이 표현할 수 있다.

$$\widetilde{c_n} = \sum_{k=1}^{K} (\log \widetilde{S_k}) \cos\left[ n(k-\frac{1}{2})\frac{\pi}{K} \right], \quad n = 1,2,\ldots,L \qquad \cdots\cdots (3)$$

### 2.2.2 포만트 추출

폐에서 시작된 공기의 흐름이 성대(vocal cord)를 거치면서 만들어진 음원(speech source)은 성도(vocal tract)와 여러 가지 조음 기관들의 조음(articulation) 동작에 의해 결정되는 음향 전달 특성의 영향을 받아 음성으로 발화된다. 이 과정을 어떤 전달함수를 갖는 선형 시스템에 여기 신호를 입력한 것으로 모델링할 수 있으며[10], 이 때 전달함수는 포만트라 불리는 주파수에서 공명을 일으킬 수 있다. 따라서 주기성을 갖는 음원을 가진 모음이나 반모음 및 비음들은 각각 고유한 공명 주파수인 포만트들을 갖게 된다. 본 논문에서는 각 모음에 대한 포만트를 찾기위해 Welling과 Ney 등이 제안한 2차 디지털 병령 공명기를 모델링하는 방법[7][11]을 이용하였다. 이 방법에서는 포만트 추적[2] 알고리즘 등에서 필요로 하는 peak-picking 과정이 필요치 않으므로 효율적으로 포만트를 찾을 수 있다.

그림 3은 포만트를 추출하는 과정과 이를 이용해 "fast"의 포만트 4개(F1~F4)를 프레임별로 구한 결과이다.

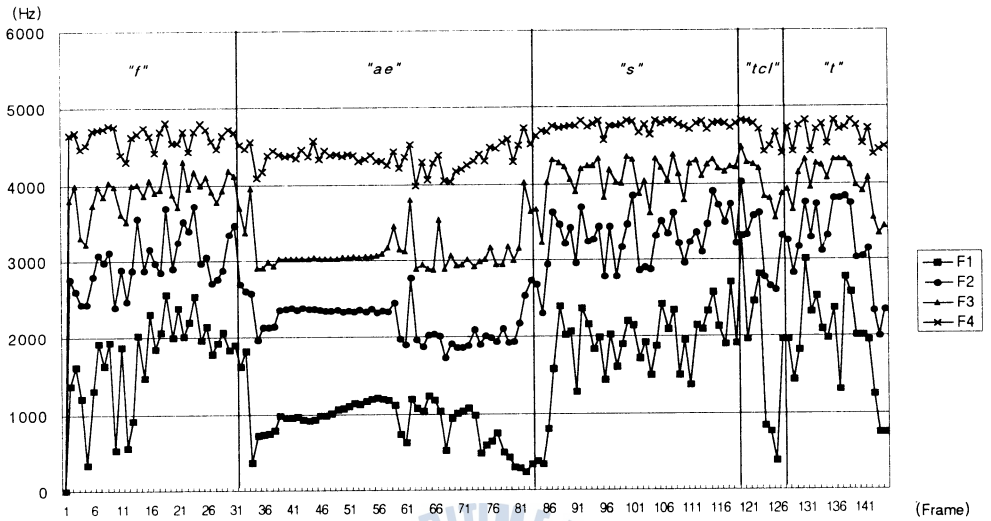


그림 3. "fast"의 포만트

# 3. 모음 구간의 추출

## 3.1 모음 전체 구간과 모음의 안정된 구간 설정

음소 단위의 세그먼테이션에서 각 음소들은 인접 음소에 영향을 받은 불안정한 구간과 영향을 받지 않는 안정된 구간으로 이루어져 있다. 본 논문에서는 각 모음에 대하여 불안정한 구간을 제외한 부분을 적용한 인식률과 불안정한 구간을 포함한 모음 구간 전체를 적용한 인식률을 비교한다.

모음의 특성이 불안정한 구간을 제외시키기 위하여 분할된 음성 데이터에서 모음 구간을 찾아낸 다음 LBDP 알고리즘[12]을 사용하여 모음 앞쪽 음소의 영향을 받아 모음으로서의 특징이 불안정한 부분과 모음 뒷쪽 음소에 의해 모음적 특징이 불안정안 부분의 경계를 구하여 모음 구간을 세 부분으로 다시 분할한다. 그런데, 본 논문에서 사용한 음소들 중에 'ey'와 같은 이중 모음은 두 가지의 음성적 특징을 포함하고 있으므로 모음의 특징이 불안정한 구간을 찾을 때 오인할 확률이 높다. 이를 보완하기 위해 모음 구간을 세 부분으로 나눌 때 불안정한 구간의 범위를 미리 제한하는 방법을 제안하여 사용하였는데 그림 4와 같다.

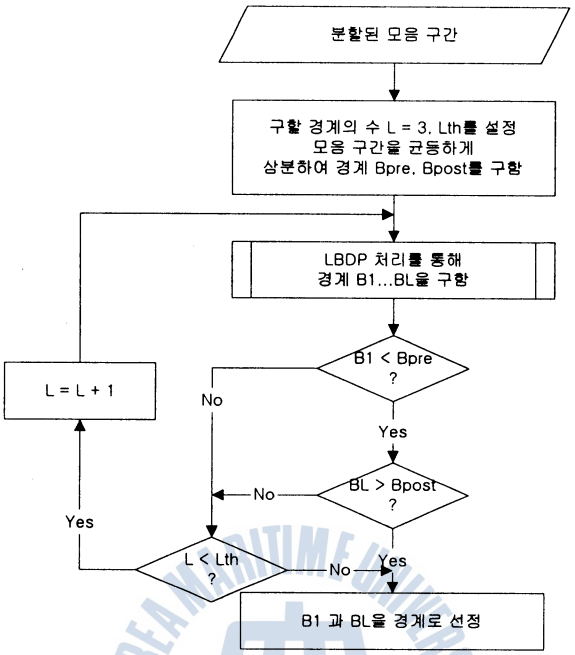


그림 4. 모음 구간 분할 알고리즘

그림 5는 이 알고리즘을 이용해 구한 'fast'에 있는 모음 'ae'에 대한 예이다.

분할된 모음의 세 부분에 전체에 대한 프레임들의 포만트와 MFCC를 추출한다. 모음 분할 과정을 통해 추출한 포만트와 MFCC에 대한 정보들을 모음별로 그룹화한다.

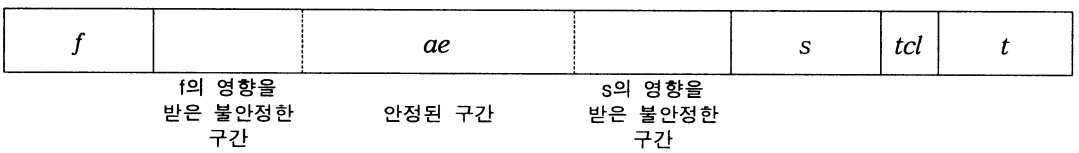


그림 5. fast의 모음 ae에 모음 구간 분할 알고리즘을 적용한 결과

## 3.2 주요 성분 분석법(PCA; principal components analysis)

모음 구간에 해당하는 프레임들의 포만트와 MFCC를 모두 적용하지 않고 특징 벡터의 차원을 줄여 근사화한 값을 이용하면 계산량을 줄일 수 있는데 본 논문에서는 다음과 같이 주요 성분 분석법을 사용하여 모음 구간에 해당하는 특징 벡터를 근사화시킨 고유 벡터(eigen vector)를 구하였다[13][15].

먼저, N개의 프레임으로 구성된 음성 특징 벡터들의 시퀀스 $\boldsymbol{X}$를 정의한다.

$$\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N\}$$

각 프레임의 특징벡터 $x_i$ ($i=1, 2, \cdots, N$)를 P차원의 랜덤벡터 $x_i = [x_{i1}, x_{i2}, \cdots, x_{ip}]$로 간주하면 각 특징벡터 $x_i$는 고유성분분석법을 이용하여 다음과 같이 표현할 수 있다.

$$x_i = \Phi h_i + x_M \quad \cdots\cdots (4)$$

여기서, $\Phi$는 X의 공분산 행렬인 Cx의 고유벡터들(eigenvectors) $\varphi_j$를 열 벡터로 하는 행렬이며, 이 열벡터들은 서로 정규직교(orthonormal)한다.

$$\Phi = \begin{bmatrix} | & | & & | \\ \varphi_1 & \varphi_2 & \cdots & \varphi_p \\ | & | & & | \end{bmatrix} \quad \cdots\cdots (5)$$

$$\varphi_j^{*T} \varphi_k = \begin{cases} 1, & j=k \\ 0, & j=/k, \end{cases} \text{ for } 1 \le j,\ k \le P \quad \cdots\cdots (6)$$

$x_M$은 $X$의 무게중심 (평균)이며,

$$x_M = E\{x_i\}, \quad i=1, \cdots, N$$

$h_i$는 $x_i$의 확장계수 벡터로, 다음의 식 (8)로부터 구해질 수 있다.

$$\mathbf{h}_i = \begin{bmatrix} h_{i1} \\ h_{i2} \\ . \\ . \\ h_{iP} \end{bmatrix} \quad \cdots\cdots (7)$$

$$h_i = \Phi^{*T}(x_i - x_M) \quad \cdots\cdots (8)$$

일반적으로 고유성분 분석법은 주어진 데이터 집합 X의 차원을 줄여 근사화 하기 위해 이용될 수 있다. 식 (5)에 보인 행렬의 열벡터 $\varphi_j$ ($j=1, 2, \cdots, P$)는 공분산 Cx의 고유벡터이므로 $\varphi_j$에 대응하는 고유값(eigenvalue) $\lambda_j$가 클수록 $\varphi_j$ 방향의 분산이 크다는 것을 의미한다. 따라서 가장큰 고유값에 상응하는 K개 (주요고유성분)를 취한 다음, 특징벡터 xi를 고유벡터들에 투영한 K차원의 벡터 $\hat{x}_i$로 표현하면 (원래의 좌표계에서 $\hat{x}_i$는 ( $\hat{x}_1$, $\hat{x}_2$, $\cdots$, $\hat{x}_P$)로 표현된다) P차원의 $\mathbf{x}_i$를 가장 작은 오차를 갖는 K차원 (P > K)의 벡터로 근사화할 수 있다. 행렬 $\Phi$가 $\lambda_1 > \lambda_2 > .. > \lambda_P$ 의 순서대로 고유벡터 $\varphi_j$를 정렬한 행렬이라 할 때 $x_i$를 K차원의 특징벡터로 근사화한 벡터 $\hat{x}_i$는 다음 식 (9)와 같이 나타낼 수 있다.

$$\hat{x}_i = \sum_{j=1}^{K} \varphi_j h_{ij} + x_M \quad \cdots\cdots (9)$$

## 3.3 모음의 특징 벡터들의 범위 설정

앞에서 구성한 모음 그룹에 포함되어 있는 모음 요소들에 대하여 먼저, 모음 요소별로 포만트와 MFCC를 이용해 평균 벡터와 고유 벡터을 구하였다. 다음으로, 모음 요소 각각에 대해 구한 특징 벡

터들을 특징 벡터의 각 요소들에 대하여 비교해 가장 큰 값과 작은 값 그리고 평균값을 구하였다.

각 모음 그룹별로 구하여진 특징 벡터의 각 요소들의 최소값과 최대값을 그 모음의 특징 벡터의 범위로 설정하였다.

# 4. 실험 및 결과

## 4.1 실험 환경 및 방법

단일 화자에 대한 모음 발성을 분석하기 위하여 영어를 모국어로 하는 여자 1명이 발음한 단어 120개를 실험용 데이터로 이용하였다. 녹음된 데이터의 샘플링 주파수는 22050Hz이고 8bit의 모노로 녹음된 웨이브 신호이다. IBM PC Pentium Ⅱ 266㎒를 이용하여 수작업으로 음성의 웨이브 신호와 주파수 분석 도구를 이용해 음소단위로 분할을 하였다. 그리고 음성 데이터 전체에 대하여 5msec 길이의 프레임으로 나눈 다음 각 프레임에서 MFCC와 포만트를 추출하였다. 분할된 음소 정보를 이용해 모음을 분할하였으며, 분할한 모음의 정보를 저장할 파일에 모음의 인접 음소들을 알 수 있도록 표시하였다. 모음을 LBDP를 통해 다시 세부분으로 나누었으며, 모음 구간에 대하여 추출한 특징 벡터들을 가지고 PCA[18]를 통해 고유벡터를 구하고 동시에 평균과 최대와 최소값을 구하였다. 실험의 신뢰성을 위해 실험용 데이터를 이용해 구한 모음들 중에서 모음의 개수가 10개 이상인 모음들만을 선정하여 실험에 사용하였는데 선정한 모음은 다음과 같다.

**ih iy eh ey ae aa aw ay ah er**

모음 분석을 통한 모음의 안정성을 검증하기 위해 선정한 모음 10 가지에 대하여 비교할 특징 벡터로 모음 구간 전체를 이용해 구한 고유 벡터와 평균 벡터, 그리고 모음의 안정된 구간에서 구한 고유 벡터와 평균 벡터를 사용하였으며 고유 벡터와 평균 벡터를 같이 적용을 해 보았다. 실험에 사용한 모음들에 대하여 앞에서 선정한 여섯 가지의 비교 특징 벡터들을 적용하여 각 모음의 특징 벡터들에 대한 허용 범위에 들어갈 경우 정인식으로 결정하게 하였다.

## 4.2 실험 결과 및 결과 분석

10 가지의 모음에 대하여 구한 MFCC와 포만트 벡터들의 범위 중에 포만트의 경우를 표 1과 그림 6에 나타내었다. 모음 'ah' 경우에 F2와 F3이 겹쳐지는 부분이 나타나는데 이것은 실험용 데이터 중의 'channel'에서 다른 단어와는 달리 'ah'의 앞에 파열 마찰음인 'ch'가 있기 때문이고 'er' 경우는 실험용 데이터의 'alert'에서 다른 단어와 달리 'er' 뒤에 바로 입을 닫는 부분에 해당하는 'tcl'이 오기 때문이다.

표 1. 각 모음별 포만트 범위

| | ih | iy | eh | ey | ae | aa | aw | ay | ah | er |
|---|---|---|---|---|---|---|---|---|---|---|
| F1_Min | 308 | 260 | 412 | 405 | 473 | 609 | 838 | 509 | 293 | 313 |
| F1_Max | 542 | 868 | 1095 | 733 | 1123 | 965 | 1189 | 936 | 1301 | 716 |
| F2_Min | 1948 | 2161 | 1991 | 2355 | 2073 | 1243 | 1843 | 1659 | 1669 | 1385 |
| F2_Max | 2799 | 3044 | 2786 | 2790 | 2770 | 2277 | 2211 | 2394 | 2982 | 1920 |
| F3_Min | 3063 | 3258 | 2981 | 3175 | 2992 | 2829 | 2964 | 2986 | 2830 | 1950 |
| F3_Max | 3926 | 4100 | 3726 | 3884 | 3542 | 3543 | 3353 | 3507 | 3823 | 3680 |
| F4_Min | 4229 | 4343 | 4181 | 4334 | 4088 | 3912 | 4075 | 4055 | 4024 | 3947 |
| F4_Max | 4632 | 4703 | 4556 | 4652 | 4452 | 4386 | 4457 | 4474 | 4546 | 4586 |

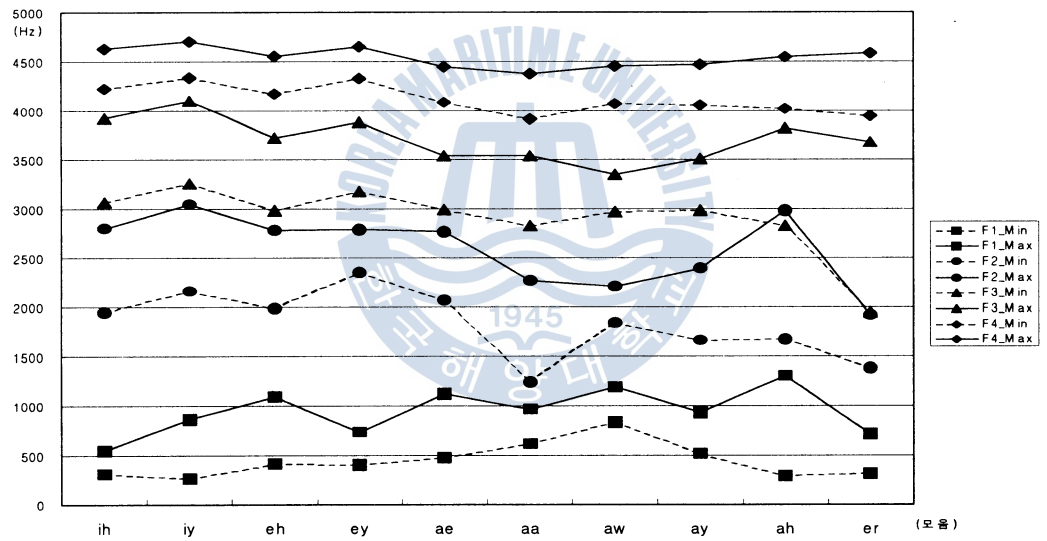


그림 6. 각 모음별 포만트 범위

이와 같이 같은 모음이라도 모음의 전후에 있는 음소의 영향 때문에 신뢰성있는 모음의 특징을 찾기는 어렵다. 본 논문에서는 화자 종속 시스템을 기본으로 한 인식률을 향상 시킬 수 있도록 각 모음에 대한 포만트와 MFCC의 범위를 구하여 적용하였다.

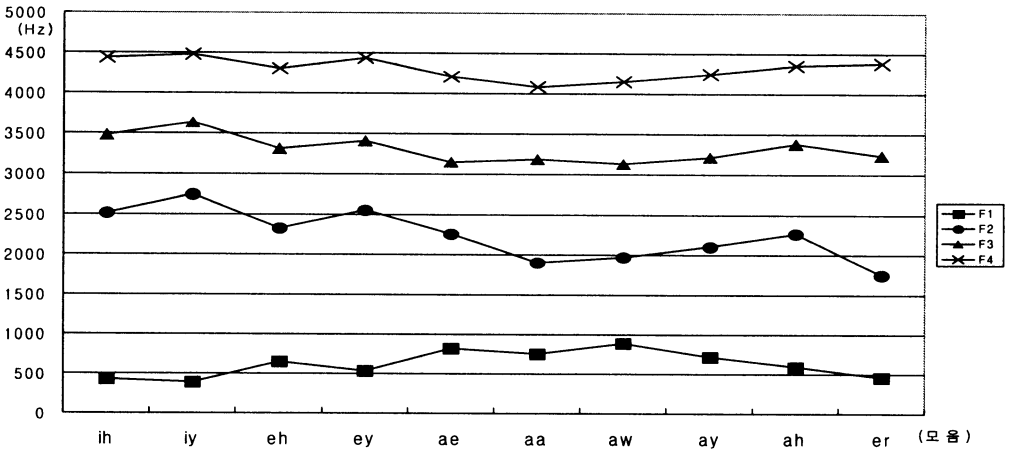


그림 7. 각 모음별 포만트의 평균

표 2. 포만트와 MFCC를 각각 적용한 실험 결과

| 적용 데이터 / 모음/개수(개) | | 모음구간 전체의 포만트의 평균 벡터 | | 모음구간 전체의 포만트의 고유 벡터 | 모음구간 전체의 MFCC의 평균 벡터 | 모음구간 전체의 MFCC의 고유 벡터 |
|---|---|---|---|---|---|---|
| | | F1~F4 적용 | F1~F3 적용 | | | |
| ih | 17 | 16 | 14 | 15 | 11 | 15 |
| iy | 23 | 22 | 22 | 19 | 21 | 21 |
| eh | 18 | 16 | 18 | 12 | 14 | 11 |
| ey | 14 | 10 | 11 | 10 | 8 | 6 |
| ae | 19 | 15 | 16 | 17 | 15 | 10 |
| aa | 10 | 8 | 8 | 6 | 7 | 2 |
| aw | 9 | 6 | 6 | 7 | 7 | 2 |
| ay | 15 | 13 | 12 | 12 | 10 | 8 |
| ah | 21 | 18 | 17 | 19 | 17 | 16 |
| er | 14 | 13 | 14 | 12 | 10 | 5 |
| 계 | 160 | 125 | 138 | 117 | 110 | 91 |
| 정인식률(%) | | 85.63 | 86.25 | 80.63 | 75.0 | 60.0 |

표 3. 포만트와 MFCC를 같이 적용한 실험 결과

| 적용 데이터 / 모음/개수(개) | | 모음구간 전체의 평균 벡터 | 모음구간 전체의 고유 벡터 |
|---|---|---|---|
| ih | 17 | 17 | 17 |
| iy | 23 | 23 | 23 |
| eh | 18 | 18 | 16 |
| ey | 14 | 10 | 11 |
| ae | 19 | 18 | 19 |
| aa | 10 | 9 | 7 |
| aw | 9 | 8 | 7 |
| ay | 15 | 14 | 12 |
| ah | 21 | 21 | 20 |
| er | 14 | 14 | 13 |
| 계 | 160 | 152 | 145 |
| 정인식률(%) | | 95.0 | 90.63 |

그림 7은 각 모음별 포만트의 평균을 나타낸 것으로 모음에 따라 F2가 가장 많은 차이를 보이고 있다. F4의 경우에는 모음에 따른 구별이 용이하지 않음을 알 수 있다. 표 4에 모음구간 전체에 대한 포만트를 F1~F4 까지 적용한 것과 F1~F3 까지 적용한 것을 비교해 놓았는데 포만트를 네 개 적용한 것보다 세 개를 적용한 경우가 더 좋은 결과를 보였다. 이것은 화자 종속 음성 인식에 있어서는 비교할 특징 벡터를 적당하게 사용하는 것이 계산량도 줄이고 인식률도 높일 수 있음

을 나타낸다.

본 논문에서 적용한 모음의 특징 벡터의 범위를 검증하기 위하여 실험용 데이터를 선정하여 적용한 결과는 모음의 안정된 구간을 적용한 것보다 모음 구간 전체에 대한 특징 벡터를 적용한 경우가 더 좋은 결과를 보였다. 모음 부분 전체에 대한 포만트와 MFCC를 각각 적용한 결과는 표 2와 같고 포만트와 MFCC를 같이 적용한 결과는 표 3과 같다.

포만트를 적용한 경우에 모음 전체 구간의 평균 벡터를 적용하여 인식을 한 경우가 가장 좋은 결과를 나타내고 있으며 MFCC를 이용한 경우에도 모음 전체 구간의 평균 벡터를 적용한 경우가 고유 벡터를 적용한 것보다 좋은 인식률을 보였다. 모음의 안정된 구간에 대해서는 평균 벡터를 적용한 경우에는 포만트와 MFCC를 이용한 두 가지 모두 다 모음 전체 구간을 이용한 것보다 인식률이 떨어졌다. 모음의 안정된 구간에 대해 고유 벡터를 적용한 경우에는 모음 전체 구간을 적용한 경우와 비슷한 인식률을 보였다.

평균 벡터를 안정된 구간에 적용했을 때 인식률이 떨어지는 원인은 LBDP를 이용하여 안정된 구간을 찾을 때 하나의 음소로 분류된 모음이라도 두 가지의 특성을 가지는(예를 들면, 'ey'는 'e'와 'y'의 두 가지의 특성을 지님) 이유로 불안정 구간이 경계로 선택되지 않고 이중모음의 한 부분에 대한 경계가 선택되는 경우가 발생하기 때문으로 보인다. 이 문제점에 대한 해결책으로 불안정 구간의 경계로 설정할 수 있는 영역에 제한을 두는 방법을 적용하였으나 이중 모음에 대한 문제를 근본적으로 해결할 수는 없었다.

## 5. 결 론

본 논문에서는 단일 화자가 발성한 모음에 대한 특징 벡터들 중 포만트와 MFCC의 변화량을 화자 종속 음성인식의 특징 벡터로 적용하는 방법을 제안하였다. 본 논문에서 제안한 특징 벡터를 화자 종속 모음 인식에 적용한 결과 95%의 인식률을 나타내었다. 각 특징 벡터들의 요소들을 분석하여 모음의 구분에 영향이 큰 벡터 요소에 대하여 가중치를 부여한다면 인식률을 향상시킬 수 있을 것이다. 단일 화자에 의해 발성된 모음의 특징을 분석한 본 연구에서는 모음에 대한 특징 벡터들의 범위를 결정하여 적용할 벡터의 크기를 적절히 선택하여 사용하면 보다 적은 계산량으로 인식의 신뢰성을 높일 수 있다는 것을 알 수 있었으며 이 결과는 화자 종속 음석 인식을 이용하는 분야나 컴퓨터를 이용한 외국어 발음 학습 시스템 등에 적용할 수 있을 것이다. 각 모음에 대하여 많은 데이터를 적용한다면 보다 정확한 모음의 특징 벡터들의 범위를 결정할 수 있을 것이다. 화자에 대한 모음의 특징 벡터들의 범위 정보를 서로 비교하여 차이점을 보상해 준다면 화자 독립적인 시스템에도 적용이 가능할 것이다.

## 참고 문헌

[1] Lawrence Rabiner and Biing Hwang Juang, Fundamentals of Speech Recognition,

Prentice Hall, 1993.

[2] W. H. Press et al, Numerical Recipes in C, 2nd ed. Cambridge.

[3] Leonard Janer, Juan Jose Bonet, Eduardo Lleida-Solano, "Pitch detection and voiced/uncoiced decision algorithm based on wavelet transforms", ICLSP, 3-6, vol. 2., October, 1996.

[4] John Makhoul, "Linear prediction: a tutorial review", Proc. IEEE, vol. 63, No. 4, April 1975

[5] Hynek Hermansky, "Perceptual linear predictive analysis of speech", JASA, 87(4), pp. 1783~1752, April 1990.

[6] Steven B. Davis, Paul Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences", Proc. IEEE, vol. 81 No. 9, pp 1215~1247, 1993

[7] L. Welling, H. Ney, "A model for efficient formant estimation", ICASS 96, pp.797~800, May, 1996.

[8] http://speech.sogang.ac.kr/splib/lib-ref.html, RTSA library 0.9.3 Reference.

[9] 이채욱, 디지털 신호처리, 청문각, pp.100~114, 1995년 4월.

[10] 오영환, 음성언어 정보처리, 홍릉과학 출판사, pp. 214, 1998년 1월.

[11] L. Welling, H. Ney, "Formant estimation for speech recognition", Proc. IEEE, vol. 6, No. 1, pp. 36~48, Jan. 1998.

[12] Manish Sharma, Richard Mammone, "Blind speech segmentation: automatic segmentation of speech without linguistic knowledge", Proc. ICLSP, vol. 2, October 3~6, 1996.

[13] C. W. Therrien, Discrete Random Signals and Statistical Signal Processing, Prentice Hall, 1992

[14] John S. Garofolo, Lori F. Lamel, etc., DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus, Feb. 1993.