

성도 면적 함수를 이용한 음성 인식에 관한 연구

송제혁 · 김동준* · 박상희

=Abstract=

A Study on Speech Recognition using Vocal Tract Area Function

Jei-Hyuck Song, Dong-Jun Kim*, Sang-Hui Park

The LPC cepstrum coefficients, which are an acoustic features of speech signal, have been widely used as the feature parameter for various speech recognition systems and showed good performance.

The vocal tract area function is a kind of articulatory feature, which is related with the physiological mechanism of speech production. This paper proposes the vocal tract area function as an alternative feature parameter for speech recognition. The linear predictive analysis using Burg algorithm and the vector quantization are performed. Then, recognition experiments for 5 Korean vowels and 10 digits are executed using the conventional LPC cepstrum coefficients and the vocal tract area function. The recognitions using the area function showed the slightly better results than those using the conventional LPC cepstrum coefficients.

Key words : Speech recognition, vocal tract area function, vector quantization, linear predictive

서 론

인간의 언어에 대한 기계의 인식을 목적으로 하는 음성 인식에 대한 연구는 지금까지 국내외에서 많은 학자들에 의하여 연구되어 왔으며, 최근에는 멀티미디어에 대한 관심이 고조되어 중요성이 더욱 강조되고 있다. 음성 인식 시스템을 개발하는데 있어서 DTW, HMM, 신경망, 벡터 양자화 등의 인식 알고리즘의 개발도 중요하나, 음성의 특징을 잘 반영할 수 있는 매개변수의 추출 역시 매우 중요한 문제이다.

많은 음성 특징 추출 방법이 제안되어 왔으나, 그들 중 대부분은 음성 파형으로부터 직접 얻어지는 음향학적 매개 변수에 의존하며, 그 중 LPC cepstrum 계수가 음성 인

식에 가장 널리 이용되어 왔다¹⁾. 음성의 효과적인 특징 추출을 위하여 음성 발생 메커니즘에 대한 연구는 기본적인데 매우 중요하다. 음성 생성 과정에 대한 해석을 통하여 음성으로부터 성도 면적(vocal tract area)을 추출함으로써 성도에서의 조음적(articulatory) 특징을 구하려는 연구는 Schroeder²⁾와 Mermelstein³⁾ 등에 의하여 처음 시도된 후, Paige와 Zue⁴⁾, Gopinath와 Sondhi⁵⁾ 등의 많은 학자들에 의하여 수행되어 왔다. Atal 등⁶⁾은 성도를 무손실의 음향 튜브로 가정하고 선형 예측 계수로부터 직접 성도의 단면적을 추정하는 방법을 제시하였으며, Wakita⁷⁾는 Atal과는 대치되는 경계 조건을 제시하여 성도 면적을 추정하였는데, 이는 실제 성도의 모양과 상당히 유사한 결과를 얻었다. 이와 같이 추정된 성도 면적 함수는 음성 생성에 직접

연세대학교 전기공학과

Dept. of Electrical Engineering, Yonsei University

* 청주대학교 정보통신공학과

* Dept. of Computer Communication and Engineering, Chongju University

통신저자: 송제혁, (120-752) 서울시 서대문구 신촌동 134, Tel. (02) 361-2768, Fax. (02) 312-7735

적으로 연계되어 있는 음성의 조음적 특징으로서 음성 인식에 적용할 수 있는 효과적인 매개변수라고 볼 수 있다.

본 연구에서는 음성 인식을 위한 특징 벡터로서 성도 면적 함수를 이용하고, 벡터 양자화(vector quantization, VQ)를 적용하여 한국어 5개 단모음 및 10개 숫자음에 대한 인식 실험을 수행하여, 기존의 널리 이용되는 특징 벡터인 LPC 켈스트럼 계수에 의한 인식 결과와 비교함으로써 음성 인식을 위한 특징 벡터로서 성도 면적 함수의 이용 가능성을 검증하여 보고자 한다.

성도 면적 함수의 추정

1. Burg 알고리즘

1960년대초 개발된 Burg의 방법은 Levinson recursion 과 격자(lattice) 구조에 기초하고 있고, 반복의 각 단계에서 반사계수(reflection coefficient)가 얻어지는 반복적인 방법이다. P차의 반사계수를 구하기 위해서는 식 (1)과 같이 전향과 후향 예측오차의 합이 최소화되도록 선택된다.

$$\alpha = \sum_{n=1}^{N-1} (|\epsilon_p^+[n]|^2 + |\epsilon_p^-[n]|^2) \quad (1)$$

식 (1)에 쓰이는 오차항들은 격자 방정식에 의하면 식 (2)와 (3)을 만족한다.

$$\begin{bmatrix} \epsilon_p[p] \\ \epsilon_p[p+1] \\ \vdots \\ \epsilon_p[N-1] \end{bmatrix} = \begin{bmatrix} \epsilon_{p-1}[p] \\ \epsilon_{p-1}[p+1] \\ \vdots \\ \epsilon_{p-1}[N-1] \end{bmatrix} - \gamma_p^* \begin{bmatrix} \epsilon_{p-1}^-[p-1] \\ \epsilon_{p-1}^-[p] \\ \vdots \\ \epsilon_{p-1}^-[N-2] \end{bmatrix} \quad (2)$$

$$\begin{bmatrix} \bar{\epsilon}_p[p] \\ \bar{\epsilon}_p[p+1] \\ \vdots \\ \bar{\epsilon}_p[N-1] \end{bmatrix} = \begin{bmatrix} \epsilon_{p-1}^-[p-1] \\ \epsilon_{p-1}^-[p] \\ \vdots \\ \epsilon_{p-1}^-[N-2] \end{bmatrix} - \gamma_p^* \begin{bmatrix} \epsilon_{p-1}[p] \\ \epsilon_{p-1}[p+1] \\ \vdots \\ \epsilon_{p-1}[N-2] \end{bmatrix} \quad (3)$$

위의 격자 방정식은 다음의 벡터를 정의함으로써 간단하게 표현될 수 있다.

$$\vec{\epsilon}_p = \begin{bmatrix} \epsilon_p[p] \\ \vdots \\ \epsilon_p[N-1] \end{bmatrix}; \quad \vec{\bar{\epsilon}}_p = \begin{bmatrix} \bar{\epsilon}_p[p] \\ \vdots \\ \bar{\epsilon}_p[N-1] \end{bmatrix} \quad (4)$$

$$\vec{e}_p = \begin{bmatrix} \epsilon_p[p+1] \\ \epsilon_p[p+2] \\ \vdots \\ \epsilon_p[N-1] \end{bmatrix}, \quad \vec{\bar{e}}_p = \begin{bmatrix} \bar{\epsilon}_p[p] \\ \bar{\epsilon}_p[p+1] \\ \vdots \\ \bar{\epsilon}_p[N-2] \end{bmatrix} \quad (5)$$

라 놓으면, 식 (2)와 (3)은

$$\begin{aligned} \vec{\epsilon}_p &= \vec{e}_{p-1} - \gamma_p^* \vec{\bar{e}}_{p-1} \\ \vec{\bar{\epsilon}}_p &= \vec{\bar{e}}_{p-1} - \gamma_p^* \vec{e}_{p-1} \end{aligned} \quad (6)$$

가 된다. 따라서 식 (1)은

$$\begin{aligned} \alpha_p &= \|\vec{\epsilon}_p\|^2 + \|\vec{\bar{\epsilon}}_p\|^2 \\ &= (1 + \gamma_p \gamma_p^*) (\|\vec{e}_{p-1}\|^2 + \|\vec{\bar{e}}_{p-1}\|^2) \\ &\quad - 2\gamma_p^* (\vec{e}_{p-1}^+)^* \vec{\bar{e}}_{p-1} - 2\gamma_p (\vec{\bar{e}}_{p-1}^-)^* \vec{e}_{p-1} \end{aligned} \quad (7)$$

로 표현될 수 있다. 여기서 복소 gradient를 이용하면

$$\nabla \gamma_p \alpha_p = \gamma_p (\|\vec{e}_{p-1}^+\|^2 + \|\vec{\bar{e}}_{p-1}^-\|^2 - 2(\vec{e}_{p-1}^+)^* \vec{\bar{e}}_{p-1}^-) = 0 \quad (8)$$

$$\gamma_p = (2(\vec{e}_{p-1}^+)^* \vec{\bar{e}}_{p-1}^-) / (\|\vec{e}_{p-1}^+\|^2 + \|\vec{\bar{e}}_{p-1}^-\|^2) \quad (9)$$

가 되어 반사계수가 구해진다.

2. 성도 면적 추정 절차

Burg 알고리즘은 음성 신호에 대한 성도의 전달 함수를 구하여 성도를 디지털 필터로 모델링하는데 사용될 수 있다. 이는 디지털 필터의 계수, 즉, 반사계수는 성도를 무손실 음향튜브로 모델링하였을 때의 단면적을 특징지을 수 있기 때문이다.

음성 신호로부터 성도의 단면적을 추정하기 위해서는 성도를 구간마다 면적이 다른 음향 튜브의 연속으로 가정 한 비균일 음향 튜브 모델을 많이 이용하며, 이는 그림 1 과 같다.

여기서, Δl 은 한 구간의 길이, $U_m^+(t, dm)$ 과 $U_m^-(t, dm)$ 은 구간 m에서 성문에서 입술로 진행되는 음파와 그 반대로 진행되는 음파에 의한 체적 속도(volume velocity), dm 은 성문에서 구간 m과 m+1의 접점까지의 거리이다.

디지털 역필터와 음향 튜브가 일치하려면 샘플링 주파수 F_s 와 구간의 수(분석 차수) $M = l/\Delta l$ 의 관계에서는 $F_s = Mc/2l$ 이 만족되어야 한다. 여기서, c 는 음속, Δl 은 구간

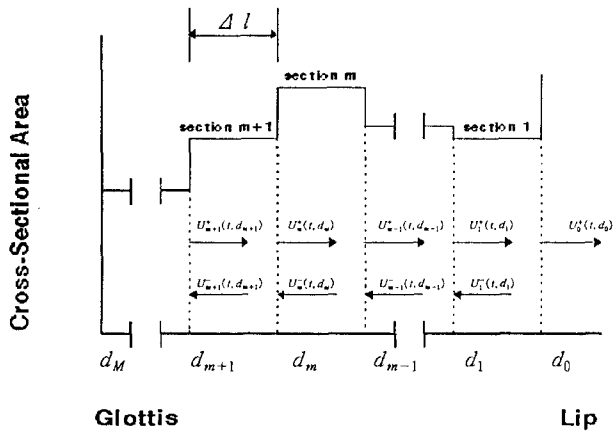


그림 1. 성도의 음향 튜브 모델
Fig. 1. Acoustic tube model for the vocal tract

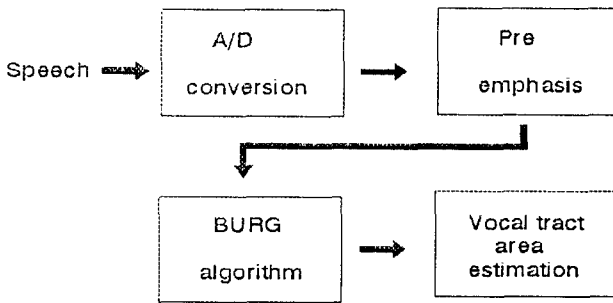


그림 2. 성도 면적 추정 블록 선도
Fig. 2. Block diagram of the vocal tract area estimation

분할된 성도 길이, l 은 전체 성도 길이를 나타낸다. 성도 면적 함수를 추정할 때 Atal 등과 Wakita가 제시한 2가지의 경계 조건이 있으며, Wakita의 방법이 더 정확하다고 알려져 있으므로, 본 연구에서는 Wakita의 경계 조건을 이용한다⁷⁾. 그림 2는 성도 면적 추정의 블록 선도를 나타낸 것이다.

A/D 변환된 음성 데이터에 대하여 프리엠퍼시스를 거쳐 Burg 알고리즘을 이용하여 반사계수를 구한다. 추출된 반사계수와 음성 신호의 관계는 식 (10)과 식 (11)로 표현될 수 있고, 식 (11)에 의하여 성도 면적을 구할 수 있다.

$$k_m = \frac{S_m}{S_{m+1}} \quad (10)$$

$$S_m = S_{m+1} \frac{1+k_m}{1-k_m} \quad (11)$$

여기서, k_m 은 반사계수이고, S_m 은 구간 m 에서의 성도 면적이다. 비교를 위한 특징 벡터인 LPC 켈프스트럼 계수는 식 (12)에 의하여 구해진다.

$$c_m = -a_m - \frac{1}{m} \sum_{k=1}^{m-1} k c_k a_{m-k}, \quad 1 \leq m \leq p \quad (12)$$

$$c_0 = \ln \sigma^2$$

여기서, a_m 은 선형 예측 계수이고, σ^2 은 residual power이다.

벡터 양자화

1. 벡터 양자화

벡터 양자화는 연속 혹은 떨어진 벡터들을 코드북(code book)으로 사상(mapping)시켜서 적절한 디지털 순열(digital sequence)로 부호화하는 방법으로서, 원래의 벡터 양자화는 데이터를 압축하기 위하여 개발되었으나, 어떠한 형태의 패턴이라도 그 패턴으로부터 샘플들의 벡터를 구성함으로써 그 패턴들을 표현할 수 있으므로, 벡터들을 미리 정해져 있는 패턴들과 매칭시킴으로써 벡터 양자화를 음성 인식에 적용시킬 수 있다.

K 차원, 레벨 N 의 벡터 양자화기는 각 입력 벡터 $X = \{x_0, x_1, \dots, x_{k-1}\}$ 에 재합성 알파벳 $\hat{A} = \{y_i; i=1, \dots, N\}$ 을 할당하는 사상이라고 할 수 있다. 벡터 양자화기는 재합성 알파벳(또는 코드북)에 의하여 표현된다.

벡터 양자화기는 source 벡터 X_0 에 가장 가까운 재합성 벡터 \hat{X}_0 을 찾아낼 수 있도록 X_0 을 양자화한다. 이것이 잘 되었는지를 알기 위한 왜곡치 척도(distortion measure)로 평균 왜곡치 $E \{d(X, \hat{X})\}$ 를 이용한다. 벡터 양자화에서 평균 왜곡치 $E \{d(X, \hat{X})\}$ 이 최소화되는 경우 이를 최적 벡터 양자화기라고 한다.

벡터 양자화기의 성능은 사용되는 코드북에 의하여 결정된다. 본 연구에서는 이를 위하여 Linde등이 제시한 LBG 알고리즘⁸⁾을 이용하며, 이는 알고리즘 1과 같다.

2. 벡터 양자화를 이용한 음성 인식 시스템

벡터 양자화기를 이용하여 음성 인식을 할 때는 다음과 같은 과정을 거친다. 먼저, 훈련 음성 데이터로부터 특징 벡터들을 추출하고, 추출된 특징 벡터를 이용하여 각 발음에 대하여 벡터 양자화를 수행하여 코드워드를 만든다. 코드워드를 y_j 로 표시하고, 코드북을 C 라 표시하면 $C = \{y_1, y_2, \dots, y_N\}$ 가 된다. S_j 를 부호화한 음성의 j 번째 프

알고리즘 1. LBG 알고리즘
Algorithm 1. LBG Algorithm

step 1. Initialization
 : 훈련 데이터 $\{x_i; i=0, \dots, n-1\}$, 코드북의 크기 K, 임계치 값 ϵ 이 주어졌을 때 $m=0$ 으로 놓고 초기 코드북을 추정한다. $M=1$ 로 놓는다.

step 2. Splitting
 : M개의 코드워드를 가진 코드북 $\{y_i; i=1, \dots, M\}$ 에서 각 코드워드 y_i 를 두 개의 매우 근접한 벡터 $y_i+\epsilon$ 과 $y_i-\epsilon$ 으로 분할하면 코드북은 2M개의 코드워드들을 가지게 된다. M을 2M으로 바꾼다.

step 3. Classification
 : 훈련 데이터를 아래와 같은 최소 거리 법칙에 따라 최소의 왜곡치를 갖는 영역으로 구분한다.
 $x_l \in S_i$ if $d(x_l, y_i) \leq d(x_l, y_j)$ for all l . compute MSE
 $D_m = MSE(\hat{A}_m, P(\hat{A}_m)) = n^{-1} \sum_{j=0}^{n-1} \min_{y \in A_m} d(x_j, y)$

step 4. : $((D_m - D_{m-1})/D_m) \leq \epsilon$ 이 아니면 step 5로 가고 이 조건도 만족하고 $M=K$ 조건도 만족하면 멈추고, 만족하지 않으면 step 2로 간다.

step 5. : 각 영역내의 훈련 데이터의 중심점을 계산하여 모든 영역내의 코드 워드들을 갱신하고 step 3로 간다.

레이스에서 추출한 음성 특징 벡터라고 하면, S_i 는 최소 거리 법칙에 의하여 코드북 내의 한 코드워드 C_i 로 부호화된다.

$$d(S_i, y_i) = \min d(S_i, y_j) \quad (13)$$

미지의 음성과 각각의 코드북사이의 평균 거리를 구한 뒤 가장 최소의 거리를 가지는 코드북의 음성을 인식된 음성으로 한다. 음성 특징 벡터사이의 거리 계산을 위하여 많이 사용되는 척도는 제곱 오차 거리척도(squared error distance measure), 수정 이타쿠라-사이토 거리척도(modified Itakura-Saito distance measure), 대수 스펙트럼 거리척도(log spectrum distance measure) 등이 있으며, 본 연구에서는 대수 스펙트럼 거리척도를 이용한다.

벡터 양자화기를 이용한 음성 인식 시스템의 블록 선도는 그림 3과 같다.

실험 및 결과 고찰

1. 실험

그림 4는 본 연구에서 구성한 전체 실험 과정에 대한 블록 선도이다.

실험에 사용한 음성은 연세대학교 교육방송국 스튜디오

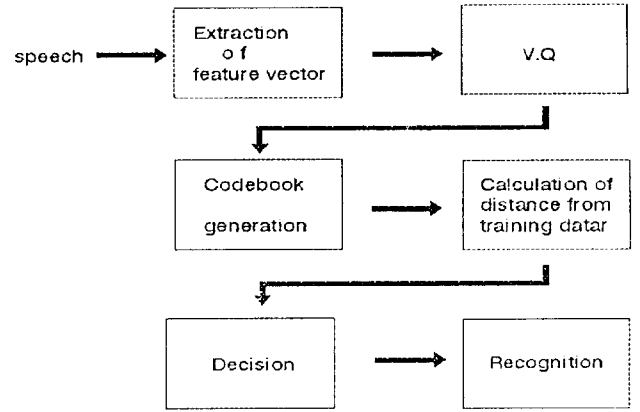


그림 3. 벡터 양자화를 이용한 음성 인식의 블록 선도
Fig. 3. Block diagram of speech recognition using vector quantization

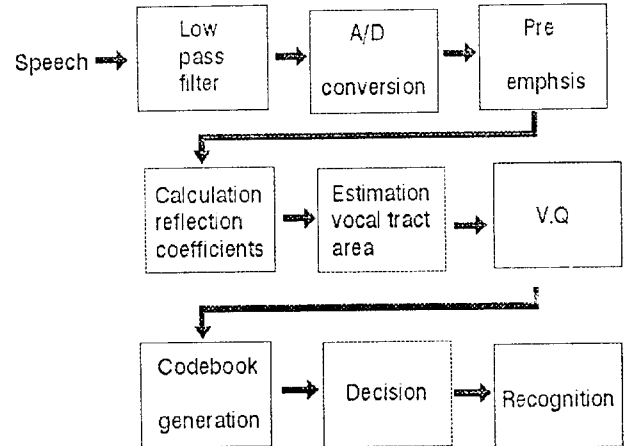


그림 4. 전체 실험 과정에 대한 블록 선도
Fig. 4. Block diagram of overall experiment procedure

오에서 녹음한 후, 카세트로 재생시키고, 저역 통과 필터를 통과시킨 후 A/D변환하였다. 저역 통과 필터는 차단 주파수 7.8[kHz], 2차 Butterworth 필터를 제작하여 사용하였다. A/D변환은 미국 Data Translation사의 12bit DT2821을 사용하여 20[kHz]로 표본화하였다. 프리엠퍼시스는 $(1 - \mu z^{-1})$ 에서 μ 를 0.95로 고정시켰다. 선형 예측 분석을 하여 반사계수를 추출하고, 이를 이용하여 성도 면적을 구하였다. 성도 면적을 특징 벡터로 하여 벡터 양자화를 하고, 코드북을 생성하여 거리 측정에 의하여 인식하게 된다. 인식 대상 음성은 5개의 한국어 단모음과 10개의 숫자음이고, 비교를 위한 특징으로는 기존에 널리 이용되던 LPC 켈스트럼 계수를 구하여 같은 실험을 수행하고 인식율을 비교하였다.

2. 결과 고찰

1) 모음 인식 실험

모음 인식 실험에서는 4인의 남성 화자가 단모음 /아/, /에/, /이/, /오/, /우/를 5회 발음한 총 100개의 데이터를 이용하였다. 각 화자의 첫번째 발음을 훈련 데이터로 이용하여 각 모음에 대한 5개의 코드북을 4명의 화자에 대하여 작성하여 모두 20개의 코드북을 작성하였다. 그림 5와 그림 6은 5개 단모음에 대하여 추출된 성도 면적과 스펙트럼을 나타낸다.

모음에서 특히 오인식이 잘 발생하는 /오/, /우/의 경우에, 그림 5의 성도 면적에서는 형태의 차이가 비교적 구별이 가능한 반면, 그림 6의 스펙트럼에서는 매우 유사한 형태를 보인다. 따라서 스펙트럼에 근거한 LPC 캡스트럼 계수 역시 /오/, /우/의 경우에 구별이 애매할 것이다. 벡터 양자화를 수행할 때 코드워드 개수는 4개를 선택하였다. 화자 종속 인식 실험은 4(화자) × 5(모음) × 4(회) = 80개의 데이터로 실험을 수행하였다. 그림 7에는 성도 면적을 특징 벡터로 벡터 양자화를 하여 얻은 코드북을 나타내었다.

5개 한국어 단모음에 대한 화자 종속 인식 실험 결과, 성도 면적을 이용한 경우에는 100%의 인식율을 얻었으

나, LPC 캡스트럼 계수를 이용한 경우에는 /오/를 /우/로 오인식한 경우가 있어서 98.8%의 인식율을 얻었다. 이러한 결과는 그림 5와 6에서 살펴본 바와 같이 성도 면적에서의 /오/, /우/의 패턴 구분 능력에 기인하는 것으로 생각된다. 화자 독립 실험은 4화자에 대하여 발음한 화자를 제외한 나머지 3인의 데이터에 대하여 인식 실험을 하였다. 성도 면적을 이용하였을 때는 96.3%의 인식율을 얻을 수 있었고, LPC 캡스트럼 계수를 이용하였을 때는 96.0%의 결과를 얻을 수 있었다. 화자 독립 실험에서도 화자 종속 실험과 마찬가지로 /아/, /에/, /이/에 있어서는 성도 면적과 LPC 캡스트럼 계수 모두에서 오인식이 발생하지 않았으나, /오/, /우/에 있어서 오인식이 발생하였다.

2) 숫자음 인식 실험

숫자음 인식 실험에서는 4인의 남성 화자가 10개의 숫자음 /영/, /일/, /이/, /삼/, /사/, /오/, /육/, /칠/, /팔/, /구/를 5회 발음한 총 200개의 데이터를 이용하였다. 각 화자의 첫번째 발음을 훈련 데이터로 이용하여 각 발음에 대한 10개의 코드북을 만들고 나머지 4번씩의 발음을 실험 데이터로 이용하였다. 숫자음 인식 실험에서도 코드북의 코드워드 개수는 4로 고정시켜 놓고 코드북을 만들었다. 그림 8은 성도 면적을 이용하여 만든 10개의 숫자음에 대한 코드워드 4개의 코드북을 나타낸 것이다.

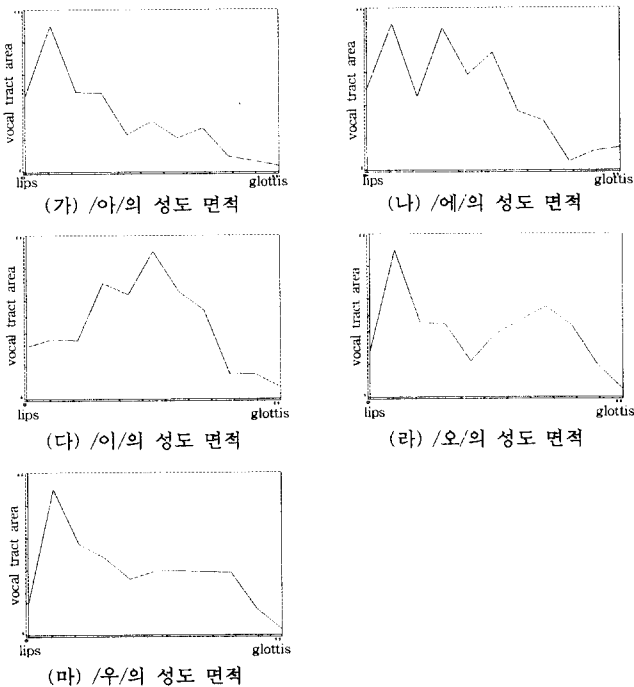


그림 5. 5개의 모음에 대한 성도 면적
Fig. 5. Vocal tract area functions for 5 vowels

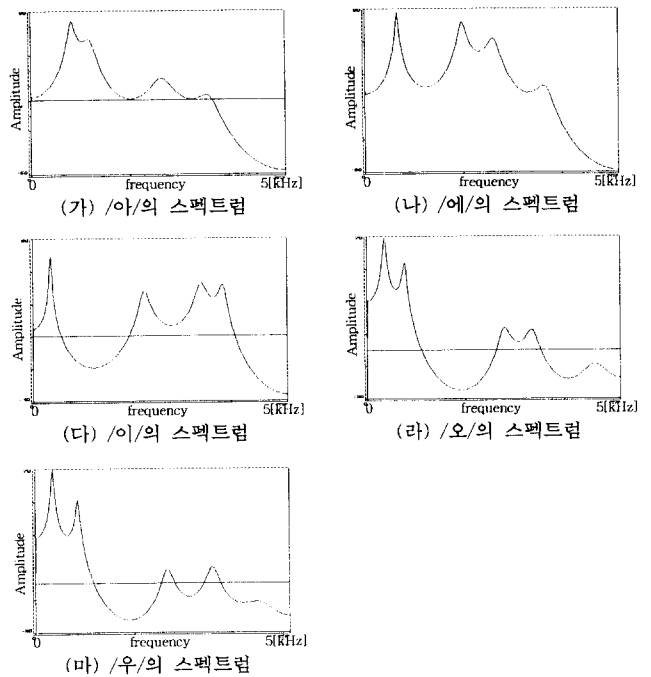


그림 6. 5개의 모음에 대한 스펙트럼
Fig. 6. Spectra for 5 vowels

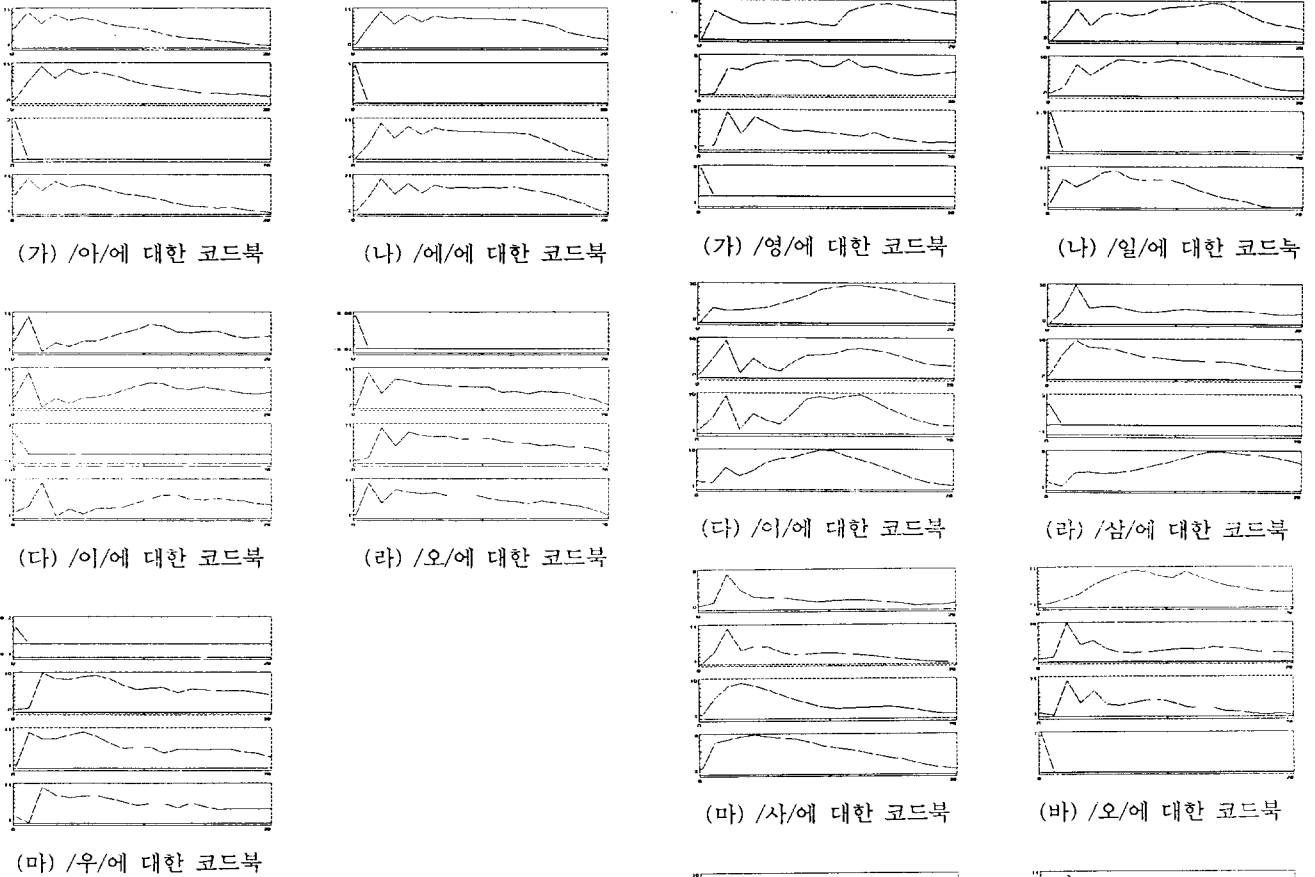


그림 7. 5개의 모음에 대한 코드북
Fig. 7. Codebooks for 5 vowels

가. 화자 종속 인식 실험

10개의 숫자음에 대한 화자 종속 인식 실험은 4명의 화자가 5회씩 발음한 10개의 숫자음 가운데 코드북을 만드는데 사용하지 않은 4(화자)×10(숫자음)×4(회)=160개의 데이터에 대하여 실험을 수행하였다. 숫자음에 대한 화자 종속 인식 실험 결과를 표 1에 나타내었다.

10개의 숫자음에 대한 화자 종속 인식 실험에서 성도 면적을 이용한 결과는 화자 KDJ의 /일/을 /이/로 1회 오인식하여 인식율 99.3%를 얻었고, LPC 켈스트럼 계수를 이용한 결과는 RKK의 /삼/을 /사/로 2회 오인식하고, KDJ의 /사/를 /삼/으로 1회 오인식하여 모두 3회의 오인식이 발생하여 인식율 98.1%를 얻어서 성도 면적을 이용한 결과가 더 우수한 결과를 얻었다.

나 화자 독립 인식 실험

화자 독립 인식 실험은 기준 데이터의 화자 이외의 3화

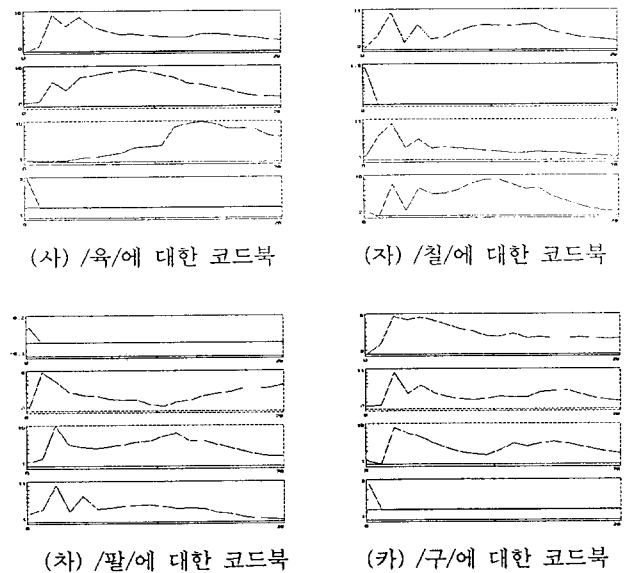


그림 8. 숫자음 인식에 사용된 성도 면적의 코드북
Fig. 8. Codebooks of vocal tract area in recognition of digits

자에 대하여 총 600회의 실험을 하였다. 10개의 숫자음에 대한 성도 면적과 LPC 켈스트럼 계수를 이용한 화자 독립

표 1. 10개의 숫자음에 대한 화자 종속 실험 결과

Table 1. Results of speaker dependent experiments for 10 digits

숫자음	숫자음에 대한 오인식 갯수							
	성도 면적				LPC 켈스트럼			
	KDJ	SJS	RKK	JSW	KDJ	SJS	RKK	JSW
/영/	0	0	0	0	0	0	0	0
/일/	1	0	0	0	0	0	0	0
/이/	0	0	0	0	0	0	0	0
/삼/	0	0	0	0	0	0	2	0
/사/	0	0	0	0	1	0	0	0
/오/	0	0	0	0	0	0	0	0
/육/	0	0	0	0	0	0	0	0
/칠/	0	0	0	0	0	0	0	0
/팔/	0	0	0	0	0	0	0	0
/구/	0	0	0	0	0	0	0	0
인식율 (%)	99.3				98.1			

표 2. 10개의 숫자음에 대한 화자 독립 실험 결과

Table 2. Results of speaker independent experiments for 10 digits

숫자음	숫자음에 대한 오인식 갯수							
	성도 면적				LPC 켈스트럼			
	KDJ	SJS	RKK	JSW	KDJ	SJS	RKK	JSW
/영/	3	2	4	2	2	1	2	0
/일/	3	5	2	0	3	1	3	2
/이/	3	2	3	2	2	5	2	1
/삼/	2	0	1	0	2	0	2	1
/사/	1	1	3	3	3	2	4	0
/오/	0	0	1	0	1	1	0	3
/육/	6	9	8	7	4	8	7	6
/칠/	1	2	1	2	1	2	1	2
/팔/	0	0	1	0	0	2	4	2
/구/	0	2	1	1	1	0	1	1
인식율 (%)	86.2				86.0			

립 인식 실험 결과를 표 2에 나타내었다.

10개의 숫자음에 대한 화자 독립 실험 결과를 살펴 보면, 성도 면적을 이용한 결과는 86.2%의 인식율을 얻었고, LPC 켈스트럼 계수를 이용한 결과는 86.0%를 얻어서 성도 면적을 이용한 결과가 약간 더 우수함을 알 수 있다. 오인식은 /영/과 /육/사이에 오인식이 많이 발생하였고, /일/과 /이/, /삼/과 /사/에서도 오인식이 많이 발생하였다.

3) 계산량의 고찰

본 연구에서는 특징 벡터 추출을 제외하면 모두 같은 실험 과정을 이용한다. 일반적으로 선형 예측 분석은 방법에 따라 선형 예측 계수나 반사계수를 추출하고, 이들은 쉽게 상호 변환된다. 성도 면적함수는 반사계수에서 추출되고, LPC 켈스트럼 계수는 선형 예측 계수에서 계산되므로, 선형 예측 분석 방법의 선택에 따라 변환 과정의 유무가 결정된다. 따라서 변환 과정은 계산량에 고려하지 않기로 하고 식 (11)과 식 (12)의 특징 벡터 추출식을 보면, 성도 면적 함수에 비하여 LPC 켈스트럼 계수를 추출할 때 계산량이 훨씬 많음을 알 수 있다.

면적 함수의 이용 가능성을 검증하기 위하여, 인식을 위한 특징 벡터로서 성도 면적 함수를 이용하고, 벡터 양자화를 적용하여 한국어 5개 단모음 및 10개 숫자음에 대한 화자 종속 및 화자 독립 인식 실험을 수행하였다. 또한 기존에 널리 이용되는 LPC 켈스트럼 계수에 의하여 동일한 과정의 실험을 수행하여 그 결과를 비교하였다.

언어진 결과는 다음과 같다. 첫째, 5개의 단모음에 대한 인식 실험에서 성도 면적을 특징 벡터로 이용할 때 화자 종속 실험과 화자 독립 실험 모두에서 LPC 켈스트럼 계수를 이용할 때보다 더 나은 인식율을 보였다. 둘째, 10개의 숫자음에 대한 인식 실험에서도 화자 종속 실험과 화자 독립 실험에서 성도 면적을 이용한 경우에 더 나은 인식율을 얻을 수 있었다. 셋째, 특징 벡터 추출 과정의 계산량을 비교해 볼 때, 성도 면적 함수에 비하여 LPC 켈스트럼 계수를 추출할 때 계산량이 훨씬 많음을 알 수 있었다. 끝으로, 성도 면적을 이용한 인식 실험에서 입력 패턴과의 거리를 계산할 때 유클리드 거리법을 이용하는 것보다 대수 거리법을 이용할 때 더 높은 인식율을 얻을 수 있었다.

결 론

본 연구에서는 음성 인식을 위한 특징 벡터로서 성도

참 고 문 헌

1. Panos E. Papamichalis, *Practical Approaches To Speech*

- Coding*, Prentice Hall, 1987.
2. M. R. Schroeder, "Determination of the geometry of the human vocal tract by acoustical measurement," JASA, vol. 41, pp. 1002-1010, 1967.
 3. P. Mermelstein, "Determination of the vocal tract shape from measured formant frequencies," JASA, vol. 41, pp. 1283-1294, 1967.
 4. A. Paige, V. W. Zue, "Computation of vocal tract area function," IEEE Trans. Audio Electroacoustics, Vol. AU-18, pp. 7-18, 1970.
 5. B. Gopinath, M. M. Sondhi, "Determination of the shape of the human vocal tract from acoustical measurement," Bell System Journal, Vol. 49, pp. 1195-1214, 1979.
 6. B. S. Atal, S. L. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave," JASA, Vol. 50, No. 2, pp. 637-655, Aug. 1971.
 7. H. Wakita, "Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveform," IEEE Trans. Acoust., Speech, Signal Processing, Vol. AU-21, No. 5, Oct. 1973.
 8. Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," IEEE Trans. Comm., Vol. Com-28, pp. 84-95, Jan. 1980.
 9. 이명국, 김동준, 윤태성, 박상희, "선형 예측 오차와 ARMA 모델을 이용한 음성 발생에 관한 연구," 1992년도 제5회 신호처리 합동학술대회, Vol. 5, No. 1, pp. 23-27, 1992. 9.
 10. G. Fant, *Acoustic Theory of Speech Production*, Mouton, 1970.