

# PTM 모델을 사용한 HMM 음성인식기에서 효율적인 디코딩을 위한 가우시안 선택기법

## Gaussian Selection in HMM Speech Recognizer with PTM Model for Efficient Decoding

손 종 목\*, 정 성 윤\*, 배 건 성\*

(Jong-Mok Son\*, Sung-Yun Jung\*, Keun-Sung Bae\*)

\*경북대학교 전자공학과

(접수일자: 2003년 11월 3일; 수정일자: 2003년 12월 10일; 채택일자: 2004년 1월 7일)

가우시안 선택기법은 연속 확률분포를 갖는 HMM 음성인식기에서 인식성능을 저하시키지 않으면서 관측확률을 구할 때 계산되는 가우시안의 수를 줄여 효율적인 디코딩을 하기 위해 많이 이용되는 방법이다. 본 논문에서는 PTM 구조를 갖는 HMM에서 관측확률을 계산하는데 필요한 가우시안 함수의 부분집합을 구하는 새로운 가우시안 선택기법을 제안한다. PTM 모델에서는 음성신호의 음향특성에 따라 구분되는 클래스별 가중치와 공통적인 가우시안 집합을 이용하여 각 상태를 나타내는데, 제안한 방법에서는 PTM 구조가 갖는 이러한 특성을 이용하여 인식성능의 저하없이 관측확률 계산에 소요되는 적은 수의 가우시안 부분집합을 구한다. 실험결과 기존의 가우시안 선택기법이 가우시안 선택기법을 적용하지 않았을 경우에 비해 20~30% 계산량을 필요로 하는데, 제안한 기법은 16.41%의 가우시안 함수 계산만으로도 별다른 인식성능 저하없이 인식 과정을 수행할 수 있었다.

**핵심용어:** 가우시안 선택, 음성인식, HMM, PTM

**투고분야:** 음성처리 분야 (2,5)

Gaussian selection (GS) is a popular approach in the continuous density hidden Markov model for fast decoding. It enables fast likelihood computation by reducing the number of Gaussian components calculated. In this paper, we propose a new GS method for the phonetic tied-mixture (PTM) hidden Markov models. The PTM model can represent each state of the same topological location with a shared set of Gaussian mixture components and context dependent weights. Thus the proposed method imposes constraint on the weights as well as the number of Gaussian components to reduce the computational load. Experimental results show that the proposed method reduces the percentage of Gaussian computation to 16.41%, compared with 20~30% for the conventional GS methods, with little degradation in recognition.

**Keywords:** Gaussian selection, Speech recognition, HMM, PTM

**ASK subject classification:** Speech signal processing (2,5)

## I. 서론

음성인식 기술은 인간과 기계 사이의 가장 자연스러운 의사 소통 수단을 제공할 수 있으므로 이와 관련된 연구가 꾸준히 이루어져 왔고, 일부 분야에서는 성공적으로 적용되고 있다. 하지만, 보다 많은 분야에서 음성인식 기

술이 적용되기 위해서는 다양한 시스템 환경에서 구현될 수 있도록 적은 계산량으로도 높은 성능을 얻을 수 있는 음성인식시스템이 요구된다. 현재, 많은 음성인식시스템들이 높은 인식성능을 얻기 위해서 연속분포 은닉마르코프 모델 (Continuous Density Hidden Markov Model; CDHMM)을 사용하고 있는데, 이러한 시스템들은 각 상태에서 관측확률을 얻기 위한 계산량이 전체 인식과정에서 소요되는 계산량의 대부분을 차지한다. 이는 CDHMM에서 인식성능을 높이기 위해 각 상태에서의 음향특성

분포를 많은 수의 가우시안 함수를 사용하여 모델링 하기 때문이다. 이러한 관측확률 계산과정에서의 계산량을 줄이기 위해서 특징 파라미터의 차수를 낮추는 LDA (Linear Discriminant Analysis) 기법이나[1], PTM (Phonetic Tied-Mixture) 모델과 같이 비슷한 특성을 갖는 파라미터를 공유하여 사용하는 파라미터 결합 (parameter-tying, parameter-sharing) 기법에 관한 연구가 많이 이루어져 왔으며[2,3], 또한 가우시안 선택기법과 같이 인식과정에서 계산되는 가우시안 함수의 수를 줄이고자 하는 기법도 많이 연구되어 왔다[4-7].

가우시안 선택기법은 CDHMM에서 관측확률을 계산할 때, 입력 음성신호와 유사도가 높은 일부의 가우시안 함수만을 사용함으로써 계산량을 줄이고자 하는 기법이다. 이때, 각 입력음성에 따라 선택되는 가우시안 함수의 부분집합 (shortlist)에 따라 인식성능과 계산량이 영향을 받게 되므로 적절한 부분집합을 구하는 것이 가우시안 선택기법에서 중요한 문제이다. 따라서, 본 연구에서는 PTM 구조를 갖는 HMM 음성인식시스템에서 보다 효율적으로 계산량을 줄이기 위해 가우시안 함수의 부분집합을 구하는 새로운 방법을 제안하고, 그에 따른 실험 결과를 제시한다.

본 논문의 구성은 다음과 같다. 1장 서론에 이어 2장에서는 음성인식시스템에 사용되는 PTM 구조에 대해 설명한다. 3장과 4장에서 기존 가우시안 선택기법과 PTM 모델에서의 가우시안 선택기법에 대해 각각 설명하고, 5장에서는 실험환경을 기술하며 각 기법에 따른 실험 결과를 제시하고 검토한다. 마지막으로 6장에서 결론을 맺고 향후 연구 방향을 제시한다.

## II. Phonetic Tied-Mixture 모델

일반적으로 유사음소를 기본단위로 하는 음성인식시스템에서는 음향특징을 보다 잘 나타내기 위하여 biphone 이나 triphone과 같은 문맥 종속 유사음소 모델을 많이 사용한다. 하지만, 주변 음소에 따라 모델을 다르게 설정하는 문맥 종속 모델의 경우 파라미터 수가 많아지기 때문에 충분한 훈련 데이터의 확보가 어려워질 뿐만 아니라 인식 과정에서의 계산량도 많아지기 때문에, 비슷한 특성을 가지는 모델 파라미터를 공유하여 전체 파라미터의 수를 줄이고자 하는 파라미터 결합기법이 많이 사용된다. 현재 많이 사용되는 파라미터 결합기법으로는 음향학적 유사도에 따라 각 상태의 파라미터를 공유하는 상태결합

기법과 기본단위에서의 위치에 따라 가우시안 집합을 공유하는 PTM 기법 등이 있으며[2,3], 제한된 훈련 데이터를 사용하여 유사한 크기의 모델 파라미터를 훈련하는 경우 PTM 기법이 우수한 성능을 보이는 것으로 보고되고 있다[3].

PTM 기법은 문맥 종속 모델의 기초가 되는 문맥 독립 모델에서의 위치가 같은 상태에서는 하나의 가우시안 함수 집합을 공유하고, 문맥에 따른 조음효과를 가우시안 함수의 가중치로서 표현하는 기법이다. 따라서, 문맥에 따른 조음효과를 반영하는 triphone 모델의 앞/뒤 위치에 따른 각 상태에 대해 각각 다른 가우시안 함수 집합을 가지는 상태결합기법에 비하여, PTM 기법에서는 천이특성을 나타내는 상태의 가우시안 집합이 음향공간에서 상대적으로 넓은 영역을 나타내게 된다. 그림 1에 일반적인 상태결합기법과 PTM 모델의 구성상의 차이를 나타내었다.

음성인식시스템에 사용한 PTM 구조는 각 상태에서 문맥에 따른 가우시안 함수의 가중치로써 조음효과를 표현하고, 가우시안 집합을 공유하기 때문에 모델 파라미터의 수를 크게 줄일 수 있다. 또한, 주변 유사음소의 특성에 따라 비슷한 특징을 가지는 가중치 파라미터를 공유함으로써 파라미터 수를 더욱 줄일 수 있다. 본 연구에서는 주변 유사음소를 그 특성에 따라 10개의 클래스 (Class)로 구분하고, 각 클래스별 가중치를 사용하여 조음효과를 고려하고자 하였다.

## III. 기존 가우시안 선택기법

CDHMM 알고리즘 기반의 음성인식시스템에서는 가우시안 함수를 사용한 관측확률 계산 과정이 전체 인식 과정에서 소요되는 계산량의 많은 부분을 차지한다. 관측

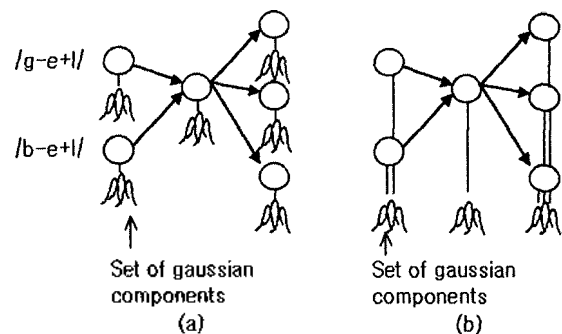


그림 1. 모델의 구조: (a) 상태 결합 HMM (b) PTM HMM  
Fig. 1. Model Structure: (a) Shared-state triphone HMM (b) PTM HMM.

확률은 입력음성 데이터에 가까운 가우시안 함수에서 주로 영향을 받기 때문에, 인식 과정에서 입력음성 데이터에 가까운 가우시안 함수만을 사용하여 관측확률을 계산하여도 큰 인식성능의 저하없이 계산량을 줄일 수 있다. Bocchieri[4]에 의한 가우시안 선택 기법의 구현에서는 훈련과정에서 벡터 양자화 (Vector Quantization; VQ)를 통하여 음향영역을 여러 개의 양자화된 영역으로 나누고, 각 가우시안 함수를 하나 이상의 코드워드 (codeword)에 할당하였다. 이때, 하나의 코드워드에 할당된 가우시안의 부분집합을 shortlist라 부른다. 인식 과정에서는 벡터 양자화 과정을 통하여 입력데이터에 가장 가까운 코드워드를 찾고, 이 코드워드에 연결된 shortlist의 가우시안 함수만을 계산하여 관측확률을 계산한다. 따라서, 벡터 양자화 기반의 가우시안 선택기법에서는 효율적인 shortlists를 구하는 것이 중요한 문제이며, 인식성능의 큰 저하를 유발하지 않으면서 작은 크기를 가지는 것이 요구된다.

### 3.1. Standard Gaussian Selection (SGS)

벡터 양자화에 기반한 표준 가우시안 선택기법의 적용 과정은 다음과 같다. 우선, HMM의 가우시안 함수를 사용하여 코드북을 생성한 후, 각 가우시안 함수들을 생성한 코드북의 코드워드 ( $c_\phi$ )에 할당하여 shortlist ( $v_\phi$ )를 작성한다. 이때, 하나의 가우시안 함수는 여러 개의 shortlists에 동시에 포함될 수 있으며, 가우시안 함수가 특정 shortlist에 포함되기 위해서는 식 (1)의 거리조건을 만족해야 한다.

$$G(m) \in v_\phi \text{ iff } D(G(m), c_\phi) \leq \Theta, \\ \text{where } D(G(m), c_\phi) = \frac{1}{K} \sum_{k=1}^K \frac{(c_\phi(k) - \mu_m(k))^2}{\sqrt{\sigma_{avg}^2(k)} \sigma_m^2(k)} \quad (1)$$

여기서,  $G(m)$ 은  $m$ 번째 가우시안 함수이고,  $K$ 는 특징 파라미터의 차수,  $\sigma_{avg}^2$ 는 전체 가우시안 함수의 평균 공분산,  $\mu_m, \sigma_m^2$ 은  $m$ 번째 가우시안 함수의 평균과 공분산. 그리고  $\Theta$ 는 문턱치를 나타낸다.

인식과정에서는 생성된 코드북을 사용하여 입력음성 데이터와 가장 가까운 코드워드를 구한 뒤, 코드워드에 할당된 shortlist에 포함되어 있는 가우시안 함수들만을 사용하여 관측확률을 계산한다.

### 3.2. State-Based Gaussian Selection (SBGS)

Shortlist의 크기는 식 (1)에서 문턱치  $\Theta$  값에 따라

달라진다.  $\Theta$  값을 줄이면 인식 과정에서 계산되는 가우시안의 수가 줄어들기 때문에 계산량의 측면에서 보다 효율적일 수 있지만, 대부분의 경우 인식성능을 떨어뜨리게 된다. 만약 인식 과정에서 특정 상태의 관측확률 계산에 큰 영향을 끼치는 가우시안 함수가 선택된 shortlist에 포함되지 않을 경우 인식성능은 감소하게 되며, 특히 특정상태의 가우시안 함수가 선택된 shortlist에 하나도 포함되지 않았을 경우 인식성능에 있어서 많은 감소를 일으키게 된다. 때문에, 큰 인식성능의 감소를 일으키지 않기 위해서는 특정상태에 대하여 관측확률 계산에 큰 영향을 끼치지 않는 가우시안 함수의 수는 줄이는 반면, 큰 영향을 끼칠 수 있는 가우시안 함수는 하나 이상 포함시키는 것이 바람직하다. 상태정보를 사용하지 않는 표준 가우시안 선택기법에서는 이러한 사항들이 고려될 수 없으므로, 상태정보를 도입하여 shortlists를 작성하는 상태 기반 가우시안 선택기법이 도입되었다[5,6]. 상태 기반 가우시안 선택기법은 특정 상태에서 하나의 shortlist에 포함되는 가우시안의 수를 제한하는 방법이다.

HMM의 특정상태에서의 관측확률은 입력음성 데이터에 가까운 몇 개의 가우시안 함수에 의해 주로 결정된다. 그러므로, 상태 기반 가우시안 선택기법에서는 특정 코드워드의 shortlist를 구할 때, 하나의 상태 내에서의 가우시안 수를 제한함으로써 인식성능은 크게 저하시키지 않으면서 계산량을 줄이고자 하는 기법이다. 상태 기반 가우시안 선택기법에서 가우시안 함수가 특정 shortlist에 포함될 조건은 식 (2)와 같다.

$$G_j(m) \in v_\phi \text{ iff } G_j(m) \in \arg \min_m (n) \{ D(G_j(m), c_\phi) \}, \\ \text{where } D(G_j(m), c_\phi) \leq \Theta \quad (2)$$

여기서,  $G_j(m)$ 은  $j$ 번째 상태에서  $m$ 번째 가우시안 함수이고,  $\arg \min_m (n) \{ \cdot \}$ 은 주어진 거리측정 방법에 따라 최소  $n$ 개 내에 포함됨을 나타낸다.

코드워드에 가까이 있는 가우시안에서 그 코드워드에 할당되는 데이터를 많이 발생시킨다고 가정하면, 코드워드와 가까운 가우시안 함수를 가지는 상태에 보다 많은 가우시안 함수를 할당하고 그 반대의 경우에는 적은 가우시안 함수를 할당하여 음향특성의 분포를 효율적으로 반영함으로써 보다 적절한 shortlists를 구할 수 있다. 이러한 과정은 멀티 링 기법에 의해 쉽게 구현되는데, 특정 코드워드 ( $c_\phi$ )에 관계된 shortlist에  $j$ 번째 상태에서 포함시킬 수 있는 가우시안의 최대 개수 ( $n$ )를 구하는 알고리즘은 다음과 같다[6].

if  $\min_m \{D(G_j(m), c_\phi)\} \leq \Theta_1$ , then  $n = N_1$   
 else if  $\min_m \{D(G_j(m), c_\phi)\} \leq \Theta_2$ , then  $n = N_2$   
 else  $n = 0$

여기서,  $\Theta_1 < \Theta_2 \leq \Theta$  이고,  $N_1 \geq N_2 \geq 0$ 이며, 본 연구에서는  $\Theta_2$  와  $\Theta$  를 같게 하여 실험하였다 ( $\Theta = \Theta_2$ ).

## IV. PTM 모델에서의 가우시안 선택기법

### 4.1. State Topology-Based Gaussian Selection (STBGS)

상태기반 가우시안 선택기법의 효율적인 적용을 위해서는 문턱치 설정에 있어 각 상태에서의 가우시안 분포 특성이 고려되어야 한다. 일반적인 상태결합기법에서는 각 상태가 독립적인 가우시안 집합을 가지게 되지만, PTM 모델의 경우 문맥 종속 모델의 기초가 되는 문맥 독립 모델에서의 위치가 같은 상태에서는 같은 가우시안 집합을 사용하기 때문에 조음효과를 고려하여 천이특성을 나타내는 상태의 가우시안 집합이 음향공간에서 상대적으로 넓은 영역을 나타내게 된다. 반면, 문맥에 따른 조음효과의 영향을 크게 받지 않는 정상상태에서의 가우시안 분포는 상태 결합 모델의 가우시안 집합과 비슷한 분포를 나타낸다. 즉, PTM 모델에서 문맥에 따른 조음효과를 고려하는 천이상태의 가우시안 집합이 문맥에 상관하지 않는 정상상태의 가우시안 집합에 비해 음향공간에서 넓은 공간을 나타낸다. 이와 같은 특성을 고려하였을 때, 효율적인 shortlists의 작성을 위해서는 PTM 모델에서 천이상태와 정상상태에서의 문턱치 설정을 다르게 해 줄 필요가 있다. 즉, 천이상태의 가우시안 분포가 정상상태의 가우시안 분포에 비해 넓은 영역을 표현하기 때문에, 천이 상태의 문턱치를 정상상태의 문턱치보다 크게 잡아줄 필요가 있다. 본 연구에서는 이를 고려하여 PTM 구조의 음성인식시스템에서 천이상태와 정상상태에서 shortlist 작성을 위한 문턱치 값을 다르게 설정해주는 상태 위치 기반 가우시안 선택기법을 제안하고, 이를 STBGS (State Topology-Based Gaussian Selection) 라고 하였다. 사용한 HMM은 하나의 기본 유사음소에 대해 3 상태 Bakis (Left-to-Right) 모델을 사용하였기 때문에, PTM 구조의 첫번째와 세번째 상태는 가중치에 따라서 조음효과를 고려하는 천이상태로 두번째 상태는 조음효과를 고려하지 않는 정상상태로 정의하였다.

### 4.2. Class-Based Gaussian Selection (CBGS)

PTM 모델의 천이상태에서는 하나의 가우시안 집합에 대하여 클래스별 가우시안 함수의 가중치를 통하여 조음효과를 고려하므로, 천이상태에 대한 코드워드의 shortlist를 구할 때, 상태 단위로 가우시안 수를 제한하는 것보다는 클래스 단위로 가우시안 수를 제한하는 것이 더욱 타당하다고 할 수 있다. 또한, PTM 모델의 천이상태에서는 주변에 올 수 있는 모든 유사음소에 대해 하나의 가우시안 집합을 사용하므로, 실제 하나의 클래스에 의한 조음효과를 고려하면 보다 적은 가우시안 함수를 사용하게 된다. 따라서, 본 연구에서는 이와 같은 점들을 고려하여 클래스 기반 가우시안 선택기법 CBGS (Class-Based Gaussian Selection)을 제안한다.

제안한 클래스 기반 가우시안 선택기법은 특정 코드워드의 shortlist를 구할 때, 하나의 클래스 내에서의 가우시안 수를 제한함으로써 인식성능은 크게 저하시키지 않으면서 계산량을 줄이고자 하는 기법이다. PTM 모델의 천이상태에서는 식 (3)과 같이 클래스에 따른 가우시안 함수의 가중치 합으로써 관측확률이 계산된다.

$$P_{jk}(O_t) = \sum_m w_{jk}(m) N(O_t, \mu_j(m), \sigma_j(m)) \quad (3)$$

여기서,  $P_{jk}(\cdot)$ 는  $j$ 번째 상태의  $k$ 번째 클래스에 대한 관측 확률이고,  $O_t$ 는 시간  $t$ 에서의 입력 데이터이다.  $w_{jk}(m)$ 는  $j$ 번째 상태의  $k$ 번째 클래스에서  $m$ 번째 가우시안에 대한 가중치,  $\mu_j(m)$ ,  $\sigma_j(m)$ 은 각각  $j$ 번째 상태의  $m$ 번째 가우시안의 평균과 표준편차,  $N(\cdot)$ 은 가우시안 분포 함수이다. 식 (3)에서 볼 수 있듯이 클래스별 관측확률이 클래스에 따른 가우시안 함수의 가중치 합으로써 표시되기 때문에, 가중치 값이 낮게 나타나는 가우시안 함수에 의한 영향은 작다 할 수 있다. 그러므로, 천이상태에서 클래스에 따라 유용한 가우시안 집합을 재정의하였다. 각 상태에서의 가우시안 함수가 재정의된 가우시안 집합에 포함될 조건은 식 (4)와 같다.

$$G_j(m) \in G_{jk} \text{ iff } w_{jk}(m) > \Theta_c \quad (4)$$

여기서,  $G_j(m)$ 은  $j$ 번째 상태에서  $m$ 번째 가우시안이고,  $G_{jk}$ 는  $j$ 번째 상태에서  $k$ 번째 클래스에 유용한 가우시안 집합을,  $\Theta_c$ 는 문턱치를 각각 나타낸다. 본 연구에서는  $\Theta_c = 4.54 \times 10^{-5}$ 을 사용하여 실험하였다. 클래스 기반 가우시안 선택기법에서는 클래스별로 재정의된 가우시안 집합을 사용하여 shortlist를 작성한다. 클래스 기반 가우시안 선택기법에서 가우시안 함수가 특정 shortlist

에 포함될 조건은 식 (5)와 같다.

$$G_{jk}(m') \in v_{\phi} \text{ iff } G_{jk}(m') \in \arg \min_m \{D(G_{jk}(m'), c_{\phi})\},$$

where  $D(G_{jk}(m'), c_{\phi}) \leq \Theta$  (5)

여기서,  $G_{jk}(m')$ 은  $j$ 번째 상태에서  $k$ 번째 클래스에 대해 재정의된 가우시안 집합에서  $m'$  번째 가우시안이다.

## V. 실험 및 검토

실험 환경은 다음과 같다. 가우시안 선택기법에 따른 인식능과 계산량의 차이를 보기 위해 PTM 모델을 사용한 가변어휘 인식을 구현하였으며, 인식시스템의 기본 단위로는 한국어 자소에 기반한 46개의 유사음소를 사용하였다. 발음사전의 생성을 위해서는 한국어 읽기 규칙에 기반하여 단어를 유사 음소열로 표기한 후 유사음소 사이에 나타나는 부가음운을 확률적으로 첨가하였다. PTM을 위해서 10개의 클래스를 정의하였고, 하나의 음소를 모델링하기 위해 3 상태 Bakis (Left-to-Right) 모델을 사용하였으며, 기본 유사음소 중 길이가 짧은 유사음소를 나타내기 위하여 상태의 생략을 허용하였다. 각 상태에서 관측분포를 나타내기 위해 사용한 가우시안의 수는 16개이다. 인식 알고리즘으로는 비터비 빔 탐색 (Viterbi Beam Search) 기법을 사용하였으며, 프루닝 (pruning) 기법을 적용하였다. 기본 시스템의 훈련을 위해서 ETRI (Electronics and Telecommunications Research Institute)의 445 DB 중 훈련용 데이터를 8 KHz로 다운 샘플링하여 사용하였다. 각 기법의 적용에 따른 인식률과 계산량의 변화를 보기 위한 평가용 DB로는 445 DB의 테스트용 데이터를 사용하였다.

음성신호의 특징추출을 위한 분석 조건은 표 1과 같다. 음성 데이터를 전처리 계수 0.97로 전처리 한 후, 20 ms의 해밍 윈도우를 10 ms마다 취하여 구간 분석하였다. 각 구간에서 1차의 에너지와 12차의 멜 캡스트럼을 구하고, 현재 구간을 포함한 전후 6구간 (전체 13구간)의 정보를 이용하여 1차의 차분에너지와 12차의 차분 멜 캡스트럼, 그리고 차분-차분 값을 구하였다. 여기에, LDA를 적용하여 최종적으로 16차의 특징 파라미터를 추출하였다.

각 가우시안 선택기법의 성능은 인식능과 가우시안 함수 계산량의 감소 정도로 평가하였으며, 가우시안 함수 계산량의 감소 정도는 가우시안 선택기법을 적용하기 전의 가우시안 계산량에 대한 비교지수 (Computation

표 1. 음성데이터 분석 조건

Table 1. Analysis condition of speech data.

Sampling Freq.	8 kHz
Preemphasis factor	0.97
Analysis window size	20 ms (160 points)
Frame rate	10 ms (80 points)
Feature parameters	MFCC based (38) → LDA (16)

Fraction; CF)를 사용하였다. 표 2에 표준 가우시안 선택기법과 상태 기반 가우시안 선택기법의 인식률과 계산량 비교지수를 나타내었으며, 생성된 shortlists의 크기를 비교하기 위하여 shortlists에 포함되어 있는 가우시안 인덱스의 수를 함께 나타내었다. 가우시안 선택기법을 적용하기 전의 기본 인식시스템의 어려움은 3.38%이며, 가우시안 선택기법에 따른 어려움 증가가 0.1% 내에 드는 경우에 대하여 각 기법 적용에 따른 비교지수를 나타내었다. 제시된 비교지수는 여러 문턱치에 대한 실험을 통하여 얻은 국부 최소값이다. 가우시안 선택기법의 적용 전에 비해 표준 가우시안 선택기법 (SGS)의 경우 28.96% 가우시안 함수 계산만으로도 큰 인식능의 저하가 발생하지 않음을 볼 수 있다. 상태 기반 가우시안 선택기법 (SBGS)의 경우 20.66%의 가우시안 함수 계산만으로도 비슷한 성능을 보여, shortlist 작성 시 상태 정보를 고려해줌이 유용하다는 것을 보여주고 있다.

PTM 구조를 고려하여 제안한 상태 위치 기반 가우시안 선택기법 (STBGS)과 클래스 기반 가우시안 선택기법 (CBGS)의 인식률과 계산량 비교지수를 표 3에 나타내었다. 또한, 상태기반 가우시안 선택기법과 상태 위치 기반 가우시안 선택기법의 비교를 위해 상태 위치 기반 가우시안 선택기법에서 사용한 파라미터값에 대한 상태기반 가우시안 선택기법의 성능을 함께 나타내었다. 표에서  $\Theta'$ ,  $N'$ 는 조음효과를 고려하는 천이상태에서의 파라미터이고,  $\Theta$ ,  $N$ 는 조음효과를 고려하지 않는 정상상태에서의 파라미터를 나타낸다.

PTM 모델의 특성을 고려한 상태 위치 기반 가우시안 선택기법과 클래스 기반 가우시안 선택기법의 경우 각각 18.29%와 16.41%의 가우시안 계산만으로도 큰 인식 성능의 저하를 일으키지 않는 것을 볼 수 있다. 특히, 클래스 기반 가우시안 선택기법의 경우 큰 인식 성능의 저하를 일으키지 않으면서도 표준 가우시안 기법에 비해서 12.55%, 상태 기반 가우시안 선택기법에 비해서 4.25%의 계산을 추가적으로 줄일 수 있었다.

표 2. 표준 가우시안 선택기법과 상태기반 가우시안 선택기법의 에러율과 비교지수  
Table 2. Error rate and computation fraction of SGS, and SBGS.

Selection scheme	Thresholds				# of Index (in shortlists)	C.F. (%)	Error (%)
	$\theta_1$	$N_1$	$\theta_2$	$N_2$			
Without GS	-	-	-	-	-	100	3.38
SGS	2.4	-	-	-	51,313	28.96	3.47
SBGS	1.4	6	3	1	38,375	20.66	3.45

표 3. 상태 위치 기반 가우시안 선택기법과 클래스 기반 가우시안 기법의 에러율과 비교지수  
Table 3. Error rate and computation fraction of STBGS, and CBGS.

Selection scheme	Thresholds				# of Index (in shortlists)	C.F. (%)	Error (%)
	$\theta_1^a$	$N_1^a$	$\theta_2^a$	$N_2^a$			
Without GS	-	-	-	-	-	100	3.38
SBGS	1.6	5	2.6	1	34,557	18.68	3.59
	2.0	4	2.6	1	34,449	18.36	3.77
STBGS	2.0	4	2.6	1	34,349	18.29	3.41
	1.6	5	2.6	1			
CBGS	1.4	3	2.6	1	30,484	16.41	3.45
	1.0	5	2.4	1			

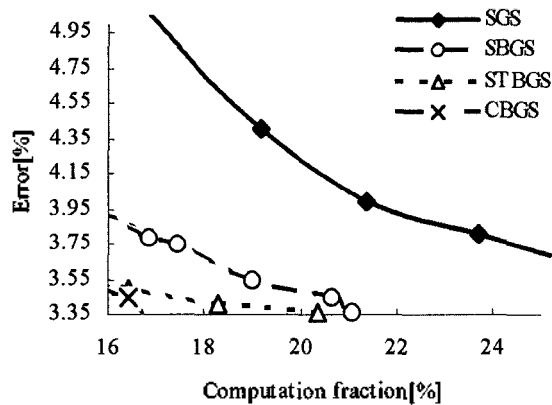


그림 2. SGS, SBGS, STBGS, CBGS의 비교  
Fig. 2. Comparison of SGS, SBGS, STBGS, and CBGS.

각 가우시안 선택기법의 성능을 비교하기 위하여 가우시안 비교지수별 에러율 곡선을 그림 2에 나타내었다. 비교지수별 에러율 곡선을 살펴보면 에러율이 크게 늘지 않는 범위 (< 0.1%)에서 클래스 기반, 상태 위치 기반, 상태 기반, 표준 가우시안 선택기법의 순으로 성능이 우수하게 나타나므로, PTM 구조의 인식시스템에서 상태 정보 보다는 클래스 정보를 고려하여 shortlist를 작성하는 것이 바람직하다는 것을 알 수 있다.

## VI. 결론

본 연구에서는 가우시안 선택기법을 적용한 PTM 구조의 음성인식시스템에서 효율적인 shortlists 작성을 위한 방법을 제안하였다. PTM 모델에서는 문맥에 의한 조음 효과를 나타내는 천이상태에서 하나의 가우시안 집합을 공유하고 클래스별 가중치로써 조음효과를 고려하므로, 각각의 가우시안 집합을 가지는 일반적인 상태결합기법에 비해 가우시안 집합이 음향공간에서 더욱 넓은 영역을 나타내게 된다. 이를 고려하여 상태 위치 기반 가우시안 선택기법과 클래스 기반 상태 공유기법을 제안하고, 실험을 통하여 기존의 가우시안 선택기법과 그 성능을 비교하였다.

각 기법별로 에러율이 크게 증가하지 않는 범위 내 (< 0.1%)에서 가우시안 계산량의 비교지수를 구한 결과, 표준 가우시안 선택기법과 상태 기반 가우시안 선택기법의 경우 28.96%, 20.66%의 가우시안 계산만으로도 인식 성능에 큰 저하가 없음을 보였다. 특히, 본 연구에서 제안한 PTM 모델의 특성을 고려한 가우시안 선택기법인 상태 위치 기반 가우시안 선택기법 및 클래스 기반 가우시안 선택기법에서는 각각 18.29%, 16.41%의 가우시안 계산만으로도 인식 성능에 큰 저하를 일으키지 않으므로, 제안한 방법의 유용성을 확인할 수 있었다.

## 참고 문헌

1. R. Haeb-Um bach and H. Ney, "Linear discriminant analysis for improved large vocabulary continuous speech recognition," *In Proc. International Conference on Acoustic, Speech and Signal Processing*, 1, 13-16, 1992.
2. Akinobu Lee, Tatsuya Kawahara, Kazuya Takeda, and Kiyohiro Shikano, "A new phonetic tied-mixture model for efficient decoding," *In Proc. International Conference on Acoustic, Speech and Signal Processing*, 3, 1269-1272, 2000.
3. Mei-Yuh Hwang and Xuedong Huang, "SubPhonetic modeling with Markov States-SENONE," *In Proc. International Conference on Acoustic, Speech and Signal Processing*, 1, 33-36, 1992.
4. E. Bocchieri, "Vector quantization for efficient computation of continuous density likelihoods," *In Proc. International Conference on Acoustic, Speech and Signal Processing*, 2, 692-695, 1993.
5. K. M. Knill, M. J. F. Gales, and S. J. Young, "Use of gaussian selection in large vocabulary continuous speech recognition using HMMs," *In Proc. International Conference on Spoken Language Processing*, 1, 470-473, 1996.
6. Mark J. F. Gales, Katherine M. Knill, and Stephen J. Young, "State-based gaussian selection in large vocabulary continuous speech recognition using HMM's," *IEEE Trans. on Speech and Audio Processing*, 7 (2), 152-161, 1999.
7. Akinobu Lee, Tatsuya Kawahara, and Kiyohiro Shikano, "Gaussian mixture selection using context-independent HMM," *In Proc. International Conference on Acoustic, Speech and Signal Processing*, 1, 69-72, 2001.
8. Mosur K. Ravishankar, "Efficient algorithms for speech recognition," Ph. D. thesis, Carnegie Mellon University, 1996.
9. 손종목, 정성윤, 김민성, 배건성, "PTM 모델을 적용한 HMM 시스템

에서 효율적인 디코딩을 위한 가우시안 선택기법에 관한 연구," 한국음향학회 추계학술대회 논문집, 22 2(s), 15-18, 2003.

## 저자 약력

## ● 손종목 (Jong-Mok Son)



1997년 2월: 경북대학교 전자공학과 (공학사)  
 1999년 2월: 경북대학교 전자공학과 (공학석사)  
 1999년 3월~현재: 경북대학교 전자공학과 박사과정 재학 중  
 ※ 주관심분야: 디지털 신호처리, 음성신호처리, 음성인식

## ● 정성윤 (Sung-Yun Jung)



1991년 2월: 경북대학교 전자공학과 (공학사)  
 1994년 2월: 영남대학교 전자공학과 (공학석사)  
 2000년 3월~현재: 경북대학교 전자공학과 박사과정 재학 중  
 ※ 주관심분야: 디지털 신호처리, 음성신호처리, 음성인식

## ● 배건성 (Keun-Sung Bae)



1977년 2월: 서울대학교 전자공학과 (공학사)  
 1979년 2월: 한국과학기술원 전기 및 전자공학과 (공학석사)  
 1989년 5월: University of Florida (공학박사)  
 1979년 3월~현재: 경북대학교 전자공학과 교수  
 ※ 주관심분야: 음성분석 및 인식, 디지털 신호처리, 음성 부호화, 웨이브렛 분석 등