

# 코드워드 의존 거리 정규화와 거리에 기반한 코드워드 가중을 이용한 은닉마르코프모델의 파라미터 추정

## Estimation of HMM parameters Using a Codeword Dependent Distance Normalization and a Distance Based codeword Weighting by Fuzzy Contribution

최 환 진\*, 오 영 환\*  
(Hwan Jin Choi\*, Yung Hwan Oh\*)

### 요 약

본 연구에서는 견고한 이산형 은닉마르코프모델의 파라미터를 얻기 위한 방법으로 CDDN(codeword dependent distance Normalization)과 거리에 기반한 코드워드 가중방법을 제안한다. 제안된 방법에서 FVQ(fuzzy vector quantization)에 기반한 방법에서 코드워드에 대한 출력확률 계산 시, 코드워드의 분포특성과 상태 의존적인 코드워드의 특성을 반영하여 거리를 계산하고, 이 거리값에 퍼지목적함수를 적용하여 코드워드별 기여도를 계산한다.

실험결과, 제안된 방법이 기존의 FVQ기반한 방법에 비해서 4.5%정도의 인식율 향상이 있음을 할 수 있었다. 특히, 거리 가중치를 사용하여 출력확률 평활화를 적용한 경우가 단순히 코드워드별 가중을 적용한 경우에 비해서 2.5%의 성능향상을 보였다.

### ABSTRACT

In this paper, we have proposed the robust estimation of HMM parameters which is based on CDDN(codeword dependent distance normalization) and codeword weighting by distance.

The proposed method has used a distance normalization based on the characteristics of a codeword dependent distribution and have computed fuzzy contributions of codewords to a input vector with a fuzzy objective function.

From experimental results, we have shown the effectiveness of the proposed method in that the correction rate of the proposed method is improved 4.5% over the conventional FVQ based method. Especially, the application of distance weighting to smoothing of output probability is improved the performance of 2.5% compared to distance based codeword weighting.

### I. 서 론

1950년에 음성인식을 위한 연구가 시작된 이래, 현재까지 인간이 발성한 음성으로부터 인간이 이해할 수 있는 언어적인 정보를 추출하는 음성인식에 관한 다양한 연구들이 진행되고 있다. 음성인식을 위한 방법[1]으로는 동적정합법(DTW:dynamic time warping), 지식처리에 의한 방법(knowledge based approach), 신경회로망(neural network)에 의한 방법과 통계적인 방법인 은닉마르코프모델(HMM: hidden Markov model)에 의한 방법들이 주로 사

용되고 있으며, 통계적인 방법에 기반한 HMM이 다른 방법에 비해서 널리 사용되고 있는 실정이다.

HMM[2]은 기존의 Markov chain에 기반한 방법으로, Markov모델의 경우 한 상태는 하나의 의미를 갖으나 HMM에서는 한 상태가 확률적으로 여러개의 의미들을 가질 수 있는 특징을 갖고 있다. 따라서, 임의의 시점에서 한 상태가 가질 의미를 명확하게 알 수 없으므로, 은닉(hidden)이라는 말을 덧붙여 사용한다. 은닉마르코프모델은 음성신호의 흐름을 몇가지 상태(state)로 나누고, 그러한 상태간의 전이(transition)와 각 상태에서 심벌(symbol)의 출력이라는 2가지 랜덤 프로세스들로 모델링한다. 여기서, 심벌이란 은닉 마르코프모델의 입력으로 들어가는 코드북의 코드워드 색인을 의미하며, 상태에서 심벌의 출력이란 입력으로 들어온 코드워드가 해당 상태에서 출

\* 한국과학기술원 전산학과/인공지능연구센터  
Department of Computer Science/CAIR(Center for Artificial Intelligence Research)

접수일자: 1996년 4월 29일

력될 확률을 의미한다. 이러한 마르코프모델을 이용해서 음소, 음절, 단어등의 다양한 인식단위의 모델링이 가능하다.

HMM은 상태별 출력확률을 모델링하는 방법[3]에 따라 연속형과 이산형으로 나누어진다. 연속형의 경우 ellipse형 함수들의 선형결합을 통해서 출력확률을 모델링함으로써, 특정 파라미터의 양자화가 불필요하나 상태별로 보다 상세한 모델링을 위해서는 많은 학습자료들을 필요로 한다는 단점이 있다. 이산형의 경우 모델링이 단순하며, 적은 학습량으로도 일정한 수준의 인식율을 보이며, 응답속도 또한 연속형에 비해 빠른 특성을 가지고 있다. 그러나, 이산형 HMM은 코드북에 따른 양자화 오류발생, unseen데이터의 처리, 모델간의 분별력 향상등에 관한 문제들이 있다. 코드북을 사용해서 입력된 음성신호를 양자화함으로써 오류가 발생하게 되는데, 이러한 문제를 해결하기 위해서 코드북 작성 시에 입력과 최소 거리를 갖는 표준패턴만을 고려하기보다는 다수의 표준패턴들을 함께 고려하여 사용한다[4]. 그리고, 학습 시에 학습이 되지 않은 자료가 실험 시에 나타나는 unseen데이터 처리문제가 발생되는데, 이를 위해서 출력확률을 평활화하는 것이 필요하게 된다[5, 6].

본 연구에서는 이산형 HMM을 대상으로 입력패턴의 양자화 오류를 줄이면서, 상태 내에서 심벌의 출력확률 향상시키기 위한 방법을 제안하고자 한다. 기존의 FVQ(fuzzy vector quantization)를 적용한 심벌의 출력확률 모델링 방법과는 달리 코드워드의 분포특성과 상태 내에서의 코드워드의 가중치를 반영한 입력패턴에 대한 코드워드와의 거리를 기반으로 한다. 이러한 각 코드워드에 대한 퍼지목적함수를 사용해서 얻어진 퍼지기여도를 출력확률의 가중치로 적용하여 상태별 출력확률을 계산하고자 한다.

본문의 구성은 다음과 같다. 2장에서는 FVQ를 이용한 상태내의 출력확률모델링에 대해서 기술하고, 3장에서는 본 논문에서 제안한 CDDN과 거리기반 코드워드 가중치 유도과정에 대해서 설명하고자 한다. 4장에서는 실험 및 결과에 대해서 기술하고, 마지막 5장에서는 결론과 향후 연구방향에 대해서 논하고자 한다.

## II. FVQ에 기반한 은닉마르코프모델

본 장에서는 은닉마르코프모델의 구성과 학습방법, FVQ에 대한 기본원리, 그리고 FVQ를 은닉마르코프모델의 학습과정에서 출력확률 계산과정에 반영하는 방법에 대해서 설명하고자 한다.

### 2.1 은닉마르코프모델을 이용한 음성인식

은닉마르코프모델은 음성의 전이과정을 상태들의 전이와 상태에서 심벌의 출력이라는 2가지 랜덤 프로세스로 모델링하는 방법이다. 1980년대 초반 은닉마르코프모

델을 음성인식연구에 적용한 이래 현재는 음성인식의 대표적인 방법이 되고 있다.

음성신호의 통계적인 특성을 이용한 이러한 모델링 방법은 언어모델이나 통계적 파싱(parsing)과 같은 음성인식 후처리과정과의 통합이 용이하며, 성능면에서도 다른 방법들에 비해 우수한 특성을 가지고 있어 현재 많은 음성인식시스템들이 이 방법을 적용하고 있다. 본 절에서는 실험에 사용된 이산형 은닉마르코프모델을 중심으로 설명하고, 이러한 모델이 갖는 출력확률의 평활화에 따르는 문제점에 대해서 설명하고자 한다.

### 2.2 이산형 은닉마르코프모델

은닉 마코프모델[2,7]은 상태의 의미가 미리 알려진 기존의 마코프모델과는 달리, 상태의 의미가 확률적으로 정해지며, 상태출력확률을 모델링하는 방법에 따라서 이산형과 연속형으로 나뉘어지게 된다. 본 연구에서는 이산형 HMM을 대상으로 모델링하였다. 일반적으로 이산형 은닉마르코프모델은 아래의 5가지의 구성요소들로 되어 있다.

$$\lambda = \{S, V, \pi, A, B\}$$

여기서,  $S$ 는 은닉마르코프모델을 구성하는 상태들인  $s_1, s_2, \dots, s_n$ 의 집합을 나타내며,  $V$ 는 모델에서 사용될 심벌  $v_1, v_2, \dots, v_m$ 의 집합을 나타낸다.  $\pi$ 는 상태의 초기확률을 나타내며,  $A$ 는  $\{a_{ij}; i, j \in N(S)\}$ 로 상태  $i$ 에서 상태  $j$ 간의 전이확률(transition probability)을 나타낸다.  $N(S)$ 는 상태  $S$ 의 개수를 나타낸다. 그리고  $B$ 는  $\{b(i); j \in N(S), i \in N(V)\}$ 로, 상태  $j$ 에서 심벌  $i$ 가 출현할 확률을 나타낸다. 여기서,  $N(V)$ 는 모델에서 사용되는 심벌의 개수를 나타낸다. 은닉마르코프모델의 일반적인 형태는 그림 1과 같다.

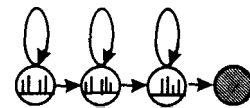


그림 1. 이산형 은닉 마르코프모델

은닉마르코프모델의 학습을 위해서 전향-후향(forward-backward)학습식이나 Viterbi학습식이 사용된다. 전향-후향방법의 경우, 입력패턴에 대해서 가능한 모든 경로들이 최대출력확률을 갖는 학습 방법이며, Viterbi학습의 경우는 입력패턴에 대한 최적경로에 대한 출력확률을 최적화하도록 학습한다. 모두 MLE(Maximum Likelihood Estimation)에 기반한 방법으로 두 방법간에 인식율 차이가 거의 없는 것으로 알려져 있다.

은닉마르코프모델을 이용한 음성인식의 경우, 위에서 정의한 모델을 이용해서 음소, 단어, 음절등의 인식단위

를 정의하여 사용하게 된다. 단어나 유절의 경우는 하나의 인식모델로 구성이 가능하나, 음소모델의 경우 단어나 음절을 위한 기본단위로 사용된다. 따라서, 단어모델이나 음절모델은 음소모델들의 결합형태로 구성된다. 본 연구에서는 문맥 독립 PLU(phone like unit)에 기반한 음소단위를 사용하여 단어모델을 구성하였으며, 이러한 단어모델에 대해서 최대출력확률을 갖는 모델을 최종인식단어로 출력한다. 아래 그림 2는 PLU들을 접속하여 단어모델을 구성하는 방법을 보여주고 있다.

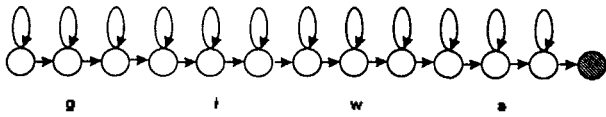


그림 2. 유사음소모델들의 접속으로 구성된 단어모델(기어)

2.3 FVQ에 기반한 은닉마르코프모델

FVQ(fuzzy vector quantization)은 퍼지클러스터링(clustering)을 위해서 사용되던 방법으로, 하나의 입력패턴에 대해서 다수의 대표패턴과의 거리를 기반으로 입력패턴과의 관계 혹은 기여도(contribution)를 구하는 방법이다. 이러한 방법은 패턴인식에서 단순히 입력패턴과 가장 가까운 하나의 대표패턴만을 선정하는 방법과는 달리, 인접한 대표패턴들을 함께 고려한다. 결과적으로, 두번째나 세번째등 부차적으로 가까운 패턴들을 함께 고려함으로써 최적의 것만을 고려한 경우의 단점을 해결하고 있으며, 성능면에서도 최적의 것만을 고려한 경우에 비해서 좋은 결과를 나타내는 것으로 알려져 있다. 이러한 방법은 벡터양자화를 위해서 사용될 뿐만 아니라, 이미 구해진 코드북을 대상으로 적용될 수 있다.

LBG나 FVQ과정을 통해서 구해진 코드북을 대상으로, 실험시나 학습시에 입력으로 들어온 입력패턴에 대해서 코드북의 모든 코드워드와의 거리를 계산한 후, 계산된 거리값들을 사용하여 퍼지목적함수(fuzzy objective function)를 최소화하는 입력패턴에 대한 각 코드워드의 최적 퍼지기여도(fuzzy contribution)를 구하게 된다. 계산과정은 다음과 같다.

먼저, 코드워드를  $v_1, v_2, \dots, v_M$ 이라고 하고, 입력벡터들의 열을  $x_1, x_2, \dots, x_T$ 라고 가정하자. 그리고, 입력패턴과 코드워드패턴간의 거리를  $d(x_i, v_j)$ 라고 하면, FVQ는 아래의 (1)과 같이 입력 벡터  $x_i$ 를 출력벡터  $o_i = (m_{i1}, m_{i2}, \dots, m_{iM})$ 로 사상한다[8].

$$m_{ij} = \left[ \sum_{k=1}^M [d(x_i, v_j)/d(x_i, v_k)]^{(F-1)} \right]^{-1} \quad (1)$$

여기서,  $F > 1$ 이며, 퍼지 정도(degree of fuzziness)를 나타낸다. 벡터  $o_i$ 는 아래 식 (2)의 퍼지목적함수를 최소화

하는 값으로 선택된다.

$$\sum_{i=1}^T \sum_{j=1}^M m_{ij}^F \times d(x_i, v_j) \quad (1)$$

2.4 FVQ를 이용한 HMM의 학습

앞서 설명한 바와 같이, 입력패턴  $x_i$ 에 대해서 구해진 퍼지기여도 벡터  $o_i$ 를 은닉마르코프모델의 출력확률을 추정하는데 사용한다. 은닉마르코프모델은 2가지 방법으로 학습이 가능하다. 첫 번째로 전향-후향방법에 기반하여 학습을 수행하는 것으로, 입력패턴에 대한 모든 가능한 경로들에서의 입력패턴열에 대한 출력확률의 값이 최대화되도록 학습한다. 다른 방법으로는 Viterbi탐색에 기반하여 입력패턴열에 대한 최적경로를 구하고, 그러한 경로에 대해서 출력확률을 최대화할 수 있도록 학습한다. 본 연구에서는 Viterbi방법에 기반하여 학습을 수행하였다. FVQ를 이용한 출력확률추정[4]은 다음과 같다.

먼저, 매 시점에서 얻어진 특징파라미터인  $x_t$ 에 대해서 코드워드에 대한 퍼지기여도 벡터인  $o_t$ 를 계산한다. 기존의 방법의 경우, 매 시점에서 하나의 최적코드워드의 출력확률을 고려하였으나 FVQ는 퍼지기여도벡터를 사용하여 상태  $i$ 에서의 출력확률을 다음과 같이 변환한다.

$$\omega_t(i) = \sum_{j=1}^M m_{ij} b_{ij} \quad (3)$$

여기서,  $\omega_t(i)$ 는 시점  $t$ 에서의 상태  $i$ 에서의 출력확률을 나타내며,  $b_{ij}$ 는 상태  $i$ 에서 심벌  $j$ 의 출력확률을 의미한다. Viterbi학습시, 각 상태에서의 최적출력확률은  $\phi_t(i)$ 라고 정의한다. 여기서,  $t$ 는  $t$ 번째 프레임을 그리고  $i$ 는  $i$ 번째의 상태를 나타낸다. 재귀적인 반복식에서 초기확률  $\phi_1(i) = \pi_i \omega_1(i)$ 이고, 나머지처리과정은 식(4)과 같으며, 입력패턴열에 대한 모델의 best 출력확률( $P^*$ )은 식(5)와 같다.

$$\phi_t(j) = \max_{1 \leq i \leq N} \{ \phi_{t-1}(i) a_{ij} | \omega_t(j) \} \text{ where, } 2 \leq t \leq T, 1 \leq j \leq N \quad (4)$$

$$P^* = \max_{1 \leq i \leq N} \phi_T(i) \quad (5)$$

III. 코드워드 의존 분포 정규화와 거리가중예 기반한 파라미터 추정

본 장에서는 기존의 FVQ를 이용한 HMM 파라미터 추정에 기반한 문제점을 중심으로 기술하고, 이러한 문제점들을 해결하기 위한 방법으로 코드워드 의존 분포정규화와 상태별 코드워드의 가중치를 적용하는 방법을 제안한다. 제안된 방법에서는 기존의 FVQ가 갖는 동일한 거리척도를 사용하는 방법과는 달리, 코드워드별 분포특성을 이용해서 코드워드에 의존적인 거리를 계산하고 이들을 이용해서 입력패턴에 대한 코드워드벡터의 기여도를 계산한다.

### 3.1 FVQ에 기반한 모델 파라미터 추정시의 문제점

앞서 설명한 바와 같이, FVQ는 입력패턴에 대한 코드워드들과의 퍼지기여도를 구하는데 사용된다. 기존의 FVQ에서 사용하는 거리는 코드워드의 분포특성을 반영하기보다는 입력패턴과 해당코드워드벡터간의 상대적인 거리에 기반하고 있다. 이러한 거리척도는 모든 코드워드에서 동일하게 적용된다. 결과적으로, 입력패턴과 대표패턴과의 동일한 거리척도가 대표패턴의 분포특성을 반영하고 있지 못하므로, 정확한 코드워드의 기여도 값을 계산하기 어렵다. 만일, 코드워드의 분산과 평균을 고려한다면 가까운 거리가 유클리디언 거리에서는 먼 것으로 될 수 있다. 즉, 각 상태에서 코드워드의 출력확률과 곱하게 될 코드워드의 기여도 값이 적어지므로, 분포특성을 반영한 거리값을 적용하는 것이 바람직하다. 본 연구에서는 코드워드 의존 거리 정규화(CDDN: codeword dependent distance Normalization)방법을 사용하여 입력패턴과 코드워드벡터 간의 거리를 코드워드의 분포특성을 반영하여 분포의존적인 거리로 변환하는 방법을 적용하고자 한다. 이와 아울러, 상태별 코드워드의 중요성을 반영하고자 한다. 기존의 FVQ의 경우, 코드워드의 거리에 대해서 동일한 가중치로 반영되고 있다. 상태에 존재하는 코드워드의 경우 그러한 코드워드마다 가중치가 다르므로, 반영되는 거리에 대한 가중이 반영될 필요가 있다. 이러한 가중치를 위해서 입력패턴에 대한 최적 코드워드와의 정규화된 거리값을 누적하고, 누적 값의 평균치에 대해서 퍼지목적함수를 사용하여 상태별 코드워드의 가중치를 유도하는 방법을 적용하고자 한다.

### 3.2 코드워드 의존 거리 정규화

본 연구에서 제안한 코드워드 의존 분포특성을 구하기 위해서 코드북의 생성시 계산된 각 코드워드의 평균과 분산값을 이용해서 입력패턴에 대한 각 코드워드의 출력확률을 계산하고, 이 값을 거리로 변경하여 FVQ에서의 입력패턴과 코드워드와의 거리값으로 변환한다. 양자화된 코드벡터의 분포특성을 나타내는 코드워드의 평균거리와 분산 식은 아래와 같다.

$$\begin{aligned} \mu_i &= \frac{1}{N_i} \sum_p^{N_i} x_p(i) \\ \sigma_i^2 &= \frac{1}{(N_i-1)} \sum_p^{N_i} (x_p(i) - \mu_i) \Sigma^{-1} (x_p(i) - \mu_i) \end{aligned} \quad (6)$$

여기서,  $i$ 는  $i$ 번째 해당되는 코드워드를 나타내며,  $N_i$ 는  $i$ 번째 코드워드에 대응되는 학습패턴의 개수를,  $x_p(i)$ 는  $i$ 번째 코드워드에 대응되는  $p$ 번째 벡터형태의 학습패턴을 나타낸다. 그리고, 공분산을 나타내는 행렬( $\Sigma^{-1}$ )은 diagonal행렬로 정의된다고 가정한다. 식 (6)에 의해서 구해진 코드워드의 분산과 평균을 사용하여 입력패턴  $x_i$ 에 대해서 코드워드 벡터와의 거리를 계산하고, 이를 입력으

로 하여 코드워드에서의 출력확률을 계산한다.

$$P(x_i, i) = \frac{(x_i - \mu_i)}{\sigma_i} \quad (7)$$

이렇게 구해진 출력확률을 거리로 변환한다. 본 연구에서는 확률값에 대한 서리변환 식으로 아래의 식 (8)을 사용하였다.

$$D(x_i, i) = -1.0 \times \log_e(P(x_i, i)) \quad (8)$$

입력패턴  $x_i$ 에 대해서 계산한 거리  $D(x_i, i)$ 는 기존의 FVQ에서 사용되던 단순한 유클리디언 거리가 아니라, 그 입력패턴이 코드워드에 속한다는 가정 하에서 그 코드워드의 평균거리와 분산을 고려해서 계산된 값이다. 식 (8)에서 계산된 코드워드별로 정규화된 거리값을 사용하여 식 (1)에서의 입력패턴에 대한 각 코드워드벡터와의 거리를 대치하여 각 코드워드의 퍼지기여도를 계산하게 된다.

### 3.3 퍼지기여도에 기반한 상태별 코드워드 가중치의 유도

기존의 FVQ에 기반한 HMM 파라미터 추정의 경우, 퍼지목적함수를 사용하여 코드워드에 대한 퍼지기여도 계산시 모든 거리에 대해서 가중치를 부여하지 않고 있다. 상태마다 코드워드의 중요성이 다르므로, 그러한 특성을 반영하기 위해서 본 연구에서는 퍼지기여도에 기반한 상태별 코드워드 가중방법을 제안한다. 이를 위해서, 각 상태별로 분할된 학습자료를 대상으로 Viterbi 학습과정에서 얻어진 모델에 대한 Viterbi 경로를 이용하여 상태별로 입력패턴에 대한 분할을 수행하고, 상태별 입력패턴에 대한 최적 코드워드별로 거리값을 누적한다. 누적행렬  $D$ 는  $M \times M$ 의 행렬을 갖는다. 여기서  $M$ 은 코드워드 개수이다. 상태에 대한 모든 입력패턴에 대해서 거리에 대한 누적행렬을 구한 뒤, 최적코드워드별로 평균을 내고( $\bar{D}$ ), 최적코드워드별 모든 코드워드의 거리값을 사용하여 퍼지목적함수를 사용하여 식 (9)과 같이 퍼지기여도를 구한다. 이렇게 구한 퍼지기여도값에 반도를 곱하여 식 (10)과 같이 최종적인 기여도를 계산하게 된다.

$$m_i(s) = \left[ \sum_{k=1}^M [D(i, j) / \bar{D}(i, k)]^{1/(K-1)} \right]^{-1} \quad (9)$$

$$m_i^*(s) = \sum_{j=1}^M [P(j) \times m_{ji}(s)] \quad (10)$$

여기서,  $m_i(s)$ 는 상태  $s$ 에서  $i$ 번째 코드워드를 최적으로 하는  $j$ 번째 코드워드의 퍼지기여도 값을 나타내며,  $m_i^*(s)$ 는  $i$ 번째 코드워드에 대한 최종적인 기여도이다. 이 값은 거리에 대한 가중치로 사용될수 있으며, 상태별 출력확률을 대치하는 평활화된 상태별 출력확률로도 사용가능하다. 상태별 코드워드에 대한 가중치의 경우  $m_i^*(s)$ 의 정

규화된 값인  $\bar{m}_i^*(s)$  값을 사용하여, 식 (11)과 같은 최종적인 거리값을 계산하게 된다.

$$\bar{D}(i, s) = D(i, s) \times \bar{m}_i^*(s) \quad (11)$$

식 (11)에서 얻어진 새로운 코드워드에 대한 거리값을 식(1)에 적용하여 입력패턴에 대한 코드워드별 퍼지기여도를 구하며, 이 값을 이용해서 최종적인 상태별 출력 확률값을 계산하게 된다. 상태별 출력과정은 2.4절에서 기술한 처리과정과 동일하다.

식 (10)의 결과를 해당 상태에서의 평활화된 출력확률로 사용하는 경우,  $m_i^*(s)$ 를 상태에서의 코드워드에 대한 출력확률로 대치하여 사용한다. 이러한 경우에는 식(11)을 사용하지 않고, 식(8)의 정규화된 거리값만을 사용하여 코드워드에 대한 가중치를 계산한다.

#### IV. 실험 및 결과

본 절에서는 제안된 방법과 기존의 FVQ기반 방법을 적용한 비교실험을 위한 환경과 실험결과를 기술하고자 한다. 먼저, 인식에 사용된 단어와 인식단위, 특징파라미터와 이산형 은닉마르코프모델의 구성에 대해서 설명하고, 다음으로 인식실험결과를 기술하고자 한다.

##### 4.1 실험환경

본 연구에서 제안한 방법의 성능 평가를 위해서 각기 다른 남자 17명의 화자가 2회씩 발성한 115개의 단어\*<sup>1)</sup>(총 3910개)를 대상으로 실험하였다. 학습을 위해서 12명이 2회씩 발성한 자료를 사용하였으며, 실험을 위해서 나머지 5명이 2회씩 발성한 자료를 사용하였다. 실험에 사용한 특징 파라미터는 14차 멜켵스트림 계수, 14차 델타 멜켵스트림 계수와 (energy + delta-energy)를 함께 사용하였다. 멜켵스트림 계수는 선형 예측 계수를 구한 뒤 Oppenheim이 제안한 양선형 변환(bilinear transformation)[9]을 통해 구하였으며, 주파수의 warping을 위한 변수값은 0.64를 사용하였다. 델타 파라미터는 차분 멜켵스트림(differenced mel-cepstrum)과 차분 에너지(differenced energy)[10]를 사용하였으며, 차분창(difference window)의 크기는 5로 정하였다.

입력 파라미터의 벡터 양자화를 위해서 일반적으로 널리 사용하는 LBG[11]방법을 사용하였다. 멜켵스트림과 델타 멜켵스트림의 경우 각각 256의 코드워드를 사용하였으며, 에너지와 델타 에너지의 경우는 64의 코드워드를 갖는 단일 코드북을 사용하였다. 코드북의 종료 조건은  $1.0e-6$ 으로 정하였으며, HMM의 각 상태에서 관측되지 못한 심벌의 출력 확률값으로  $1.0e-5$ 를 사용하였다. 그리고, 제안된 방법에서의 퍼지정도는 다수의 실험을 통해서 가장 좋은 성능을 보인 0.5로 정하였다. 특징 파라미터 추출을 위한 분석 조건은 표 1과 같으며, 실험에 사용된 115단어 집합은 표 2와 같다.

인식 시스템은 크게 학습단계와 실험단계로 나뉘어지며, 학습단계는 다시 코드워드 작성부와 모델 학습부의 두부분으로 구성되어 있다. 코드워드 작성부의 경우 학습 데이터를 대상으로 특징 파라미터를 추출하고 LBG 방법에 따라 코드북을 작성하는 부분이다. 모델 학습부에서는 문맥독립 유사음소모델을 접속하여 단어모델을 구성하고 단어별 학습자료를 이용해서 유사음소모델을 학습하게 된다. 유사음소모델의 학습을 위해서 Segmental K-means방법[12]을 사용하였다. 학습과정을 살펴보면, 단어별 음성 자료로부터 특징을 추출한 뒤, 이를 양자화하여 코드워드의 열로 변환한다. 변환된 파라미터별 코드워드열을 사용하여 유사음소모델 파라미터의 학습이 이루어지며, 종료조건이 만족될때까지 학습과정을 반복하게 된다. 일단 모델의 학습이 이루어지면, 실험 데이터에 대한 성능 평가를 위해서 학습과 동일한 과정으로 입력 패턴을 코드 열로 변환하고 각 단어모델별 입력 패턴의 출력 확률을 계산 후, 이들 중 가장 큰 것을 해당 단어로 선정하여 출력한다.

##### 4.2 실험결과

제안된 방법의 성능평가를 위해서 기존의 FVQ에 기반한 방법과 제안된 코드워드 의존 분포정규화를 적용한 방법 그리고 거리가중방법을 부가한 실험을 수행하였다. 실험결과는 아래의 그림 3과 같다.

기존의 FVQ를 적용하여 출력확률을 계산한 경우, 94.3%의 단어 인식율을 보였으며, CDDN을 적용한 경우에는 FVQ에 비해서 1.4%의 성능향상을 보여주었다. 이러한 결과로부터, 상태의존적인 특성을 반영하는 것이 유클리디언 거리를 사용하는 것보다 바람직하다는 것을 알

\*ETRI에서 제공된 자료입니다.

표 1. 실험 환경

sampling rate	16KHz	pre-emphasis	$1.0 - 0.97Z^{-1}$
analysis	20 msec (320 points)	shift interval	10 msec (160 points)
feature parameters	mel-cepstrum(14) delta-mel-cepstrum(14) energy+delta-energy(2)	codebook	mel-cepstrum(256) delta-mel-cepstrum(256) energy+delta-energy(6)

표 2. 실험 자료 (115 단어)

가운데	가위바위보	계 획 표	고 유 하 다	공 예 품	교 과 서	교 류	교 육
교 회	귀 중 하 다	규 도	규 칙	기 억 분 수	기 외	결 대 기	꽃 벌
나 그 네	나 보 침	높 이 위 기	누 구 누 구	늪 우 차 다	다 량 주	돈 키 호 테	들 쇠
뿔 러 았 다	똥 들 아 하 다	뜨 령 하 다	뛰 어 넘 다	마 음 것	마 음 씨	우 냉 왕 흥	우 님
무 역	인 주 주의	인 요	발 칭 대	발 자 취	방 위 관	보 비	보 장 되 다
부 채 골	아 주 마 니	빠 꾸 기	빠 꾀	쓰 임 세	뻘 았 다	아 직 도	애 벌 리
예 족	예 금 액	예 보	완 두 공	울 츠	외 국 기	외 상 손	외 양 간
의 견	의 논 하 다	의 례	의 식 주	의 원	의 제	사 과 나무	사 영
석 회 수	어 연 번 다	센 터 미 터	센 어 림	숫 다	선 원	애 니 멜 션	여 모 시 오
여 집 합	확 립 하 다	원 앞 서 기	원 족	요 즘 음	용 존	용 역	원 그 래 프
원 유	인 도 네 시 아	위 로 하 다	위 족	이 웃 집	자 료 신	정 옥 면 체	좌 표 평 면
죄 송 하 다	직 속 면 체	참 외	채 색 화	치 료 받 다	테 이 프	뽀 리 바 전	토 마 도
토 마 스	최 대 공 역 수	토 요 일	특 별 히	페 인 트	페 프	하 알 디	허 여 금
해 수 욕 장	쳐 들 어 오 디	헛 빛	해 업	해 맥	여 학 생	효 과 적	훈 민 정 음
휴 지	이 산 화 탄 소	이 동 변 삼 각 형					

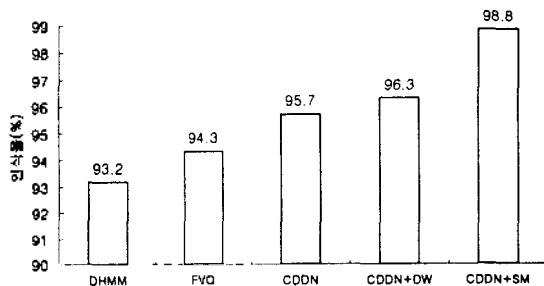


그림 3. 제안된 방법의 성능평가

수 있었다.

상태별 코드워드의 가중치를 적용한 “CDDN+DW(distance weighting)”의 경우, 유도된 가중치를 입력패턴에 대한 각 코드워드와의 정규화된 거리에 직접 반영하는 경우에 96.3%의 인식율을 보였다. 단순히 CDDN을 사용한 경우에 비해서 0.6% 정도의 인식율 향상이 있었다. 성능 향상이 크지 못한 이유는 제안된 거리 가중방법이 학습자료의 부족으로 인해 출력확률에 직접영향을 미치지 못하기 때문인 것으로 판단된다. 따라서, 이러한 거리에 기반한 가중치를 출력확률의 평활화를 위해서 사용한 “CDDN+SM(smoothing)” 경우 98.8%의 인식율을 보임으로써 기존의 FVQ방법에 비해서 4.5%의 성능향상이 있음을 알 수 있었다.

## V. 결 론

본 연구에서는 이산형 은닉마르코프모델의 학습시 견고한 상태별 출력확률을 얻기 위한 방법으로 CDDN(code-word dependent distance Normalization)과 거리에 기반한 코드워드 가중방법을 제안하였다. 기존의 FVQ에 기반한 HMM 파라미터추정에서 사용되는 방법과는 달리, 코드워드의 분포특성과 상태 의존적인 코드워드의 가중치를 사용하여 입력패턴에 대한 각 코드워드별 거리를 계산하고, 이 거리값에 퍼지목적함수를 적용하여 코드워드별 기여도를 계산하였다. 계산된 코드워드별 퍼지기여도와 상태별 코드워드의 출현빈도인 출력확률을 선형결합하여 한 상태에서의 최종적인 출력결과를 계산하게 된다.

실험결과, 제안된 방법이 기존의 FVQ기반한 방법에 비해서 4.5% 정도의 인식율 향상이 있음을 할 수 있었다. 특히, 거리 가중치를 사용하여 출력확률 평활화를 적용한 경우가 단순히 코드워드별 가중을 적용한 경우에 비해서 2.5%의 성능향상을 보였다.

추후로는 연속형이나 준연속형 은닉마르코프모델과의 성능비교실험과 연속음성을 대상으로한 실험을 통해서 제안된 방법의 유효성을 보이는 연구가 필요할 것으로 판단된다.

## 참 고 문 헌

1. L. R. Rabiner, B. H. Juang, Fundamentals of Speech Recognition, Prentice-Hall, 1993.

2. L. R. Rabiner, B. H. Juang. "An Introduction to Hidden Markov Model", IEEE ASSP Magazine, No. 4, pp. 4~16, 1986.
3. 홍형진, 최승호, 이기철, 김기철, 이황수, "한국어 숫자음 인식을 위한 이산분포 HMM과 연속분포 HMM의 성능비교", 제 11회 음성통신 및 신호처리 워크샵, 연세대학교, Vol. 11, No. 1, pp. 157~160, 10월 1994.
4. H. P. Tseng, M. J. Sabin, et al, "Fuzzy Vector Quantization applied to hidden Markov modeling", Proc. IEEE ICASSP87, Dalas, Texas, pp. 15.5.1~15.5.4, April 1987.
5. J. M. Koo and C. K. Un, "Fuzzy smoothing of HMM parameters in speech recognition", Electronic Letters, Vol. 26, pp. 743~744, 1990.
6. J. DAI, "Robust Estimation of HMM Parameters Using Fuzzy Vector Quantization and Parzen's Window", Pattern Recognition, Vol. 28, No. 1, pp. 53~57, 1995.
7. K. F. Lee, H. W. Hon, "Speaker-Independent Phoneme Recognition Using Hidden Markov Models", Proc. IEEE ICASSP 88, New York, pp. 123~126, April 1988.
8. A. Kandel, Fuzzy Techniques in Pattern Recognition, John Wiley & Sons, Inc., 1982.
9. A. V. Oppenheim, D. H. Johnson, "Discrete Representation of Signals", The Proceedings of the IEEE, Vol. 60, No. 6, pp. 681~691, June 1972.
10. K. Shikano, "Evaluation of LPC Spectral Matching Measures for Phonetic Unit Recognition", Technical Reports, Computer Science Department, Carnegie Mellon University, 1986.
11. Y. Linda, A. Buzo, R. M. Gray, "An Algorithm for Vector Quantization Design", IEEE Trans. on Communication, Vol. 28, No. 1, pp. 84~95, Jan. 1980.
12. L. R. Rabiner, J. G. Wilpon, et al., "A Segmental K-means Training Procedure for Connected Word Recognition Based on Whole Word Reference Patterns", AT&T Technical Journal, Vol. 65, No. 3, pp. 21~31, 1986.

▲최 환 진

한국음향학회 제15권 제1호(1996), pp.61~70 참조

▲오 영 환

한국음향학회 제15권 제1호(1996), pp.61~70 참조