

상태의존 소스 양자화에 기반한 음성인식을 위한 은닉 마르코프 모델 파라미터의 견고한 추정

Robust estimation of HMM parameters Based on the State-Dependent Source-Quantization for Speech Recognition

최 환 진*, 박 재 득*
(Hwan Jin Choi*, Jay Duke Park*)

요 약

최근 음성인식을 위한 대표적인 방법으로 은닉 마르코프 모델이 사용되고 있으며, 이러한 방법은 음성의 특성을 잘 표현하도록 하는 음향적인 모델링 방법에 따라서 성능이 좌우된다. 본 논문에서는 상태에서의 출력확률을 견고히 추정하기 위한 방법으로 상태에서의 출력확률들 소스들의 분포와 그들의 빈도로 가중한 출력분포로 표시하는 상태 의존 소스 양자화 모델링 방법을 제안한다. 이 방법은 한 상태 내에서 특징 파라미터들이 유사한 특성을 가지며, 그들의 변이가 다른 상태에 있는 특징 파라미터들에 비해서 작다는 사실에 기반한다. 실험결과에 의하면, 제안된 방법이 기존의 baseline 시스템보다 단어 인식율의 경우는 2.7%, 문장 인식율의 경우 3.6%의 향상을 보였다. 이러한 결과로부터 제안된 SDSQ-DHMM 이 인식율 향상면에서 유효하며, HMM에 있어서 상태별 출력확률의 견고한 추정을 위한 대안으로 사용될 수 있을 것으로 판단된다.

ABSTRACT

Recently, hidden Markov model(HMM) has become the predominant approach to speech recognition and its performance depends on the acoustic modeling techniques adopted to represent the characteristics of speech.

In this paper, we proposed the state-dependent source-quantization(SDSQ) discrete hidden Markov model in which robust output probabilities are estimated as the sum of product of the frequency of a state-dependent source and its own codeword distribution on the facts that feature vectors within the same state have similar characteristics in their acoustical aspects and their variations are lower than that of other states.

The experimental result has shown that the proposed method improved word accuracy of 2.7% and sentence accuracy of 3.6% over the baseline system. These results indicated that the SDSQ-DHMM leads to improve the recognition rate and therefore it may be used as an alternative to the robust estimation of output probabilities for HMMs.

I. 서 론

음성인식을 위한 다양한 방법들이 현재 연구되고 있으며, 그 중에서 몇몇 방법들은 우수한 인식율을 나타내는 것으로 알려져 있다. 이들 중에서 은닉 마르코프 모델(HMM: hidden Markov model)에 기반한 방법[1, 2, 3]은 다른 방법들에 비해서 향상된 인식율을 나타내고 있으며, 음성 인식결과의 후처리를 위해서 사용되는 언어 및 의미적인 정보들과도 손쉽게 확률적으로 통합될 수 있는 장점들로 인해 근간에 들어서 음성인식을 위해서 널리 사용되고 있다. 은닉 마르코프 모델은 음성의 시간적인

변화특성을 몇 개의 상태들(states)로 나누고, 각 상태들 간의 천이(transition)와 상태 내에서 음성신호의 출현을 확률적으로 모델링하고 있다. 상태를 모델링하는 방법이나 모델의 파라미터들을 학습하는 방법에 따라 다양한 형태의 모델링이 가능하다.

은닉 마르코프 모델에서 상태로부터 특징벡터의 출력 확률을 연속밀도 함수로 표현하는가, 혹은 이산밀도 함수로 표현하는가에 따라서 연속형 HMM과 이산형 HMM으로 나뉘어진다. 연속형 HMM은 상태에서 출력확률을 정규화 분포(normal distribution)의 함수 형태로 모델링하는데, 이때 사용되는 함수의 파라미터들의 부적절한 추정은 인식율을 저하시키며, 인식을 위해서 많은 계산량을 필요로 한다. 이에 반해서, 이산형 HMM은 빠른 연산 속도와 불충분한 학습자료로도 어느 정도의 인식 성능을

*시스템공학연구소 자연어정보처리연구부
접수일자: 1997년 11월 5일

언을 수 있는 장점이 있다. 그러나, 특징 파라미터의 양자화에 따른 양자화 오류가 발생하며, 아울러 실험 시에 학습되지 않은 자료들에 대해서 낮은 확률을 출력함으로써 성능이 저하되는 단점이 있다.

이러한 상태별 출력확률을 모델링하는 방법 이외에 은닉 마르코프 모델을 학습자료에 대해서 모델의 출력확률을 최대화하는 형식으로 학습할 것인가, 아니면 인식할 모델 전체와의 관계를 고려화해서 학습할 것인가에 따라 ML(maximum likelihood)나 MME(minimum mutual information expectation), MCE(minimum classification error), GPD(generalized probabilistic descent) 그리고 CT(corrective training) 등의 다양한 학습방법들이 사용된다 [4, 5, 6].

이상에서 설명하였듯이, 은닉 마르코프 모델은 상태 내에 출력확률을 추정하는 방법, 은닉 마르코프 모델의 학습 방법, 그리고 모델간의 변별력을 향상시키는 방법에 따라서 다양한 모델이 가능하지만, 실제적인 모델의 성능은 은닉 마르코프 모델을 구성하는 상태의 수, 상태간의 전이확률 및 상태 내에서의 심벌의 출력확률에 의존하게 된다. 상태 수는 시간적인 지속 시간을 모델링하는데 사용되는 것으로, 너무 적은 상태 수는 변별력있는 구간들을 하나의 상태 지속시간 안에 모델링함으로써 모델의 변별력을 저하시키고, 너무 많은 수는 모델의 지속시간을 지나치게 연장시키는 결과를 가져옴으로써 적절한 상태수를 정의하는 것이 필요하다[7]. 일반적으로 상태수는 인식 실험결과에 의존해서 결정되며, 일반적으로 음소의 경우에는 세개의 상태들-초기(start), 중앙(middle), 끝(end)-들로 모델링하거나 음소의 다양한 시간길이의 변이를 흡수할 수 있는 형태의 모델이 사용된다[2].

두 번째로, 상태와 상태간의 전이확률은 한 상태에 머무를 가능성과 지금의 상태에서 다른 상태로 전이할 확률로 나뉘어진다. 이러한 전이확률은 지속시간과 밀접한 관련이 있으며, 부적절한 모델링에 의한 성능저하가 상태 내에서 심벌의 출력확률에 대한 부적절한 모델링에 비해서 매우 적은 경향을 나타낸다.

마지막으로, 상태에서 심벌의 출력확률은 임의의 시점에서 음성신호에 대응되는 특징 파라미터가 해당 상태에서 출현할 확률을 나타내는 것으로, 입력 신호에 대한 인식모델의 유사도에 직접적으로 큰 영향을 미치는 요소이다. 음성인식 방법의 일종인 동적 정합법이 참조패턴과 입력패턴간의 특징 파라미터간의 거리에 기반하여 정합도를 계산하는 방식이라면, 은닉 마르코프 모델은 입력패턴의 양자화된 코드워드가 해당상태에서 얼마만큼의 반도로 출현하는 가를 나타내는 확률값을 정합도로서 사용한다는 점을 제외하고는 동일하다. 따라서, 실제적으로 정합과 관련된 은닉 마르코프 모델에서의 상태별 심벌의 출력 확률값의 부적절한 추정은 인식율의 저하를 가져옴으로, 은닉 마르코프 모델의 성능을 향상시키기 위해서 보다 견고한 상태별 출력확률의 추정이 중요하게 된다.

본 논문에서는 이산형 은닉 마르코프 모델에서 상태별 심벌의 출력확률을 추정하기 위한 방법으로, 상태 의존 소스 양자화(SDSQ: state-dependent source quantization) HMM 방법을 제안하고자 한다. 이 방법을 통해서 상태에서 출력 심벌의 출력분포가 기존의 학습식에 의해서 얻어진 출력심벌의 분포에 비해 보다 분별력을 가지며, 인식 성능면에서도 보다 향상된 결과를 나타내는 것을 알 수 있었다.

본 논문의 구성은 다음과 같다. 1 장에서는 상태별 출력확률 추정의 중요성에 대해서 설명하고, 2장에서는 상태별 출력확률의 추정과 관련된 기존의 연구방법들과 이들이 갖는 문제점들에 대해서 기술하고자 한다. 3장에서는 본 연구에서 제안한 상태 의존 소스 양자화방법에 대해서 설명하고, 4장에서는 실험 및 결과에 대한 검토 내용에 대해서 기술하고, 마지막 5장에서 이장의 결론을 맺고자 한다.

II. 기존의 연구방법들과 문제점

이 절에서는 이산형 은닉 마르코프 모델에서 상태별 출력확률을 모델링하는 기존의 방법들에 대해서 설명하고자 한다. 상태별 출력확률에 영향을 미치는 요소들과 기존의 방법이 가지는 문제점들에 대해서 기술하고자 한다.

2.1 기존의 연구방법들

은닉 마르코프 모델에서 상태별 출력확률의 추정은 상태에서 출력확률을 어떻게 모델링하는가, 은닉 마르코프 모델의 파라미터를 어떻게 학습하는가, 그리고 추정된 인식 모델의 출력확률 분포간에 혼동율을 줄이기 위해서 어떻게 변별력을 향상시킬 것인가에 따라 다양한 형태의 모델링이 가능하게 된다. 이러한 은닉 마르코프 모델링과 관련된 주요 측면들을 나열하면 다음과 같다. 먼저, 은닉 마르코프 모델에서 상태별 출력확률을 모델링하는 방법에 대해서 알아보자. 은닉 마르코프 모델은 모델을 구성하는 상태를 어떻게 모델링하는 방법에 따라서 연속형 HMM[8], 이산형 HMM[2], 그리고 반연속형 HMM[9] 등으로 나뉜다. 이산형과 연속형의 구분은 상태를 표현하는 특징벡터의 공간을 어떠한 방식으로 표현하는가에 의존한다. 이산형 HMM의 경우는 상태를 나타내는 특징벡터의 공간을 유한한 개수의 양자화된 공간으로 분할하고, 분할된 각각의 부공간(subspace)에 대해서 평균적인 특징벡터를 계산하고 그러한 평균적인 특징벡터가 해당 부공간을 대표하도록 한다. 이러한 부공간을 대표하는 평균적인 특징벡터를 유도하기 위해서 이산형 HMM에서는 벡터 양자화(VQ: vector quantization)과정[10, 11]이 수행된다. 입력된 음성신호에 대한 인식 모델과의 정합도는 입력 음성신호에 대한 코드북을 적용하여 얻어진 특징 파라미터의 코드워드 색인열이 대상 은닉 마르코프 인식모델에서 출력될 확률의 형태로 표현되며, 최고의

출력확률을 갖는 인식모델이 최종적인 인식결과로써 인식되게 된다. 연속형 HMM은 앞서 설명한 이산형 HMM에서 상태에서의 출력확률의 추정을 위해서 연속함수를 사용하여 추정한다는 점을 제외하고는 동일하다. 상태 내에서 특정 파라미터의 공간을 연속 함수들의 형태로 표현함으로써, 이산형 HMM에서와 같이 코드북을 이용한 특정 파라미터의 양자화에 따른 오류가 발생되지 않으므로 보다 견고한 상태별 출력확률을 구하게 된다. 그러나, 연속형 HMM의 경우는 인식을 위해서 많은 계산시간을 요하며, 아울러 상태의 특정 파라미터 부공간을 모델링하기 위해서 사용되는 자료가 불충분한 경우 부적절하게 추정된 모델 파라미터는 인식 성능을 크게 저하시키는 단점이 있다.

두 번째로, 은닉 마르코프 모델의 파라미터들의 학습방법에 따른 분류를 살펴보자. 은닉 마르코프 모델의 학습을 위해서 전향-후향(forward-backward) 학습방법[2]과 Viterbi 학습방법[13]이 일반적으로 사용된다. 이들 방법은 모두 ML(maximum likelihood)방법의 일종으로, 인식모델의 파라미터를 학습을 위한 자료들에 대해서 최대의 출력값을 갖도록 학습하는 방법이다. 성능 면에서 전향-후향방법이나 Viterbi 방법이 실제적인 인식 실험에서 서로 유사한 성능을 나타내는 것으로 보고되고 있다[14].

마지막으로, 인식모델간의 혼동율을 줄이기 위해서 다양한 변별력 학습방법들이 사용되고 있다. 앞서 설명한 최대우도에 기반한 학습의 경우 인식 모델별로 학습자료를 분할하고, 해당 학습자료에 대해서 인식모델의 출력확률이 최대화되도록 학습하게 된다. 이 방법의 단점은 임의의 한 인식모델의 학습에 사용된 자료에 대해서 다른 인식모델의 출력결과가 높게 나타날 수 있다는 점이다. 즉, 주어진 자료에 대응되는 인식모델의 출력확률이 다른 인식모델의 출력확률보다 낮게 나타남으로써, 오인식이 될 수 있는 가능성이 있다는 의미이다. 따라서, 혼동 가능한 모델에서 해당 심벌의 출력확률을 낮춤으로써 인식율을 향상시키는 방법이 필요하다. 이러한 방법들로는 오류보정 방법의 일종인 CT(corrective training)나 LVQ(learning vector quantization)에 기반한 방법, 오류함수에 기반한 GPD(generalized probabilistic descent)나 MCE(minimum classification error)등의 방법들이 사용된다. 이러한 과정을 통해서 변별력을 가진 모델의 상태별 출력 확률 분포는 최초의 학습 자료로부터 얻어진 출력 확률 분포와는 다른 형태를 가지며, 실제적인 인식실험에서 성능 향상을 가져오는 것으로 알려졌다.

본 연구에서는 이러한 다양한 방법들에 의해서 영향을 받는 상태별 출력확률의 추정을 위해서, Viterbi 학습방법에 기반한 이산형 은닉 마르코프 모델의 상태별 출력확률을 추정하는 방법을 중심으로, 개선된 상태별 출력확률의 추정방법인 상태의존 소스 양자화(SDSQ: state-dependent source-quantization)방법을 제안하고, 이 방법을 기존의 이산형 HMM의 학습식에 적용하고자 한다. 기존

의 상태별 출력확률 추정과 관련된 문제점들에 대해서 아래의 2.2절에서 기술하고자 한다.

2.2 문제점들

은닉 마르코프 모델을 구성하는 두 개의 통계적인 프로세스 중에서 상태간 전이 프로세스보다는 상태별 심벌의 출력분포를 추정하는 프로세스가 인식율에 커다란 영향을 미친다. 즉, 은닉 마르코프 모델을 이용한 음성의 모델링에 있어서, 음성신호의 특성을 보다 잘 표현하는 것이 신호의 변화특성보다는 중요함을 나타낸다. 기존의 방법들이 갖는 상태 모델링의 문제점들에 대해서 알아보도록 하자.

은닉 마르코프 모델을 구성하는 파라미터들을 견고히 추정하는 것은 실제적인 인식성능과 밀접한 연관관계를 가지고 있으며 매우 중요하다. 앞서 설명하였듯이, 추정된 은닉 마르코프 모델의 파라미터들 중에서 상태별 전이관계를 나타내는 확률보다는 임의의 시점에서 입력된 특정 파라미터가 상태 내에서 출현할 확률의 추정이 인식율에 보다 영향을 미치게 된다. 임의의 시점에서 입력된 특정 벡터가 하나의 상태에서 출현할 가능성이 상태별 출력확률의 형태로 은닉 마르코프 모델에서 표현되기 때문이다. 상태별 심벌의 출력확률 모델링하는 방법은 학습방법에 따라 달라지게 된다. 일반적으로, 인식 단위로 수집된 학습자료에 대해서 인식단위 모델의 해당 학습자료에 대한 출력확률을 최대화하도록 학습하는 최대우도(maximum likelihood)에 의한 학습방법이 널리 사용되는데, 이 방법의 단점은 다른 모델의 학습을 위해서 사용되는 학습자료에 대해서도 인식모델이 최고의 출력확률을 갖는다는 보장이 없다. 이 문제를 해결하기 위해서, 인식모델간의 변별력을 향상시키기 위한 다양한 방법들이 사용되고 있다.

본 연구에서는 최대우도에 기반한 인식방법 중에서 Viterbi 학습방법에 의해서 학습된 상태별 출력확률분포의 방법의 문제점을 개선하여 보다 견고한 상태별 출력확률을 추정하는 방법을 제안하고자 한다. 일반적으로 Viterbi 방법에 의해서 학습된 출력확률의 분포는 상태단위로 모델링된다. 은닉 마르코프 모델에서 상태는 음성 신호 상에서 유사한 특성을 가지는 시간구간을 모델링하는 단위로, 한 상태 내에서 신호의 변화특성은 다른 상태에서의 신호변이보다는 작은 특성을 가지고 있다. 하나의 상태 내에서의 신호의 변화특성을 기존의 방법의 경우는 상태단위로 모델링하는데, 이러한 방법에는 다음과 같은 두 가지 문제점이 있다.

첫 번째로, 상태 내에서 출력확률이 해당 상태 내에서의 코드워드의 빈도에 대한 정규화된 값으로 표시된다. 이러한 경우, 출력확률은 지속길이가 긴 코드워드의 빈도에 의존하게 됨으로써, 여러 학습자료에서 빈도가 높은 코드워드에 비해서 높은 확률을 가질 수 있게 된다. 실제적으로, 적은 빈도로 출현하면서 긴 지속길이를 갖으며

로써 높은 출력확률을 갖는 코드워드에 비해서 여러 학습자료에 걸쳐 나오는 높은 빈도의 코드워드가 보다 가중되어야 하는 것이 바람직하다.

두 번째로, 하나의 상태를 대표하는 대표패턴이 있다고 가정하면, 그러한 대표패턴은 해당 상태에서의 코드워드 벡터들과 그들의 존재 가능성을 나타내는 확률의 곱에 대한 합의 형태로 표시할 수 있다. 이러한 가정 하에서 상태에서의 코드워드 색인의 출력 확률분포는 하나의 상태에 대응되는 대표패턴을 표시하기 위해서 각 코드워드가 하나의 상태에서 존재할 가능성을 나타낸다고 할 수 있다. 이러한 관계를 이용해서, 상태를 나타내는 학습자료상에서의 대표패턴들과 이에 대응되는 코드워드 색인의 분포들을 이용해서 보다 분별력을 갖는 상태별 코드워드의 출력 확률분포를 추정하는 것이 바람직하다.

앞서 기술한 문제점들을 해결하기 위해서, 본 연구에서는 인식모델의 상태별로 수집된 자료에 대해서 상태를 대표하는 평균적인 벡터를 구하고, 이들의 빈도와 각각의 대표벡터에 대응되는 코드워드 색인의 분포특성을 이용해서 상태별 출력확률을 추정하는 상태 의존 소스 양자화(SDSQ: state-dependent source-quantization)방법에 기반한 상태별 출력확률의 추정방법을 제안하고자 한다. 상세한 내용은 다음의 3장에서 기술하고자 한다.

III. 상태의존 소스 양자화 모델링

이 장에서는 본 연구에서 제안한 견고한 상태별 출력확률을 추정하기 위한 방법인 상태 의존 소스 양자화에 따른 상태별 출력확률 추정방법을 기술하고, 아울러 이러한 추정방법을 기존의 HMM의 학습식에 적용한 SDSQ-DHMM의 유도과정에 대해서 설명하고자 한다.

3.1 상태 의존 소스(SDS : State Dependent Source)의 정의

상태 의존 소스의 유도에 앞서서, 상태의존 소스에 대해서 정의하고자 한다. 본 논문에서 상태 의존 소스(SDS)란 상태수준의 평균 특징벡터와 그의 빈도, 그리고 대응되는 코드워드의 분포를 말한다.

SDS = (평균 특징 벡터, 빈도, 코드워드의 분포)

이러한 정의는 하나의 상태에는 그 상태를 표시할 수 있는 유일한 코드워드의 분포가 있고, 그러한 코드워드의 분포로써 대표되는 특징벡터가 있다는 사실에 기반한다. 상태별로 대표하는 특징벡터와 그에 대응되는 코드워드의 분포를 실제적으로 추정할 수 없으므로, 학습 자료로부터 얻어진 상태의존 소스들의 선형결합형태로써 추정하고자 한다.

상태 의존 소스들을 얻기 위해서, 학습 자료는 우선 인식모델의 상태단위로 정렬된다. 이를 위해서 학습문장에

대응되는 triphone모델들을 접속하고, 이렇게 해서 만들어진 문장 단위모델에 학습 자료를 입력으로 해서 segmental K-means방법[12]을 적용하여 인식모델인 triphone 모델의 상태별로 특징 파라미터의 코드워드 색인열을 정렬한다. Triphone모델의 상태별로 정렬된 코드워드 색인열과 특징 파라미터 열을 사용하여, 한 상태를 소스벡터와 그에 대응되는 소스 분포형태로 표현하게 된다.

소스벡터(SV: source vector)는 상태에서 정렬된 특징 파라미터들의 평균벡터로 정의되며, 이러한 평균벡터가 상태의 특성을 대표한다고 가정하고 이 벡터를 상태의존 소스벡터(SDSV: state-dependent source vector)라고 부르며, 그 유도식은 아래의 식과 같다.

$$\vec{c}_{s,i} = \frac{1}{T_s(i)} \sum_{t=1}^{T_s(i)} \vec{x}_{s,i}(t) \quad (1)$$

여기서, i 는 상태 s 로 정렬된 특징 벡터열들 가운데에서 i 번째 해당되는 특징 벡터열을 나타내며, $\vec{x}_{s,i}(t)$ 는 상태 s 에서의 i 번째 특징 벡터열에서 t 번째 특징벡터를 표시하며, $T_s(i)$ 는 i 번째 특징 벡터열의 프레임 길이를, 그리고 $\vec{c}_{s,i}$ 는 상태 s 에서 i 번째 특징 벡터열에 대한 평균벡터로, SDSV를 나타낸다. 구해진 SDSV의 빈도를 추정하기 위해서, SDSV는 해당 특징파라미터의 코드북을 이용해서 양자화되며, 대응되는 코드워드의 빈도를 상태 s 에서의 SDSV 빈도로서 사용한다. 하나의 SDSV에 대응되는 코드워드의 색인을 구하는 식은 아래와 같다.

$$r_{s,i} = f(\vec{c}_{s,i}), \quad r_{s,i} \in M \quad (2)$$

여기서, $r_{s,i}$ 은 상태 s 에서 i 번째 특징 벡터열에 대응되는 SDSV의 양자화된 코드워드 색인이며, $f(\cdot)$ 는 코드북 사상함수(mapping function)이다. 코드북 사상함수는 입력된 특징벡터와 해당 코드북내의 모든 코드워드들간의 Euclidean 거리를 계산해서, 그 중에서 가장 작은 거리를 가진 코드워드의 색인값을 출력하게 된다. 이러한 SDSV에 대응되는 해당 코드북의 코드워드 색인을 구하는 것과 마찬가지로, 상태 s 에서 i 번째 특징 벡터열에 속하는 각각의 특징벡터에 대한 코드워드 사상함수는 아래와 같다.

$$o_{s,i}(t) = f(\vec{x}_{s,i}(t)), \quad o_{s,i}(t) \in M \quad (3)$$

여기서, $o_{s,i}(t)$ 는 특징벡터 $\vec{x}_{s,i}(t)$ 의 양자화된 코드워드 색인을 나타낸다. 상태별로 정렬된 특징 벡터열에 대한 SDSV와 특징벡터에 대한 코드워드사상과는 그림 1과 같다.

다음으로, 각 상태별로 소스의존 코드워드 분포를 구하기 위해서, 구해진 특징 벡터열에 대응되는 SDSV별로 해당 특징 벡터열에 대응되는 코드워드 색인열을 이용해서 SDSV별 코드워드의 빈도를 계산하고 누적한다. 조건부 확률 $\delta_s(i|j)$ 은 상태 s 에서 상태의존 소스벡터의 양자

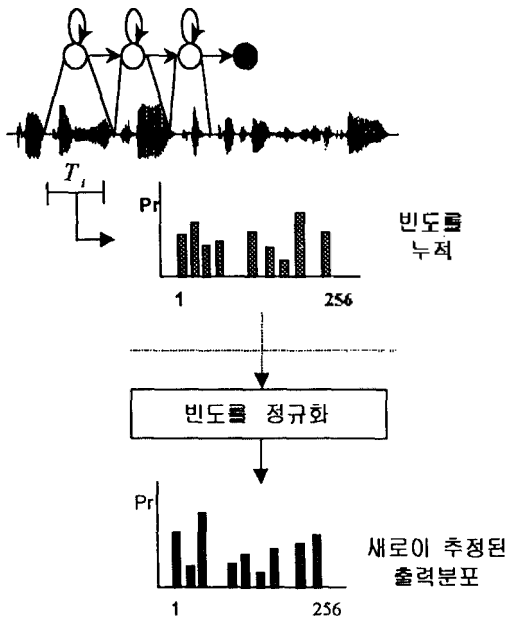


그림 1. 상태 s에서 SDSV와 특징 파라미터들

화된 색인이 j 라고 가정하였을 경우, 특징벡터의 양자화된 코드워드가 i 의 정규화된 빈도를 나타내며, 그 유도식은 아래와 같다.

$$b_s(i|j) = \frac{1}{N_s(j)} \sum_{k=1}^{N_s(j)} \sum_{t=1}^{T(k)} I(o_s, k(t), j), \quad \forall i, j \in M \quad (4)$$

$$I(o_s, k(t), j) = \begin{cases} 1, & \text{if } h_s(o_s, k(t)) = j \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

여기서, $N_s(j)$ 는 상태 s 에서 상태의존 소스벡터 j 의 빈도를 나타낸다. 함수 $I(\cdot)$ 는 단위함수으로써, 함수의 인수들이 서로 같은 경우에 1의 값을 출력하고 그렇지 않은 경우 0의 값을 출력한다. 이 함수값은 입력인수에 대해서 상태의존 소스벡터에 대응되는 코드워드의 색인을 반환하는 함수인 $h_s(\cdot)$ 에 의해서 결정된다.

각 상태에 대한 모든 특징 벡터열들에 대한 처리가 완료된 후, 한 상태에서의 최종적인 코드워드 출력 분포는 소스벡터에 대응되는 코드워드 색인의 정규화된 빈도와 소스 벡터별로 구해진 특징 벡터들에 대한 코드워드의 분포들간의 선형 결합 형태로써 추정된다.

$$\hat{b}_s(i) = \sum_j^{C(s)} b_s(i|j) \times \bar{F}_s(j) \quad (6)$$

$$\bar{F}_s(j) = \frac{N_s(j)}{\sum_{k=1}^{C(s)} N_s(k)}, \quad 1 \leq j \leq C(s) \quad (7)$$

여기서, $C(s)$ 는 상태 s 에서 정의되는 유일한 상태의존 소스들의 갯수를 나타내며, $\hat{b}_s(i)$ 는 상태 s 에서 추정된 i 번째 코드워드의 출력확률을, 그리고, $\bar{F}_s(j)$ 는 상태의존 소

스 코드워드 j 의 정규화된 빈도를 나타낸다. SDSQ에 의한 상태별 출력확률의 추정과정은 아래의 그림 2와 같다.

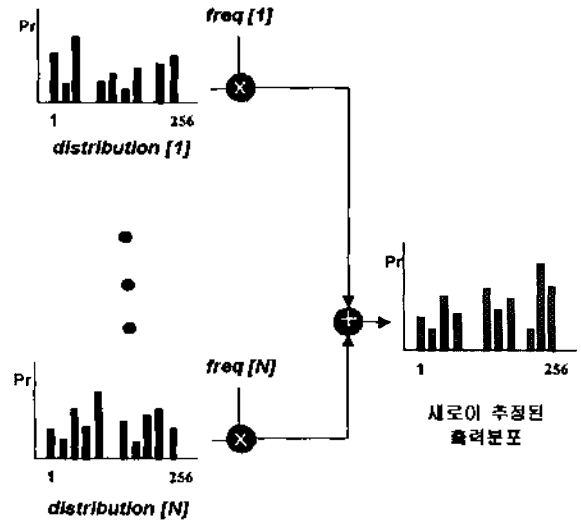


그림 2. 상태의존 소스 모델링

3.2 상태의존 소스 양자화 이산형 은닉 마르코프 모델

이 절에서는 기존의 DHMM을 상태의존 소스 모델링으로 확장한 SDSQ-DHMM 방법을 중심으로 기술하고자 한다. SDSQ-DHMM의 학습식의 유도과정은 다음과 같다. 모델 λ 이 주어졌을 때, $P_\lambda(X)$ 는 입력 벡터열 X 가 주어졌을 때의 출력확률을 나타내며, $P_\lambda(O)$ 은 입력 벡터열 X 에 대한 양자화과정을 통해서 얻어진 관측 심벌열 $O = o_1, o_2, \dots, o_T$ 이 주어졌을 때의 출력확률을 나타낸다. 특징 벡터열 $X = \vec{x}_1, \vec{x}_2, \dots, \vec{x}_T$ 이 주어졌을 때, 은닉 마르코프 모델 λ 에서의 출력 확률식은 아래와 같다.

$$P_\lambda(X) = \sum_S P_\lambda(X|S) \cdot P_\lambda(S) \quad (8)$$

$$P_\lambda(X|S) = \prod_{t=1}^T P_\lambda(\vec{x}_t | s_t) \quad (9)$$

$$P_\lambda(S) = P_\lambda(s_1) \prod_{t=2}^T P_\lambda(s_t | s_{t-1}) \quad (10)$$

여기서, $S = s_1, s_2, \dots, s_T$ 는 상태열을 나타내며, s_t 는 시점 t 에서의 상태를 의미한다. 상태열 S 에서의 합은 모델 내에서의 가능한 모든 상태열들의 합을 나타내는데, 코드워드 심벌의 집합 V 가 상태와 한 모델내에서 독립이라고 가정하면, 임의의 시점 t 에서의 상태 s_t 에서 입력 특징 벡터의 출력확률의 아래와 같이 표현할 수 있다.

$$P_\lambda(\vec{x}_t | s_t) = \sum_{o_t \in V} P_\lambda(\vec{x}_t | o_t) \cdot P_\lambda(o_t | s_t) \quad (11)$$

이상의 수식은 반연속 HMM에서 사용되는 수식과 동

일한 형태를 취하고 있다. 더 나아가서, 특징벡터의 공간이 서로 분할되었다고 가정하면, 위의 수식은 아래와 같이 변경될 수 있다.

$$P_{\lambda}(\vec{x}_i | s_i) = P_{\lambda}(\vec{x}_i | o_i^*) \cdot P_{\lambda}(o_i^* | s_i) \quad (12)$$

$$o_i^* = \operatorname{argmax}_{o_i \in \nu} P_{\lambda}(\vec{x}_i | o_i) \quad (13)$$

수식 (13)은 입력된 특징벡터 \vec{x}_i 의 양자화된 코드워드 색인을 나타낸다. 여기서, o_i^* 는 특징파라미터의 코드북에서의 최적 코드워드 색인이다. 수식 (8), (9) 그리고 (10)을 사용해서, 아래와 같은 관계가 이루어진다.

$$P_{\lambda}(X) = P_{\lambda}(X | O^*) \cdot P_{\lambda}(O^*) \quad (14)$$

where,

$$P_{\lambda}(X | O^*) = \prod_{i=1}^T P_{\lambda}(\vec{x}_i | o_i^*)$$

$$P_{\lambda}(O^*) = \sum_{S_i} P_{\lambda}(O^* | S_i) \cdot P_{\lambda}(S_i)$$

수식 (14)에서, $O^* = o_1^*, o_2^*, \dots, o_T^*$ 은 입력 특징 벡터열에 대응되는 최적 코드워드열을 나타낸다. 양자화된 코드워드열에 대한 모델 λ 에서의 특징 벡터열의 출력확률인 $P_{\lambda}(X | O^*)$ 이 현재의 모델 λ 와 무관하므로, 코드워드열의 생성확률인 $P_{\lambda}(O^*)$ 만을 계산할 수 있게 된다. 이러한 수식이 이산형 HMM에서 출력확률을 계산하는 방법이다.

SDSQ-DHMM을 유도하기 위해서, $P_{\lambda}(o_i^* | s_i)$ 의 추정 시에 상태 의존적인 소스의 모델링이 포함되도록 기존의 학습식을 확장하였다. 그 수식은 아래와 같다.

$$P_{\lambda}(o_i^* | s_i) = \sum_{c_i=1}^{C(s)} [P_{\lambda}(o_i^* | c_i, s_i) \cdot P_{\lambda}(c_i | s_i)] \quad (15)$$

Bayesian규칙을 이용해서, $P_{\lambda}(o_i^* | s_i)$ 을 모든 서로 다른 소스들에 대한 출력확률 $P_{\lambda}(o_i^* | c_i, s_i)$ 과 소스벡터에 대응되는 코드워드 색인의 빈도와의 $P_{\lambda}(c_i | s_i)$ 의 곱에 대한 합의 형태로 재정의할 수 있다. 수식 (15)를 SDSQ를 위한 기호로 재 작성하면, 재정의된 수식은 다음과 같다.

$$P_{\lambda}(o_i^* | s_i) = \sum_{j=1}^{C(s)} [b_{\lambda}(o_i^* | j) \cdot \bar{F}_{\lambda}(j)] \quad (16)$$

수식 (16)에서 알 수 있듯이, $P_{\lambda}(o_i^* | s_i, c_i)$ 와 $P_{\lambda}(c_i | s_i)$ 은 $b_{\lambda}(i | j)$ 로 대체되며, $\bar{F}_{\lambda}(j)$ 은 식(13)에 의해서 유도된다. SDSQ-DHMM에서의 학습식은 기존의 상태별 출력확률을 위의 식 (16)에 의해서 추정된다는 점을 제외하고는 기존의 Viterbi방법을 이용한 인식과정과 동일하다.

IV. 실험 및 결과

4.1 실험환경 및 음성 인식 시스템의 구성

본 연구에서는 제안된 방법들의 성능평가를 위해서 이산형 은닉 마르코프 모델에 기반한 연속 음성인식 시스템

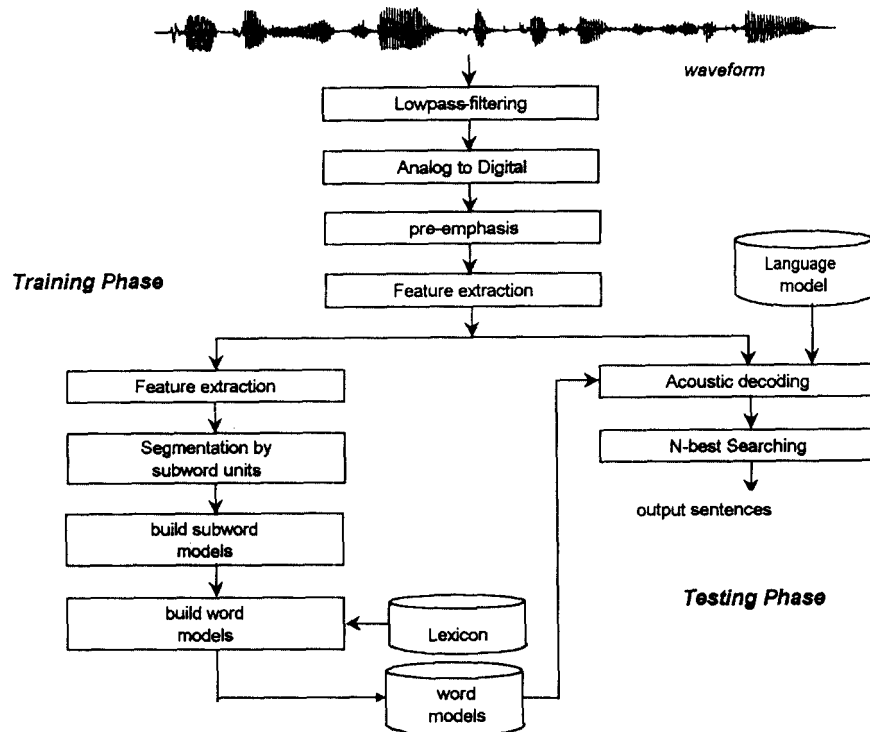


그림 3. Baseline 음성인식 시스템의 구성도

탐을 구성하였으며, 그 개요도는 아래의 그림 3과 같다.

이산형 은닉 마르코프 모델에 기반하여 음성인식을 수행하기 위해서는 크게 세가지 과정이 필요하게 된다. 첫 번째가 특징 파라미터들에 대한 양자화된 코드워드의 사상을 위해서 사용하는 코드북 생성과정이며, 두 번째로 음성인식에 사용된 인식모델을 학습하고, 학습된 인식모델들을 접속해서 단어모델로 구성하는 학습단계가 있다. 그리고 마지막으로 실험단계가 있는데, 실험단계에서는 생성된 코드북과 학습과정에서 얻어진 인식모델을 이용해서 실제적인 인식을 수행하게 된다.

첫 번째 과정인 코드북 생성은 이산형 HMM에 기반한 음성 모델링을 위해서 반드시 필요한 단계로, 이산형 HMM을 구성하는 각 상태에서의 출력확률이 코드북을 구성하는 코드워드들의 이산형 확률 밀도 함수 형태로 표시된다. 생성된 코드북으로부터 주어진 특징 벡터에 대응되는 최적 코드워드의 색인(index)을 얻기 위해서 주어진 특징 벡터와 코드워드 벡터간의 거리 계산을 위해 Euclidean 거리 척도가 사용된다[15]. 본 연구에서는 특징 파라미터에 대한 코드북 생성을 위해서 일반적으로 사용되는 방법인 LBG방법[16]을 사용하였으며, 코드북의 생성을 위한 최대 반복횟수(epoch)는 100, 종료조건은 이전 단계에서 얻어진 코드북과 총 특징 벡터들간의 평균 거리값과 현재 시점에서 얻는 평균 거리값과의 차이가 0.00001 이하일 경우에 종료하였다. 음성인식을 위한 특징 파라미터로서 14차 멜캡스트럼, 14차 차분 멜 캡스트럼, 에너지와 차분 에너지의 2차 특징 파라미터등 총 세가지의 특징 파라미터들을 사용하였으며, 벡터 양자화를 위한 코드북의 크기는 멜캡스트럼과 차분 멜캡스트럼의 경우는 256씩을 사용하였으며, 에너지와 차분 에너지를 병합한 특징 파라미터의 경우에는 128개를 각기 사용하였다.

두 번째 과정인 학습 과정에서는 인식에 사용될 인식단위모델을 학습하고, 학습이 완료된 인식모델을 접속하여 인식 어휘모델을 구성하게 된다. 인식단위로는 triphone 모델을 기본 인식단위로 사용하였다[2, 17]. 인식단위 모델을 학습하기 위해서 segmental K-mean방법[12]을 사용하여 학습자료를 인식단위로 분할하였다. 인식단위별로 학습이 완료되면, 음성인식을 위해 어휘사전(lexicon)을 사용하여, 어휘별 인식모델을 생성하게 된다. 본 기본 시스템에서는 어휘사전에 들어있는 각 어휘의 발음표기를 사용하여 기본 인식단위인 triphone 모델을 접속해서 어휘모델을 구성하였다.

세 번째 과정인 실험단계는 학습된 인식모델을 이용해 실제적인 인식을 수행하는 단계로 특징 파라미터별 코드북과 실제적으로 인식할 단어형태의 어휘모델들, 그리고 인식대상 어휘간의 접속관계에 대해 제약을 위한 언어모델을 사용하여 화자에 의해서 발생된 음성에 대해서 단어열 형태의 인식결과를 얻게 된다. 기본 인식 시스템을 위해서 언어모델로는 bigram이 사용되며, 화자의 발성음에 대한 최적 단어열을 구하기 위해서 onepass DP

(dynamic programming) 탐색방법[18]을 사용하여 가장 큰 출력확률을 가지는 경로를 구하고, 이로부터 최적 단어열을 얻으므로써 최종적인 문장인식이 완료되게 된다. 인식 시스템의 성능평가를 위해 인식된 단어열과 인식할 단어열과의 정합을 통해서 삽입(insertion), 삭제(deletion), 대치(substitution) 및 정인식된 단어의 빈도를 구하고, 그러한 값들에 기반하여 단어 정확도(word accuracy)나 문장 인식율(sentence accuracy)등을 계산하였다[19].

본 논문에서는 제안된 방법들의 평가를 위해서 연속 발생된 문장 단위의 데이터를 사용하였으며, 90명의 화자에 의해서 발생된 1074단어로 구성되어 있다. 무역상단에 관련된 회화체의 문장들을 대상으로 수집된 초기 문장집합에 대해서 단어 조성, 문장의 발생 및 화자별 문장 배분등의 절차를 통해서 재구성된 문장들을 남성화자 100명이 평균적으로 98.3문장씩 발성하였으며, 녹음은 조용한 사무실 환경에서 Sennheiser HMD224X headset 마이크로폰을 사용하여 Ariel ProPort 656로 16KHz, 16bit선형 PCM으로 A/D변환하였다. 음성자료의 구성은 표 1과 같다.

표 1. 음성자료의 구성

항 목	항목치	항 목	항목치
샘플링율	16 KHz	분석창 이동길이	10 msec
샘플당 할당비트	16 bits	인식 어휘수	1074
pre-emphasis	$1.0 - 0.97Z^{-1}$	단어당 평균 음소수	6.04
프레임 길이	20 msec	음소 모델수	224 triphone

본 연구에서는 이러한 데이터베이스 중에서 빈도수가 높은 1074단어를 추출하고, 이러한 단어들로 구성된 학습문장과 실험문장을 대상으로 인식실험을 수행하였다. 음성인식을 위한 기본적인 인식단위로는 음소모델을 사용하며, 음소의 좌/우 문맥을 반영하기 위해서 triphone을 사용하였다. 부적절한 triphone 모델링에 의한 성능저하를 방지하기 위해서, triphone 축소 규칙(reduction rule)[20]을 이용하여 triphone 수를 224개로 줄여서 사용하였다. 본 논문에서 사용된 음성 데이터베이스의 구성은 표 2와 같다.

표 2. 학습 및 실험자료의 구성

자료집합	문장 수	화자 수	복잡도(perplexity)
학습자료	2981	55	5.6
실험자료	1322	35	4.0
총 합	4203	90	6.2

4.2 상태별 양자화 소스벡터의 빈도

아래의 그림 4는 상태별 소스 벡터의 빈도를 나타내는 것으로, 세 개의 상태들로 이루어진 음소 모델에서 어느 상태에서 소스벡터의 변이가 가장 큰지를 표시하고 있다. 그림에서 알수 있듯이, 음소모델의 처음상태가 다른

상태들 보다 가장 적은 빈도를 가지며, 가운데 상태에서 소스벡터의 빈도가 가장 큰 특성을 갖는다. 이러한 결과로부터 음소모델에서 처음 상태가 마지막 상태보다 적은 변이를 가지며, 음소모델의 변별력적인 측면에서 가운데 상태와 마지막 상태가 처음 상태보다 큰 변별력을 갖는 것을 알 수 있다.

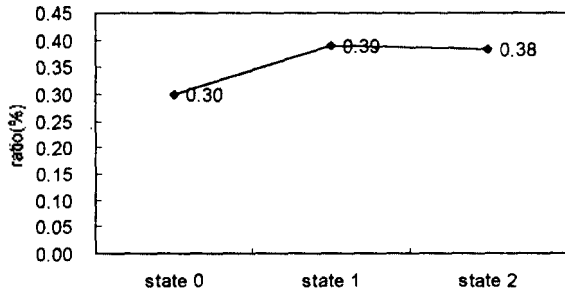


그림 4. 상태별 양자화 소스벡터의 빈도비율

다음으로, 양자화된 소스벡터의 상태별 출현빈도를 살펴보고자 한다. 실험결과에 의하면, 휴지부를 나타내는 음소군을 제외하고는 평균적으로 음소군 당 8.2번의 빈도로 하나의 양자화된 소스벡터가 출현하며, 빈도의 변이는 2.6~16.3회인 것으로 나타났다. 상태별로 평균한 양자화된 소스벡터의 출현빈도는 아래의 그림 5와 같다.

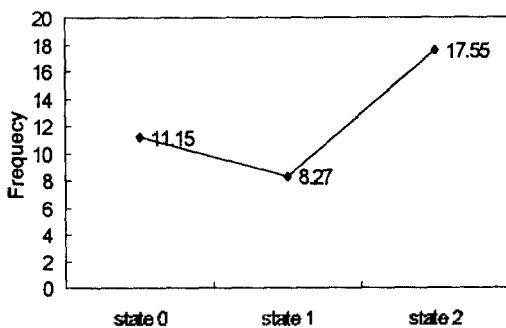


그림 5. 양자화된 소스벡터별 평균값

위의 그림에서 알 수 있듯이, 차분 델타스트림이 다른 특징 파라미터들에 비해서 가장 적은 빈도를 가지며, 에너지와 차분 에너지를 병합한 특징 파라미터가 가장 큰 빈도를 갖는다. 그림 5에서 알 수 있듯이, 중간 상태와 마지막 상태에서 소스벡터의 평균 빈도 비율이 유사하므로, 마지막 상태에서의 출력확률 추정이 중간상태나 처음 상태에 비해서 견고하게 추정될 것으로 판단된다. 다음으로, 제안된 SDSQ-DHMM의 성능 평가를 위한 실험결과에 대해서 기술하기로 한다.

4.3 인식실험 및 성능평가

제안된 SDSQ-DHMM의 성능평가를 위해서, 이산형 은닉 마르코프 모델에 기반한 Baseline 음성인식 시스템과 성능 비교를 수행하였다. 단순히 이산형 HMM에 기반한 음성 인식기의 경우, 인식을 위한 기본인식 단위인 triphone모델이 부족한 학습자료에 의해서 적절하게 학습되지 않을 수 있으므로, 이러한 문제를 해결하기 위해서 예비실험을 통해서 문맥독립 음소단위 인식모델을 구성하고, 그러한 문맥독립음소모델과 triphone모델간의 선형 결합을 통해서 Baseline 시스템에 사용될 음성 인식 모델을 생성하였다.

인식 실험을 위해서, 55명의 남성화자에 의해서 발생한 2981문장으로 학습된 triphone 모델을 사용하여 단어 형태의 인식어휘모델을 구성하였으며, 문장 단위의 인식을 위해서 학습문장으로부터 확률언어모델인 bigram을 구하였다. 이러한 triphone을 기본단위로 하는 인식어휘모델과 학습에서 얻어진 언어모델을 사용하여, 학습에 사용되지 않은 35명의 남성화자가 발생한 1322문장을 대상으로 인식 실험을 수행하였다. Baseline시스템과 제안된 SDSQ-DHMM과의 인식실험 결과는 다음과 같다.

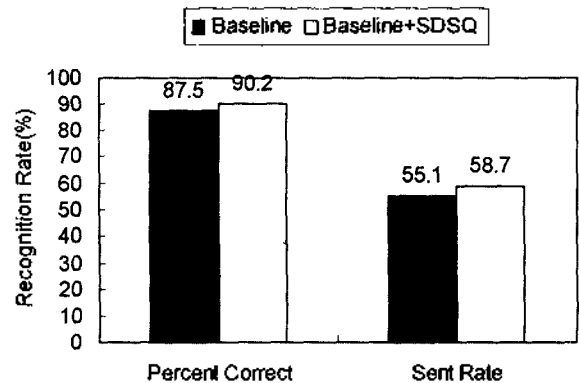


그림 6. 인식성능의 비교

위의 실험결과에서 알 수 있듯이, 제안된 SDSQ-DHMM이 단어 인식율에서 2.7%의 성능향상을 보였으며, 문장 인식율 면에서도 3.6%의 성능향상을 보였다. 이러한 성능향상은 상태에서의 출력확률이 기존의 방법에 비해서 보다 변별력을 가지며, 상태를 대표하는 출력확률의 분포가 그렇지 않은 출력확률의 분포에 비해 보다 가중됨이 인식에 유효함을 나타낸다고 할 수 있다. 이와 아울러, 시간적으로 긴 길이를 갖는 출력분포보다는 학습자료에 자주 출현하는 코드워드의 빈도가 보다 가중되는 것이 인식에 유효함을 나타낸다고 할 수 있다.

4.4 상태별 특징 파라미터의 출력분포의 특성비교

제안된 방법의 유효성을 살펴보기 위해서, Baseline 인

식 시스템과 SDSQ-DHMM에서 학습된 출력확률분포간의 비교를 수행하고자 한다. 이를 위해서, 각각 학습된 triphone 모델들에서의 특징 파라미터별 entropy를 구하고, 모든 모델들에 대한 평균적인 entropy값을 계산하였다. "엔트로피(Entropy)"란 통신이나 정보이론에서 사용되는 용어로, 통신채널상에서 평균적으로 요구하는 정보의 양(amount of information)을 나타낸다. 엔트로피가 높다는 의미는 통신채널 상에서 전달되는 심벌이 매우 모호함을 나타내며, 엔트로피 값이 작다는 것은 통신채널 상에서 사용되는 심벌들간에 서로 변별력이 크다는 것을 의미한다. 출력심벌과 그에 해당되는 확률분포가 주어졌다는 가정 하에, 해당 심벌들에 대한 엔트로피는 다음과 같이 정의한다.

$$E = - \sum_{k=1}^M P(k) \cdot \log_2 P(k)$$

여기서, E는 엔트로피를 나타내며, M은 출력심벌의 개수를, 그리고 P(k)는 k번째 심벌의 출력확률을 나타낸다. 엔트로피 값 E는 양의 실수값으로 가지며, 정보량을 나타낸다. 이러한 엔트로피 값에 기반하여 Baseline 인식 시스템과 제안된 SDSQ-DHMM을 구성하는 인식모델간의 엔트로피 값을 계산하였으며, 그 결과는 아래의 그림과 같다.

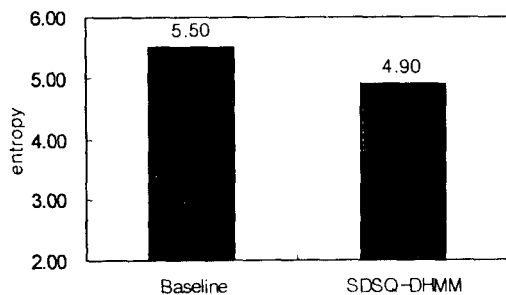


그림 7. Baseline과 SDSQ-DHMM 인식모델에서의 출력분포의 평균 엔트로피

그림 7에서 알 수 있듯이, 기존의 Baseline 인식 시스템의 경우 평균적으로 5.5의 엔트로피 값을 나타내며, 제안된 SDSQ-DHMM의 경우는 4.9의 엔트로피 값을 나타내었다. 즉, 제안된 방법으로 학습한 모델의 출력확률분포가 Baseline 시스템에서 학습된 모델에서의 출력확률분포에 비해서 0.6정도 보다 변별력을 갖는 것을 의미한다. 가장 모호한 경우가 6.7임을 감안하며, 제안된 SDSQ-DHMM은 Baseline 시스템에 비해서 약 50%정도 엔트로피 값을 감소시켰음을 알 수 있다. 이러한 독립적인 형태의 각 인식 시스템간의 엔트로피 비교방법 이외에, Kullback-Liebler에 의해서 제안된 두 출력분포간의 차이를 계산하는 방법에 따른 결과는 아래의 그림 8과 같다.

그림 8은 상태별 모델간의 출력분포의 차이를 나타내

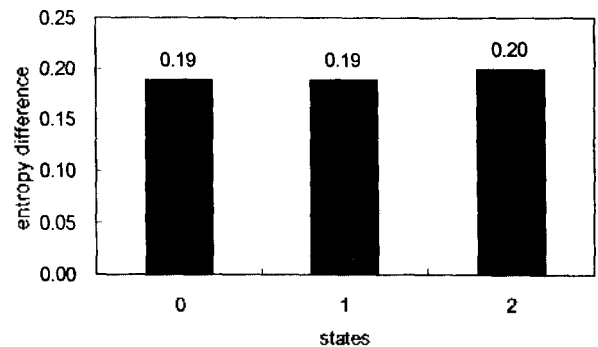


그림 8. Kullback-Liebler 거리 척도에 의한 출력분포의 차

는 것으로, SDSQ-DHMM에 의해서 학습된 상태별 출력확률분포가 Baseline 시스템의 상태별 출력 확률분포와 평균적으로 0.2의 차이를 나타냄을 표시한다. 상태별로 유사한 차이특성을 나타냄을 알 수 있으며, 마지막 상태가 처음 상태보다 차이가 있음을 의미한다. Baseline에서의 상태별 출력확률에 비해 SDSQ-DHMM의 경우가 보다 변별력을 갖는 것을 알 수 있다.

V. 결 론

본 장에서는 이산형 은닉 마르코프 모델에서 상태별 출력확률의 추정을 위한 방법으로, 상태의존 소스 양자화에 기반한 상태별 출력확률을 추정하는 방법을 제안하였다. 제안된 방법은 상태에서 출력확률의 분포와 그에 대응되는 특징 파라미터의 관계를 이용하여, 상태별 출력확률의 분포를 추정한다. 이를 위해서, 한 상태에서 그 상태를 대표하는 표준패턴을 코드워드벡터들과 그들의 출력확률의 곱의 형태로 표시할 수 있다는 가정 하에, 상태를 대표할 수 있는 대표패턴의 추정을 실제적인 학습자료들로부터 상태를 대표하는 대표패턴들과 대표패턴 각각의 빈도, 그리고 이들에 대응되는 코드워드의 분포로 표현하였으며, 대표 패턴 각각에 대응되는 코드워드의 분포와 그러한 분포의 상태 내에서의 정규화된 빈도를 이용해서 최종적인 상태별 출력확률의 추정을 수행한다. 실험결과, 제안된 상태 의존 소스 양자화 방법은 기존의 방법으로 추정된 출력 확률분포에 비해 상태에서의 분포특성이 보다 강조된 특성을 가지며, 인식율 면에서 기존의 Baseline시스템에 비해서 단어 인식율의 경우는 2.7%의 향상을 보였으며, 문장 인식율 면에서는 3.6%의 성능향상을 나타냄으로써 제안된 방법이 유효함을 알 수 있었다.

향후 연구방향으로는 상태 가중을 위한 연구가 수행되어야 할 것으로 생각된다. 인식모델에서 상태의 유효성은 음성신호의 특성에 따라 달라지게 된다. 은닉 마르코프 모델에 기반한 음성인식의 경우, 모델은 상태들로 구성되며 그러한 모델을 구성하는 상태들이 인식에 있어서

동일한 기여를 하는 것은 아니다. 모델링 대상의 음성 신호의 특성에 따라, 인식에 유효한 상태들이 있으며, 그러한 상태들이 그렇지 않은 상태들에 비해서 인식에 보다 기여함으로 그들의 출력확률이 보다 가중되어야 한다. 이러한 상태가중은 혼동 가능한 인식모델들간의 변별력을 향상시켜, 인식기의 인식율을 향상시킬 것으로 기대됨으로써, 인식 모델에서의 상태별 가중을 위한 연구가 필요할 것으로 판단된다.

참 고 문 헌

1. J. Picone, "Continuous speech recognition using hidden Markov models," IEEE ASSP Magazine, pp. 26-41, July, 1990.
2. C. H. Lee, "On the use of some robust modeling techniques for speech recognition," Computer Speech and Language, Vol. 3, No. 1, pp. 35-52, 1989.
3. L. R. Rabiner, J. G. Wilpon, et al., "A segmental k-means training procedure for hidden Markov models with continuous mixture densities," AT & T Technical Journal, Vol. 65, pp. 21-31, 1986.
4. L. R. Bahl, P. F. Brown, et al., "Acoustic Markov Model used in the Tangora speech recognition system," Proc. of ICASSP, pp. 497-500, 1988.
5. K. F. Lee, S. Mahajan, "Corrective and reinforcement learning for speaker independent continuous speech recognition," Computer Speech and Language, Vol. 4, pp. 231-245, 1990.
6. A. Ljolje, Y. Ephraim, et al., "Estimation of hidden Markov model parameters by minimizing empirical error rate," Proc. of ICASSP, pp. 709-712, 1990.
7. M. J. Russell, R. K. Moore, "Explicit modeling of state occupancy in hidden Markov models for automatic speech recognition," Proc. of ICASSP, pp. 5-8, 1985.
8. L. R. Rabiner, B. H. Juang, et al., "Recognition of isolated digits using hidden Markov models with continuous mixture densities," AT&T Technical Journal, Vol. 64, pp. 1211-1234, 1985.
9. X. D. Huang, M. A. Jack, "Semi-continuous Hidden markov models for speech signals," Computer Speech and Language, Vol. 3, pp. 239-251, 1989.
10. R. M. Gray, "Vector quantization," IEEE ASSP Magazine, Vol. 1, pp. 4-28, Apr., 1984.
11. J. D. Mark, A. H. Gray, Linear prediction of speech, Springer-Verlag, 1976.
12. J. G. Wilpon, L. R. Rabiner, "A modified K-means clustering algorithm for use in isolated word recognition," IEEE trans. on ASSP, Vol. 33, pp. 587-594, Feb., 1985.
13. A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," IEEE trans. on Inform. Theory, Vol. 13, pp. 260-269, Apr., 1967.
14. S. E. Levinson, L. R. Rabiner, et al., "An Introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition," Bell Syst. Tech. J., Vol. 62, No. 4, pp. 1035-1074, 1983.

15. R. O. Duda, P. E. Hart, Pattern Classification and Scene Analysis, John Wiley & Sons, 1973.
16. Y. Linde, A. Buzo and R. M. Gray, "An algorithm for vector quantizer design," IEEE trans. on Comm., Vol. 28, pp. 84-95, 1980.
17. L. R. Rabiner, J. G. Wilpon, et al., "High Performance connected digit recognition using Hidden Markov Models," IEEE trans. on ASSP, Vol. 37, No. 8, pp. 1214-1225, 1989.
18. H. Ney, "The use of a one-stage Dynamic Programming algorithm for connected word recognition," IEEE trans. on ASSP, Vol. 32, No. 2, pp. 236-271, 1984.
19. F. Jelinek, L. R. Bahl, et al., "Design of a linguistic statistical decoder for the recognition of continuous speech," IEEE trans. on Information Theory, IT-21, pp. 404-411, 1975.
20. E. P. Gialchin, et al., "On the use of inter-word context-dependent units for word juncture modeling," Computer Speech and Language, Vol. 6, pp. 197-213, 1992.

▲최 환 진(Hwan Jin Choi)



1990년 2월: 고려대학교 전산학과(학사)
 1992년 2월: 한국과학기술원 전산학과(석사)
 1997년 8월: 한국과학기술원 전산학과(박사)
 1997년 9월~현재: 시스템공학연구소 자연어정보처리연구부 (선임연구원)

※주관심분야: 음성인식, 화자인식, 신경회로망, 퍼지이론, 사용자 모델링

▲박 재 득(Jay Duke Park)



1983년 2월: 서울대학교 계산통계학과(학사)
 1985년 2월: 한국과학기술원 전산학과(석사)
 1994년 2월: 한국과학기술원 전산학과(박사)
 1989년 2월~1995년 2월: 삼성전자 멀티미디어 연구소, 선임연구원

1995년 3월~현재: 시스템공학연구소 자연어정보처리연구부, 실장

※주관심분야: 한국어정보처리, 지식표현 및 추론, 기계학습, 인지 언어학