

한국어 연속음성중 키워드 인식을 위한 반연속 은닉 마코브 모델과 One-Pass 알고리즘의 개선방안

최관선*, 한민홍
고려대학교 공과대학 산업공학과

Improvement of Semicontinuous Hidden Markov Models and One-Pass Algorithm for Recognition of Keywords in Korean Continuous Speech

Kwan-Seon Choi* and Min-Hong Han
Department of Industrial Engineering, Korea University

Abstract

This paper presents the improvement of the SCHMM(Semicontinuous Hidden Markov Models) using discrete VQ(vector quantization) and One-Pass algorithm for keywords recognition in Korean continuous speech. The SCHMM using discrete VQ is a simple model that is composed of a variable mixture gaussian probability density function with dynamic mixture number.

One-Pass algorithm is improved such that recognition rates are enhanced by fathoming any undesirable semisyllable with the low likelihood and the high duration penalty, and computation time is reduced by testing only the frame which is dissimilar to the previously tested frame.

In recognition experiments for speaker-dependent case, the improved One-Pass algorithm has shown recognition rates as high as 99.7% and has reduced computation time by about 30% compared with the currently available One-Pass algorithm.

1. 서 론

간단한 음성명령어로 자율이동로봇 동작을 통제하거나 전화 및 주소를 안내해 주는 음성인식 시스템을 구축하기 위해서는 연속음성을 모두 인식하는 것보다 중요한 키워드(keywords)를 인식하는 것이 효과적이다[1]. 이는 연속음성중 인식된 키워드의 조합으로 연속음성의 내용을 인식할 수 있기 때문이다. 예로써 "좌로 백팔십도 돌아"라는 연속음성 명령어에서 키워드의 조합은 '좌로(방향)'·'백팔십'(숫자)가 될 수 있다. 이러한 키워드만의 정확한 인식으로 자율이동로봇은 명령어대로 움직이게 된다. 키워드가 아닌 다른 음성들은 명령어의 인식에는 직접적인 영향을 주지 않고, 단지 명령어의 자연스러운 발음과 다양한 형태를 제공해 준다.

음성인식모델로는 DTW(Dynamic Time Wrapping) 모델과 신경망 모델, HMM(Hidden Markov Model)들이 매우 성공적으로 이용되고 있다[2][3][4][5]. 특히, HMM은 이산확률분포(discrete probability distributions)와 연속확률밀도함수(continuous probability density function)를 이용하는 매우 좋은 통계적인 방법이다. SCHMM(Semicontinuous Hidden Markov Models)은 이산 HMM과 연속 HMM의 단점을 보완하는 방법으로 이산형과 연속형이 혼합된 모델로 Huang과 Jack[6]에 의해 소개되었다. SCHMM에서 코드북의 코드기호는 이산적인 중심벡터의 색인이 아니라, 연속확률밀도함수의 색인이다. 음성신호의 프레임은 여러 개의 코드기호를 사용하여 반연속출력확률(semicontinuous output probability)로 나타냄으로써 인식의 정확도를 높일 수 있었다[6][7].

본 논문에서는 기존 SCHMM과 달리 이산 중심벡터 코드북(이하 이산 VQ라 칭한다.)을 그대로 사용하여 모수적인 연속확률밀도함수로 모델화하는 단순화된 SCHMM을 음성인식모델로 이용한다. 즉, 인식모델은 이산 VQ의 중심벡터를 혼합확률밀도함수의 평균벡터로 공유하며, 동적인(dynamic) 혼합수(mixture number)를 갖는 단일변수 혼합가우시안 확률밀도함수(single variable mixture gaussian probability density function)로 단순하게 모델화된다.

연속음성인식에서 인식단위열(recognition unit sequence)(단어열, 음절열, 음소열 등)의 탐색방법(search method)으로는 Level Building 알고리즘과 One-Pass 알고리즘이 일반적으로 이용되고 있다[2][3][8]. 특히, One-Pass 알고리즘은 탐색시간면에서 가장 좋은 성능을 나타내며, 상대적으로 작은 기억장소만을 이용하여 탐색이 가능하다. 그리고 One-Pass 알고리즘은 인식단위(단어, 음절, 음소 등)의 경계 검출, 시간축에서의 비선형 정렬, 인식의 과정이 시간의 흐름에 따라 프레임 동기적(frame-synchronous)으로 진행된다. 즉, 시간축상의 입력된 음성신호의 프레임들(관측프레임열)을 모든 참조패턴들의 프레임과 동기화시키면서 경로를 생성하여 나

간다. 관측프레임열과 매칭되는 참조패턴열은 연결된 경로로 나타나는데 경로분기는 진행중인 경로에서 새로운 참조패턴에 대한 경로의 시작으로 발생된다. 이러한 경로분기는 관측프레임열에 대한 검사(test)가 한 단계로(One Pass) 진행되면서 이루어짐으로 실시간처리가 가능하다. 그러나 One-Pass 알고리즘의 프레임 동기적인 방법은 스트링 길이(string length)를 통제하는 기능이 부족하다는 문제점이 있다[4]. 이에 대한 개선방안으로 Lee와 Rabiner[9]는 인식단위(단어) 및 상태(state)에 대한 기간(duration)과 천이(transition)의 벌점(penalty)을 반영하는 방법을 제시하였다.

기존 One-Pass 알고리즘의 또다른 문제점으로, 최적 경로(path)는 누적 Likelihood (또는 누적 거리)로 결정되는데, 누적값은 부분적인 특성보다는 전체적인 특성을 반영함으로 부분적인 특성이 전체적인 누적값에 의하여 상실될 수 있다는 점이다. 예로 초성부분의 프레임수가 아주 작고, 중성과 종성부분의 프레임수가 큰 관측음절을 참조음절들과 비교할 경우, 초성부분의 특성치의 차이는 중성부분과 종성부분의 누적된 차이에 의해서 상실될 수 있다. 이러한 경우, 음절 초성부분의 지역 Likelihood가 아주 작은 부적합한 음절에 대해서도 경로를 분기시킬 수 있고, 그만큼 경로들은 복잡해지므로 최적의 경로중에는 다른 음절 또는 동일한 음절의 첨가로 인한 오인식이 많이 발생하게 된다. 그러므로 음절의 초성부분값을 판단하여 임계치보다 작은 음절은 더이상 경로로 고려하지 않는 것이 오인식을 감소시킬 수 있는 방법이 된다. 이 문제점에 대한 해결방안은 음절의 시작부분 상태(state)에 대한 지역(local) Likelihood가 임계치보다 작은 음절에 대해서는 경로분기를 중단(fathom)하고, 경로분기점에는 음절의 기간벌점을 반영하는 것이다. Likelihood의 임계치는 Viterbi 알고리즘에 의한 학습결과로부터 얻을 수 있다.

또한 One-Pass 알고리즘에서 탐색은 관측프레임열(observation frame sequence)의 모든 프레임에 대하여 검사(test)하게 된다[4][5]. 관측프레임열중 인접 프레임들은 서로 유사한 특성치를 갖는 경우가 많다. 그러므로 한 프레임이 바로 전에 검사된 프레임과 유사하다면 검사하지 않고, 바로 전에 검사된 프레임의 탐색결과를 이용한다면 탐색의 계산량을 현저히 감소시킬 수 있게 된다. 이러한 유사한 프레임에 대해서 검사를 생략방법은 바람직하지 않는 경로(path)를 가지치기(pruning)하는 빔탐색(beam search)과는 다른 방법으로 빔탐색과 병행하여 사용될 수 있다.

제한된 방법의 검증은 인식실험을 통하여 기존 One-Pass 알고리즘과 개선된 One-Pass 알고리즘을 비교평가로 이루어진다. 본 논문의 구성은 다음과 같다. 2장에서는 한국어 연속음성중 키워드(keywords) 인식방법을 소개하고, 3장에서는 실험결과를 제시한다. 4장에서는 결론 및 추후 연구사항을 정리한다.

11. 한국어 연속음성중 키워드(keywords) 인식방법

그림 1은 연속음성중 키워드를 인식하는 절차를 보여준다.

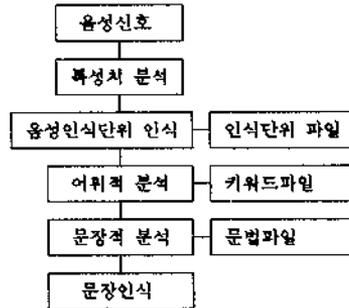


그림 1 연속음성중 키워드 인식절차

2.1. 특성치 분석(feature analysis)

본 연구실험에서는 PCM(Pulse Code Modulation)방식으로 샘플링 주파수를 11 KHZ, 양자화 정밀도(resolution)를 16 bit로하여 음성신호를 녹음하였다.

음성신호에서 일관성있는 프레임의 특성치를 추출하기 위해서 프레임을 분할하는 방법은 피치가 발견된 부분에서는 피치점(pitch point)를 기준으로 윈도우 길이(window length)를 27ms로 하는 프레임으로 분할하였고, 피치가 정확하게 발견되지 않는 음성시작점(voice start point)부터 최초 피치점(pitch point)까지는 10개의 균등하게 분할된 시점에서 피치기준 프레임과 동일한 길이를 갖는 프레임으로 분할하였다. 특히, 이 10개의 프레임들은 무성음부분에 해당되는 것으로 프레임특성치에 대한 이산 VQ(vector quantization)를 생성할 때 유성음부분과 구분되어 생성된다. 무성음부분에 대한 색인은 탐색시 부분적으로 테스트를 실시하는 방법의 한 조건으로 이용된다.

프레임의 특성치는 선형예측법(LPC : Linear Predictive Coding)/Cepstral 분석으로 Cepstral 계수를 추출하고, Cepstral 계수에 가중치를 가한 계수(weighting parameter) 벡터와 일치 미분한 계수 벡터를 구하였다. 이때 선행강조(Premphasis)의 상수는 0.95로 하였고, LPC분석의 차수는 10차로 하였으며, Cepstral 계수의 차수는 15로 하였다. 결론적으로 프레임의 특성치 벡터는 Cepstral 계수의 가중치 계수와 일치미분계수로 30차수의 벡터이다[2][10].

2.2 음성인식단위(speech recognition unit)

한국어는 특별히 음절이 잘 밀달된 언어이다[4]. 그러므로 한국어 연속음성중 키워드를 인식하기 위해서는 음절단위를 기본 인식단위로 하고, 조음현상이 심하여 구분

이 어려운 음절들은 한 음절로하는 준음절(semisyllable) 인식단위가 효과적일 것이다. 준음절을 인식단위로 한다면 비록 음소보다 많은 참조준음절이 필요하다는 단점은 있지만 음소단위의 인식을보다는 좋은 인식을 기대할 수 있게 된다. 본 실험에서 인식단위를 준음절로 하였다.

2.3. 음성인식모델

2.3.1. 이산 VQ(discrete vector quantization)를 이용한 SOM(SOHM-이산 VQ)

2.3.1.1. 모델의 단순화

이산 HMM에서는 연속적인 음향적 공간이 이산 벡터 양자화(discrete vector quantization)되어 중심코드북(centroid code book)으로 나타난다. 만약 중심코드북의 크기를 충분히 갖고, 관측기호의 출력확률(output probability)을 연속 HMM처럼 확률밀도함수로부터 구한다면, 이산화에 의한 정보손실을 극복할 수 있을 것이다. 또한 연속 HMM에서는 관측기호의 확률밀도함수를 다수변수 혼합 가우시안 확률밀도함수(multiple variable mixture gaussian probability density function)로 가정할 수 있다[10]. 확률밀도함수의 변수가 많은 경우, 정확한 공분산을 예측하는 것은 어렵다. 이러한 복잡한 모델은 단순화하여 평균분산을 공분산으로 이용하여도 좋은 결과를 얻을 수 있다.

본 연구에서 관측기호의 확률밀도함수는 모델의 단순화를 위하여 단일변수 혼합 가우시안 확률밀도함수로 하였다. 관측기호의 혼합확률밀도함수(mixture probability density function)의 평균벡터 및 분산, 혼합수, 혼합이득은 한계분산(limit variance)에 의한 군집화(clustering)방법으로 예측되었다. 군집(cluster)의 분산은 평균 벡터로부터 평균거리로 하였으며, 군집화는 군집의 분산이 한계분산이내가 되도록 군집수(cluster number, mixture number)를 증가시키면서 반복된다. 그러므로 군집화가 완료된 모든 군집(cluster)의 분산은 한계분산이하가 된다. 이와같은 한계분산에 의한 혼합 확률밀도함수 수의 동적인 결정은 모델의 단순화에 따른 정보손실을 보상하는 의미가 있다. 한계분산은 실험에 의해 결정될 수 있는데, 한계분산을 크게 하면 혼합수(mixture number)는 작아지지만 반복력이 떨어지게 되고, 반면에 한계분산을 작게 하면 반복력은 좋아지나 혼합수(mixture number)의 증가로 처리계산량이 증가하게 되는 상반관계가 있다. 만약 한계분산을 0으로 하면 이산 HMM과 동일해진다. 확률 밀도함수의 분산은 한계분산의 절반(1/2)으로 상수화하였다. 분산의 상수화 의미는 불충분한 분포에서 부정확한 분산을 추정하는 것보다 분산을 모든 모델에 동일하게 상수화함으로써 평균벡터와의 거리만으로 상대적인 유사성을 판단할 수 있기 때문이다. 즉, 적용된 확률밀도함수는 관측기호에 대해서 절대적으로 정확한 분포가 아니라 상

대적으로 유사성을 판단하기 위한 함수가 되는 것이다.

확률밀도함수의 평균벡터는 공유화된 이산중심코드북인 이산 VQ의 기호(symbol)로 나타내진다. 그러므로 단순화된 모델은 모수(parameter)의 상수화 및 공유화로 계산 처리량과 기억장소를 단축시키고, 확률밀도함수의 적용으로 부족한 학습자료를 이용한 모델의 모수예측에 대해서도 좋은 인식을 기대할 수 있다는 장점이 있다.

그리고 확률밀도함수는 관측기호벡터와 평균벡터에 대한 거리의 함수이므로 크기가 L인 이산 VQ의 기호집합 $V = \{v_1, v_2, \dots, v_L\}$ 에 대한 거리메트릭스(distance matrix), $dist[v_i, v_j], 1 \leq i, j \leq L$ 를 만들어 놓으면 매우 효과적으로 사용할 수 있다. 예로 학습이나 인식시 입력되는 관측열에 대해서 이산 VQ를 통하여 관측기호열(observation symbol sequence)을 구하면, 참조모델과 비교할 때 관측기호와 참조모델의 확률밀도함수의 평균벡터를 나타내는 기호를 색인으로하여 사전에 구축된 거리메트릭스에서 거리를 선택함으로써 모든 참조모델마다 벡터 거리를 계산해야하는 부하를 감소시킬 수 있다.

본 연구에서 상태전이모델(state transition model)은 left-right 모델을 이용하였다. 그러므로 초기 상태에 있을 확률은 $\pi = (1, 0, 0, \dots)$ 으로 고정된다.

2.3.1.2. 모델의 예측(estiimation)

초기 모델, $\lambda = (A, B, \pi)$ 의 모수(parameter)는 학습(training)을 통하여 예측(estiimation)된다. 모델의 모수는 모델의 상태정보(state information)인 혼합 확률밀도함수와 혼합이득(mixture gains), 상태(state)의 평균기간(mean duration)의 예측으로부터 산출된다.

예측(estiimation) 방법은 각 상태(state)에서 군집수(cluster number)인 혼합수가 한계분산(limit variance)에 의해 결정됨으로 변형된 Segmental K-Means Segmentation 방법을 사용하였다. 학습에 필요한 준음절자료는 연속음성신호로부터 수작업으로 얻었으며, 준음절별로 구분되어 저장되었다. 각 준음절모델에 대한 학습은 다음 절차로 이루어 진다.

단계 1. 초기화(initialize) : 주어진 준음절모델에 대한 학습음절자료들(training syllable datas)을 상태수(state number)만큼 균일하게 구분(segment)한다.

단계 2. 군집화(clustering) : 주어진 준음절모델의 i 상태, $1 \leq i \leq N$ 에 대해서 다음을 수행한다.

- a) i 상태의 초기 군집수 M_i 을 1로 한다.
- b) i 상태에 속하는 모든 학습관측벡터들을 M_i 개로 군집화(clustering)한다.
- c) 군집화(clustering)된 M_i 개의 군집중 분산이 한계분산보다 크면 M_i 을 1 증가시키고 b)로 돌아간다. 만약 M_i 개의 모든 군집의 분산이 한계분산이하이면, 다음 상태($i=i+1$)에 대해서 a)부터 c)까지 반복한다. 이러한 군집화(clustering)은 주어진 준음절모델의 모든 상태가 한계분산이하로 군집화(clustering)가 완료될 때까지 반복한다.

단계 3. 예측(estimate) : 주어진 준음절모델의 i 상태, $1 \leq i \leq N$ 의 k 군집, $1 \leq k \leq M_i$ 에 대한 혼합확률밀도함수의 평균백터(mean, μ_{ik})와 분산(variance, σ_{ik}^2), 혼합이득(mixture gain, c_{ik})을 예측하고, i 상태의 평균기간(mean duration)을 구한다. 이러한 예측은 주어진 준음절모델의 모든 상태에 대해서 반복한다.

단계 4. 구분(segmentation) : Viterbi 알고리즘의 Decoding을 통해서 수정된 HMM의 모수를 가지고 학습 준음절자료들을 상태단위로 다시 구분한다

단계 5. 반복(iteration) : 모델의 전체적인 Likelihood가 더이상 증가되지 않을 때(수렴될 때)까지 단계 2부터 단계 4까지 반복한다.

2.3.2. One-Pass 알고리즘의 개선

One-Pass 알고리즘에서 경로분기란 새로운 참조패턴에 대한 경로시작을 의미하며 매칭후보로 등록됨을 말한다. 기존 One-Pass 알고리즘의 경로분기는 경로중 참조패턴의 종료 상태에 대한 누적 Likelihood가 새로운 참조패턴의 시작 상태에 대한 누적 Likelihood보다 큰 경우 발생한다. 이는 시작 상태의 지역 Likelihood가 아주 작은 참조패턴도 경로로 진입될 수 있음을 의미한다.

One-Pass 알고리즘의 개선방안은 먼저 음절의 시작부분에 대한 지역(local) Likelihood가 임계치보다 작은 음절에 대해서는 경로분기를 중단(fathom)하고, 분기시점에 음절의 기간별점을 반영하는 것이다. 즉, 분기시점부분(본 연구에서는 관측프레임일중 현시점(t) 프레임 α_t 과 현시점부터 음절의 첫번째 상태의 평균기간($m_duration[1]$)이 경과된 시점의 프레임 $\alpha_{(t+m_duration[1])}$)에 대해서 음절의 상태 Likelihood중 최대값을 갖는 상태가 시작부분(본 연구에서는 첫번째 상태부터 두번째 상태까지)이고, 시작부분의 Likelihood가 임계치보다 작은 음절은 경로분기를 중단한다. 이 방법의 효과는 Likelihood의 임계치 결정에 종속되지만, 임계치가 적절히 결정될 수 있다면, 부적합한 음절들의 경로(path)중으로 난입되는 것을 방지하여 헛거로 인한 오인식을 감소시킴으로써 인식률을 향상시킬 수 있는 좋은 방법이다. Likelihood의 임계치를 구하는 간단한 방법은 참조패턴인 인식단위모델을 학습시킨 후, 학습자료(인식단위모델에 속하는 학습자료) 중 인식단위모델에 가장 거리가 먼, 즉 인식단위 Likelihood가 가장 작은 학습자료에 대해서 프레임의 Likelihood를 산출하여 사용할 수 있다.

추가적으로 경로(path)의 분기시점을 통제하기 위해서, 본 연구에서는 준음절의 기간별점(duration penalty)을 반영하였다. 기간별점(duration penalty)은 경로(path)의 분기시점에서 종료되는 준음절의 경로중의 기간(path duration)과 평균기간(mean duration)에 대한 기간 유사율(duration likelihood rate)을 이용하였다.

두번째 개선사항은 인식시간을 단축시키기 위해서 탐색(search)시 관측프레임(observation frame)들의 검사(test)를 부분적으로 생략(skip)하는 것이다. 관측프레임

일중 인접하는 유사한 경우가 많다. 그러므로 프레임이 바로 전에 검사된 프레임과 유사하다면 검사를 하지않고, 바로 전에 검사된 프레임의 검사결과를 이용하면 탐색의 계산량을 현저히 감소시킬 수 있다. 탐색검사를 생략하는 방법은 바로 전에 검사된 프레임을 기준프레임이라 하고, 그 색인을 bt라고 할 때, 기준프레임 α_k 과 t번째 프레임 α_t 간의 거리 $d(bt, t)$ 을 구한다. 그리고 $d(bt, t)$ 이 거리임계치이상인 경우만 프레임 α_t 에 대해서 검사하고, 그렇지 않으면 α_t 의 검사결과를 이용한다. 단, α_t 가 무성음부분(어산 VQ는 무성음부분과 유성음부분으로 구분되어 생성될 수 있고, 무성음부분은 어산 VQ의 거로도 판단이 가능하다.)이면, 무성음들간에는 판별거리가 미세함으로 생략으로 인한 인식을 저하를 방지하기 위해서 검사를 실시한다. 또한, 생략수(skip number)의 임계치(3회)를 두어 과다한 생략에 의한 인식을 저하를 방지한다.

2.4. 어휘적 분석 및 문장적 분석

실험에 사용된 키워드(keywords)는 자율이동로봇 통제명령어로 표 1에 나타나 있다. 개선된 One-Pass 알고리즘에 의해서 인식된 최적 패스상의 음절들은 순차적으로 조합되어 키워드를 만든다. 표 2와 같이 연속음성명령어 문장은 표 1의 키워드들의 조합이 된다. 문법파일에는 이러한 키워드들의 조합규칙이 정의되어 있다.

표 1 키워드(keywords) 목록

종 류	키워드(keywords) 명	키워드수
자율이동로봇 통제명령어	좌로, 우로, 전진, 후진, 속도	5
숫자	십부터 천까지 십단위씩의 숫자(90)	90
총 키워드(keywords) 수		95

표 2 연속음성명령어 및 키워드 조합 예

문 장 예	키워드
좌로/우로(방향) 백십삼(숫자)	(방향)+(숫자)
속도 전진/후진(방향) 백십(숫자)	속도+(방향)+(숫자)
전진/후진(방향) 이백오십(숫자)	(방향)+(숫자)

III. 실험결과

음성인식 실험장비는 주장비로 67MHz의 PC486과 음성 신호처리용으로 사운드 블라스터 16(SB16)을 사용하였다. 한 프레임의 특성치는 30 차수의 벡터로 이루어졌으며, 어산 VQ(중심코드북)는 300개의 프레임벡터를 갖는 신경망인 SOM층으로 이루어졌다. 어산 VQ를 생성할 때 높은 판별력을 위하여 무성음부분과 유성음부분을 구분하여 학습을 시켰다. 어산 VQ중 100개는 무성음에 대한 특성치벡터들이고, 200개는 유성음에 대한 특성치벡터들이다. 실험

업에 사용된 준음절모델은 표 1의 키워드를 구성하는 준음절에 대한 모델로 총 19개이다. 각 준음절모델의 학습용 준음절자료는 남성 2인 화자에 의해 키워드(keywords)들의 조합으로 이루어진 연속음성에서 수작업으로 준음절당 10개씩 얻었다. 준음절모델의 상태수 N은 4로 하였다.

기존 One-Pass 알고리즘과 개선된 One-Pass 알고리즘을 비교하기 위해 다음과 같이 인식실험을 실시하였다.

방법 1. 기존 One-Pass 알고리즘만 사용한다.

방법 2. 기존 One-Pass 알고리즘에 준음절의 기간벌점(duration penalty)만 적용한다.

방법 3. 기존 One-Pass 알고리즘에 준음절의 Likelihood 임계치에 의한 경로분기 중단(fathom)만 적용한다.

방법 4. 기존 One-Pass 알고리즘에 준음절의 Likelihood 임계치에 의한 경로분기 중단(fathom)과 기간벌점(duration penalty)만 적용한다.

방법 5. 방법 4에 탐색검사 생략방법(거리 임계치 0.1)을 적용한다.

방법 6. 방법 4에 탐색검사 생략방법(거리 임계치 5.0)을 적용한다.

인식실험은 2인 남성실험자에 의해서 화자종속으로 이루어졌으며, 표 2의 키워드들을 조합으로 하는 일의의 연속음성으로 각 방법마다 160회 실시하였다. 인식을 실험 결과는 표 3과 같다. 표 3의 단위계산량은 관측프레임열에서 최적 경로를 탐색하는데 소요된 계산량(DP 매칭의 grid 수)을 관측프레임수로 나눈 값으로 단위 관측프레임당 소요 계산량이다. 단위계산량란의 괄호안의 값은 방법 1의 단위계산량에 대한 백분율(%)이다.

표 3 인식을 실험결과

방법	첨가 수	삭제 수	치환 수	단위 계산량	인식률
1	13	9	5	463(100)	96.6
2	0	1	3	463(100)	99.5
3	0	0	1	463(100)	99.8
4	0	0	1	463(100)	99.8
5	0	0	2	331(71.5)	99.7
6	0	0	3	326(70.4)	99.6

(방법당 실험회수: 160회, 방법당 총음절수: 800)

실험결과를 분석하여 보면 기존의 One-Pass 알고리즘을 적용한 경우(방법 1)보다 준음절의 기간벌점(duration penalty) 적용한 경우(방법 2)와 준음절의 Likelihood 임계치에 의한 경로분기 중단(fathom)을 적용한 경우(방법 3)에 첨가 및 삭제 오인식수는 거의 0으로 나타나 첨가 및 삭제로 인한 오인식의 문제점은 거의 해결되었고, 치환 오인식수도 기존 One-Pass 알고리즘경우(방법 1)보다 현저히 감소된 결과를 보였다. 그리고 부분적인 탐색검사 생략방법에서는 거리 임계치를 0.1로 적용한 경우(방법

5). 인식이 거의 저하되지 않는 수준에서 관측프레임당 계산량이 검사를 생략하지 않는 경우(방법 4)보다 29%정도 감소하는 결과를 보였다. 이러한 인식실험 결과는 소수의 화자에 의한 인식실험결과임으로 절대적인 인식시스템의 인식률로 결정하는데는 부족하지만, 방법간의 상대적인 인식률로 기존 One-Pass 알고리즘에 비교하여 개선된 One-Pass 알고리즘의 개선효과를 보여주는데는 충분하다고 본다.

IV. 결 론

본 논문에서는 한국어 연속음성중에서 키워드(keywords)를 인식하기 위해서 이산 VQ를 이용한 SCMM과 One-Pass 알고리즘의 개선방안에 대해서 연구하였다. 이산 VQ를 이용하는 SCMM은 동적인 혼합수를 갖는 단일변수 혼합 가우시안 확률밀도함수로 단순하게 모델화되었다. 이 단순화된 모델은 예측 모수(parameter)의 수를 감소시키고, 모수들을 공유화하였으며, 거리매트릭스(distance matrix)를 이용하여 계산처리량과 저장공간을 감소시켰다. 그리고 기존 One-Pass 알고리즘은 다음과 같이 개선되었다. 준음절의 Likelihood 임계치와 기간벌점(duration penalty)을 이용하여 준음절 시작부분의 Likelihood가 임계치보다 작은 부적절한 준음절들에 대해서는 경로분기를 중단(fathom)하는 방법으로 인식을 향상시켰으며, 탐색시 바로 전에 검사된 프레임과 유사하지 않는 프레임만 검사함으로써 탐색시간을 단축시켰다.

시스템의 인식률실험은 2인 화자에 의거 화자종속으로 실시되었으며, 인식률실험결과 개선된 One-Pass 알고리즘을 이용한 경우 기존의 One-Pass 알고리즘을 이용한 경우보다 첨가와 삭제에 의한 오인식문제는 거의 해결되었고, 치환에 의한 오인식수도 현저히 감소되었다. 그리고 탐색시 부분적인 검사생략방법은 인식을 거의 저하시키지 않는 수준에서 기존 방법보다 관측프레임당 계산량(DP 매칭시 grid 수)을 약 30%정도 감소시키는 효과를 보였다. 시스템 인식률은 기존 One-Pass 알고리즘을 이용한 경우 96.6%였고, 개선된 One-Pass 알고리즘으로 탐색시 부분적인 검사생략방법(거리임계치 0.1)을 이용한 경우 99.7%의 높은 인식률을 보였다.

참 고 문 헌

- [1] Jay G. Wilpon, Lawrence R. Rabiner, Chin-Hui Lee, E. R. Goldman, "Automatic Recognition of Keywords in Unconstrained Speech Using Hidden Markov Models", IEEE Trans. Acoust., Speech, Signal Processing, Vol. 38, No.11, pp. 1870-1878, Nov. 1990
- [2] Lawrence Rabiner, Bing-Hwang Juang, "Fundamentals of Speech Recognition", Prentice-Hall, New Jersey, pp. 112-117, pp. 321-481, 1993
- [3] John R. Deller, John G. Proakis, John H. L.

- Hansen, "Discrete-Time Processing of Speech Signals", Macmillan, New York, pp. 623-740, 1993
- [4] 정 홍, "신경망을 이용한 음성인식", 정보과학회 논문지, Vol. 10, No. 2, pp. 49-59, Apr. 1992
- [5] 김동국, 정창균, 정홍, "한국어 음소인식을 위한 시간 지연신경망", 한국정보과학회 논문지, Vol.18, No.3, pp. 300-313, May 1991
- [6] X. D. Huang, "Phoneme Classification Using Semicontinuous Hidden Markov Models", IEEE Trans. on Signal Processing, Vol. 40, NO. 5, pp. 1062-1067, May, 1992
- [7] Jerome R. Bellegarda, David Nahamoo, "Tied Mixture Continuous Parameter Modeling for Speech Recognition", IEEE Trans. Acoust., Speech, Signal Processing, Vol. 38, No. 12, pp. 2033-2045, Dec. ,1990
- [8] 김도영, 박용규, 권오욱, 은종관, 박성현, "한국어 연속 음성 인식시스템 개발", 제 10회 음성통신 및 신호처리 워크샵 논문집(제 SCAS-10권 1호), pp. 238-242, 1993
- [9] Chin-Hui Lee, Lawrence R. Rabiner, "A Frame-Synchronous Network Search Algorithm for Connected Word Recognition", IEEE Trans. on Acoustics, Speech, and Signal Procession, Vol.37, No. 11, pp. 1649-1658, Nov., 1989
- [10] Lawrence R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", Proc. IEEE, Vol. 77, No. 2, pp. 257-286, Feb. 1989