은닉 마코프 모델 기반 병렬음성인식 시스템
(A Parallel Speech Recognition System based on Hidden Markov Model)

정 상 화 † 박 민 욱 ††
(Sang-Hwa Chung) (Min-Uk Park)

요 약 본 논문의 병렬음성인식 모델은 연속 은닉 마코프 모델(HMM; hidden Markov model)에 기반한 병렬 음소인식모듈과 계층구조의 지식베이스에 기반한 병렬 문장인식모듈로 구성된다. 병렬 음소인식 모듈은 수천개의 HMM을 병렬 프로세서에 분산시킨 후, 할당된 HMM에 대한 출력확률 계산과 Viterbi 알고리즘을 담당한다. 지식베이스 기반 병렬 문장인식모듈은 음소모듈에서 공급되는 음소열과 지식베이스간의 예측-활성화 알고리즘을 이용하여 입력음성에 대한 인식결과를 출력한다. 본 논문에서 제안하는 병렬 음성인식 알고리즘은 분산메모리 MIMD 구조의 다중 트랜스퓨터와 Parsytec CC 상에 구현되었다. 실험결과, 병렬 음소인식모듈을 통한 실행시간 향상과 병렬 문장인식모듈을 통한 인식률 향상을 얻을 수 있었으며 병렬 음성인식 시스템의 실시간 구현 가능성을 확인하였다.

Abstract This paper presents a speech recognition model composed of a parallel phoneme recognition module based on continuous hidden Markov model (HMM) and a parallel sentence recognition module based on hierarchical knowledge base model. The phoneme recognition module distributes thousands of HMMs to multiprocessors, computes output probabilities in parallel, and improves the Viterbi search. The parallel sentence recognition module generates recognition results using a prediction-activation algorithm, applied to phonetic code sequence from the phoneme recognition module and knowledge base. The algorithm is implemented on a Multi-Transputer system and a Parsytec CC, which are distributed-memory MIMD multiprocessors. In the result of experiments, we obtained improved speedup in the phoneme recognition module and improved recognition rate in the sentence recognition module. Furthermore, the experimental results show the feasibility of the real-time parallel speech recognition system based on continuous HMM.

1. 서 론

최근까지 연속음성인식 시스템에 대한 연구는 꾸준히 진행되고 있지만, 한정된 도메인내에서의 제한된 음성인식에 그치고 있다. 현재까지 가장 많이 사용되는 음성인식 접근방법은 은닉 마코프 모델(HMM) [9,12,13,14]을 이용하는 것이며, HMM중에서도 연속 확률분포를 사용하는 연속 HMM [7]을 기반으로 하는 음성인식 시스템 [8]이 VQ 처리를 기반으로 하는 이산 HMM보다 효과적이다. 그러나, 연속 HMM기반 음성인식은 많은 시간이 소요되는 출력확률(output probability) 계산량으로 인해 실시간 음성인식 시스템의 구현에 어려움이 있다. 그러므로, 현재의 음성인식 시스템은 HMM기반 인식알고리즘으로 사용되는 Viterbi 알고리즘 [10]에 빔(beam)을 적용함과 동시에 유한 상태 네트워크(FSN; finite state network) [1]에 기반한 언어모델을 설정함으로써 계산량을 줄이고 적정한 인식률을 유지하는 방법을 사용한다.

음성인식 시스템의 FSN 적용은 입력음성의 인식시간을 효과적으로 줄일 수 있지만, 도메인내의 발생 가능한 음소열에 대하여 문법제한을 미리 설정함으로 인해 인식률 향상에 제약이 따른다. 그러므로 본 논문에서는

· 본 논문은 정보통신부 대학기초연구지원사업 연구비지원 및 부산대학교 기성회재원 학술연구 조성비에 의한 연구결과임.
† 종신회원 : 부산대학교 컴퓨터공학과 교수
shchung@hyowon.pusan.ac.kr
†† 비 회 원 : 부산대학교 컴퓨터공학과
ukui@hyowon.pusan.ac.kr
논문접수 : 2000년 3월 2일
심사완료 : 2000년 10월 18일

기존의 음성인식 시스템을 FSN을 제외한 음소인식모듈과 지식베이스를 기반으로 하는 문장인식모듈로 분리하고 각 모듈을 병렬화 함으로써 인식시간을 단축시킴과 동시에 인식률 향상을 도모한다. 병렬 음소인식 모듈에서는 HMM을 병렬 프로세서에 분산시킴으로써 많은 시간이 요구되는 출력확률계산을 효과적으로 수행한다. 또한, 발생가능한 음소열 선정에 사용되는 Viterbi 알고리즘에 빔 적용시 발생하는 병렬 프로세서의 부하 불균형 문제를 해결하는 부하균등법을 제시한다. 음소인식 모듈 실행후 수반되는 문장인식모듈에서는 음소, 단어, 문장구조, 의미 등의 지식원을 효과적으로 결합할 수 있는 계층구조의 지식베이스를 구축하고, 음소인식모듈에 의해 생성되는 음소열와 지식베이스간의 예측-활성화(prediction-activation) 알고리즘을 이용하여 목표문장을 인식한다. 또한, 실험을 통하여 음성인식 시스템의 병렬화에 의한 실시간 음성인식 가능성과 병렬 음성인식 모델을 효과적으로 구성할 수 있는 적정 규모의 분산메모리 다중 처리 시스템을 제시한다.

본 논문은 2장에서 병렬 음성인식 시스템에 관한 관련연구를 소개하고, 3장에서 본 논문에서 제안하는 병렬 음성인식 시스템의 구조 및 알고리즘을 제시한다. 또한, 4장에서는 본 논문을 기반으로 하는 인식시스템의 성능을 보여주고, 5장에서 결론을 제시한다.

2. 관련연구

현재 음성인식 시스템의 병렬화에 관한 연구는 많이 진행되지 않았으며, 관련된 연구는 다음과 같다. Gijsbert Huijsen [6]은 <그림 1>과 같이 이산 HMM을 기반으로 하여 nCube2 시스템에서 모델링과 음성인식을 시도함으로써 병렬화를 통한 음성인식의 모델링과 인식에 있어서의 시간 단축을 보여주었다. 그러나, 이산 HMM은 많은 계산량을 필요로하는 연속 HMM 보다 인식속도는 빠르지만, 인식률이 상대적으로 낮으므로 인해 대용량 음성인식 시스템 구현에는 잘 사용되지 않는다. 한편, AT&T [15]에서는 인식하려는 도메인내의 문장을 유한 상태 변환기(FSM; finite state transducer) 기반 리스트로 구성한 후, 그 리스트들을 각 병렬 프로세서에 분산하여 처리하는 음성인식 시스템을 구현하였다. 그러나, 이러한 유한상태변환기 또는 유한상태 네트워크를 기반으로 하는 언어모델의 사용은 발생가능한 음소열을 문법 네트워크로 제한함으로써 인식률의 성능향상에 제약을 준다.

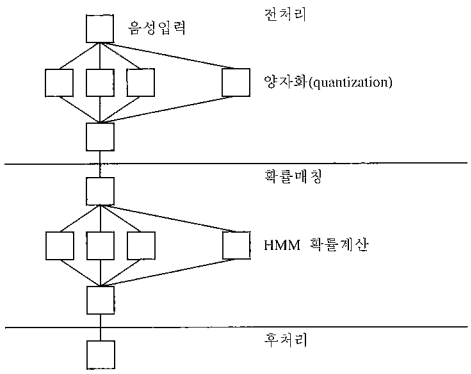


그림 1 Gijsbert Huijsen의 병렬 음성인식 과정

한편 병렬 언어처리에 관한 연구를 살펴보면, 대부분의 연구가 구어에 비해 구현하기가 쉬운 문어에 집중되었다. Huang과 Guthrie [5]는 통사와 의미의 통합처리에 기본을 둔 자연어 파싱을 위한 병렬 모델을 제안하였으며, Waltz와 Pollack [16]은 connectionist 모델을 적용한 MPP 모델을 개발하였다. 구어에 관한 연구로는 Giachin과 Rullent [4]가 트랜스퓨터 기반 분산구조상에 개발한 병렬 파서가 있다. 또한 Chung과 Moldovan [2]은 AI 응용을 위하여 개발된 SNAP MPP 시스템상에 음성과 자연어의 통합처리를 위한 병렬 파서를 개발하였다.

3. HMM기반 병렬음성인식 시스템

3.1 병렬 음성인식 시스템 개요

본 논문에서 제안하는 병렬 음성인식 시스템은 <그림 2>에 제시된 바와 같이 병렬 음소인식 모듈과 병렬 문장인식 모듈로 구성된다. 병렬 음소인식 모듈은 훈련된 HMM과 병렬 Viterbi 알고리즘으로 구성되며, 병렬 문장인식 모듈은 예측-활성화 알고리즘 및 계층구조의 지식베이스(KB)로 구성된다. 병렬 음소인식 모듈은 입력되는 음성과 훈련된 HMM에 기반한 병렬 Viterbi 알고리즘을 진행한다. 병렬 Viterbi 알고리즘은 HMM을 각 프로세서에 분산 적재한 후, 매 프레임 발생되는 음

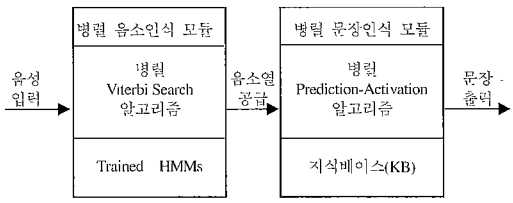


그림 2 병렬 음성인식 시스템 구조

성특징벡터와 훈련된 HMM을 기반으로 유사도(likelihood ratio)를 계산한다. 음성입력이 종료되면 역추적(backtracking) 기법을 통해 음소열을 생성하여 병렬 문장인식 모듈의 입력으로 제공한다. 또한, 병렬 문장인식 모듈은 공급된 음소열과 음소, 단어, 문장구조, 의미등 지식원들 사이의 밀접한 상호 연관성을 지원하는 지식베이스를 기반으로 예측-활성화 알고리즘을 수행하고 결과 문장을 형성한다.

3.2 연속 HMM기반 음소모델

Hidden Markov model은 관찰값을 가지는 상태열(state sequence)를 찾기 위한 Markov chain으로부터 확장된 통계적 모델이며 HMM의 매개 변수는 (1)과 같다.

$$\lambda = \{ A, B, \pi \} \quad (1)$$

$A = \{ a_{ij} \}$ 상태 i로부터 j로의 전이확률

$B = \{ b_j(k) \}$ 상태 j에서의 관측확률

$\pi = \pi_i$ 상태 i의 초기확률

(1)에서 λ는 HMM 집합, A는 HMM내부의 상태간 전이확률, B는 각 상태에서의 관측확률을 나타내며 π는 각 상태가 초기에 나타날 확률이다. 한편, 본 논문에서 사용하는 연속 HMM에서는 (1)에서의 $b_j(k)$가 (2)와 같이 표현되며 연속확률분포를 기반으로 한다.

$$b_j(x) = \sum_{k=1}^{M} c_{jk}\, b_{jk}(x) = \sum_{k=1}^{M} c_{jk} N(x, \mu_{jk}, \sum_{jk}) \quad (2)$$

(2)에서 x는 입력음성이 모델링되는 관측벡터이고, c_{jk}는 상태 j의 k번째 혼합밀도에 해당되는 계수이며, N은 상태 j에서 k번째 혼합밀도의 평균(μ_{jk})과 분산(Σ_{jk})을 가지는 가우시안 확률밀도함수이다. (2)에서 보는 바와 같이 각 상태에서의 출력확률 $b_j(x)$의 계산량은 혼합밀도와 상태의 개수에 비례하여 증가한다. 사실상, 매 프레임의 매 상태에서 계산되어져야 하는 출력확률 $b_j(x)$의 계산은 연속 HMM을 기반으로 하는 음성인식 시스템에서 대부분의 시간을 차지한다 [18].

HMM을 기반으로 하는 음소인식의 성능은 HMM의 상태수, 혼합밀도 수, 상태의 구조와 같은 HMM의 요소와 HMM 매개 변수를 기반으로 가장 높은 확률의 음소열을 찾는 Viterbi 알고리즘에 의해 결정된다. <그림 3>은 본 논문에서 사용되는 3-상태 좌우 HMM의 구조를 보여준다.

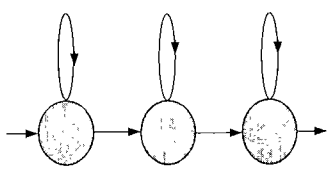


그림 3 3-state left-right HMM모델구조

HMM의 특징파일은 Entropic사의 HMM기반 음성인식 프로그램인 HTK(HMM Tool Kit) V2.0을 사용하여 1차적으로 46개의 문맥독립형 음소모델(context independent model)을 구성한 후, 2000여개의 연속 HMM기반 문맥종속형 음소 모델(context dependent model)을 생성하였다. 음성특징분석은 청각기관의 주파수 선택특성을 고려하여 대역필터군 (bandpass filter bank)을 이용한 MFCC(mel-frequency cepstral coefficients) [3]를 사용하였다.

3.3 Viterbi 빔 알고리즘의 병렬화

Viterbi 알고리즘은 입력음성의 음성특징벡터와 훈련된 HMM을 기반으로 가장 확률이 높은 음소열을 제공한다. Viterbi 알고리즘 [12]은 다음과 같다.

1. 초기화

$$\delta_1(i) = \pi_i b_j(o_1) \quad 1 \le i \le N \quad (3)$$

$$\psi_1(i) = 0$$

2. 반복

$$\delta_t(j) = \max_{1 \le i \le N} [\, \delta_{t-1}(i)\, a_{ij} \,]\, b_j(o_t) \quad 2 \le t \le T,\ 1 \le j \le N \quad (4)$$

$$\psi_t(j) = \arg\max_{1 \le i \le N} [\, \delta_{t-1}(i)\, a_{ij} \,] \quad 2 \le t \le T,\ 1 \le j \le N$$

3. 종료

$$P^* = \max_{1 \le i \le N} [\, \delta_T(i) \,]$$

$$q_T(j) = \arg\max_{1 \le i \le N} [\, \delta_T(i) \,]$$

4. 역추적

$$q_t^* = \psi_{t+1}(q^*_{t+1}), \qquad t = T-1,\ T-2, \ldots, 1$$

위의 알고리즘에서 i, j는 상태 변수이며 t는 프레임 변수를 나타낸다. 연속 HMM을 기반으로 하는 본 논문을 위해 (3)의 $b_j(o_1)$는 $b_j(x_1)$으로 대치되며, (4)의 $b_j(o_t)$는 (2)의 $b_j(x_t)$로 대치된다. 위의 Viterbi 알고리즘은 계산시간을 줄이기 위해 모델의 대수(logarithm)값을 이용하여 구현될 수 있으며, 매 프레임마다 어떤 특정한 한계치 이상의 유사도를 가지는 HMM의 상태만 평가하는 빔 알고리즘을 구현하여 불필요한 출력확률 계산을 줄일 수 있다. Viterbi 빔 알고리즘은 모든 상태가 가지는 (4)의 $\delta_t(j)$중에서 가장 높은 유사도를 획득하여 그 값으로부터 일정한 한계치(threshold)내에 포함되는 상태만을 빔에 포함하고, 다음 프레임이 발생되면 직전 프레임에서 생성된 빔에서만 다음 상태로의 전이를 진행한다. 즉, 매 프레임마다 모든 상태에서 계산

되어져야 하는 출력확률계산량을 빔에 포함되는 상태에서만 발생하게 함으로써 전체 시스템의 인식시간을 단축시킨다. 그러나, Viterbi 빔 알고리즘은 한계치의 설정으로 인해 반드시 평가되어야 하는 상태를 계산에서 제외시킴으로써 인식률에 영향을 미칠 수 있다.

본 논문에서는 Viterbi 빔 알고리즘을 병렬화함으로써 인식시간을 단축함과 동시에 빔을 최대한 확장함으로써 적절한 인식률을 유지하는 병렬 음소인식 알고리즘을 구현하였다. <그림 4>는 3.2절에서 언급한 HMM구조를 기반으로 하는 병렬 Viterbi 알고리즘의 전이구조를 을 보여준다.

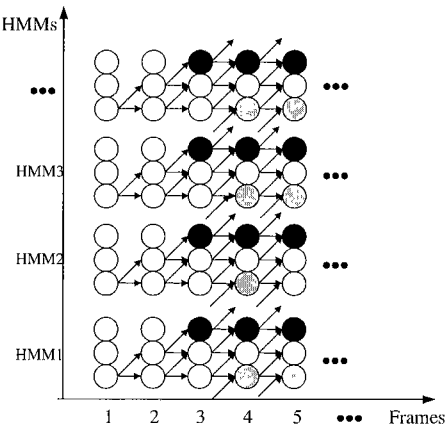


그림 4 병렬 Viterbi 알고리즘 진행중 발생되는 상태간 전이 구조

<그림 4>에서 각 HMM의 마지막 상태(●)는 자신의 프로세서 및 다른 병렬 프로세서로 전이를 진행시키며, 첫 번째 상태(○)는 다른 상태로부터 전송되는 전이확률을 평가한다. HMM이 가지는 상태의 구조적 특성상 세 번째 프레임까지는 단일 HMM내에서만 전이가 발생하며, 그 이후로는 HMM간의 전이가 발생한다. 그러므로, 병렬 음소인식 알고리즘은 세 번째 프레임부터 전이값을 효과적으로 전송할 수 있는 효과적인 메시지 전송 메카니즘을 필요로 한다.

3.4 병렬 음소인식 알고리즘

본 논문에서는 3.3에서의 Viterbi 알고리즘을 병렬화하였다. Viterbi 알고리즘은 초기화, 반복, 종료, 역추적 순으로 진행되는데, 이를 기반으로하는 병렬 Viterbi 알고리즘은 루트 프로세서와 병렬 프로세서의 역할로 나뉘어진다. 초기화는 모든 병렬 프로세서에서 발생하며, 반복에서는 병렬프로세서에 유사도 계산 및 역추적을 위한 메모리가 나누어 할당되며 다른 프로세서가 보유하고있는 HMM의 유사도가 루트 프로세서를 통해 모든 병렬 프로세서로 전송되어야한다. 그러므로, 반복과정에서의 병렬알고리즘은 다음과 같이 수정된다.

$$\delta_t(j) = \max_{1\le i\le N}[\ \delta_{t-1}(i)a_{ij}\ ]\ b_j(x_t)\quad 2\le t\le T,\ 1\le j\le M \quad (5)$$

$$\psi_t(j) = \arg\max_{1\le i\le N}[\ \delta_{t-1}(i)\ a_{ij}\ ]\quad 2\le t\le T,\ 1\le j\le M \quad (6)$$

(5), (6)에서의 M은 [모든 상태의 개수 / 병렬프로세서의 개수]를 나타내고, Viterbi 알고리즘의 복잡도는 *O(상태수×프레임수)*에서 *O(상태수/프로세서수×프레임수)*로 줄어듬에 따라 병렬 음소인식시스템의 성능향상이 이루어진다. 한편, 종료 및 역추적에서의 병렬화는 전체 알고리즘의 성능에 거의 영향을 미치지 않는다. 이러한 Viterbi 알고리즘은 병렬화에도 불구하고 불필요한 계산량을 줄이기 위해 인식률에 영향을 미치지 않는 범위내에서의 빔적용을 필요로 한다. 그러므로 병렬 Viterbi 빔 알고리즘의 복잡도는 기존 Viterbi 빔 알고리즘의 *O(빔에 포함되는 상태수×프레임수)*에서 *O(빔에 포함되는 상태수/프로세서수×프레임수)*로 줄어든다.

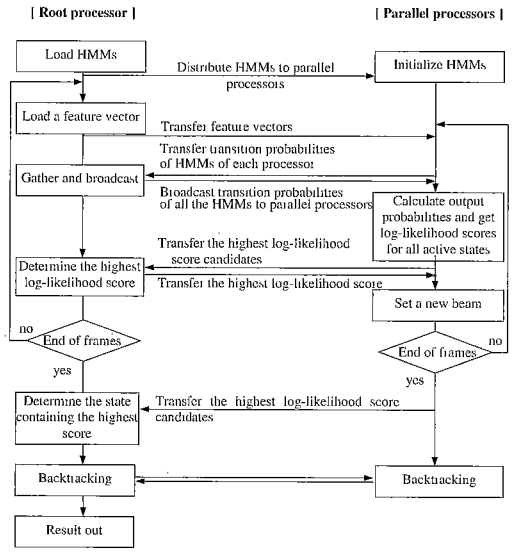


그림 5 병렬 음소인식 알고리즘

이를 기반으로 하여 본 논문에서 제안하는 병렬 음소인식 알고리즘은 <그림 5>와 같다. 병렬 프로세서들은 루트 프로세서에서 전송된 HMM을 초기화 한 후, 각

프레임마다 생성되는 음성특징벡터를 이용하여 Viterbi 빔 알고리즘을 진행한다. 병렬 프로세서는 자신이 가진 HMM중에서 가장 높은 유사도를 루트 프로세서로 전송하고, 루트 프로세서는 가장높은 유사도값과 HMM 번호를 각 병렬 프로세서로 전송한다. 또한, 병렬 프로세서는 전송된 유사도 값으로부터 일정한 한계치내에 포함되는 상태를 활성(active)으로 설정한 후 출력확률을 계산하여 새로운 빔을 확정한다. 입력음성이 종료되면 역추적을 통해 결과를 출력한다.

최종적인 음성인식을 위해서는 <그림 5>의 음소인식모듈의 결과로써 제공되는 음소열을 기반으로 실행되는 효과적인 문장인식모듈이 필요하다. 본 논문에서는 문장인식을 위해 음소, 단어, 문장구조, 의미 등의 지식원을 효과적으로 결합할 수 있는 계층구조의 지식베이스와 지식베이스 상에서 예측-활성화에 기반한 문장인식알고리즘을 개발하였다.

3.5 계층구조 지식베이스 구축

본 논문에서는 여러 계층의 지식원들 사이의 밀접한 상호연관성을 지원하기 위하여 계층구조의 지식베이스를 사용하였다. 이러한 계층구조하에서 예측-활성화 알고리즘을 수행하기 위하여 단어열(WS; word sequence)에 기반을 둔 블록(block)을 구성하였다. WS는 하나의 단어열의 시작(WSR; word sequence root)과 하나 이상의 단어열의 원소(WSE; word sequence element) 노드로 구성되어 있다. 이는 문장과 어절을 이루는 단위가 되며, WSE 노드는 first, next, last link로 연결되어 있다. 이와 유사하게 음소열(PS)는 단어를 이루는 단위가 되는 것으로, 하나의 단어 노드에 부착되어 p-first, p-next, p-last link로 연결되어 있다. 해당하는 단어 노드에 부착된 각 PS는 계층구조의 가장 하위 계층을 형성한다.

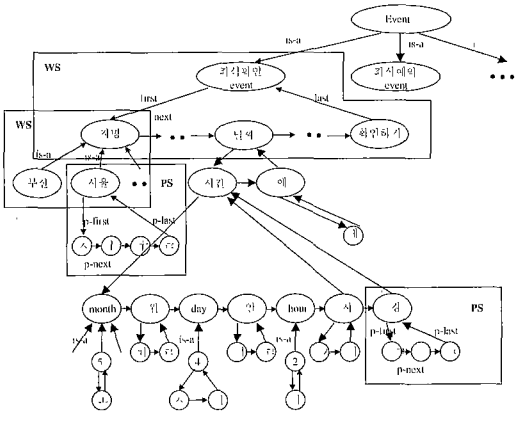


그림 6 계층구조의 지식베이스 예

<그림 6>은 본 연구에서 사용하고 있는 컴퓨터 바서 에이전트 도메인의 지식베이스 일부를 예로서 보여주고 있다. 여기서 좌석확인 사건(event)은 WSR 노드이며 지명, 날짜, 확인하기 등은 이에 대응되는 WSE 노드들이다. 또한 그림에서 볼 수 있듯이 이들 WSE 노드는 하위 레벨의 WSE를 가질 수 있으며 WSE 노드에는 대응되는 PS들이 배치되어 있다

3.6 예측-활성화 기반 문장인식 모듈

본 논문에서는 3.5절에서 구현한 계층구조 지식베이스를 기반으로 예측-활성화 기술을 사용하여 문장인식 모듈을 구현하였다. 예측-활성화에 기반한 인식과정은 다음과 같다. 우선 지식베이스를 이루는 지식원(단어, 어절)들은 위에서 보여준 바와 같이 WS로 메모리에 저장된다. 예측-활성화 과정은 기본적으로 하향 예측(top-down prediction)과 상향 활성화(bottom-up activation)를 통해 이루어진다. 예측은 지식베이스의 상위 계층을 구성하는 노드로부터 하위 계층으로 전파된다. 처음에는 모든 WSR과 그것이 포함하는 하위 계층의 첫 번째 WSE와 PSE가 가설로서 예측되지만, 음소 후보해(phoneme candidate)가 루트 프로세서로부터 들어오게 되면, 예측된 노드들 중 그 입력과 일치하는 노드만 활성화된다. 한 노드에서의 예측과 활성화의 충돌은 노드에 연결된 link를 통해 다음 노드로 예측이 전파되게 한다. 초기에는 모든 문장이 가설로서 예측되어도 실제 활성화가 일어나는 가설에는 한계가 있으므로 가설의 범위는 점차 줄어든다. 이러한 과정을 거치면서 WS가 인식되면 이로부터 출력문장을 생성한다. 예측-활성화모듈은 지나온 경로에 대한 점수를 계산하고 운반하며, 최종적으로 도착한 가설들에 대해 지나온 경로에 대한 점수를 비교한 후 선택된 가설을 역추적하여 인식된 문장을 생성한다.

3.7 병렬 문장인식 알고리즘

본 논문의 제안하는 병렬 음성인식 알고리즘은 <그림 7>과 같다. 음소모듈을 통해 다수개의 후보 음소열이 생성되면, 루트 프로세서는 병렬 음성인식을 위한 지식베이스를 병렬 프로세서에게 전송하여 분산 적재시킨다. 루트 프로세서는 후보해의 음소열을 음소인식 시간을 기준으로 정렬하여 병렬 프로세서에 전송하고, 각 프로세서에는 적재된 지식베이스를 이용하여 병렬 예측-활성화를 수행한다. 병렬 프로세서는 루트 프로세서로 결과를 전달하고, 루트 프로세서는 이 값을 받아 후보해중에서 가장 점수가 높은 것을 선택한다. 루트 프로세서

는 선택된 문장을 구성하는 단어를 인식하기 위해 병렬 프로세서로 메시지를 전송한다. 루트 프로세서는 병렬 프로세서에서 전송된 결과를 이용하여 인식된 문장을 출력한다.

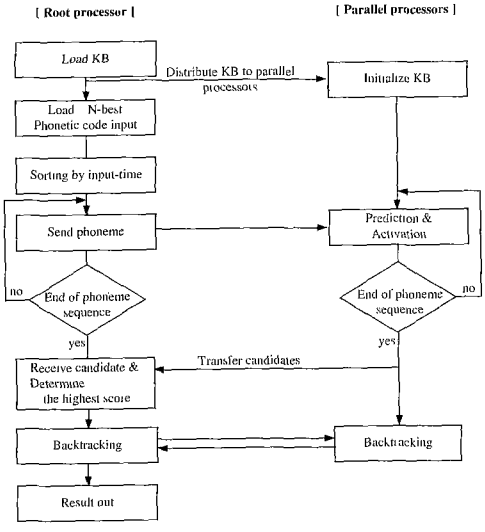
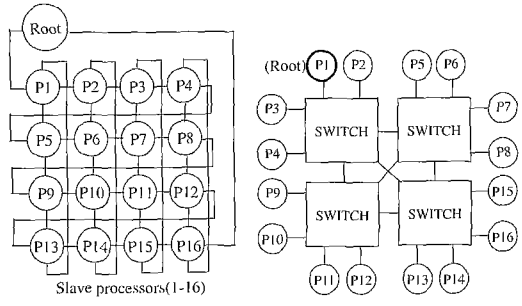


그림 7 지식베이스 기반 문장인식 알고리즘

3.8 병렬 음성인식 시스템의 구성

본 논문의 병렬 음성인식 시스템은 루트 프로세서와 병렬 프로세서로 구성된다. 루트 프로세서는 음성인식에 사용되는 입력 데이타를 공급하고 전체 인식 결과를 사용자에게 제공하며 전체 시스템을 관리하는 역할을 담당한다. 병렬 프로세서는 병렬음소인식 단계에서 출력확률 계산을 담당하고, 문장인식 단계에서 루트 프로세서로부터 전송되는 음소열에 대한 예측-활성화를 수행하는 역할을 한다.

본 논문에서는 이상과 같은 시스템을 구성하기 위해 분산 메모리 MIMD 구조의 병렬 컴퓨터로써 가격 대 성능비가 우수하고 병렬 프로그램 개발환경이 뛰어난 다중 트랜스퓨터 시스템을 사용하였다. 다중 트랜스퓨터는 1개의 루트 프로세서와 16개의 병렬 프로세서로 구성되어 있다. 각 병렬 프로세서는 Inmos사의 25MHz T805 32bit 트랜스퓨터로써, 다른 프로세서와 통신의 지원하기 위해 4개의 양방향 interprocessor 통신 링크를 가지고 있으며 20Mbits/sec로 background 통신이 가능하다. Interconnection 네트워크는 2차원 메쉬 구조의 끝노드들을 모두 연결한 <그림 8>(a)의 Illiac 메쉬를 사용한다. 한편, 본 논문에서는 2000여개의 HMM



(a) 트랜스퓨터 (b) Parsytec CC-16

그림 8 병렬 음성인식 시스템의 platform

을 기반으로 하는 음소인식 시스템의 인식시간 및 성능 향상도을 측정하기위해 위해 독일 Parsytec사의 CC-16 [17]을 이용하여 실험하였다. <그림 8>(b)의 구조로 된 CC-16 시스템은 IBM사의 AIX 4.1 운영체제를 기반으로 하고 있으며 병렬환경으로 Parsytec사의 EPX (Embeded Parix)를 사용한다. 각 노드에는 PowerPC 604 100MHz CPU와 64MB의 메인메모리가 있다. 각 노드와 스위치, 스위치와 스위치를 연결하는 interconnecton 네트워크는 Parsytec사의 HS-link이며 1Gbps의 속도로 동작한다.

4. 실험 및 결과

병렬음소인식 알고리즘 및 병렬 문장인식 알고리즘 수행을 평가하기 위해, 3.2절의 모델을 이용하여 1, 2, 4, 8, 16개의 프로세서하에서 실험이 진행되었다. 또한, 본 논문에서는 인식률 검증을 위해 Entropic사의 HMM기반 음성인식 프로그램인 HTK(HMM Tool Kit) V2.0을 이용하여 병렬 음성인식 시스템의 인식률을 비교하였다.

본 논문의 실험을 위한 음성 데이타베이스로는 KOREAN SPEECH DB(ETRI Wonkwang SPEECH DB) [11]중 컴퓨터 비서 에이전트 도메인의 데이타가 이용되었다. KOREAN SPEECH DB에는 100명분의 훈련 데이타와 10명분의 평가용 데이타가 있으며 16KHz및 16bits로 샘플링 되었다. 또한, HMM 모델 생성을 위해 남성화자 56명이 1인당 77문장씩 발음한 데이타를 이용하였으며, 평가용 음성을 위해 남성화자 6명이 1인당 77문장씩 발음한 데이타를 사용하였다.

4.1 인식률

본 실험에서는 HTK를 이용한 FSN 기반 음성인식과 본 연구의 지식베이스 기반 음성인식의 인식률을 비교하였다. 표 1에서 지식베이스를 이용한 시스템이 FSN을 이용하는 시스템보다 높은 인식성능을 가지고 있음

표 1 음성인식시스템의 인식률 비교

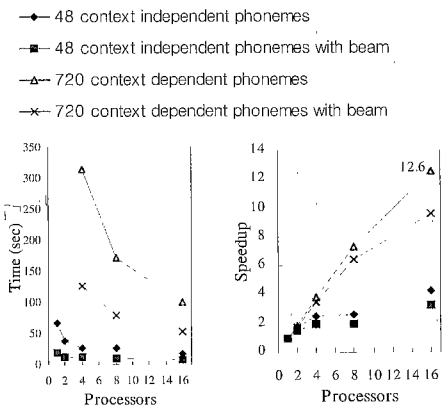
음성인식 방법	인식률
FSN 기반 음성인식	82.2%
지식베이스 기반 음성인식	87.4%

을 보여준다. Viterbi 빔 알고리즘에 FSN이 적용되는 인식 시스템에서는 인식되는 문장에 대한 제약이 음성의 입력과 동시에 진행되지만, 지식베이스를 이용한 인식시스템은 생성가능한 모든 문장을 대상으로 인식이 진행되므로 더욱 향상된 인식률을 보여준다.

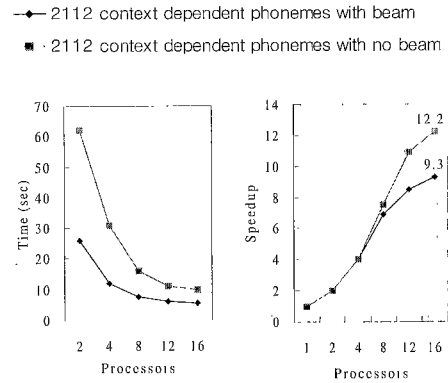
4.2 실행시간 및 성능향상도

병렬 음소인식 알고리즘은 다중 트랜스퓨터와 Parsytec CC-16상에서 구현되었으며, 시스템의 성능향상도는 최대 12.6을 나타내었다. 이는 음성인식 시스템의 병렬화가 효과적임을 보여준다. <그림 9>는 다중 트랜스퓨터와 Parsytec CC-16상에서의 인식시간 및 성능향상도 실험을 보여준다.

또한, 문장인식 모듈의 성능향상도는 <그림 10>과 같으며, 16개의 프로세서를 사용하였을 때 4.8배의 향상을 보인다. 프로세서 수가 8개에서 16개로 증가됨에도 불구하고 성능향상도가 거의 직선에 가까운 것은 지식베이스의 규모가 프로세서수에 비해 충분히 크지 않음에 따른 것으로 분석된다.



a) 트랜스퓨터 상에서의 실험결과



b) Parsytec CC-16상에서의 실험결과

그림 9 병렬 음소인식 시스템의 인식시간 및 성능향상도

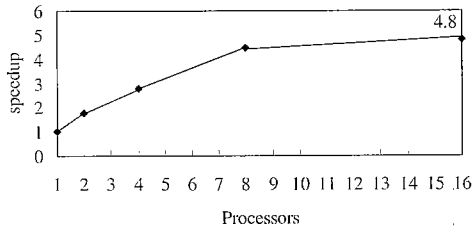


그림 10 지식베이스 기반 음성인식 시스템의 성능향상도

4.3 성능향상을 위한 실험결과

본 논문에서는 음소인식 모듈과 문장인식모듈의 개별적인 성능향상을 위한 실험결과를 제시한다. 먼저, 병렬 음소인식 모듈에 적용된 병렬 Viterbi 빔 알고리즘은 전체 유사도를 기준으로 빔을 설정하기 때문에 각 병렬 프로세서가 보유하는 빔의 크기(active state의 수)는 프로세서마다 달라진다. 그로 인해 프로세서마다 계산되어지는 출력확률의 계산시간은 프로세서마다 다르게 된다. 그로 인한 부하 불균형은 <그림 9>에서 나타난 바와 같이 빔을 적용하지 않은 Viterbi 알고리즘에 비해 Viterbi 빔 알고리즘의 성능향상도를 떨어지게 하는 중요 원인이 된다.

본 연구에서는 Viterbi 빔 알고리즘의 병렬화로 발생되는 부하 불균형문제를 해결하기 위해 매 프레임마다

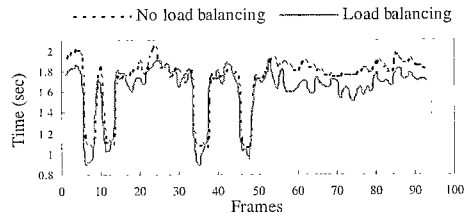


그림 11 부하균등법을 통한 실행시간 개선
(전체인식시간: [부하균등법 비적용시: 172초],
[부하균등법 적용시 : 162초],
트랜스퓨터 16 노드, 720 문맥의존음소)

모든 프로세서의 활성 상태수를 파악하여 활성 상태수의 차이가 특정 한계치 이상되는 프로세서들 간에 부하를 균등히 하는 기법을 사용하였으며, <그림 11>에서와 같이 인식 시간을 줄일 수 있었다.

또한 문장인식모듈의 지식베이스는 각 프로세서에 분산 적재된다. 이는 자주 사용되는 단어들이 각 병렬 프로세서에 분산 적재됨으로 인해 문장인식 모듈에 불필요한 메시지 발생을 유발할 수 있다. 그러므로, 본 논문에서는 지식베이스의 분산 적재방식을 다르게 함에 따른 성능향상도를 측정하였다.

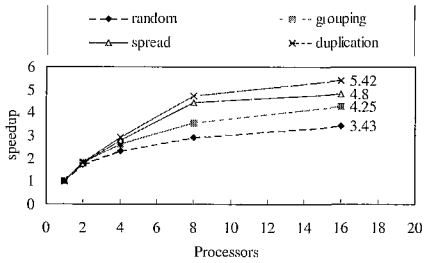


그림 12 지식베이스의 분산 적재 방법에 따른 성능향상도

<그림 12>는 지식베이스의 분산 적재 방법에 따른 성능향상도를 보여준다. 각 경우는 지식베이스를 무작위로 적재시켰을 때(random), 단어 그룹(word group)을 한 노드에 적재시켰을 때(grouping), 단어 그룹내의 각 단어를 인접한 다른 노드에 적재시켰을 때(spread), 그리고 자주 사용되는 단어 그룹을 모든 노드에 중복시켰을 때(duplication)를 나타낸다. 그림에서 보는 바와 같이 프로세서를 16개 사용하면 중복의 경우가 무작위로 적재한 경우보다 1.58배의 성능향상을 보여준다. 또한 한 노드에 같은 단어 그룹의 단어를 모두 적재시켰을 때보다 단어 그룹의 각 단어를 분리시키는 것이 더 좋은 성능을 보였다.

5. 결론 및 향후과제

본 논문에서는 병렬 음소인식 모듈과 병렬 문장인식 모듈로 구성된 연속 음성인식 시스템을 제안하였다. 음소인식 모듈은 연속 HMM에 기반한 병렬 Viterbi 알고리즘를 이용하였고, 문장인식 모듈은 계층적 지식베이스에 기반한 예측-활성화 알고리즘을 사용하였다. 트랜스퓨터 및 Parsytec CC 상에서의 실험을 통해 Viterbi 빔 알고리즘의 병렬화는 최대 12.6의 높을 성능향상도를 보여주었고, 빔 적용에 따른 프로세서간 부하불균형 문제는 부하균등 알고리즘을 적용하여 해결하였다. 또한, 본 논문의 병렬 음성인식 시스템의 인식률은 HTK를 이용한 FSN 기반 음성인식 시스템보다 높게 나타났으며, 각 모듈의 성능향상 실험들을 통하여 병렬 음성인식 시스템의 실시간 구현 가능성을 확인하였다. 앞으로의 연구방향은 병렬 음소인식 모듈에 부분 역추적 기술을 적용하여 음소인식 모듈과 문장인식 모듈이 밀결합된 실시간 음성인식 시스템을 개발하는 것이며, 음성 데이타베이스의 확장을 통해 도메인에 제약이 없는 음성언어를 기반으로 하는 대용량 HMM의 개발방안도 필요하다.

참 고 문 헌

[1] P. Baggia, E. Gerbino, et al, "Efficient representation of linguistic knowledge in continuous speech understanding," *Proceedings of IJCAI 91*, 1991.

[2] S.-H. Chung, D.I. Moldovan, and R.F. DeMara, A Parallel Computational Model for Integrated Speech and Natural Language Understanding, *IEEE Transactions on Computers*, vol. 42, no. 10, pp. 1171-1183, 1993.

[3] S. Davis, P. Mermelstein, "Comparison of parametric representation for monosyllabic word recognition in continuously spoken sentences," in *IEEE Trans. ASSP*, pp 357-266, 1980.

[4] E.P. Giachin and C.Rullent, A Parallel Parser for Spoken Natural Language, *Proceedings of IJCAI*, pp. 1537-1542, 1989.

[5] X. Huang and L. Guthrie, "Parsing in Parallel," *Proceedings of COLING-86*, pp. 140-145, 1986.

[6] G. Huijsen, Parallel Implementation of Hidden Markov Models on the nCUBE2, M.Sc. thesis, Alparon report, nr. 96-03, Delft University of Technology, 1996.

[7] M.Y. Hwand and X.D Huang, Shared-Distribution Hidden Markov Models for Speech Recognition, *IEEE Trans.* on SAP, vol.l, no.4 pp.414-420, 1993.

[8] K.F. Lee, H.W. Hon, Speaker-Independent Phone Recognition Using Hidden Markov Models, in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Vol. 37, No 11, Nov. pp. 1641-1648, 1989.

[9] K.F.Lee,Context-dependent phonetic hidden Markov models for speaker-independent continuous speech recognition, *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 4, Apr. pp. 599-609, 1990.

[10] K.F.Lee, F. Alleva, Continuous Speech Recognition, in *Advances in Speech Signal Processing*, pp. 623-650, 1992.

[11] Y.J. Lee, Design and Construction of Korean Speech Database, Final report, ETRI, 1996.
[12] L. Rabiner and B. H. Juang, Fundamentals of Speech Recognition, Prentice-Hall, New Jerscy, 1993.
[13] R. M. Schwartz, Y. L. Chow, S.Roucos, M. Krasner, and J. Makhoul, Improved hidden Markov modeling of phoneme for continuous speech recognition, in *IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 1984.
[14] F.J. Smith, J. Ming, P. O'Boyle, A.D. Irvine, A Hidden Markov Model with Optimized Inter-Frame Dependence, *Proc. of ICASSP*, pp. 209-212, 1995.
[15] Steven Phillips, Anne Rogers. Parallel Speech Recognition, In *Preceedings of EUROSPEECH-97*, pp. 135-138, 1997.
[16] D. L. Waltz and J. B. Pollack, "Massively Parallel Parsing: A Strong Interactive Model of Natural Language Interpretation," *Cognitive Science* **9**, pp. 51-74, 1985.
[17] http://www.parsytec.de/top/products/cc-pp.htm.
[18] 박민욱, 정상화, 김형순, "HMM기반 음소인식의 병렬 구현", 병렬처리시스템 학술발표회 논문집, 제9권, 제1호, pp. 117-125, 1998.

정 상 화

1985년 서울대학교 전기공학 학사. 1988년 Iowa State University 컴퓨터공학 석사. 1993년 University of Southern California 컴퓨터공학 박사. 1993년 ~ 1994년 University of Central Florida 전기 및 컴퓨터공학과 조교수. 1994년 ~ 현재 부산대학교 컴퓨터공학과 부교수. 관심분야는 클러스터 시스템, 병렬처리, VOD, 정보검색

박 민 욱

1996년 부산대학교 영어영문학 학사. 1998년 부산대학교 인지과학협동과정 석사. 1998년 ~ 현재 부산대학교 컴퓨터공학과 박사과정. 관심분야는 음성인식, 병렬처리, 자동통역