

# HMM과 GA를 이용한 한국어 음성의 음소단위 인식

Phoneme-based Recognition of Korean Speech Using  
HMM(Hidden Markov Model) and Genetic Algorithm

°박 준 하    조 성 원  
°Junha Park,    Seongwon Cho

홍익대학교 전자·전기공학부

School of Electronic and Electrical Engineering, Hong Ik University  
72-1 Sangsu-dong, Mapo-gu, Seoul, Korea

**요 약 :** 현재에 주로 개발되어 상용화가 시작되고 있는 음성인식 시스템의 대부분은 단어인식을 기본으로 하는 시스템으로 적용 단어수를 늘려줌으로서 인식범위를 늘릴 수 있으나, 그에 따라 검색해야하는 단어수가 늘어남으로서 전체적인 시스템의 속도 및 성능이 저하되는 경향이 있다. 이러한 단점의 극복을 위하여 본 논문에서는 HMM(Hidden Markov Model)과 GA(Genetic Algorithm)를 이용한 한국어 음성의 음소단위 인식 시스템을 구현하였다. 음성 특징으로는 LPC Cepstrum 계수를 사용하였으며, 인식시는 인식시는 인식대상이 되는 단어에 대하여 GA(Genetic Algorithm)를 통하여 각 음소를 분리하고, 음소단위로 학습된 HMM 파라미터를 적용하여 인식함으로써 각각의 음소별 인식이 가능하도록 하는 방법을 제안하였다.

## 1. 서 론

최근 컴퓨터 관련 기술의 발전과 더불어 정보산업 및 멀티미디어 산업의 발전이 급속도로 이루어지고 있다. 이러한 기술의 중요성이 날로 커지고 있는 가운데 화상이나 음성등을 통해 이루어지는 사람과 기계사이의 보다 편리한 인터페이스기술에 관한 다양한 측면에서의 연구가 활발히 이루어져 왔다. 특히 이들 중에서 음성은 인간의 핵심적인 의사소통 수단으로서, 음성인식 기술의 연구는 날로 중요성을 더해가고 있다. 하지만 많은 관심과 연구에도 불구하고 기계의 음성인식 정도는 음성의 유동성에 대처할 수 있는 획기적인 방법의 결여로 기술의 상용화는 아직 매우 단순한 수준에 불과하며 화자독립, 대용량 어휘인식, Noise 환경에서의 인식등 아직도 많은 부분에서의 연구의 소지를 가지고 있다.

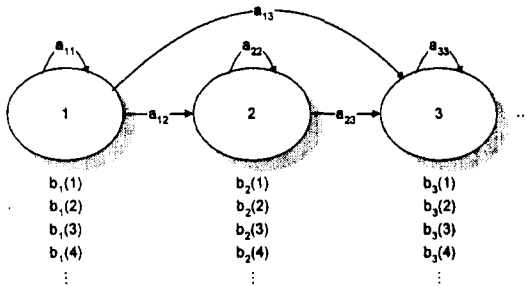
## 2. 본 론

음성인식을 위해서는 우선 화자로부터 얻은 음성 데이터로부터 음성에 대한 특징벡터를 추출하는 전처리부와 이 추출된 특징들을 이용하여 실제 음성인식을 수행하는 인식부로 구성된다. 본 논문에서는 음성의 특징으로 LPC Cepstrum 계수를 사용하였다. 또한, 음성의 특징들을 음소별로 분리하여 HMM의 Baum-Welch 방정식을 통하여 학습함으로써 음소별 파라미터를 생성한 후 인식이 이루어지도록 하는 구조를 갖으며 이때 인식부에는 유전 알고리즘을 적용하여 음소별 분리를 수행하였다.

### 2.1. 이론적 배경

#### 2.1.1 HMM(Hidden Markov Model)

시간  $t$ 에서 상태  $X(t)$ 가 주어져 있는 경우  $t$ 이후의  $s(>t)$ 의 시점에서의 상태  $X(s)$ 에 대한 확률 법칙이  $t$  이전 시점  $u(<t)$ 에서의  $X(u)$ 의 값에 관계없는 확률 과정을 Markov Process라 하는데, 이때 각 state가 하나의 관측심벌로 표현되는 것이 아니라 각 state 별로 또 하나의 관측 심벌(symbol)에 대한 확률과정을 포함하는 process를 Hidden Markov Model이라 하며 기본적인 구조를 도시하면 다음의 <그림 1>과 같다[1].



(단,  $a_{ij}$  : 상태전이 확률

$b_j(k)$  : 관측심벌의 확률 )

<그림 1> Hidden Markov Model의 기본구조

### (1) HMM 파라미터의 학습

HMM 파라미터  $\lambda(A, B, \pi)$ 는 다음의 Baum-Welch 방정식에 의하여 학습한다.

$$\begin{aligned} \xi_t(i, j) &= \frac{P(q_t = i, q_{t+1} = j, O | \lambda)}{P(O | \lambda)} \\ &= \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{P(O | \lambda)} \\ &= \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)} \end{aligned} \quad (1)$$

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) \quad (2)$$

여기서,  $\sum_{i=1}^N \gamma_t(i)$ 는 관측열  $O$ 에서 상태  $i$ 에

머무른 횟수를 의미하고,  $\sum_{i=1}^N \xi_t(i, j)$ 는 관측열  $O$ 에서  $i$ 에서  $j$ 로 전이한 횟수를 나타낸다.

### (2) 최대확률 갖는 경로 탐색 :

최대확률을 갖는 관측심벌의 경로는 다음과 같은 Viterbi 알고리즘으로 표현된다[1][2].

$$P^* = \max_{1 \leq i \leq N} [\delta_T(i)] \quad (3)$$

$$q^*_t = \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i)] \quad (4)$$

$$q^*_t = \psi_{t+1}(q^*_{t+1}), \quad t = T-1, T-2, \dots, 1 \quad (5)$$

(단,  $\delta_t(i)$  : 시간  $t$ 에서 state  $i$ 로 끝나는 단일 경로를 따른 최대 확률)

### 2.1.2 유전 알고리즘 (Genetic Algorithm)

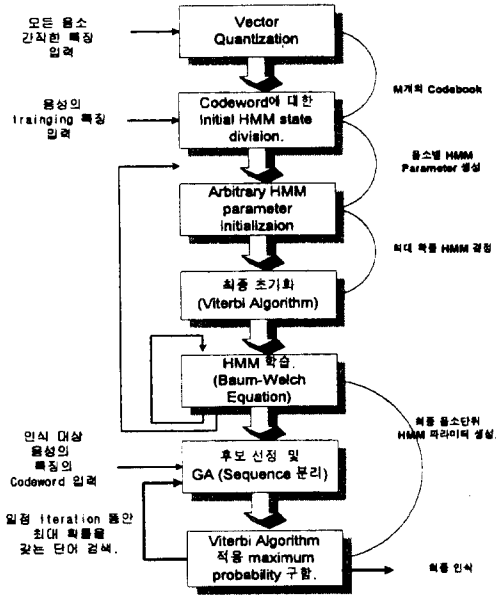
유전 알고리즘은 자연선택과 유전 메커니즘에 입각한 자연세계의 진화 현상을 바탕으로 하는 전역적(global)인 최적화 알고리즘이다. 본 논문에서는 SSGA(Steady State Genetic Algorithm)를 사용하였다.

SSGA의 과정은 다음과 같다[3][4].

- step 1. 초기 개체군(population)을 생성.
- step 2. 개체군 내에서 적합도가 가장 높은 개체 두 개를 선택.
- step 3. 선택된 두 부모개체에 대해서 균일 교배(uniform crossover) 수행.  
(자식개체의 생성)
- step 4. 생성된 자식개체에 돌연변이(mutation) 연산자 수행.
- step 5. 개체군 내에서 가장 낮은 적합도를 갖는 두 개의 개체를 개체군에서 제외.
- step 6. 생성된 자식개체를 개체군에 삽입.
- step 7. 정해진 세대수(generation)만큼 또는 목적하는 값이 될 때까지 step 2에서 step 6의 과정을 반복.

## 2.2 음성인식 시스템의 전체과정

본 논문에서 구성한 인식기의 구조는 다음의 <그림 2>와 같다.



<그림 2 > 전체 시스템의 구조도

본 논문에서 음성특징으로 사용한 LPC 과정은 위의 <그림 3>과 같다[1][5].

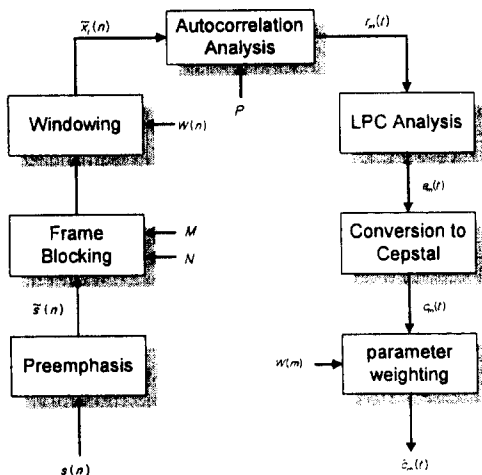
### 2.2.2 음소 학습과정

본 논문에서 인식을 위해 사용된 인식 음소의 범위는 다음과 같다.

- 초성 (13개) : ㄱ, ㄴ, ㄷ, ㄹ, ㅁ, ㅂ, ㅅ, ㅈ, ㅊ, ㅋ, ㆁ, ㆆ, ㆏
- 중성 (7개) : ㅏ, ㅑ, ㅓ, ㅕ, ㅗ, ㅛ, ㅜ, ㅠ
- 종성 (4개) : ㄴ, ㄹ, ㅁ, ㅇ

중성에서는 이중모음을 제외한 기본 단모음을 기본 인식어로 사용하였다. 또한, 종성에서는 “ㄱ, ㄴ, ㄷ, ㄹ, ㅁ, ㅂ, ㅅ, ㅇ, ㅈ, ㅊ, ㅋ, ㆁ, ㆆ, ㆏”에 대한 실제 발생어인 “ㄱ, ㄴ, ㄷ, ㄹ, ㅁ, ㅂ, ㅇ”의 7개의 음중에서 무성음으로 발음되지 않는 “ㄱ, ㄷ, ㅂ”은 다음 음절의 초성으로 간주될 수 있다는 가정하에 “ㄴ, ㄹ, ㅁ, ㅇ”의 네가지의 종성만을 인식 대상으로 정하였다.

### 2.2.1 전처리 과정



<그림 3 > LPC Process

#### ■ 초성 추출 :

가 거 게 고 구 구 기  
나 너 네 노 누 느 니  
⋮  
하 히 후 헤 호 흐 히

: [7×13]의 요소에서 가지고 공통된 음소인 “ㄱ”, “ㄴ”, ... “ㅎ” 요소추출.

#### ■ 중성 추출 :

가 나 다 라 마 바 사 아 자 차 카 타 파 하  
거 너 더 러 머 버 서 어 저 처 커 터 피 허  
⋮  
기 니 디 리 미 비 시 이 지 치 키 티 피 히

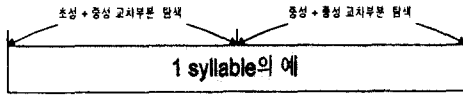
: [14×7]을 가지고 “ㅏ”, “ㅑ”, ... “ㅜ” 분리

■ 중성 추출 :  
 안 언 윤 은 온 인  
 ⋮  
 양 영 용 앵 응 잉

: [7×4]을 가지고 “ㄱ”, “ㄴ”, … “ㅇ” 분리

### 2.2.3 인식시의 GA를 이용한 음소분리

한 음절에 있어서의 초성과 중성, 중성과 중성부분간의 분리점(seperate point)의 탐색구간은 다음 <그림 4>와 같이 설정하였다.



<그림 4> 탐색 구간

분리점(seperate point)의 설정을 위한 적합도 함수는 학습된 음소별 HMM 파라미터  $\lambda$ 의  $B$  파라미터 중에서 “초성+중성”의 경우에는 중성의 “중성+중성”의 경우에는 중성의 first-state 관측심볼확률을 해당 음소의 5~10 관측횟수 만큼을 적용하여 그 상태열이 최대의 확률을 갖을 경우 그 위치를 “초성+중성” 혹은 “중성+중성”의 분리점으로 간주하였다.

즉, 적합도 함수는 (식 6)과 같이 표현가능하다.

$$f(t) = b_1(t) b_1(t+1) b_1(t+2) \cdots b_1(t+Q) \quad (6)$$

### 3. 실험 결과

본 연구에서는 우선 초성, 중성, 중성으로 추출되어진 데이터에 대하여 각각의 HMM 파라미터를 학습시킨후 그 음소별 인식율을 살펴보고, ‘초성+중성’, ‘중성+중성’ 으로 구성되어 있는 음절에 대해서 유전 알고리즘을 통하여 자동 분리한 후 학습된 음소별 HMM 파

라미터를 Viterbi 알고리즘에 적용하여 실험을 수행하였다.

### 3.1 음소별 데이터에 대한 인식실험

다음의 <표 1>은 단순 음소별 HMM 인식율을 나타낸다. 모음인 중성의 인식율은 초성이나 중성에 비해 높은 인식율을 보여주었다.

음소종류	인식율
초 성	77.0%
중 성	98.0%
종 성	85.7%

<표 1> 단순 음소별 HMM 인식율

### 3.2 초성+중성에 대한 음소분리별 인식 실험

앞선 2.2.2절에서 소개한 초성 13개와 중성 7개와의 조합을 통하여 구성된 13×7개의 발음에 대하여 각각 5개씩의 음에 대해서 인식율을 실험하였다(<표 2> 참조).

	분리된 음소	인식율(%)
아	‘아’와 결합된 초성	75.0
	아	93.9
어	‘어’와 결합된 초성	90.0
	어	100
에	‘에’와 결합된 초성	71.0
	에	98.5
오	‘오’와 결합된 초성	82.7
	오	98.5
우	‘우’와 결합된 초성	75.0
	우	98.5
으	‘으’와 결합된 초성	73.1
	으	100
이	‘이’와 결합된 초성	65.4
	이	96.9

<표 2> 초성+중성에 대한 음소별 인식율

### 3.3 중성+중성에 대한 음소분리별 인식 실험

역시 앞선 2.2.2절에서 소개한 중성 7개와 중성 4개와의 조합을 통하여 구성된 7×4개의 발음에 대하여 각각 5개씩의 음에 대해서 인식을 실험하였다(<표 3>참조).

	분리된 음소	인식율(%)
중성	'니은'과 결합된 중성	74.3
니은	아	85.7
중성	'리을'과 결합된 중성	97.2
리을	어	88.6
중성	'미음'과 결합된 중성	64.3
미음	에	100
중성	'이응'과 결합된 중성	71.4
이응	오	91.4

<표 3> 중성+중성에 대한 음소별 인식율

### 4. 결론 및 향후 과제

앞선 실험 결과에서 보는 바와 같이 중성에 관한 인식율은 초성과 결합한 경우나 혹은 중성과 결합된 경우에 큰 영향없이 상당히 좋은 인식율을 보인다. 특히 초성과 결합된 경우는 거의 100%로에 가까운 인식율을 보였다. 그러나, 초성이나 중성의 경우 약간 인식율이 떨어지는 것은 사실이다. 하지만 본 실험에서의 인식율은 여타의 다른 실험 결과에 비해서도 어느정도의 우위를 지니고 있으며, 실질적으로 단어 인식이나 문장인식등으로 확장시 문법적인 요소의 가미(즉, 어법에 맞지 않고 의미가 드러나지 않는 단어를 인식에서 제외시키는 방법)를 통하여 충분히 인식율을 보상할 수 있으리라 본다. 결과적으로 보았을때는 음소단위 인식의 확장을 통하여 기존의 단어 별로 구성되어 있는 모델의 경우보다 그 속도 면에서 우위를 유지하면서 인식율에서도 뒤지 않는 시스템의 개발이 가능하리라 본다.

향후과제가 있다면 HMM의 기본적인 문제인 노이즈 환경에서 강인한 Codebook의 작성 및 음소분리 알고리즘의 개발을 들 수가 있겠다.

### 5. 참고 문헌

- [1] Lawrence R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *Proceedings of the IEEE*, VOL. 77, NO. 2, February 1989.
- [2] Robert J. Elliott, Lakhdar Aggoun, John B. Moore, *Hidden Markov Models*, Springer-Verlag, 1994.
- [3] Philip D. Wasserman, *Advanced Methods in Neural Computing*, Van Nostrand Reinhold, 1993.
- [4] David E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley Publishing Company, Inc., 1989.
- [5] L.R. Rabiner/ R.W. Schafer, *Digital Processing of Speech Signals*, Prentice Hall, 1978.