

GAVQ를 이용한 음성인식에 관한 연구

A study on Speech Recognition using GAVQ(Genetic Algorithms Vector Quantization)

이 상 희* 이 재 곤** 정 호 균***, 김 용 연***, 남 재 성***

Lee, Sang-Hee Lee, Jae-Kon Jeong, Ho-Kyoun Kim, Yong-Yun Nam, Jae-Sung

Abstract

In this paper, we proposed a modified genetic algorithm to minimize misclassification rate for determining the codebook. Genetic algorithms are adaptive methods which may be used solve search and optimization problems based on the genetic processes of biological organisms. But they generally require a large amount of computation efforts. GAVQ can choose the optimal individuals by genetic operators. The position of individuals are optimized to improve the recognition rate. The technical properties of this study is that prevents us from the local minimum problem, which is not avoidable by conventional VQ algorithms.

We compared the simulation result with Matlab using phoneme data. The simulation results show that the recognition rate from GAVQ is improved by comparing the conventional VQ algorithms.

키워드 : 유전자 알고리즘, 벡터 양자화, 코드북

Keywords : genetic algorithm, vector quantization, codebook

1. 서론

컴퓨터 및 정보통신 시스템의 급속한 발전은 인간에게 기계와 밀접한 정보교환의 기회를 제공하고 있으며, 이에 따라 기존의 방법보다 인간에게 더 자연스럽게 친숙한 정보교환의 수단이 필요하게 되었다. 이러한 추세는 인공지능(artificial intelligence), 신경회로망(neural network), 그리고 퍼지이론(fuzzy theory)에 대한 연구가 활발해지면서 인간의 시각과 청각을 이용한 문자, 음성, 화상등의 인식능력을 컴퓨터에 부여하고자하는 패턴인식분야의 연구가 폭넓게 진행되어 왔으며 부분적으로 실용화되고 있다. 특히, 인간과 기계간의 음성에 의한 맨-머신 인터페이스를 실현하는 핵심 기술로서의 음성인식기술은 그 범위가 광범위하며 응용분야와 정보산업에 미치는 파급효과가 매우 크기 때문에 선진국에서는 이미 오래 전부터 국가적 규모의 과제로 수행되어 상당한진

전이 이루어지고 있다. 이러한 선진국의 연구 성과와 비교해 볼 때, 우리 나라에서의 연구 실태 및 성과는 아직 초보 단계라 할 수 있다.

음성인식분야에 있어서 음율인식 연구는 큰 어휘의 단어, 화자독립 그리고 연속적인 음성의 기계인식을 실현하기 위한 중요한 연구과제로 제시되고 있으며, 통계적 방법이나 지식에 기반을 둔 방법 등의 다양한 접근에서 연구가 진행되고 있다. 특히, HMM과 신경회로망의 대표적인 통계적 학습방법을 사용한 음율인식 연구는 높은 인식성능을 보임으로써 많은 연구가 진행되어 있다. 그러나 이와같은 인식모델들은 벡터양자화에 기반을 둔 것으로 초기 코드북의 설정방법에 따라 인식성능에 있어서 중요한 요인으로 나타나고 있다[1].

유전자 알고리즘(genetic algorithms)[8][9][10]은 1975년 미국의 John Holland에 의해 제안된 것으로 자연진화의 법칙인 적자생존과 자연도태의 원리를 토대로 하여 정립된 최적화 알고리즘이다. 이것은 평가함수(evaluation function), 유전연산자 그리고 문제의 해를 표현하는 염색체(chromosome)로 구성되며, 이 염색체를 점진적으로 개선시켜 나감으로서 최적

* 강원대학교 전자공학과 교수, 공학박사

** 한림정보산업대학 전자통신과

*** 강원대학교 전자공학과, 석사과정

의 해를 얻게 된다. 따라서 다른 종류의 탐색 알고리즘과는 달리, 방대하고 복잡한 함수에 대하여 전역적인 최적화(global optimization)를 할 수 있는 장점을 가지고 있기 때문에, 기존의 알고리즘으로는 해결하기 어려운 여러 분야에 적용되고 있다.

본 논문에서는 유전자 알고리즘에 기초하여 오분류의 수를 최소화하는 벡터양자화 방법의 하나로 변형된 유전자 알고리즘(GAVQ)을 제안하였다. 변형된 유전자 알고리즘은 유전자 알고리즘의 특성에 기초하며, 음성인식 과정에서 코드북 생성시 오인식률에 크게 영향을 주는 단어들을 교차 또는 도태시켜 최적의 코드북을 갖는 단어 모델을 구성하는데 목적이 있다. 본 논문의 인식실험에서는 한국어 음소 데이터에 대하여 첫 번째로, 기존의 VQ(vector quantization)방법과 제안된 유전자 알고리즘을 이용하여 생성된 코드북을 사용하여 NN(nearest neighbor) 분류기를 가지고 성능평가를 하였다. 또한 VQ를 기반으로 하는 인식모델들에 대하여 위 방법들을 전처리 단계인 VQ 입력값으로 사용되었을 때의 인식모델에 따른 VQ방법에서의 성능평가 실험을 하였다.

본 논문의 구성은 2장에서 일반적인 음성인식 시스템에 관하여 기술하고, 3장에서 유전자 알고리즘과 제안한 GAVQ를 이용한 벡터 양자화 방법에 대하여 알아본다. 4장에서 한국어 모음에 대한 음소 인식 실험에 대한 결과를 고찰하고, 5장에서 결론을 맺는다.

2. 음성인식 시스템의 개요

음성인식 시스템은 기본적으로 그림 1과 같이 몇 개의 서브 유닛이 연결되어 구성된다. 음성 검출부(endpoint detection)에서는 신호 속에서 음성을 검출하며, 특징 추출부(feature extraction)에서는 검출된 음성으로부터 그 음성의 성질을 잘 표현해 주는 특징 벡터를 추출한다. 유사도 측정(distance measure) 부분에서는 이미 저장되어있는 기준 모델과 검출된 임의의 음성신호간의 유사도를 측정하고 마지막으로 최종 인식결정부분에서는 인식된 결과의 타당성을 검증한 후 인식 결과를 출력한다.

2.1 음성신호의 분석

음성신호는 일련의 분절음으로 구성되어 있고, 인간은 이러한 분절음을 언어학적 규칙에 따라 배열하여 서로 의사소통을 하기 위한 정보를 교환하게 된

다. 음성 신호처리나 음성인식 분야는 공학적인 영역

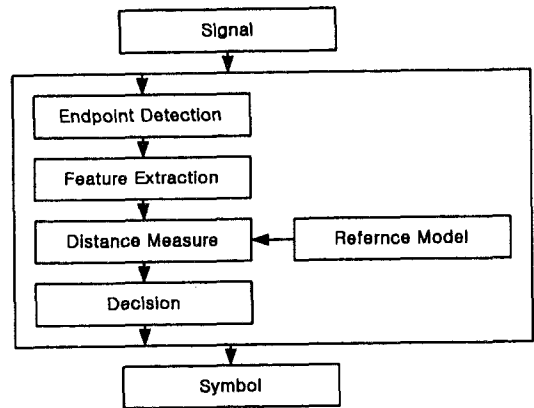


그림 1. 일반적인 음성인식 시스템

에 속하지만 음성신호로부터 그 속에 포함되어 있는 언어적 정보를 추출해 내기 위해서는 음성신호의 구조, 다시 말해 음성신호에 어떤 형태로 정보가 암호화되어 있는가에 대해 가능한 많은 지식을 가질 필요가 있다. 음성신호는 의미를 가지는 최소단위인 음소의 결합으로 이루어진다. 각각의 음소는 따로 발음될 수 없고 이웃 음소와 상호 영향을 받아 동화(assimilation)되거나 조음(coarticulation)화 된다. 그러므로 청취자는 각각의 음소를 구별하여 의미를 파악하는 것이 아니라 몇 개의 음소가 결합된 음절이나 그 이상의 단위로 의미를 파악한 다음 구성된 음소를 구별해 낸다. 그러나 대부분의 인식 시스템은 인간의 뇌와 같은 기능이 없어 음성 신호만으로 음소를 구별하기 때문에 인식을 더욱 어렵게 한다. 이러한 한계속에서 음소를 인식하기 위해서는 각 음소가 가지는 특징을 알아야 한다.

2.2 특징파라미터 추출[2]

음성인식에 있어서의 전처리과정은 음성분석을 통한 음성에 포함된 특성 즉, 특징 파라미터를 추출하는 것인데, 특징 파라미터 추출법에는 기본적으로 관측된 음성 신호의 시간적 변화를 그대로 이용하는 방법과 음성신호가 포함하고 있는 주파수 성분을 이용하는 방법이 있다. 시간 영역의 분석에서는 상관 함수에서 선형예측 계수(linear prediction coefficient)나 편자기 상관 계수(partial autocorrelation coefficient, PARCOR coefficient)를 추출하거나, 피치, 영 교차율(zero crossing rate) 등을 구하는 것이고, 주파수 영역의 분석에는 스펙트럼(spectrum)분석과 폴만트(formant) 기본 주파수 추출 등이 있으며, 또 선형

예측 계수에서 스펙트럼을 구하거나 소나그램(sonagram)에 의한 분석 등이 있다.

(1) 에너지 영교차율(zero crossing rate, ZCR)

영교차율은 한 프레임 내에서 음성 파형의 영점과 교차하는 횟수를 말하며 화자의 성량에 대해서는 독립적이다. 음성생성 모델에 의하면 유성음은 음성의 에너지가 스펙트럼의 3kHz 정도의 낮은 주파수에 밀집해 있으나 무성음은 에너지가 높은 주파수에서 주로 발견된다. 이때 높은 주파수는 많은 영교차율을 갖게 되며, 낮은 주파수에는 적은 영교차율을 갖게 되므로 영교차율과 주파수에 의한 에너지 분포는 밀접한 관계가 있다. 일반적으로 음성신호에서의 에너지 크기는 무성음 부분 보다 유성음부분에서 크게 나타나며 에너지 E_n 은 다음과 같이 주어진다.[2]

$$E_n = \sum_{m=m}^{n+N-1} x^2(m) \quad (1)$$

여기서 N은 한 프레임의 길이를, n은 프레임의 색인(index)을 나타낸다..

(2) 선형예측 계수(linear predictive coefficients)

선형예측에 의한 계수척도는 시간 영역상의 분석법에서 가장 대표적인 것으로 기본적인 음성발생 모델에 기초한다. 여기서 음성 발생까지의 시변(timevarying) 디지털 필터의 전달함수는 성대진동펄스, 성도 및 구강으로부터의 음성출력 등의 복합적 요인에 의한 것으로 모든 Pole이 단위원(unit circle)안에 존재하는 p 차 all-pole 필터로 표현될 수 있다고 가정한다. 이러한 방식에 따라 음성 발생 기구를 선형 시스템으로 볼 수 있으므로 음성 신호의 현시점 표본치 $\hat{s}(n)$ 은 다음과 같이 과거의 표본 치들의 선형결합으로 근사될 수 있다.

$$\hat{s}(n) = \sum_{i=1}^m a_i \cdot s(n-1) \quad (2)$$

여기서, a_i 를 선형계수라 하며, 선형계수를 구하는 방법은 n개의 표본만 고려하는 공분산법(covariance method), 음성신호가 윈도우 함수에 의해 잘리는 자기 상관법 (autocorrelation method), 편자기상관법(partial correlation method)등으로 나눌 수 있다. 보

통 음성신호 처리를 위해서는 자기 상관법인 Levinson-Durbin 방법을 사용하여 예측계수를 구하여 사용한다.

(3) 켈스트럼(cepstrum)과 차분 켈스트럼(delta cepstrum) 켈스트럼은 입력 음성신호의 로그 파워스펙트럼의 역 푸리에 변환을 취함으로써 얻어지며 식 3과 같다.

$$c(k) = \int_{-\pi}^{\pi} \log |s(e^{jw})| e^{jwk} dw \quad (3)$$

또한, 시간에 따른 스펙트럼의 변화가 음성 인식에 있어 중요한 역할을 한다. 따라서, 이러한 특성을 반영한 차분켈스트럼을 특징 벡터로 많이 사용하며 식 7과 같다.

$$dcep(m) = c(m+d) - c(m-d) \quad (7)$$

여기서 c는 cepstrum vector이며 d는 해당되는 프레임의 수이다.

2.3 음성인식 알고리즘

현재 음성인식 시스템은 음성의 형태, 화자의 수, 인식대상 어휘, 인식환경 등에 의하여 제한된 성능을 갖고 있다. 음성인식의 경우 1970년대 이후 동적 프로그래밍을 이용하여 기준이 되는 신호의 패턴과 입력된 음성신호를 비교한 후 가장 유사도(similarity)가 높은 것을 찾아서 인식하는 DTW (dynamic time warping)[3] 방법과 단어 혹은 임의의 단위 음성 신호에 포함된 특성들의 발생 과정을 확률적으로 모델화한 HMM(hidden markov model), 그리고 통계적 학습방법에 기초한 신경회로망등을 이용한 음성인식 모델들이 많이 이용되고 있으며 계속적으로 활발한 연구가 진행되고 있다[3][4].

3. 유전자 알고리즘을 이용한 벡터 양자화

3.1 유전자 알고리즘

유전자 알고리즘은 자연 선택과 자연계 진화과정에서 적자생존을 원칙으로 하는 적응적인 탐색 기법으로서 최적해에 대한 효과적인 탐색 도구로 널리 사용되고 있다. 기존의 탐색 방법 및 최적화 방법과 다른 점은 파라미터를 코딩한 것을 직접 이용하며, 점(single point)이 아닌 다점(multi points) 탐색 방

법을 취하는 것이다. 또한 탐색에 적합도를 이용하고 다른 부가적인 지식을 요구하지 않으며 결정론적인 규칙없이 확률적 연산자를 사용하여 수행된다.

유전자 알고리즘을 실제 응용에서 사용하기 위해 먼저 문제를 유전자형에 대응하는 문자열로 변환한다. 그리고 이진수 문자열의 나열인 염색체를 표현하고, 이 염색체들 모임인 개체집단을 생성한다. 이 문자열을 평가하여 평가치가 높은 집단을 선택하여 남도록 하는데, 이것은 자연계에 있어서 "도태"에 대응하는 것이다. 이렇게 선택된 집단에 대하여 연산자를 적용함으로써 새로운 문자열을 생성한다. 유전자 알고리즘은 선택, 교배, 돌연변이등 3개의 기본적인 연산자가 있다.

선택(selection) 연산자는 문자열을 적합도에 따라 개체집단에서 두 개의 염색체를 선택하는 역할을 하는데, 개체집단내 모든 염색체의 적합도의 합과 각 염색체의 적합도의 상대적 값, 즉 선택될 확률을 구하게 된다. 교배(crossover) 연산자는 선택 연산자에 의해 선택된 두 염색체의 인자값을 서로 맞바꾸어 새로운 염색체를 생성한다. 이때 무작위로 선택된 교배 위치와 개수에 따라 교배 알고리즘이 달라지게 된다. 또한 돌연변이(mutation) 연산자는 염색체내의 인자를 무작위로 선택하고, 그 값을 임의대로 바꾸어 새로운 염색체를 만드는 기능을 담당한다. 이것은 개체 집단의 특성을 다양하게 변화시키며, 국부적인 최소 상태를 벗어날 수 있게 한다. 위에서 설명한 연산자의 사이클을 반복함으로써 환경에 대응하는 평가치가 높은 문자열을 만들어 내어 문자열의 집단 전체의 평가치를 향상 시켜 나간다.

유전 연산자를 이용하여 새로운 세대를 생성하였을 때 부모 세대보다 더 낮은 적합도를 갖는 염색체가 생성될 수도 있는데 세대를 거듭하면서 적합도가 낮은 것들은 염색체끼리의 경쟁에서 도태되어 사라지게 되므로 낮은 적합도를 갖는 염색체가 다음 세대에 생기더라도 문제가 되지 않는다. 기본적인 단순 유전자 알고리즘(simple genetic algorithm : SGA)의 흐름도는 그림 3과 같다[8][9][10].

3.2 GAVQ를 이용한 벡터 양자화

(1) 변형된 알고리즘 개요

제안된 알고리즘은 음성 입력기로부터의 입력된 음성에서 추출한 특징벡터와 제안된 GAVQ 코드북 학습 알고리즘을 사용하여 최적화된 음성 특징 코드북(codebook)과의 패턴 매칭을 통하여 입력음성과 가장 잘 맞는 의미를 찾아내는 구조이다.

변형된 유전자 알고리즘은 방대하고 복잡한 함수에 대하여 전역적인 최적화를 할 수 있는 장점을 갖는 유전자 알고리즘과 신경회로망에서 Bayesian 결정경계면을 근사화하여 오인식의 수를 최소화하는 LVQ 알고리즘의 결합된 형태에 기초를 두고 있다 [5]. 제안된 유전자 알고리즘에서는 이러한 각 모델들의 특성을 이용하여 LVQ에서의 학습조건 즉, 오분류의 수를 최소화하는 코드북의 학습과 전체 오분류에 가장 크게 영향을 미치는 코드북의 도태와 생성을 통한 유전자 알고리즘을 이용한 오분류수를 최소화하는 알고리즘 구현에 기초한다.

(2) 타입(type)의 정의

변형된 유전자 알고리즘에서는 오인식에 영향을 주는 코드북의 교차 및 도태 생성하기 위하여 오인식되는 경우의 상태를 정의하여 선별적인 학습을 하도록 정의된다. 이것은 학습과정에서 주어진 패턴에 대한 가장 가까운 코드북(Ref1)과 두 번째로 가까운 코드북(Ref2)을 계산하여 정의하였다. NN에 근거한 분류시 Type들은 오인식 형태를 나타낸다. 변형된 유전자 알고리즘에서는 오인식 타입에서 대부분의 유전자 조작들이 이루어지며, 이와 같은 타입이 나타나는 패턴들을 최소화하는데 목적이 있다.

- ▶ Type I : 학습패턴과 Ref1은 다른 클래스, Ref2는 같은 클래스.
- ▶ Type II : 학습패턴과 Ref1, Ref2가 다른 클래스 .

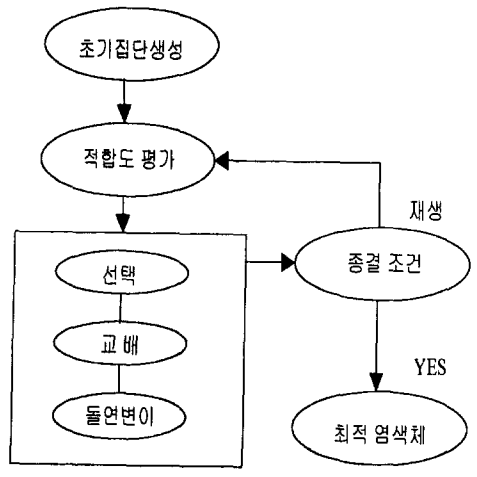


그림 3. 단순 유전자 알고리즘 흐름도

(3) 변형된 유전자 알고리즘 연산자

본 논문에서 적합도 함수는 정인식률로 정의하며 교차에서 필요한 두 염색체는 오인식된 경우에 학습 패턴과 가까운 자기 코드북을 선택하여 교차하게 하는데, 식 8과 같이 윈도우를 정의하여 윈도우 안에 놓여 있을 때 교차 연산을 수행하게 한다.

$$\min\left(\frac{d_i}{d_j}, \frac{d_j}{d_i}\right) > s \quad (8)$$

$$d_i = |m_i - x| = \sqrt{\sum_{j=1}^n (m_{ij} - x_j)^2} \quad (9)$$

여기서, s 는 윈도우 폭을 결정하는 파라미터이며, d_i , d_j 는 주어진 패턴 x와 가장 가까운 코드북 m_i 두 번째 가까운 코드북 m_j 와의 유클리드 거리(Euclidean distance)이다. 교차점은 이진수의 LSB에서 MSB까지의 70% 범위에서 랜덤한 위치를 갖는다. 이것은 MSB의 보호차원에서 생성된 숫자라고 할 수 있다.

돌연변이 및 생성은 기존의 유전자 알고리즘에서는 무작위로 선택하는 것에 대하여 Type I 조건에 있는 코드북 중에서 오분류에 영향을 많이 준, 즉, 에러 카운트가 일정 개수이상 발생하면, 그 때의 학습 패턴과 Ref2의 두 점을 이은 선분의 내분점중 계수 α 에 따르는 점을 구해 GA 알고리즘의 생성점으로 삼는다. 식 10은 생성되는 코드북의 생성범위를 나타내며 그림4에 도식적으로 나타내었다.

도태는 생성과 밀접한 관계를 갖는 것으로 오인식률에 가장 크게 영향을 미치는 코드북을 선택하여

$$q_1 = q_{x_1, y_1} = q_{x, y} \pm \alpha \cdot \text{dist}(p_{x, y}, q_{x, y}) \quad (10)$$

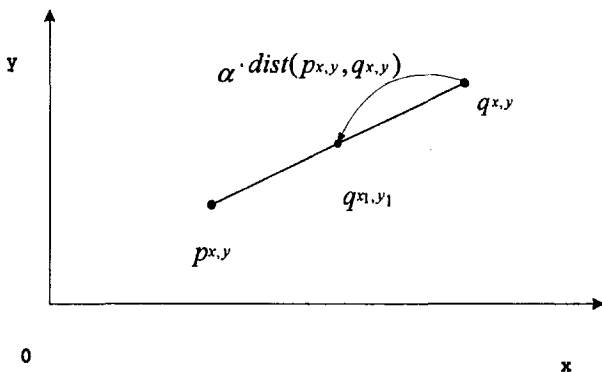


그림 4. 코드북 생성

도태시킨다. 도태 조건으로는 Type 정의에서 I, II의 요인으로 동작되는 코드북에 인덱스 카운터를 증가시킨다. 그리고 반복 수행되는 동안 학습이 더 이상 수행되지 않고 평가가 수렴되었을 경우 인덱스된 코드북을 도태시킨다.

(4) 변형된 유전자 알고리즘 수행과정

유전자 알고리즘은 여러 종류의 최적화 문제에 뛰어난 성능을 보이는 반면 양질의 해를 얻는데는 많은 시간이 걸린다는 단점이 있다. 또한 분류기로서 왜곡도를 최소화하는 것보다는 오인식률을 최소화하는 것이 인식성능에서 더 효율적이라는 것이 증명되고 있다. 이에 본 논문에서는 초기 양질의 학습조건을 갖도록 양자화 방법을 이용하여 코드북을 설정하고 유전자 알고리즘이 갖는 뛰어난 최적화 기술에 선택적인 학습을 통한 알고리즘을 제안한다. 학습기능을 갖는 변형된 유전자 알고리즘은 다음과 같다.

- 단계 1. 학습 데이터로부터 초기 코드북 및 파라미터를 초기화한다.
- 단계 2. GA의 반복횟수 만큼 단계 3-5를 반복한다.
- 단계 3. 주어진 학습패턴과 코드북들과의 유클리드인 거리를 구한 후 가장 가까운 코드북(Ref1)와 두 번째 가까운 코드북(Ref2)를 구한다.
- 단계 4. 타입의 정의에서 Type을 판정한다.
 - 1) Type I의 경우 :
 - ① Ref2의 코드북의 에러 인덱스에 1을 카운트한다.
 - ② 윈도우 안에 떨어진 경우 학습패턴과 Ref2와 일점교차를 수행한다.
 - 2) Type II의 경우 : 주어진 패턴은 학습 및 교차시 제외시킨다.
- 단계 5. 코드북의 에러 인덱스가 임의의 값 이상이고, 학습이 더 이상 수행되지 않을 시 Ref2와 주어진 패턴 사이에 가중치를 둔 내분점 구한 후 새로운 코드북을 생성한다

4. 음소 인식 실험

4.1 한국어 음소 데이터

(1) 낭독음성 데이터 베이스

실험에 사용된 음소 데이터 원광대학교 낭독음성 DB 보급판의 음성 중 단독 숫자만을 선택하여 모음 음소 부분을 추출하였다. 낭독음성 DB 보급판은 72

명을 대상으로 방음 부스에서 Senheizer HMD224X 를 사용하여 녹음하였으며, 발생된 데이터는 디지털 오디오 테이프에 저장하였다. A/D는 PC 환경에서 실시하였으며, AD/DA Module 은 KAY CSL 4300B 를 사용하였다. 그리고 16KHz 로 sampling 하고 16Bit로 양자화 하였다.

단독숫자 발성 데이터 양은 일반일 70명 × 4회, 아나운서 2명 × 2회씩 발음하여 녹음하였다. 음성 데이터를 저장하고 있는 각각의 파일들은 음성 데이터에 대한 정보를 유지하기 위하여 헤더를 갖는다. 헤더 구조는 미국의 DARPA-TIMIT Speech Database의 헤더 구조를 참조하여 상호 호환성을 유지하도록 구성되어 있다. 헤더의 크기는 1024 바이트로 고정되어 있으며, 헤더 첫 부분의 고정 영역을 제외하고는 필드의 추가, 삭제 주석의 첨가 등이 자유롭게 이루어질 수 있도록 되어있다. 음성 데이터 베이스의 대상 발성자는 20대 ~ 40대의 남녀로 한정되어 있으며, 음소 단위로 Segmentation 하여 음소 인식을 위한 음소 데이터 베이스를 구축하였다. 발성 화자의 성별, 연령별 기초 통계자료는 표 5.2와 같다.

표 5.2 발성화자의 성별 연령별 통계

	남	여	계
10 대	3	1	4
20 대	27	25	52
30 대	5	6	11
40 대	4	1	5
계	39	33	72

표 5.3 실험에 사용된 단모음

단 모 음	ㅏ ㅑ ㅓ ㅕ ㅡ ㅣ ㅗ ㅛ
-------	-----------------

(2) 한국어 모음 음소 추출 데이터

한국어 음소인식 실험을 위하여 낭독음성 데이터 베이스에서 특정 단어에서 제한된 모음 부분을 추출하였다. 한글 발음상의 특징과 구조에 의해서 한글을 음소별로 분리가 가능함에 착안하여, 자음과 모음으로 구성된 단음을 자음과 모음인 음소로 분리하고, 분리된 각각의 음소를 선형 예측법과 캡스트럼 추출법에 의해 분석하고, 분석된 데이터를 이용하여 단음

을 음소에 의하여 인식하는 실험을 행하였다.

본 논문에서는 한글이 단음을 자음과 모음인 음소로 분리하는 일반적인 방법을 사용하여 처리하여, 작은 데이터 양으로 처리 시간을 단축시켜 한글의 단음을 분석하고인식할 수 있게 하였다. 실험에 사용된 모음은 발음을 계속하여도 입 모양 또는 발음이 변화하지 않는 단모음을 사용하였다. 표 5.3은 실험에 사용된 모음 분류 표이다. 실험에 사용한 음소 데이터는 양자화된 숫자음성 데이터를 대상으로 자동 끝점 추출 알고리즘을 사용하여 필요한 모음 데이터를 찾아내고, 불량 데이터의 제거, 수정 데이터의 삽입, 음성 데이터의 앞뒤에 일정한 길이의 무음 구간의 확보등 전반적인 사항을 수정하여 음소 데이터를 DB화하였다. 그림 5.1의 파형은 낭독음성 DB의 숫자음 음성 파형과 각 숫자 음에서 모음 부분만을 추출한 음소 파형을 보여주고 있다.

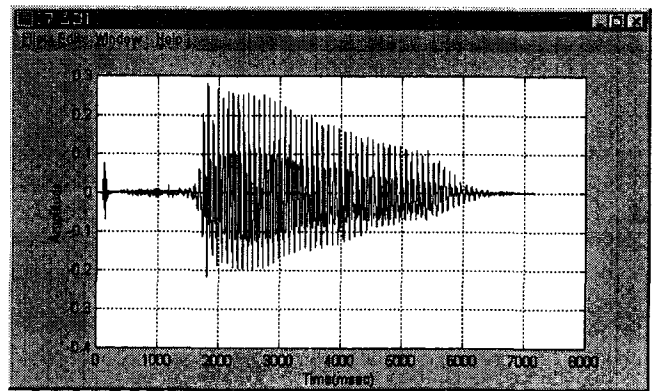


그림 5.1 낭독음성 "구"의 숫자음 음성 파형

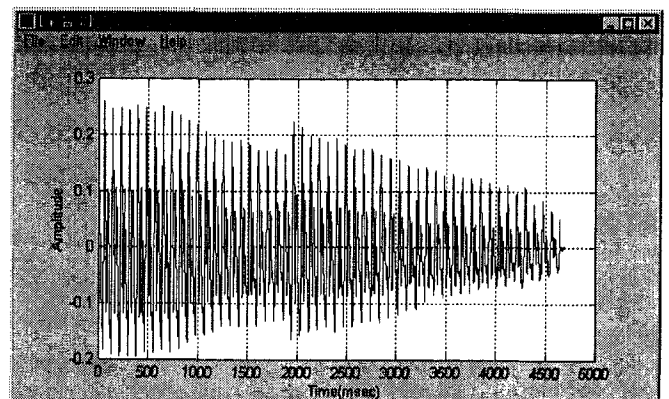


그림 5.2 낭독음성 "ㅓ"의 음소추출 파형

(3) 특징 파라미터 추출

낭독 음성에서 특정 모음 부분을 추출하여 전처리 과정을 통하여 LPC cepstrum 파라미터를 구하여 특징 벡터로 사용하였다. 추출된 모음 부분은 $H(z) = 1 - 0.97z^{-1}$ 의 Pre-emphasis 과정을 거침으로써 성도 특성을 부각시킨다. LPC 계수를 구하는 방법은 자기상관 함수를 이용한 Durbin 알고리즘을 사용했고 LPC 계수로부터 cepstrum 계수를 추출하였다. 표 3은 본 논문에서 사용한 모음 부분의 특징 벡터 생성과정의 음소데이터 분석 조건을 나타내었다.

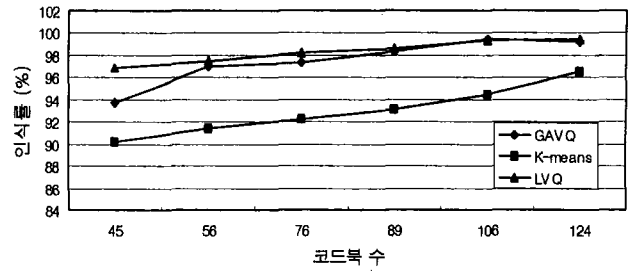
표 5.4 음소 데이터 분석 조건

Pre-emphasis	$H(z) = 1 - 0.97z^{-1}$
프레임 크기	2 frame (100 samples 중첩)
프레임 간격	300samples
Window	Hamming Window
LPC	10차
Cepstrum	10차

4.2 인식 결과

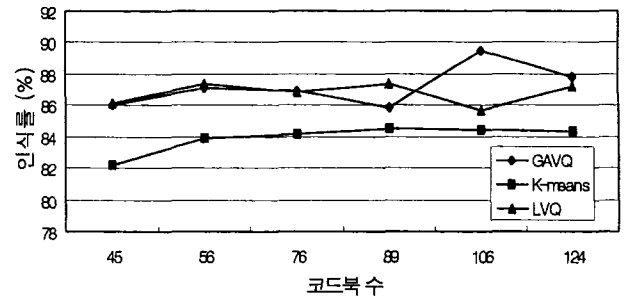
한국어 음소 데이터를 가지고 벡터양자화 방법 및 코드북 수 변화에 따른 인식실험 결과를 그림 6에 나타내었다. 양자화 방법에 있어 K-means는 한국어 음소 데이터의 각 클래스에 비례적으로 할당하였으며 LVQ 방법은 생성된 K-means 코드북 데이터를 가지고 인식성능 평가를 하였다. 여기서 학습주는 주어진 데이터의 100 epoch로 하였으며 윈도우 파라미터는 0.3, 학습 상수 파라미터는 0.3에서 단조 감소하도록 하였다. 또한 GAVQ는 초기의 임의로 설정된 코드북으로 양자화 과정을 거치면서 알고리즘의 특성상 고정된 코드북이 아닌 생성과 도태에 의해 최종적으로 설정된 코드북 수를 나타내어 인식평가를 하였다. 그림 6(a),(b)는 코드북 수 변화에 따른 각각의 VQ 방법에 대한 학습 데이터 및 테스트 데이터의 인식성능 실험 결과이다. 본 실험결과에서는 학습 데이터에 대한 인식성능 평가에서처럼 두 방법이 K-means 방법으로 코드북을 생성하였을 때 보다 테스트 데이터에 대한 인식성능이 좋음을 알 수 있었으며, 제안된 방법과 LVQ의 인식결과에서는 코드북 수가 적었을 때는 인식결과가 비슷하였지만 코드북 수가 증가하였을 때 GAVQ를 이용한 방법이 좋은 인식결과를 보였다.

한국어 음소(학습 데이터)



(a) 코드북 변화에 따른 학습데이터의 인식률

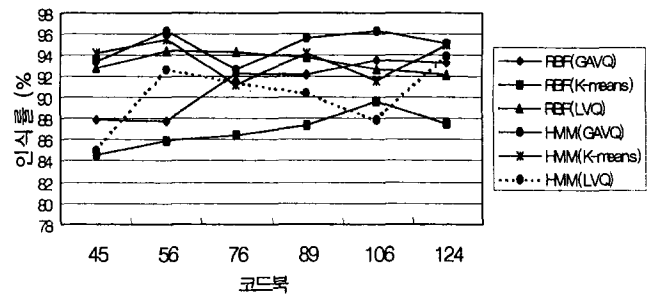
한국어 음소(테스트 데이터)



(b) 코드북 변화에 따른 테스트 데이터의 인식률

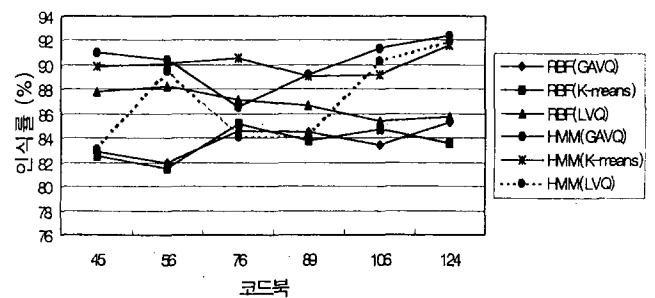
그림 6. VQ 방법에 따른 한국어 모음 인식

한국어 음소(학습 데이터)



(a) 인식모델에 따른 학습데이터 인식률 비교

한국어 음소(테스트 데이터)



(b) 인식모델에 따른 테스트 데이터 인식률 비교

그림 7. VQ와 인식모델에 따른 인식률 비교

표 4와 표 5는 VQ 방법 및 인식모델에 따른 한국어 음소인식에 대하여 코드북 수 변화에 따른 가장 좋은 인식결과를 보였을 때의 인식결과이다.

표 4. VQ 방법에 따른 한국어 음소인식(인식률/코드북수)

VQ 방법	학습 데이터	테스트 데이터
K-means	96.48% (124)	84.48% (89)
LVQ	99.32% (106)	87.39% (89)
GAVQ	99.46% (106)	89.46% (106)

표 5. 인식모델에 따른 한국어 음소인식(인식률/코드북수)

인식모델	학습데이터	테스트 데이터
RBF(GAVQ)	93.53% (106)	85.30% (124)
RBF(K-means)	89.60% (106)	85.15% (76)
RBF(LVQ)	94.45% (56)	88.26% (56)
HMM(GAVQ)	96.32% (56)	92.34% (124)
HMM(K-means)	95.50% (56)	91.54% (124)
HMM(LVQ)	93.86% (124)	91.94% (124)

5. 결론

본 논문에서는 음성인식 시스템에서의 VQ를 기반으로 하는 기존의 모델들에 대하여 일반적인 벡터양자화 방법과 자연선택과 자연계 진화과정의 적자생존을 원칙으로 하는 적용적인 탐색기법으로 최적해에 대한 효과적인 탐색도구로 널리 알려진 유전자 알고리즘을 변형한 제안된 유전자 알고리즘을 이용하여 VQ방법에 따른 인식모델들의 인식성능 평가 실험을 하였다. 실험결과에서도 알 수 있듯이 제안된 유전자 알고리즘이 일반적인 K-means 알고리즘보다 전반적으로 인식성능이 좋음을 알 수 있었으며 VQ를 기반으로 하는 모델에서도 일반적으로 인식성능이 우수함을 알 수 있었다. 결과적으로 인식시스템에서의 초기 코드북 설정 방법과 초기 코드북 수의 결정이 인식성능에 큰 영향을 주는 요소임을 알 수 있었다. 또한 각 인식모델들에 대한 실험에서 일반적으로 HMM의 인식성능이 우수함을 알 수 있었지만 학습데이터에서 각 클래스에 대한 데이터의 수 및 코드북 수에 따라 인식성능에 큰 요소로 작용됨을 알

수 있었다.

결론적으로, 제안된 유전자 알고리즘을 이용한 벡터양자화 방법이 기존의 방법보다 우수함을 실험결과를 통해 확인할 수 있었으며, 음성인식 시스템에서의 전처리 과정 및 인식모델 모든 분야에서의 연구 및 개발이 필요함을 알 수 있었다. 또한 음성인식은 학문적 측면뿐만 아니라 실제적인 응용분야에서 그 필요성이 강조되고 있다. 실제로 미국, 일본 등의 다른 선진국에서의 음성인식 분야에 대한 관심과 투자는 매우 크며, 많은 연구인력이 이 분야에 종사하고 있으나, 국내에서는 아직 그 기술의 수준과 연구성과가 미흡한 형편이며 과감한 투자와 연구가 신속하고 지속적으로 이루어져야 된다고 본다.

참고 문헌

- [1] 구명환, "음성인식 기술의 현황과 전망", *정보과학회지* PP21-34, 1993.
- [2] L. R. Rabiner and R. W. Schager, *Digital Processing of Speech Signals*, Prentice-hall, 1990.
- [3] R. P. Lippman, "Review of Neural Networks for Speech Recognition", *Neural Computation*, Vol.1, pp. 1-38, 1989.
- [4] D. P. Morgan and C. L. Scofield, *Neural Networks and Speech Processing*, Kluwer Academic Publishers, 1991.
- [5] T.Kohonen, "The self-organizing map", *Proc. IEEE*, Vol.78, No.9, pp.1464-1480, 1990.
- [6] 오영환, *패턴인식론*, 정익사, 1986
- [8] Lawrence Davis, *Handbook of genetic algorithms*, Van Nostrand Reinhold, New York, 1991.
- [9] D. Goldberg and R. Lingle, "Allecles loci, and the traveling salesman problem", *Proc. First International Conf. on Genetic Algorithms and Their Applications*, Lawrence Erlbaum, pp.154-159, 1985.
- [10] D. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning", Addison Wesley, 1989.