

# 개선된 MSVQ 인식 시스템을 이용한 단독어 인식에 관한 연구

正會員 安 泰 玉\* 正會員 金 南 中\*\* 正會員 宋 鐵\*\* 正會員 金 淳 協\*

## A Study on Isolated Word Recognition using Improved Multisection Vector Quantization Recognition System

Tae Ok AN\*, Nam Joong KIM\*\*, Chul SONG\*\*, Soon Hyeob KIM\* *Regular Members*

**要 約** 본 논문은 화자 독립의 단독어 인식에 관한 연구로 기존의 MSVQ(multisection vector quantization) 인식 시스템을 개선한 새로운 MSVQ 시스템을 제안한다. 제안된 내용은 기존의 시스템과는 달리 인식시 시험패턴의 구간 수를 표준패턴의 구간 수보다 한 구간 더 늘리는 것이다. 이 방법에 의한 시험지 인식 대상어로 146개의 DDD 지역명을 선택했으며, 특정 카라메타로는 12차 LPC cepstral(cepstrum) 계수를 사용했고, 코드북 생성시 종합전 구간의 방법으로 MINSUM과 MINIMAX 기법을 사용하였다. 시험 결과에 의하면 DTW(dynamic time warping) 패턴 매칭 방법, VQ(vector quantization)에 의한 방법은 물론 기존의 MSVQ 방법보다 개선량이 감소량과 동시에 더 높은 인식률을 얻을 수 있었다.

**ABSTRACT** This paper is a study on the isolated word recognition of speaker independent, which proposes to newly improved MSVQ(multisection vector quantization) recognition system which improve the classical MSVQ recognition system. It is a difference that test pattern has one more section than reference pattern in recognition system. 146 DDD area names are selected as recognition vocabulary, 12th LPC cepstral coefficients is used as feature parameter, and when codebook is generated, MINSUM and MINIMAX are used in finding the centroid. According to the experment result, it is proved that this method is better than VQ(vector quantization) recognition methods, DTW(dynamic time warping) pattern matching methods and classical MSVQ methods for recognition rate and recognition time.

### I. 서 론

종래의 음성인식 시스템에는 템플릿과 입력 음성간의 매칭을 위해 DTW 알고리즘을 많이 사용하였다.<sup>1)</sup> 그러나, DTW는 인식률은 높으나 계산량이 많다는 단점이 있다.

이에 따라 음성 처리에 새로운 접근이 시도되었는데 그중 하나가 VQ이다.<sup>2-4)</sup> 그러나, 이 VQ를 이용한 음성 인식기는 어떤 표준 음성이 다른 음성의 특징을 모두 포함하게 되면 오인식될 가능성이 증가된다.

그러므로, 입력 음성과 표준 음성간의 대략적인 매칭을 위하여 음성을 어떤 길이의 구간으로 나누어 구간별로 유사도를 비교하는 MSVQ 코드북이 나오게 되었는데<sup>5,6)</sup>, 이 MSVQ 코드북을 새로운 관점에서 검토하여 개선된 MSVQ 음성인식 시스템을 제안하고자 한다.

본 연구는 중규모 어휘의 음성 다이얼링(voice dialing) 시스템을 위해서 146개의 DDD 지역명을 인식 대상으로한 화자 독립의 한국어 음성 인식에 관한 연구를 행하는 것으로, 이런 중규모 이하의 단어에 있어서는 음소나 음절 단위와 같이 서브워드(Subword) 단위로 패턴 매칭을 행하는 것보다, 단어 단위로 음성인식을 행하는 것이 더 실용적이라 할 수 있다. 그런데, 이러한 단어 단위의 음성인식에 있어서 기존의 DP(

\* 光 大 大 學 校 電 子 計 算 學 科  
Dept. of Computer Science, Kwangwoon University

\*\* 朝 鮮 大 學 校 電 算 機 工 學 科  
Dept. of Computer Science, Chosun University.  
論 文 番 號 : 91 - 18 (接 受 1990. 12. 21)

dynamic programming)에 의한 인식은 인식 시간이 너무 길고 기억 용량을 많이 필요로 하는 단점이 있다. 이러한 단점을 보완하기 위해 인식률, 계산 속도, 기억 용량 등에서 우수한 개선된 MSVQ 음성인식 시스템으로, 앞으로 다가올 인간과 기계와의 통신 시대에 대비하여 기계가 우리 인간의 언어를 자연스럽게 인식할 수 있도록 하기 위한 일환으로 단국어 인식 시스템을 개발하는데 목표를 두고 이 개선된 MSVQ 음성 인식 방법을 제안한다.

## II. VQ 이론

### 1. 개요

VQ란 벡터 계열을 통신이나 디지털 채널에 저장 하기에 적당한 디지털 계열 과 매핑하기 위한 시스템이다.<sup>(23)</sup> VQ의 가장 큰 목적은 데이터 압축으로 데이터의 신뢰성을 잃지 않고, 비트율(bit rate)을 최대한으로 줄이는데 있다. 데이터 압축에 기여한 Shannon의 왜곡율(rate distortion) 이론에 의하면 스칼라 대신에 벡터를 코딩함으로써 더 좋은 성능을 얻을 수 있다는 것이다.<sup>(7)</sup> 따라서, 음성인식에 있어서 데이터 압축이라는 측면에서 표준패턴을 생성할 때 VQ를 이용한다. 즉, 음성 인식에서의 VQ는 입력된 음성의 특징 벡터를 미리 저장해 둔 특징 벡터 중에서 가장 유사도가 높은 하나의 벡터와 매핑시켜 주는 것이다.

### 2. 학습용 데이터

VQ를 이용한 음성인식 시스템에서 학습용 데이터를 구성하는 방법은 두가지가 있다. 하나는 단구간(single section) VQ 코드북을 만들기 위한 구성법이고, 또 다른 하나는 MSVQ 코드북을 만들기 위한 구성법이다.

단구간 VQ 코드북이란 한 단어당 하나의 코드북으로 표준 템플리트를 형성하는 방법이고, MSVQ 코드북이란 한 단어당 2개 이상의 코드북으로 표준 템플리트를 형성하는 방법

이다.<sup>(6)</sup> 그러므로, 매 단어의 표준패턴을 만들기 위해서는 두개 이상의 학습용 데이터가 필요하다.

### 3. 코드북 작성

VQ를 이용한 음성인식 시스템에서 코드북은 곧 표준 템플리트가 되므로 학습용 데이터의 특징이 잘 나타나도록 코드북을 만들어야 한다. MSVQ의 코드북 작성은 불특정 화자의 인식을 위해 여러 사람이 발음한 전체 데이터 분할(splitting) 기법을 사용하여 작성 하였으며<sup>(8)</sup> 중심점 설정 방법으로는 MINSUM과 MIN-IMAX 방법을 사용하였고<sup>(10)(11)</sup>, 음성 인식을 위한 특징 파라미터는 여러가지가 있지만 이들 파라미터를 상호 비교한 Shikano 등의 실험 결과에 의하면, LPC 켈스트럼 이 LPC<sup>(12)(13)</sup>에 비해 계산량이 훨씬 적고, 인식률도 우수하다는 고찰에 따라, LPC 켈스트럼을 특징 파라미터로 사용하였다<sup>(14)</sup>.

### 4. 거리값 계산

LPC 켈스트럼 벡터의 학습용 데이터 집합을  $\{C_i, i=1, \dots, I\}$ 라 하자. 이 벡터들은 어휘에서의 단어들이 다양한 화자에 의해서 발음될 때 일어나는 LPC 켈스트럼이다. VQ가 내포한 주요 개념은 주어진 M개의 벡터에 대하여 가장 가까이 있는 코드북 엔트리C에 의해 학습용 데이터 집합 벡터 C의 각각에 대한 평균 거리가 최소가 되도록 LPC 켈스트럼 벡터로 최적의 코드북의 집합을 결정하는 것이다. 공식적으로 두 LPC 켈스트럼 벡터  $C_m$ 과  $C_l$ 간의 거리를  $d(C_m, C_l)$ 라고 정의한다. 따라서 최적의 코드북 집합을 구하는 식은,

$$|D_M| = \min \{I/I \sum_{i=1}^I \min [d(C_m, C_i)]\} \quad (1)$$

로 나타낼 수 있다.  $|D_M|$ 은 벡터 양자화의 평균 거리이다. 위식 (1)에서 M(코드북 엔트리) 값에 대한 최적의 해를 발견한다. P차 LPC 켈스트럼

벡터를 사용할 경우 시스템에서 사용하는 국부적인 거리는 다음과 같다.

$$d(C_r, C_t) = w(C_{r0} - C_{t0}) + \sum_{i=1}^L (C_{ri} - C_{ti})^2 \quad (2)$$

여기서,  $w$ 는 가중치로서 본 연구에서는 0.04로 설정하였다.

### 5. 단어 인식

시험용 음성 데이터와 모든 표준패턴과의 거리를 구한 후, 시험용 음성 데이터는 가장 작은 거리를 갖는 표준 패턴의 단어를 시험용 데이터로 인식한다.

## III. MSVQ 이론

### 1. MSVQ의 개요

단어 음성 인식에서는 발성 속도에 따른 시간 변동의 제거를 위해 DP 매칭이 많이 이용되고 있다. 이 방법은 시간축의 선형변화에 따른 계산량이 증가한다. 그러므로 시간 정규화가 필요 없는 단어별로 작성된 VQ 코드북에 의해 단어들의 음향적인 특성만을 비교하는 방법을 쓴다. 그러나, 코드북에는 시간적 정보가 포함 되어 있지 않기 때문에 음향적 특성이 유사한 단어들 사이에 부정확한 인식이 일어난다. 따라서 한 단어를 발성 순서에 따라 몇개의 구간으로 나누고 구간 별로 독립된 코드북을 작성함으로써 시간적 정보를 포함시키는 MSVQ를 Burton 등이 제안하였다.<sup>(5)(6)</sup>

Burton의 MSVQ에 따르면 MSVQ 코드북 작성에 따르는 모든 음성은 발성 기간에 관계없이 일정수의 정해진 길이를 갖는 프레임으로 정규화 되어야 한다.

### 2. 기존의 MSVQ 코드북 작성

단어의 MSVQ 코드북은 그 단어를 동일 길이의 구간으로 나누고 각 구간 마다 집탄화(clustering) 알고리즘을 이용하여 작성한다.

### (1) 4 MSVQ 코드북 작성

그림 1에 4 MSVQ 코드북을 작성하는 과정이 나타나 있다. 한 단어를  $W$ 를 1회 발성된 음성을 학습용 데이터로 사용해서, 한 프레임을 LPC 분석하여 얻은 켈스트럼 벡터를  $v_k$ 라 하면 1회 발성된 음성은

$$W = \{v_{11}, v_{21}, v_{31}, \dots, v_{k1}\} \quad (3)$$

와 같이 나타낼 수 있다. 인식 대상 어휘가 모두  $L$ 개의 단어로 되어 있을 때 각 단어마다 1회 발성된 음성으로 MSVQ 코드북을 구성하기 위해 이들을  $L$ 개의 구간으로 나눈다.

$$W_l(i) = \{V_1(i), V_2(i), \dots, V_j(i)\} \quad (i=1,2,\dots,L) \quad (4)$$

연약 한 구간이  $N$  프레임으로 구성 되어 있다면

$$V_1(i) = \{v_{11}(i), v_{21}(i), v_{31}(i), \dots, v_{N1}(i)\} \quad (5-1)$$

$$V_j(i) = \{v_{1j}(i), v_{2j}(i), v_{3j}(i), \dots, v_{Nj}(i)\} \quad (5-2)$$

와 같이 각 구간을 벡터열로 표시 할 수 있다. 그림 1에서 보는 바와 같이 각 구간의 프레임 수는 다르다. 4 구간으로 되어 있으므로 4개의 독립된 VQ 코드북의 조합에 의해 MSVQ 코드북이 구성된다. 구간  $j$ 에 해당하는 학습용 데이터의 집합을  $S_j$ 라 하면

$$S_j = \{V_j(1), V_j(2), \dots, V_j(L)\} \quad (j=1,2,3,4) \quad (6)$$

이 된다. 각 구간에 대한 코드북  $C_j$ 는  $S_j$ 를 학습용 데이터로 하여서 분할 알고리즘에 의해 작성된다. 이때, 각 구간 코드북  $C_j$ 는 4개의 코드워드(codeword)로 이루어졌다. 이 과정을 통해 작성된 코드북의 개념

$$C = \{C_1, C_2, C_3, C_4\} \quad (7)$$

는 MSVQ 코드북을 의미한다.

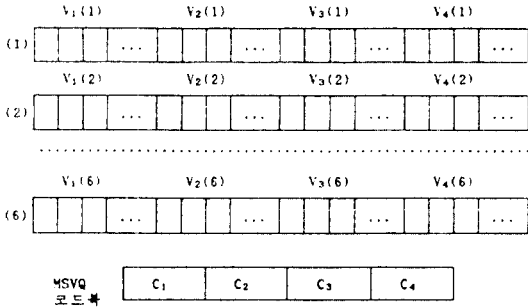


그림 1. 4 MSVQ 코드북 작성  
Fig. 1. 4 MSVQ codebook generation.

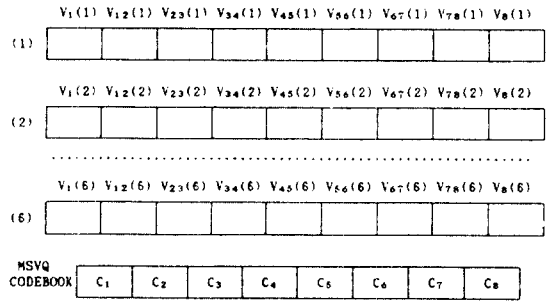


그림 2. 8 Overlapped MSVQ 코드북 작성  
Fig. 2. 8 Overlapped MSVQ codebook generation

(2) 8 MSVQ 코드북 작성 및 16 MSVQ 코드북 작성

8 MSVQ 코드북을 만드는 과정은 4 MSVQ 코드북을 만드는 과정과 동일하나 단지 구간의 수를 4에서 8로 늘리는 것만이 다르다. 그래서 8개의 각 구간 코드북이 모여서 한 단어의 전체 코드북을 이루게 된다. 각 구간 코드북  $C_j$ 는 2개의 코드워드로 이루어진다.

16 MSVQ 코드북을 만드는 과정도 8 MSVQ 코드북을 만드는 과정과 동일하나 단지 구간의 수를 8에서 16으로 늘리는 것만이 다르다. 그래서 16개의 각 구간 코드북이 모여서 한 단어의 전체 코드북을 이루게 된다. 각 구간 코드북  $C_j$ 는 1개의 코드워드로 이루어진다.

(3) 8 중첩(overlapped) MSVQ 코드북 작성

8 MSVQ 코드북을 만드는 것처럼 이 MSVQ 코드북을 작성하는 과정이 그림 2에 나타나 있다. 코드북을 작성하는 방법은 8 MSVQ와 동일하나  $V_{12}$ 번째 구간부터 중첩시키면서 코드북을 만드는 방법만이 다르다. 즉,  $V_{12}$ 는 구간 1에도 속하고 구간 2에도 속한다. 마찬가지로  $V_{28}$ 은 구간 7에도 속하고 구간 8에도 속한다. 중첩시키 나가는 이유는 구간으로 분할할때 분할될 부분의 정보 손실을 우려하여 9개의 구간으로 나눈 후 중첩 하였다. 이렇게 9구간으로 나눈후 중첩시켜 8구간으로 코드북을 만든다.

3. 개선된 MSVQ 코드북 작성

개선된 MSVQ의 시험패턴은 기존의 시험패턴과 달리 학습용 데이터로 코드북을 작성할 때보다 한 구간을 더 두는 방법으로, 첫번째와 마지막 구간은 다른 구간의 크기의 절반이며, 그 나머지 구간은 코드북 작성시 학습용 데이터의 구간의 크기와 같다. 따라서, 시험 패턴이 표준패턴과 매칭시에는 자기 구간과 같은 구간의 코드워드와 비교할 뿐만 아니라 그보다 하나 앞선 구간의 코드워드와도 비교하여 가장 거리값이 작은 것과 매칭하는 것으로 한다. 그 이유는 어떤 구간 부분의 패턴이 다른 시험패턴의 구간 부분의 패턴과 1대1로 매칭되지 않기 때문에, 이를 고려해서 인식시키는 방법이다.

(1) 4 MSVQ 코드북 작성 및 5구간에 의한 인식

그림 3에서 보는 바와 같이 코드북을 작성은 4 MSVQ 코드북을 작성 과 동일하나 인식시에는 시험패턴을 5구간으로 해서 매칭하는데 이 시험패턴의 5구간중 첫번째 구간과 마지막 구간은 나머지 3구간의 크기의 절반이며, 나머지 3구간은 코드북 작성시 학습용 데이터의 4구간의 크기와 같다. 따라서, 시험패턴의 표준패턴과의 매칭 관계는 8중첩 MSVQ 방법과 같이 중첩시켜 나가면서 인식시킨다.

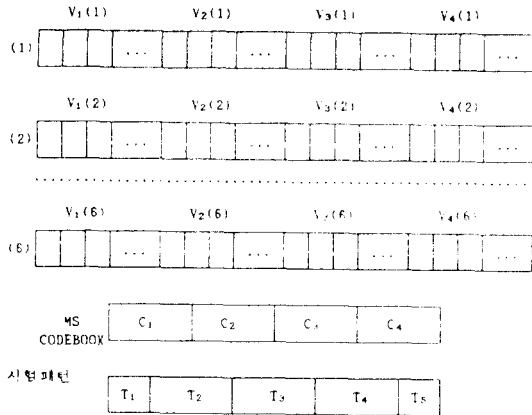


그림 3. 4 MSVQ 코드북 작성  
Fig. 3. 4 MSVQ codebook generation.

(2) 8 MSVQ 코드북 작성 및 9구간에 의한 인식

본 연구에서 제안하는 개선된 MSVQ인식 시스템에서 코드북을 만드는 과정은 8 MSVQ 코드북을 만드는 과정은 동일하나 인식시에는 그림 3에서처럼 시험 패턴을 9구간으로 해서 매칭하는데 이 9구간의 첫번째와 마지막 구간은 나머지 7구간이 크기의 절반이며, 나머지 7구간은 코드북 작성시 학습용 데이터의 8구간 크기와 같다. 인식시 시험패턴의 표준패턴과의 매칭 관계는 8 중첩 MSVQ 방법과 같이 중첩시키 나가면서 인식시킨다.

4. MSVQ에 의한 단어 인식

MSVQ 역시 일반적인 VQ 방법에 의한 거리 값 계산과 동일하나 구간별 거리값을 합하여 전체의 왜곡률(total distortion)로 최적의 코드북을 찾는 것이 다르다. 인식하고자 하는 시험용 입력 음성  $W_x$ 는 먼저 전체 프레임을 구한 후 J개의 구간으로 나뉜다. 이것을 벡터일로 나타내면

$$W_x = \{X_1, X_2, X_3, \dots, X_J\} \quad (8)$$

가 되고,  $X_j$ 는 j번째 구간을 구성하는 프레임들로

부터 LPC 분석을 통해 구한 특정 벡터의 시계열이다. 따라서  $X_j$ 는

$$X_j = \{x_{j-1N+1}, x_{j-1N+2}, \dots, x_{jN}\} \quad (9)$$

N: j번째 구간의 프레임 수

로 표현할 수 있다. 이들 각 구간에 대한 특정 벡터들을 표준 패턴의 상대 구간 코드북의 코드워드들과 거대 비교를 통해 전체 평균거리인  $D_{av}$ 를 구한다. 이면 단어 l에 대한 표준 패턴과 전체 평균 거리는

$$D_{lav} = 1/N \sum_{j=1}^J d_j(X_j, C_{lj})^2 \quad (10)$$

이며

$$d_j = \sum_{i=1}^M \min d(x_i, C_{lj}) \quad (11)$$

이다.  $C_{lj}$ 는 단어 l의 j번째 구간 코드북의 한 코드워드를 나타낸다. 이상의 과정을 모든 단어의 표준패턴에 대하여 반복하여 최종적으로 전체 평균 거리가 최소인 단어를 인식된 단어로 한다.

IV. 실험 및 결과

1. 실험조건 및 대상어

VQ에 의한 불특정 화자의 음성 인식을 하는데 있어서 대상 어휘로는 146개의 DDD 번호에 의한 지역명을 선정 하였고, 3명의 남성 화자에 의해 각각 3번째 발음된 것 중에서 각각 2번째 발음된 것으로 코드북을 작성 하였다. 그리고, 나머지 각각 1번째 발음된 것으로 인식 실험을 하였다.

본 부분에서 사용한 단어 인식 시스템은 그림 4과 같다. 마이크로를 통하여 입력된 신호는 샘플링 주파수를 8KHz로 하였으며 3.5KHz 저역 여파기를 통과한 후 12bit A/D 변환을 거쳐 음성신호를 구한 다음, 직각과 90 구간을 검출하

고 12차 LPC 켈스트럼 계수를 구한 후 코드북을 작성한다.

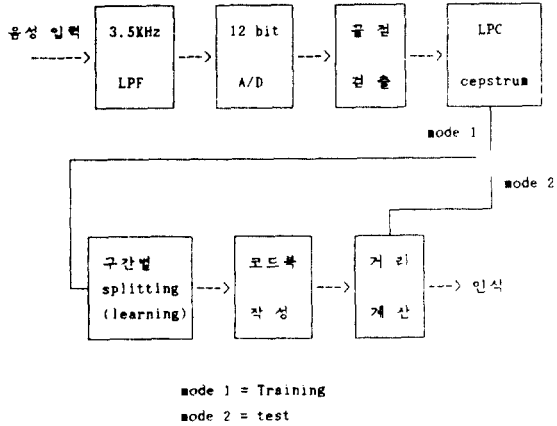


그림 4. 인식 시스템의 구성도.  
Fig. 4. Block diagram of recognition System.

## 2. 인식 실험

인식 실험시 사용된 데이터는 제안된 시스템에 의한 인식에서와 똑같이 마이크로 입력된 데이터로 3명의 화자가 전국 DDD 지역명 146개 단어를 각각 3번씩 발음한 총 1,314개로 그 중 각 화자가 두번씩 각 단어에 대해 발음한 876개는 단어의 모델을 작성하는데 이용되었으며, 나머지 각 화자가 1번씩 발음한 438개는 인식하는데 이용하였다.

### (1) DTW에 의한 인식 실험 결과

DTW 인식에 있어서 기준 패턴을 선택하는 방법으로 3가지를 선택하여 실험하였다. 첫번째 방법은 6개의 토큰 중에 하나를 C언어의 랜덤 함수를 이용하여 선택하였으며, 두번째 방법은 6개의 토큰을 집단화 기법에 의해 거리값이 가장 작은 것으로 나타난 하나의 토큰을 선택하여 기준 패턴으로 선택하였으며 마지막으로 세번째 방법은 집단화 방법에 의해서 새로운 기준 패턴을 정하는 방법으로 인식 실험을 했다. 그 인식 결과를 표 1에 나타내었다.

표 1. DTW의 인식 결과.  
Table 1. Recognition Result of DTW. (단위:%)

인식방법 / 화자	무차위에 의한 오인식 인식률	Clustering 후 1개 선택 인식률	Clustering 방법 인식률
화자 1	84.2	75.3	95.9
화자 2	73.3	64.4	90.4
화자 3	78.1	96.6	93.8
총인식률	78.5	78.8	93.4

### (2) VQ 이론에 의한 인식 실험 결과

VQ 이론에 의한 음성 인식은 단구간 VQ 방법과 MSVQ 방법으로 나눈다. 이때 MSVQ는 시간의 흐름에 대한 정보를 포함하고 있으므로 단구간 VQ 코드북에 의한 음성 인식보다 실험 결과 인식률이 좋은 것으로 나타났으며, MSVQ의 경우는 Burton이 사용한 방식에 의한 인식 방법과 개선된 방법에 의한 인식 실험으로 나누어 실험하였다.

이때, VQ에 의한 인식 시스템 모두에서 전체 코드워드의 수 16으로 하여 실험하였다. 코드워드를 구하기 위한 집단화에서 중심점을 구하는 방법으로는 MINSUM방법과 MINIMAX방법을 사용하였다.

### A. 단구간 VQ 코드북에 의한 인식실험

전체의 코드워드 즉 16개의 코드워드와 비교하여 인식한 결과를 표 2에 나타냈다.

표 2. 단구간 VQ 코드북에 의한 인식률.  
Table 2. Recognition Rate of Single Section VQ Codebook. (단위:%)

화자 / 방법	화자 1	화자 2	화자 3	총인식률
MINSUM	87.7	87.7	84.2	86.5
MINIMAX	77.4	76.0	76.71	76.7

### B. 기존의 MSVQ 코드북에 의한 인식실험

#### (B-1) 4 MSVQ의 인식실험

본 실험에서는 MSVQ 코드북을 작성하는데 있어 구간의 코드워드수를 4로해서 실험하였다. 여기서 각 구간별 코드북 크기를 4로해서 총

16개의 코드워드가 되는데, 인식시 가운데 구간을 중심으로 해서 좌우로 비교 되므로 12개의 코드워드가 비교된다. 이때의 인식결과를 표 3에 나타냈다.

표 3. MSVQ의 인식결과  
Table 3. Recognition result of 4 MSVQ. (단위:%)

회차 \ 방법	회차 1	회차 2	회차 3	총인식률
MINSUM	95.9	93.2	91.1	93.4
MINIMAX	88.1	89.7	88.1	88.8

(B 2) 8 MSVQ 및 16 MSVQ의 인식실험

본 실험에서는 MSVQ 코드북을 작성하는데 있어서 구간의 수를 8로해서 실험하였는데, 각 구간별 코드북 크기를 2로해서 총 16개의 코드워드가 되도록 하였다. 여기에서 비교되는 코드워드의 수는 구간당 2개의 코드워드가 있으므로 6개가 비교된다. 또한 구간을 16으로 해서도 실험하였는데 이 경우에는 각 구간별 코드워드의 수를 1로하여 실험하였다. 여기서도 비교되는 구간을 중심으로 해서 좌우로 각각 2개씩 비교되므로 5개의 코드워드가 비교된다. 이 방법에 의한 인식결과를 각각 표 4와 표 5에 나타냈다.

표 4. 8 MSVQ의 인식결과  
Table 4. Recognition result of 8 MSVQ. (단위:%)

회차 \ 방법	회차 1	회차 2	회차 3	총인식률
MINSUM	93.2	95.2	91.8	93.4
MINIMAX	93.2	91.1	90.1	91.6

표 5. 16 MSVQ의 인식결과  
Table 5. Recognition result of 16 MSVQ. (단위:%)

회차 \ 방법	회차 1	회차 2	회차 3	총인식률
MINSUM	91.1	86.3	83.6	86.9
MINIMAX	86.3	85.6	85.6	85.8

(B 3) 8중첩 MSVQ의 인식실험

본 실험에서는 다구간 코드북을 작성할 때

구간으로 분할되는 부분의 정보손실을 우려하여 구간을 중첩하면서 코드북을 작성 하였다. 따라서 각 구간별 코드북 크기를 2로 하여 총 16개의 코드워드를 태웠으며, 인식시 자기 구간을 포함해서 일련 구간 중 가까이 중첩되는 구간의 코드워드가 선택하여 4개의 코드워드와 비교하게 하였다. 이렇게 구성된 코드북에 의한 인식 결과는 표 6에 나타냈다.

표 6. 8 중첩 MSVQ의 인식결과  
Table 6. Recognition result of 8 overlapped MSVQ.

(단위:%)

회차 \ 방법	회차 1	회차 2	회차 3	총인식률
MINSUM	92.5	89.7	86.3	89.5
MINIMAX	80.1	81.5	80.8	78.5

C. 개선된 MSVQ 코드북에 의한 인식실험

(C 1) 4 MSVQ 코드북 작성 및 5 구간에 시험패턴에 의한 인식실험

본 실험에서는 MSVQ 코드북을 작성할때 4 구간으로 분할하였고 각 구간 별 코드북 크기를 4로 하였으므로 총 16개의 코드워드가 태워지며, 인식시에 시험패턴을 5구간하여 첫번째 구간과 마지막 구간은 4개의 코드워드와 비교되고, 나머지 구간들은 8개의 코드워드와 비교됨으로써 정보의 손실을 방지하였다. 인식결과를 아래의 표 7에 나타냈다.

표 7. MSVQ 코드북과 5구간과 시험패턴에 의한 인식결과  
Table 8. Recognition result of 4 MSVQ codebook and test patternot 5 Section. (단위:%)

회차 \ 방법	회차 1	회차 2	회차 3	총인식률
MINSUM	91.5	93.2	93.8	93.8
MINIMAX	86.3	82.9	89.0	86.1

(C 2) 8 MSVQ 코드북 작성 및 9구간의 시험패턴에 의한 인식실험

본 실험에서는 MSVQ 코드북을 작성할때

8구간으로 분할하였고 각 구간별 코드북 크기를 2로 하였으므로 총 16개의 코드워드가 택해지며, 인식시에 시험패턴의 구간을 9구간으로 하여 첫번째 구간과 마지막 구간은 2개의 코드워드와 비교되고, 나머지 구간들은 4개의 코드워드와 비교되도록 하였다. 인식결과와 아래의 표 8에 나타냈다.

표 8. MSVQ 코드북과 9구간과 시험패턴과 의한 인식결과  
Table 8. Recognition result of 8 MSVQ codebook and test pattern of 9 Section. (단위:%)

화자 \ 방법	화자 1	화자 2	화자 3	총인식률
MINSUM	93.2	95.9	92.5	93.8
MINIMAX	94.5	90.4	90.4	91.8

#### D. 종합적인 실험결과

표 1과 표 9의 인식율을 비교해 보면 표 9의 4-5구간과 8-9구간의 인식율이 더 좋음을 알 수 있다. 표 9에서 4-5구간과 8-9구간의 인식율이 동일하게 나타났지만, 8-9구간이 4개의 구간으로 나누어 VQ를 구하고 5개의 구간으로 나누어 중첩시키면서 인식하는 4-5구간의 방법보다 코드북 작성이 용이하고, 코드워드 비교시 4-5구간이 8개의 코드워드를 비교하는 반면에 8-9구간은 4개의 코드워드만을 비교하기 때문에 처리시간도 훨씬 단축됨을 알 수 있다. 한편 인식 시간면에서도 VQ나 기존의 방법보다 비교되는 코드워드의 수가 훨씬 적기때문에 시간이 단축됨을 알 수 있다.

표 9. 전체의 인식율 비교.  
Table 9. Compare of Total Recognition Rate. (단위:%)

중심점 \ 방법	기존의 MSVQ 방법				Overlapped 8 구간	개선된 MSVQ방법	
	VQ	4 구간	8 구간	16구간		4-5 구간	8-9 구간
MINSUM	86.5	93.4	93.4	86.9	89.5	93.8	93.8
MINIMAX	76.7	88.8	91.6	85.8	78.5	86.1	91.8
비교되는 코드워드수	16	12	6	5	4	8	4

## V. 결 론

본 논문에서는 VQ 방식을 이용하여 단구간 코드북과 MSVQ 코드북을 만들어 실험하였으며, 더 나아가 MSVQ에 의해 구해진 각각의 코드워드에 대해서도 중복의 개념을 도입하여 실험하는 새로운 인식 방법을 제안하였다. 또한 DTW 패턴 매칭 방법과 비교하여 개선된 MSVQ 인식 방법의 유용성을 실험하였다.

단구간 코드북일 경우 음향학적 특성이 유사한 단어들 사이에 오인식되는 경우가 빈번이 발생하여 인식율을 낮추는 결과가 나왔다. 그 이유는 단구간 코드북에는 시간정보가 포함되지 않아 '양평'이 '평택'으로 '울진'이 '진안'등으로 인식되는 것과 같은 사례가 빈번이 발생하였는데 MSVQ로 실험하여 해결할 수 있었다. 그러나, MSVQ인 경우는 인식시 처리 시간이 적게 걸리고 인식률도 단구간 코드북에 비해 좋으나, DTW 패턴 매칭 방법에 비해 인식률이 낮다. 따라서, 본 논문에서처럼 인식시 비교되는 구간의 코드워드를 중복하여 비교 매칭함으로써 인식율을 향상시킬 수 있었다. 또한 코드북의 중심점 설정에 사용된 두가지 방법의 평균 인식율을 비교해 본 결과, MINIMAX 방법의 평균 인식율은 85.61% 이고, MINSUM의 평균 인식율은 91.04%로 MINSUM 방법이 더 인식률이 높게 나타남을 부수적으로 알 수 있었다.

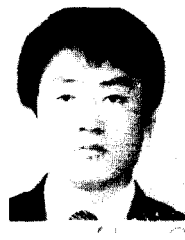


참고문헌

1. 김순철, "한국어 음성의 분석과 자동 인식에 관한 연구," 박사 논문, 연세대학교 대학원, 1982. 12.
2. R.M. Gray, "Vector Quantization," IEEE ASSP Magazine, Vol. 1, pp.4-29, Apr. 1984.
3. Y. Linde, A. Buzo, and R.M. Gray "An algorithm of Vector Quantizer Design," IEEE Trans. Commun., Vol. COM 28, pp. 84-95, Jan. 1980.
4. KUK-CHIN PAN, Frank K. Soong and L.R. Rabiner, "A Vector Quantization Based preprocessor for Speaker Independent Isolated Word Recognition," IEEE Trans. of Acoustics, Speech, Signal Processing, Vol ASSP 33, No.3, June 1985.
5. D.K. Burton, J.E. Shore, J.T. Buck "Isolated Word Speech Recognition Using Multisection Vector Quantization Codebooks," IEEE Trans. of Acoustics, Speech, and Signals, Processing, Vol ASSP 33, No 4, August 1985.
6. D.K. Burton and J.E. Shore, "Speaker Dependent Isolated Word Recognition using Speaker Independent Vector Quantization Codebooks Augmented with Speaker Specific Data," IEEE Trans. Acous. Speech, and Signal processing, Vol. ASSP 33, No. 2, pp.440-443, April 1985.
7. C.E. Shannon, "A mathematical Theory of Communication," Bell Sys. Tech. J. 27, pp.379-423, 623-656, 1948.
8. 이철진, "VQ를 이용한 DDD 지역명 인식에 관한 연구," 석사학위 논문, 연세대학교 대학원, 1989. 12.
9. L.R. Rabiner, M.M. Sondhi, S.E. Levinson, "Note on the Properties of a Vector Quantizer for LPC Coefficients," Bell Sys. Tech. J., Vol. 62, No.8, October 1983.
10. 원우진, 안태옥, 이철진, 김순철, "Modified ISODATA 방법을 이용한 불투명 화자 인공어 인식," 한국음향학회지, Vol. 6, No. 4, pp.31-43, 1987.
11. S.E. Levinson, L.R. Rabiner, A.E. Rosenberg and J.E. Wilpon, "Interactive Clustering Techniques for Selecting Speaker Independent Reference Techniques for isolated Word Recognition," IEEE Trans. on ASSP, Vol. 27, No.2, pp.131-141, Apr. 1979.
12. Manfred R. Schroeder, "Linear Predictive Coding of Speech: Review and Current Directions," IEEE Comm. Magazine, Vol.23, No. 8, August 1985.
13. J.D. Markel and A.H. Gray, Linear Prediction of Speech, Springer Verlag, Berlin Heidelberg, 1976.
14. Shukano, K., Kohda, M., "On the LPC Distance Measures for Vowel Recognition in Continuous Utterance," Institute of Electrical and Communication Engineers of Japan, Trans on D, J63 D, May 1980.



安泰玉 (Tae Ok AN) 正會員  
 1953年 6月24日生  
 1981年 2月 : 蔚山工科大学材料科 卒業  
 1987年 2月 : 光云大学校 電子計算機工學科 卒業  
 1989年 現在 : 光云大学校 電子計算機工學科 博士課程



金南中 (Nam Joong KIM) 正會員  
 1961年 10月 9日生  
 1985年 2月 : 朝鮮大學校 電算機工學科 卒業  
 1987年 2月 : 朝鮮大學校 大學院 電算機工學科 卒業 (哲學碩士)  
 1987年 9月 ~ 1996年 9月 : 朝鮮大學校 大學院 電氣工學科 博士課程 電算機專攻 修了



宋 鐵 (Chul SONG) 正會員  
1938年 2月28日生  
1961年 3月：朝鮮大學校電氣工學科卒業  
1977年 2月：朝鮮大學校 大學院 電氣工  
學科 工學碩士學位取得  
1985年 2月：全北大學校 大學院 電氣工  
學科 工學博士學位 取得  
1973年～現在：朝鮮大學校電算機工學科  
教授

※主關心分野는 신호 처리, 컴퓨터 그래픽스 等임.



金 淳 協 (Soon Hyeob KIM) 正會員  
1947年12月28日生  
1974年 2月：蔚山工科大学電氣工學科卒  
業(電子工學 工學士)  
1976年 2月：延世大學校 大學院 電子科  
卒業(工學碩士)  
1986年 8月～1987年 8月：The Univer-  
sity of Texas Austin電氣  
및 電子計算機工學科 客員  
教授

1979年 3月～現在：光武大學校 工科大学 電子計算機工學科  
副教授