

# 음소 인식을 위한 수정된 LVQ2 알고리즘의 고찰

황철준 정현열

영남대학교 전기전자공학부

## A Modified LVQ2 Algorithm for Phonemes Recognition

Cheol-Jun Hwang Hyun-Yeul Chung

School of Electrical and Electronic Engineering, Yeungnam University

### 요 약

본 논문에서는 한국어 음소를 대상으로 Kohonen이 제안한 LVQ2 방법의 결점을 보완한 MLVQ2 방법으로 인식실험을 행하고 MLVQ2 알고리즘의 유효성을 검토하고자 한다.

인식실험을 위한 음성자료는 ETRI 611단어로부터 추출한 49음소를 사용하였다.

그리고 인식실험에 있어서는 먼저 과열음을 대상으로 학습회수, 표준패턴의 수, 샘플수에 따른 인식률의 변화율 조사하였으며 이 결과 표준패턴의 수 15개, 학습회수 10회이하, 샘플 수 3000개인 경우가 가장 좋은 인식율을 보였다. 이 결과를 참고로 음소군별 인식실험 결과 모음 69.11%, 과열음 74.69%, 마찰음 및 과찰음 86.31%, 비음 및 유음 74.51%의 평균 인식율을 얻었다.

또한, 한국어 49음소 전음소에 대한 인식실험 결과 71.2%의 인식률 얻어 MLVQ2의 유효성을 확인하였다.

### I. 서 론

과학기술의 발달에 따라 정보 통신 분야의 눈부신 발전으로 인하여 인간과 기계사이의 의사소통에 대한 연구의 중요성이 증대되고 있으며 이를 위한 음성 신호처리 기술에 관한 연구가 활발하게 진행되고 있다. 음성 인식에 관한 연구가 시작된 이후로 소량의 음성 자료를 이용하는 형태에서 현재는 대용량 음성 인식기 구현을 위한 연구로 발전하고 있다. 따라서 인식의 기본 단위로 음소가 많이 이용된다. 음성 인식 기술은 음성속에 내재되어 있는 특징을 가장 효과적으로 추출하는 과정과 이 정보를 이용하여 표준 패턴을 작성하고 인식하는 과정으로 나눌수 있다.

특히 가장 안정된 음성의 표준패턴 구성은 인식시

스템에 있어 매우 중요한 부분이라 할 수 있으며 여러 가지 방법 중에서 VQ(Vector Quantization)는 음성 자료의 압축 및 표준패턴 구성과 식별에 현재까지 잘 알려져 있고 많이 이용되는 방법이라 할 수 있다.

또한, 일반적인 VQ 알고리즘의 단점을 개선한 여러 가지 VQ 방법들이 연구되어지고 있는데 이 중에서 음성인식에 사용되고 있는 LVQ2 알고리즘은 다양한 패턴을 인식하는데 있어서 계산량이 적고 인식률도 높은 장점이 있다.

LVQ2 알고리즘에서는 만약 입력 벡터에 친 번째로 가장 가까운 클래스가 들리는 반면 입력 벡터에 두 번째로 가까운 클래스가 맞다면, 동시에 두 개의 표준 벡터가 수정된다. 만약 주어진 입력 벡터가 세 번째 rank로 인식된다면 수정은 행해지지 않는다. 따라서, LVQ2 알고리즘을 이용하여 음소 인식 시스템을 구성하는 데에는 주어진 벡터의 rank가 2보다 클 때에는 혼란 알고리즘이 없다는 문제가 있다.

따라서 본 연구에서는 LVQ2 알고리즘의 문제점을 보완한 MLVQ2 알고리즘을 이용하여 한국어 음소 인식을 행하였다. 인식실험을 위한 음성자료는 ETRI 611단어로부터 추출한 49음소를 사용하였다.

그리고 인식실험에 있어서는 먼저 과열음을 대상으로 학습회수, 표준패턴의 수, 샘플수에 따른 인식률의 변화율 조사하여 음소 인식에 적당한 학습회수, 표준패턴의 수, 샘플수등을 찾고 이 결과를 바탕으로 모음, 과열음, 마찰음 및 과찰음, 비음 및 유음의 음소군별로 인식실험과 한국어 49 전음소에 대한 인식 실험을 수행하여 그 결과를 검토하여 MLVQ2의 유효성을 확인하고자 한다.

### II. 특징파라미터 추출

녹음된 음성은 BPF(75Hz - 8KHz)를 거쳐 20KHz로 샘플링되고 12비트로 양자화된 ETRI 611 단어

데이터베이스를 사용하였다. 이 음성을 29 Channel BPF를 통과시켜 분석되었다. 이때 10ms당 특징을 추출하였으며 5ms 프레임을 shift시킨 logarithmic 스펙트럼의 열로서 표현하였다. 각 프레임으로부터 power성분을 포함하는 8차의 cepstrum계수와 8차의 델타 cepstrum 계수를 구하였다. 이때 각 계수의 값은 계수의 최대 크기에 의해 정규화하였다.

### III. MLVQ2 및 인식 방법

#### 3.1 MLVQ2 알고리즘

MLVQ2 알고리즘에서는 만약 옳은 클래스가 N번째 rank안에 있다면 동시에 두 개의 표준벡터가 수정된다.

$$[m_{n-1}]^{t+1} = [m_{n-1} - \alpha_1(t)(x - m_{n-1})]^t \quad (1)$$

$$[m_n]^{t+1} = [m_n + \alpha_1(t)(x - m_n)]^t \quad (2)$$

여기서,

$$\alpha_n(t) = \alpha_0 \times (1 - \frac{t}{T}), \quad (0 < \alpha_n(t) \ll 1) \quad (3)$$

T : 샘플 수      T : 학습회수 × 샘플 수

$\alpha_0$  : 초기 학습계수로 0.02이다.

MLVQ2 알고리즘에서는 n-1번째 표준 벡터는  $\alpha_1$ 에 의하여 멀어지는 반면 n번째 표준 벡터는  $\alpha_1$ 에 의하여 더 가까워진다.

그림 1에 MLVQ2 알고리즘의 흐름도를 나타내었다.

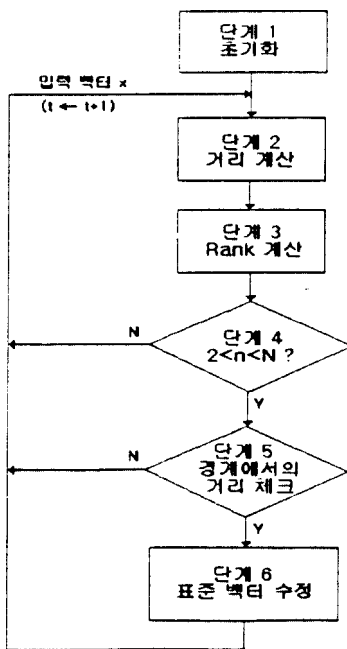


그림 1. MLVQ2 알고리즘

#### 3.2 인식 방법

인식 단계에서는 각 프레임의 입력 벡터와 각 클래스내의 가장 가까운 표준 벡터 사이의 거리를 계산하였다.

이 거리로부터 각 클래스는 다음과 같이 정의되는 activation value  $a_w$ 를 지정했다.

$$a_w(c, t) = 1 - \frac{d(c, t)}{\sum d(i, t)} \quad (4)$$

여기서 d, c, t는 거리, 클래스, 프레임 수이다.

각 프레임의 activation value는 주어진 음소 구획에 대하여 합해서 최대 activation value를 가진 클래스를 인식된 출력으로 하였다.

### IV. 인식 실험 및 고찰

#### 4.1 인식 시스템

음성 데이터는 29 Ch. BPF를 통해 분석되며 이 결과를 이용하여 각 프레임별로 8차의 cepstrum계수와 8차의 델타 cepstrum계수를 구하여 MLVQ2 알고리즘으로 표준패턴을 작성하고, 인식시에는 입력된 음성 데이터를 같은 방법으로 추출한 입력패턴과 표준패턴을 이용하여 식 (4)로서 인식결과를 출력하였다. 그림 2에 전체 인식시스템의 구성을 나타내었다.

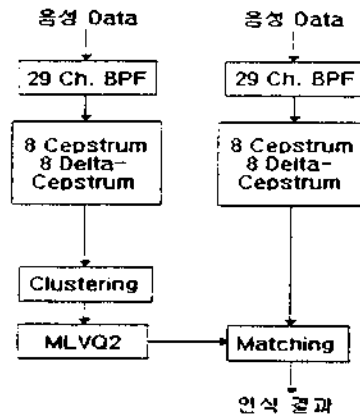


그림 2. 인식 시스템 구성도

#### 4.2 인식 실험

인식실험을 위한 음성자료는 3인이 2회 발성한 ETRI 511 단어로부터 추출한 49음소를 사용하였으며 3인의 1회 발성을 학습 데이터로 나머지 1회 발성을 테스트 데이터로한 open test를 수행하였다.

먼저 음소 인식에 적당한 표준패턴의 수, 학습회수, 샘플수를 조사하기 위하여 예비 실험으로 과일음, 모, 드, 크를 대상으로 표준패턴의 수, 학습회수,

샘플수를 변화시키면서 인식실험을 수행하고 그 결과를 참고로 음소군별 인식실험과 한국어 49음소들 대상으로 인식실험을 하였다

#### 4.2.1 표준패턴의 수에 따른 음소 인식

과열음 /표/, /리/, /리/에 대하여 각 음소당 표준패턴의 수를 10개, 12개, 15개, 17개, 20개일 경우에 인식 실험을 수행한 결과를 그림 3에 나타내었다.

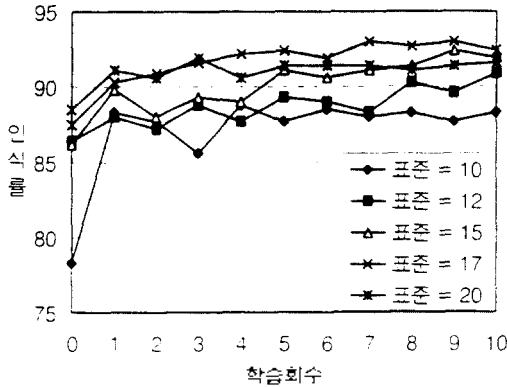


그림 3 표준패턴 수에 따른 인식 실험(close test)

이 결과로부터 표준 패턴의 수가 15개이상일 경우 비교적 높은 인식율을 보였다.

#### 4.2.2 학습 회수에 따른 음소 인식

음소당 표준패턴의 수를 15개로 하여 학습 회수수를 10회, 12회, 15회로 변화시켜 가면서 인식 실험을 수행하였다 그 결과 학습회수에 따른 인식율의 변화가 크게 없음을 알 수 있었다

그림 4에 학습회수의 변화에 따른 인식율을 나타내었다.

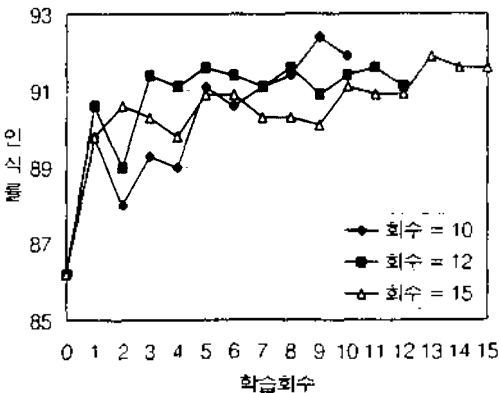


그림 4. 학습회수에 따른 인식 실험(open test)

#### 4.2.3 음소당 샘플 수에 따른 음소 인식

이상의 결과로부터 표준 패턴의 수는 15개로, 학습회수는 10회로 하여 음소당 샘플 수를 1000개, 3000개, 5000개로 변화시켜 가면서 인식 실험을 수행하였다. 그림 5에 샘플수에 따른 인식율의 변화를 나타내었다.

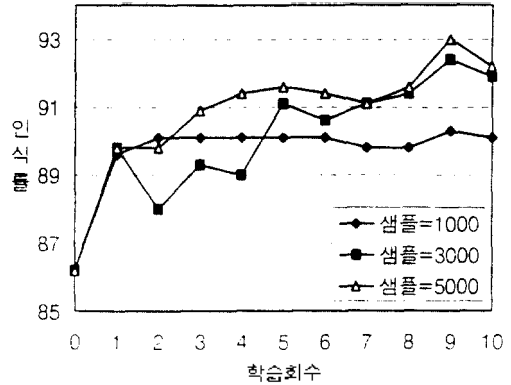


그림 5 샘플 수에 따른 인식 실험(open test)

이 결과로부터 음소당 샘플 수가 3000개 이상이 되었을 때는 인식율이 크게 변화가 없어 3000개 정도가 적당함을 알 수 있었다

#### 4.2.4 음소군별 인식 실험

이상의 실험 결과로부터 표준패턴의 수 15개이상일 때, 학습회수 10회 이하일 때, 샘플 수는 3000개 이상일 경우에 음소 인식에 적당함을 알 수 있었다. 이 결과를 참고로 하여 음소군별로 인식 실험을 하였다.

그림 6에 각 음소군별 인식결과를 나타내었다.

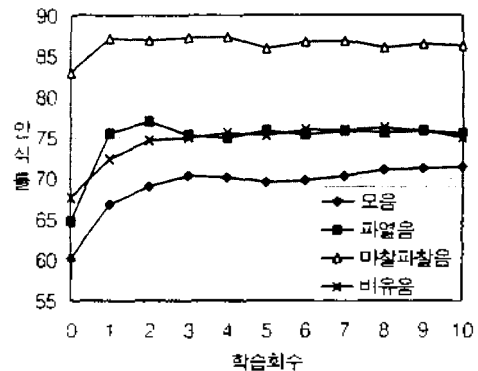


그림 6. 음소 군별 인식 실험(open test)

실험 결과로부터 모음 69.11%, 파열음 74.69%, 마

찰음 및 파찰음 86.31%, 비음 및 유음 74.51%의 평균 인식률을 얻었으며 학습회수가 5회 이상인 경우 인식률의 변화가 거의 없어 5회정도 만으로도 충분한 학습을 알 수 있었다.

#### 4.2.5 전음소 인식 실험

이상의 결과를 참고로 표준패턴의 수 15개, 학습회수 10회, 샘플 수 3000개로 하여 49 전음소에 대한 인식 실험을 수행하였다. 그 결과 최고 인식률을 71.2%로 얻었다. 결과를 그림 7에 나타내었다.

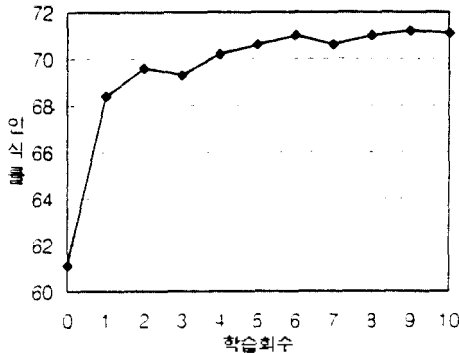


그림 7. 전음소에 대한 인식 실험(open test)

#### 4.3 결과 고찰

과열음을 대상으로 한 예비 실험으로부터 인식 실험에 있어서 음소당 표준 패턴의 수는 15개 정도가 적당한 학습을 할 수 있었고, 학습 회수는 10회, 12회, 15회로 변화시켜 가면서 실험을 하였으나 인식률의 차이는 거의 없었다. 그리고 음소당 샘플 수는 3000개 이상일 때 크게 인식률의 차이를 보이지 않았다.

이 결과를 바탕으로 음소군별 인식 실험을 수행한 결과 모음 69.11%, 파열음 74.69%, 마찰음 및 파찰음 86.31%, 비음 및 유음 74.51%의 평균 인식률을 얻었으며 학습회수가 5회 이상인 경우 인식률의 변화가 거의 없어 5회정도 만으로도 충분한 학습을 알 수 있었다.

이 결과를 전음소를 대상으로 작성한 경우 71.2%의 인식률을 얻었다.

### V. 결 론

본 연구에서는 한국어 음소를 대상으로 Kohonen 이 제안한 LVQ2 방법의 결점을 보완한 MLVQ2 방법으로 인식실험을 행하고 MLVQ2 알고리즘의 유효성을 검토하였다.

인식실험을 위한 음성자료는 ETRI 611단어로부터

추출한 49음소를 사용하였다.

그리고 인식실험에 있어서는 먼저 과열음을 대상으로 학습회수, 표준패턴의 수, 샘플수에 따른 인식률의 변화를 조사하였으며 이 결과 표준패턴의 수 15개, 학습회수 10회 이하, 샘플 수 3000개일 경우가 가장 높은 인식률을 보여 이 결과를 이용하여 음소군별 인식실험 결과 모음 69.11%, 파열음 74.69%, 마찰음 및 파찰음 86.31%, 비음 및 유음, 74.51%의 평균 인식률을 얻었다

또한, 한국어 49음소 전음소에 대한 인식실험 결과 71.2%의 인식률 얻어 MLVQ2의 유효성을 확인 하였다.

향후 이상의 결과를 바탕으로 MLVQ2 알고리즘과 HMM 알고리즘을 이용한 음소 인식시스템을 구현을 위한 실험을 진행 하고자한다.

### 참고문헌

- [1] M.Endo, S.Makino, T.Sone, K.Kido, "Phoneme Recognition Using LVQ2", Trans. IEICE, SP 59-50 (in Japanese) (1989)
- [2] M.Endo, S.Makino, T.Sone, K.Kido, "Recognition of Phonemes in Continuous Speech Using A Modified LVQ2 Method", J. Acoust. Soc. Jpn. (E) 13, 6 pp. 351-360 (1992)
- [3] 김동진, 이광순, "수정된 LVQ2 알고리즘을 이용한 음소인식", J. Acoust. Soc. Korea, Vol. 12, No. 1E, pp. 71-77 (1993)
- [4] S.Makino, Y.Okimoto, K.Kido, H.R.Kim, Y.J.Lee, "Korean Consonant Recognition Using A Modified LVQ2 Method"
- [5] 안종영, 최강민, "HMM Segmentation과 LVQ를 이용한 한국어 음성 인식에 관한 연구", 제 11회 음성통신 및 신호처리 워크샵 논문집, pp. 378-382 (1994)
- [6] 김병국, 장현연, "이론프레임 정보를 이용한 비음 및 유음의 인식", 제 13회 음성통신 및 신호처리 워크샵 논문집, pp. 314-319 (1996)