論文 96-21-3-5

계층적 자기조직화 분류기를 이용한 다수 음성자판의 생성과 레이블링

正會員 정 담*, 이 기 철*, 변 영 태*

Creation and Labeling of Multiple Phonotopic Maps using a Hierarchical Self-Organizing Classifier

Dam Chung*, Kee-Cheol Lee*, Young-Tai Byun* *Regular Members*

본 논문은 1995년도 홍익대학교 교내 연구 지원비에 의하여 연구되었음.

요 약

최근, 신경망 모델의 적응성과 학습성을 이용한 음성인식 연구가 진행되어 왔다. 그러나, 기존의 신경망 모델로는 한국어 음성의 조음결합의 처리 및 유사 음소간의 경계 분류가 용이하지 않다. 또한, 한 개의 형상지도를 이용하는 경우 이질적인 음성자료의 처리를 위한 학습속도의 급격한 증가와, 균일한 학습 및 판별방법의 적용이 갖는 부정확성이 야기될 수 있다. 이에 따라, 본 논문에서는 계층적 자기조직화 분류기(HSOC)를 이용한 신경망타자기를 설계하고, 관련 알고리즘들을 제안한다. 본 HSOC는 Kohonen의 자기조직화 형상지도(SOFM)를 이용하여, 학습시 입력되는 음소 데이타를 계층적인 구조를 갖는 다수의 형상 지도(map) 즉 음성자판에 배치한다. 또한 본 논문에서는 자판의 수효, 각 자판의 크기, 소속될 음소의 선택과 배치, 적합한 학습 및 인식기법의 자동 결정을 위한 알고리즘을 제시하고 실험하여 자기조절식인 음성자판을 구성하였다. 자판을 분류하는 방식을 언어학적 사전지식에 의존할 경우 언어학적 지식의 습득과 적용방법(예를 들면, 확장 음소의 처리) 등을 결정하는 어려움을 가지는 반면, 본 HSOC를 이용하면 주어진 입력 데이타에 적합한 다수의 음성자판을 자기 조절식으로 구성할 수 있는 장점이 있다. 제안된 방식에 따라 최종 생성된 세개의 한글 음성자판은 최적 자판과 최적 전처리기법을 갖추고 있으며, 기존의 언어학적 지식과도 부합됨을 확인할 수 있었다.

ABSTRACT

Recently, neural network-based speech recognition has been studied to utilize the adaptivity and learnability of

*홍익대학교 컴퓨터공학과
Dept. of Computer Engineering, Hong-ik University
論文番號:95401-1122
接受日字:1995年 11月 22日

neural network models. However, conventional neural network models have difficulty in the co-articulation processing and the boundary detection of similar phonemes of the Korean speech. Also, in case of using one phonotopic map, learning speed may dramatically increase and inaccuracies may be caused because homogeneous learning and recognition method should be applied for heterogeneous data. Hence, in this paper, a neural net typewriter has been designed using a hierarchical self-organizing classifier(HSOC), and related algorithms are presented. This HSOC, during its learing stage, distributes phoneme data on hierarchically structured multiple phonotopic maps, using Kohonen's self-organizing feature maps(SOFM). Presented and experimented in this paper were the algorithms for deciding the number of maps, map sizes, the selection of phonemes and their placement per map, an approapriate learning and preprocessing method per map. If maps are divided according to a prior linguistic knowledge, we would have difficulty in acquiring linguistic knowledge and how to apply it(e.g., processing extended phonemes). Contrarily, our HSOC has an advantage that multiple phonotopic maps suitable for given input data are self-organizable. The resulting three korean phonotopic maps are optimally labelled and have their own optimal preprocessing schemes, and also confirm to the conventional linguistic knowledge.

Ⅰ. 서 론

음성인식은 인공지능의 다른 여러 분야와 같이 인간의 사고, 감각, 학습 능력을 이해하고 모방하여 컴퓨터로 구현하고자 하는 응용분야중 하나이다. 음성인식은 하드웨어적 도구의 미비로 효과를 거두지 못하다가 1950년대부터 본격적으로 연구가 시작되었다. 음성인식의 궁극적인 목표는 화자 독립적이며 연속적인 음성의 효율적인 인식을 목적으로 한다. 그러나, 한국어 음성인식의 경우에는 같은 음소가 문맥에 따라 달리 발음되는 조음 결합(co-articulation) 및 음소간 음절간 단어간 경계 분류와 유사 음소들간의 경계 분리가 다른 어떤 언어보다 큰 문제점으로 남고 있다[1].

신경망 모델은 학습능력과 적응능력을 통해 이러한 단점을 극복할 수 있는 대안으로 나오게 되었다. 최근에는 기존의 단일 계층 신경망으로 해결하기 어려운 음소들간의 경계 분류 등을 위한 다계층(multilayer) 신경망 구조가 연구되었고[2], 각 계층의 분할 및 생성에 따른 전체 네트워크의 크기와 구조의 결정이 중요한 문제로 대두되었다[3,4].

본 논문에서는 새로운 계층적 자기조직화 분류기(Hierarchical Self Organizing Classifier 또는 HSOC)를 설계하고 신경망 타자기에 응용하여 자기조절식 음성자판(Self-Adaptive Phonotopic Maps)을 자동으로 생성하는 새로운 학습 알고리즘을 제안한다. 네트워크의 위상을 유지하면서 점진적 학습이 가능한 Kohonen의 자기조직화 형상지도(Self-Organizing Feature Map 또는 SOFM)를 기본 학습방법으로 이용하는 알고리즘을 제시하고, 신경망 음성타자기(Phonetic Typewriter)에 적용하여 자기조절식 음성자판을 자동으로 생성한다[5,6].

HSOC는 한국어 음소 특질상 유사 음소들간의 복잡한 경계의 분류, 임의로 주어지는 데이타에 대한 적응성을 해결하기위한 다 계층 신경망의 구조로 제시되고 있다. 또한, HSOC는 각 계층의 음성자판의 구성 음소들이 한국어 음소의 특징에 맞게 분류가 되고 전체 네트워크의 구조와 크기가 동적으로 유지되는 적응적 특징이 있으므로, 신경망 음성타자기에서 한국어 음성의 분류에 적합하게 전체 네트워크의 크기와 구조를 최적으로 자동 결정하는 자기조절식 음성자판을 생성할 수 있다. 그리고, 신경망 음성타자기의 효율을 극대화하기위한 방법으로 기본적인 음성자판의 자기적응적인 생성과 그로 인한 부음성자판의 구성 음소들의 특성 벡터를 이루는 각 주파수군의 전처리 기법을 결정하여, 자판별로 적절한 학습 및 전처리기법을 찾아낼 수 있게 하였다.

본 논문에서는 화자의 수를 3인(남성 1인, 여성 2인)으로 하여 다중 화자인식을 대상으로 실험하고 있지만, 제안된 방식이 음소단위의 인식을 기반으로 하고 있어, 앞으로 화자 독립적이고 연속적인 음성의 인식으로의 적용시 기본 모델이 될 수 있을 것으로 판단된다.

신경망 음성 타자기에서 실험되는 음성 데이타로

는 음어를 사용하였다. 음어란 군사 목적등을 위한 전송 암호로서 전송하고자하는 내용을 숫자로 음어화하여 전송하는데 사용하며, '하나', '둘', '삼', '넷', '오', '여섯', '칠', '팔', '아홉', '공'의 0~9까지의 숫자로 구성된다. 음어 음성을 인식하기위해서는 7개의 모음 음소와 12개의 자음(7개의 초성과 5개의 종성) 음소에 대한 학습이 필요하다.

Ⅱ. 자기조직화 형상지도

자기조직화 형상지도(Self-Organizing Feature Map 또는 SOFM)는 주어진 입력 패턴에 대하여 정확한 해답을 주지 않고 자기 스스로 학습할 수 있는 능력을 갖춘 네트워크이다. SOFM의 구조는 그림 1에서와 같이 일반적으로 두 개의 층으로 이루어져 있으며, 첫번째 층은 입력층이고 두 번째층은 경쟁층인데 2차원의 격자(grid)로 이루어져 있다. 입력 패턴 벡터의 차원은 특징 파라미터에 의해 결정되며 주어진 입력으로부터 얻어진 프레임 단위의 파라미터를 이용하여 신경망을 훈련시킨다. 모든 연결들은 첫번째 층에서 두번째 층의 방향으로 되어있으며, 두번째 층은 완전 연결되어있다. 전처리를 거친 특징벡터 형태로 표현된 입력데이타를 경쟁층으로 전파하여, 경쟁층의 승자 뉴런(즉 에러가 가장 적은 셀)과 그 이웃반경 안에 있는 모든 뉴런의 특징벡터값을 조정하는 경쟁학습(Competitive Learning)을 반복하여 최종의 SOFM이 작성된다[7]. 기본적인 학습 알고리즘은 다음과 같다[8].

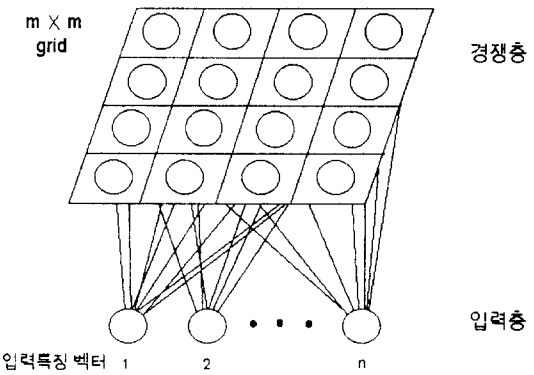


그림 1. 자기조직화 형상지도 구조
Fig 1. Self-organizing feature map structure

[단계 1] 특징 벡터를 초기화한다.
N개의 입력으로부터 M개의 출력 뉴런사이의 연결강도를 임의의 작은 값으로 초기화한다. 초기의 이웃의 범위는 모든 뉴런들이 포함될 수 있도록 충분히 크게 잡았다가 학습이 진행될수록 점차로 줄어들어 최종적으로 승자뉴런만 선택하게 한다.

[단계 2] 새로운 입력벡터 X에 대하여 모든 뉴런들간의 거리를 계산하여 최소거리의 승자뉴런을 선택한다. 식 (1)에 따라 승자뉴런을 결정하고, 척도로는 유클리디안 거리 계산 법을 사용한다.

$$\|X-W_c\| = \min_i(\|X-W_i\|) \tag{1}$$

W_i : i번째 셀의 특징벡터(Weight Vector)
c : 승자 뉴런
X : 입력벡터

[단계 3] 승자 뉴런과 그 뉴런의 이웃 범위에 속한 뉴런들의 특징 벡터를 조정한다.
식 (2)에 따라 승자 뉴런과 이웃 범위에 속하는 뉴런들에 대하여 특징 벡터를 갱신하며, 이웃 범위에 속하지 않는 뉴런들은 갱신되지 않는다. 이웃의 범위는 단계적으로 감소시키며, α는 학습 계수로서 0과 1사이의 값을 가지며 점차 감소된다.

$$W_i(t+1) = \begin{cases} W_i(t) + \alpha(t) \times (X - W_i(t)), & \text{if } i \in N_c \\ W_i(t), & \text{otherwise} \end{cases} \tag{2}$$

W_i : i번째 셀의 특징벡터(Weight Vector)
N_c : 선택된(승자 뉴런) 셀의 이웃들(Neighborhood)
t : 학습 진행시간
α : 학습계수
X : 입력벡터

[단계 4] 종료 조건이 만료될때까지 단계 2와 단계 3을 반복한다.

본 논문에서 실험되는 신경망 타자기에서는 학습

계수를 0.3에서 시작하여 학습이 진행되면서 식 (3)에 따라 감소하였다. 학습되어질 각 음성자판의 크기는 구성음소의 두배로 주고, 학습과정에서의 영향을 미치는 이웃의 결정은 초기에는 전체 음성자판의 가로와 세로의 반에 해당하는 면적에서 시작하고 점차로 줄어들어 1개의 셀이 남을때까지 식 (3)에 의해서 수행하였다. 효과적인 학습이 되기위해서 본 논문의 실험데이타로 자음의 경우는 10^7, 모음의 경우는 10^6번의 학습을 하였다.

$$\alpha(t) = \alpha_0 \times (1 - \frac{i}{np})$$
$$N(t) = d_0 \times (1 - \frac{i}{np}) \quad (3)$$

$\alpha_0 = 0.3$
d_0 = 음성자판 가로/세로의 반
i: 학습 진행 횟수
t: 학습 진행 시간
np: 총 학습 횟수

Ⅲ. 신경망 음성타자기

1. 신경망 음성타자기

신경망 음성타자기는 신경망 기술과 신호처리 및 인공지능 기술을 접목한 장치로, 입력된 음성을 음소로 변환, 출력시켜 준다.

신경망 음성타자기의 학습 및 음성타자기를 통한 인식은 먼저 음성정보의 추출을 위한 전처리 과정을 통해 시작된다. 입력음성에 대해 전처리를 수행하여 음성의 특징 벡터를 생성한 후 SOFM을 기본 학습방법으로 사용한다. SOFM은 각 음성을 구성하는 음소들의 특징 벡터에 대해 학습시켜 입력음성에 대한 분류기로의 역할을 한다. SOFM을 통해 각 음소들에 대하여 학습시 사용되었던 데이타 화일을 전파시키면 음성자판이 만들어진다. 즉, 음성자판에 대하여 가장 많이 기록된 음소들을 찾아내는 레이블링 과정을 거치면 해당 음성자판이 생성된다. 레이블을 붙인후 음성화일을 전파시키면 각 음소가 SOFM의 해당 셀을 마치 타자기의 자판을 두드리듯이 작동하여 해당 음소셀의 레이블이 음성화일의 내용으로 출력된다.

2. 음성 타자기의 레이블링

학습의 결과로 음성타자기의 각각의 음성자판이 각 음소의 특징벡터를 나타내는 형상지도(Feature Map)를 형성한다. 학습이 끝난 후에는 SOFM의 초기 음성자판의 형태인 자음자판과 모음자판의 각 셀이 대표하는 음소를 알아내야 한다. 이것을 알아내는 과정을 레이블링이라 한다. 레이블링을 하기위해서는 기본적으로 모든 음소의 데이타 화일을 각각의 음성자판에 전파한 후 SOFM학습으로 생성된 음성자판들의 각각의 셀에 대해 가장 많은 대표횟수를 기록한 음소의 이름과 음소가 자음인 경우 초성인지 종성인지의 여부를 레이블로 붙인다. 레이블을 붙인후 음성화일을 전파시키면 각 음소가 음성자판의 해당 셀을 타자기의 자판을 두드리듯이 음성화일의 내용이 출력된다.

Ⅳ. 계층적 자기조직화 분류기(HSOC)

1. HSOC의 구조 적응 과정

본 절에서는 계층적 자기조직화 분류기(Hierarchical Self Organizing Classifier 즉 HSOC)를 한국어 음소 특질상 유사 음소들간의 복잡한 경계의 분리, 임의로 주어지는 데이타에 대한 적응성을 해결하기위해 다 계층 신경망의 구조로 제시한다. 또한, 이를 적용하여 신경망 음성타자기에서 자기조절식 음성자판을 구성하는 학습 알고리즘을 제시한다.

HSOC를 이용하면 각 계층의 음성자판의 구성 음소들이 한국어 음소의 특징에 맞게 분류가되고 전체 네트워크의 구조와 크기가 동적으로 유지된다. 네트워크의 점진적 학습을 위해, 일련의 구조 적응 과정을 거치며 음성자판에서의 음소의 대표값과 음성자판의 생성시의 분리 기준값을 활성화 척도로 사용하여 각 음성자판이 자동 결정된다.

HSOC의 구조 적응 과정은 입력 데이타를 SOFM 학습 후에 먼저 각각의 음성자판에 가장 크게 활성화된 노드를 생성하고, 노드를 제거하는 제거 규칙(Deletion Rule)과 새로 생성된 계층에서 구성 요소들의 특성에 맞게 조절하는 조절 규칙(Tuning Rule)을 거쳐 완성된다.

먼저 각각의 음성자판에 가장 크게 활성화된 노드를 생성한다. 즉, 가장 많은 대표횟수(winning cell)를 기록한 노드의 레이블링을 한다.

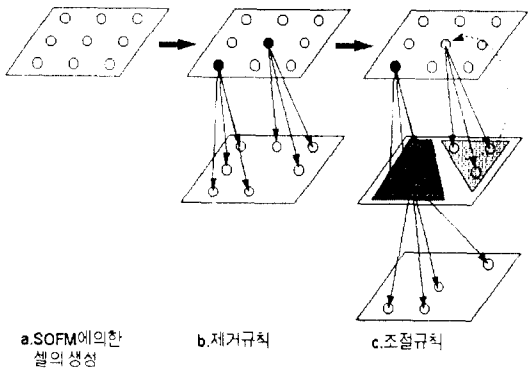


그림 2. 구조 적응 과정
Fig 2. Structure adaptation process

다음 과정은 각각의 음성사판에서 활성화 기여도가 낮은 음소들을 제거하는 제거 규칙이다. 각 음소의 제거시에는 초기음성사판에서 세종적으로 분음성자판과 부음성자판들을 생성하고 해당 음소들을 레이블링 한다. 음소의 대표값과 생성 기준값으로 음소를 제거하고 새로운 음성사판을 생성한다. 제거 규칙을 통하여 음성자판의 크기를 제한할 수 있다.

마지막으로, 생성된 부음성자판에서 각 노드들의 활성화 값의 비교를 통해 음성자판의 구조를 결정짓는 조절규칙이 있다. 즉, 부음성자판에서의 음소의 대표값이 향상되더라도, 부음성자판의 구성 음소들사이의 상관관계에서 기여도가 적으면 초기의 자음자판으로 되돌려 보내서 분음성자판과 부음성자판들 사이의 구조를 최적으로 결정한다. 이 규칙으로 각 음성자판의 구조와 동적으로 유지할 수 있다. 그림 2에서는 HSOC의 구조 적응 과정을 보여준다.

이와 같은 구조 적응 과정에 의하여 새로운 음소들이 각 음성사판에 생성되고, 활성화되지 못하는 음소들이 제거되어 다른 음성자판에 생성되고, 새로운 음성자판에서 구성 음소들 사이의 기여도를 평가하여 특성에 맞게 조절되는 과정을 반복해 나감으로서 전체 네트워크의 구조와 크기를 자동으로 결정해 나간다.

2. 활성화 척도

HSOC에서 네트워크의 구조 적응 과정을 위해 제안되는 활성화 척도를 정의하면 다음과 같다. 식 (4)에 의해서 각 음성사판의 구성음소들의 대표값을 구하고, 대표값과 식 (5)에 의해서 구해진 생성 기준값과의 비교로 수행하면 일련의 자기적응적인 과정을 통해 자기조절식 음성사판의 레이블링을 이룰 수 있다.

i: 각 음성사판
j: 각 음성사판의 구성 음소들
k: 각 음성사판에서의 음소들이 나타날 수 있는 위치
n: 각 음성사판의 크기
N: 음성사판 i의 음소들의 총 갯수

각각의 음성사판에서의 구성음소들의 대표값을 구하는 식은 다음과 같다.

$$REP_{ij} = \frac{\sum_{k=0}^{n-1} W_{ijk}^{*}}{CNT_{ij}} \tag{4}$$

REP_{ij}: 각 음성사판 i에서 j음소의 대표값
FRE_{ijk}: 각 음성사판 i의 음소 j가 위치 k에서 승자셀이 되는 경우

$$FRE_{ijk} = \begin{cases} 1, & \text{if} \quad W_{ijk} \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

W_{ijk}: 각 음성사판 i의 음소 j가 위치 k에서 승자셀이 될 때의 값
(단, 승자셀이 되지 못하는 경우에는 0을 가진다.)
W_{ijk}^{*}: 각 음성사판 i의 음소 j가 위치 k에서 승자셀이 될때의 값으로 정규화된 값

$$W_{ijk}^{*} = \alpha * W_{ijk} \qquad \left(\alpha = \frac{1}{\max(W_{ijk})}\right)$$

CNT_{ij}: 각 음성사판 i의 음소 j가 승자인 셀의 발생 빈도수

$$CNT_{ij} = \sum_{k=0}^{n-1} FRE_{ijk}$$

생성된 음성사판의 대표값을 이용하여 부음성사판을 생성시킬 수 있는 기준값을 구하는 식은 다음과 같다.

$$C_i = \overline{REP_i} - \sigma_{REP_i} \tag{5}$$

C_i: 음성사판 i를 생성시킬 수 있는 기준값
$\overline{REP_i}$: 음성사판 i의 음소들의 대표값의 평균값

$$\overline{REP_i} = \frac{\sum_{j=1}^{N} REP_{ij}}{N}$$

σ_{REP_i}: 음성자판 i의 음소들의 대표값의 표준편차

V. 자기조절식 음성자판의 레이블링

1. 제안된 HSOC의 학습 알고리즘

본 논문에서는 기존의 음성타자기의 음성자판의 레이블링을 보완하는 새로운 알고리즘을 HSOC를 이용하여 제시한다. 제안되는 알고리즘을 이용하여, 기존의 음성자판을 계층적으로 생성해나가며, 음성자판의 구성음소들이 자동적으로 최적의 배치가 되는 자기조절식 음성자판(Self-adaptive Phonotopic Maps)을 형성해 나갈 수 있다. 이러한 기법의 특징은 네트워크의 구조와 크기를 자동으로 최적으로 구성하며, 그 구성음소들의 결정이 어떠한 전문화된 지식없이도 한국어 음운의 특질에 맞는 분류가 가능하다. 또한, 생성된 부음성자판에 대하여 채널군 별로 그 전처리 기법을 교환하여 최적의 기법을 찾아내게 하였다.

각 음성자판이 활성화 척도(생성 기준값, 음소 대표값)에 의해 자기조절식 레이블(자판을 구성하는 최적의 음소들)이 되는 알고리즘은 다음과 같다. 또한 최적의 음성자판을 구성하였을 때 그에 따른 전처리 기법(FFT, Cepstrum)의 교환을 하여 효율적인 인식이 되는 방법도 기술한다. 초기 음성자판의 형태는 SOFM 학습을 거친 자음 12개음과 모음 7개음의 두 개의 음성자판으로 구성되어있고, 부음성자판의 생성시에는 본 자판의 구성 음소수의 반을 넘지 않도록 하며, 후에 기술되는 음소의 대표값과 생성 기준값을 평가하여 생성한다.

현재자판을 i라 하자;
1: /* 레이블 선언 */
'프로시저 **기준값및대표값(i)**'를 호출하여 자판의 기준값과 자판내의 각 음소의 대표값을 구한다;
/* 이제 자기조절식 음성자판을 만들어 간다 */
FOR j:=음성자판 i의 각 음소 **DO**
  IF(음소 j의 대표값 <= 자판 i의 기준값) **THEN**
    음성자판 i에서 음소 j를 제거한다;
    /* Deletion rule */
IF 위에서 제거된 음소가 없으면, 자기조절식 음성기판이 완성되었으므로 종료함;
제거된 음소들만으로 학습하여 새로운 자판 i'을 구성한다;
'프로시저 **자판조절**(i, i')'을 호출한다;
자판 i에 남은 음소들만으로 학습하여 자판 i를 새로 구성한다;
i:=i';
goto 1;

PROCEDURE 기준값및대표값(i)
  BEGIN
    FOR j:=음성자판 i의 각 음소 **DO**
    BEGIN
      음소 j의 대표 셀(발생빈도가 가장 큰 셀)을 구한다;
      음소 j의 발생 빈도수(CNT)와 정규화 값(W^*)을 통하여 대표값(REP)을 구한다;
    END;/* of **FOR** */
  대표값(REP)와 편차값(σ_{REP})을 이용하여 자판분리의 기준값(C)을 구한다;
  END;/* of 기준값및대표값 */

PROCEDURE 자판조절(i, i') /* Tuning rule */
  BEGIN
    '프로시저 **기준값및대표값(i')**'을 호출하여 자판 i'의 기준값 및 자판 소속 각 음소의 대표값을 계산한다;
    완성된 음성자판 i'에서 채널군 당 적합한 전처리 기법을 찾아 낸다;
    FOR j:=음성자판 i'의 각 음소 **DO**
      /* 새자판의 대표값과 기준값을 이용하여 본자판으로의 음소의 환원여부를 결정한다 */
      IF ((새자판i'에서 j의 대표값 < 새자판i'에서 j의 기준값) or (본자판i에서 j의 대표값 > 새자판i'에서 j의 대표값))
      THEN
        음소 j를 본 음성자판 i로 되돌려보낸다;
    되돌려진 음소가 있었으면, 학습하여 자판 i'을 다시 구성한다;
  END;/* of 자판조절 */

위의 알고리즘에서 계층적으로 생성된 부음성자판에 신경망 음성타자기의 효율을 극대화하기위한 방법으로 부음성자판의 구성 음소들의 특징 벡터를 이루는 각 주파수군의 전처리 기법을 결정하여, 자판별로 적절한 학습 및 전처리기법을 찾아낼 수 있게 하였다. 음성 특징 데이타를 SOFM 학습과정의 입력벡터로 사용하기위해서 128개의 주파수별 성분값에서 16개의 채널값으로 그룹화 시킨다. 음성신호의 저주파에 많은 정보가 있으므로 선형적 분석으로 11개의 채널을 할당하고, 고주파에는 개략적인 로그분석으로 5개의 채널을 할당하였다. 처음 레이블링 알고리즘을 적용할때에는 채널의 분리와는 상관없이 FFT와 Cepstrum의 두가지 분석을 하였고, 계층적으로 부음성자판을 생성한 후 저주파 채널군과 고주파 채널군에 대하여 각각의 분석을 적용하는(모두 4가지의 경우) 방법을 적용하였다. 결과적으로, 자판 활용도(구성 음소수/자판의 전체 크기) 값과 구성 음소들의 대표값 그리고 빈도수 등으로 레이블링이 가장 잘된 교환 기법을 채택한다.

2. 실험 데이타의 전처리

본 논문에서 사용되는 신경망 음성 타자기의 음성 데이타로는 음어를 사용한다. 음어란 군사 목적 등을 위한 전송 암호로 전송하고자하는 내용을 숫자로 음어화하여 전송하는데 사용하며, '하나', '둘', '삼', '넷', '오', '여섯', '칠', '팔', '아홉', '공'의 0~9까지의 숫자로 구성된다. 음어 음성을 인식하기 위해서는 7개의 모음음소와 12개의 자음(7개의 초성과 5개의 종성) 음소에 대한 학습이 필요하다. 또한, Kohonen이 사용했던 음성타자기와는 다르게 초기 음성자판의 구성은 자음자판(12개 자음)과 모음자판(7개 자음)으로 분할한다. 그 이유는 한국어 음소 특징상 자음과 모음간의 경계 분류가 뚜렷하며 자기조절식 음성자판(Self-adaptive Phonotopic Maps)으로의 확장이 용이하며 그에 따른 인식효율을 위해서이다.

음성 데이타의 각 음소는 주파수 대역에서의(성분과도 같은) 독특한 특징을 가지므로, 본 논문의 실험 데이타의 19개의 음소에 대해 특징을 추출해내고, SOFM을 사용한 음성타자기의 음성자판의 입력자료로 사용하여 각 음소들에 대한 특징을 학습시켜서 자기조절식 음성자판을 이루게 된다.

음성인식을 위해 사용되는 음성의 특징을 추출해내는 전처리 과정은 다음과 같다. 음성의 녹음과 전처리를 수행하기 위한 신호처리에는 Creative INC.의 Sound Blaster AWE 32를 사용하여 아날로그-디지탈 변환기의 기능을 대신했고, 샘플링(Sampling)된 신호를 화일로 저장한다. 이때, 화일은 각 샘플링된 값이 8진수 형태로 저장된 PCM(Pulse Code Modulation) 방식이다. 샘플링 비율은 앨리어싱(Aliasing)을 피하고 보다 많은 정보를 포함하기위해서 20KHz로 샘플링한다. PCM화일의 내용을 정수로 변환하여 저장한 후 샘플링된 음소의 정세값의 비연속성방지를 위해 해밍 윈도우(Hamming Window)값을 각 128포인트 샘플링 데이타 세트에 곱하여 FFT(Fast Fourier Transform)의 입력으로 사용하는데, 음성특징을 비교적 정확하게 추출하기 위해서 5ms마다 128포인트의 음성 데이타 세그먼트를 64포인트마다 중복시켜 사용하였다. FFT를 수행한 다음 음소의 각 주파수 대역의 고유의 특성을 추출하기위해서 파워 스펙트럼(Power Spectrum)방법을 사용하였다[9,10].

구해진 스펙트럼의 주파수별 값들을 로그 스케일링에 의해 스케일링(Scaling)시키고, SOFM학습의 입력벡터로 사용하기위해서 16개 채널로 그룹화 하였다. 여기에는 음성신호의 저주파에는 많은 특징이 존재하고 고주파로 갈수록 적은 특징이 존재한다는 사실에 기인하여 저주파에서는 선형적 분석으로 11개의 채널을 할당하고, 고주파에서는 상대적으로 개략적인 로그분석으로 5개의 채널을 할당한다. 이제 전처리의 마지막 과정으로 각 입력벡터의 성분에 대해 모든 벡터 성분에 대한 평균값을 빼주고 정규화(Normalization)시켰다.

전처리 과정중 네번째 과정에서 음성신호의 스펙트럼 로그값을 IDFT(Inverse Discrete Fourier Transform)로 수행하는 Cepstrum방법을 FFT방법과 함께 사용하여 저주파 채널군과 고주파 채널군 각각에 적합한 전처리 기법을 결정하였다. Cepstrum방법은 FFT, Log|·|, IFFT, Cepstrum window, FFT를 차례로 적용하는 방식으로, 본 논문에서는 현재 FFT분석의 유일한 대안으로 고려하고 있다.

Ⅵ. 실험 결과 및 분석

1. 자기조절식 음성자판의 레이블링 결과

HSOC 를 이용한 신경망 음성 타자기로 음성자판을 구성하는 실험의 음성 데이타로는 한국어 음어 음성을 사용한다. 음어란 군사 목적 등을 위한 전송 암호로서 전송하고자 하는 내용을 숫자로 음어화하여 전송하는데 사용하며, '하나', '둘', '삼', '넷', '오', '여섯', '칠', '팔', '아홉', '공'의 0~9까지의 숫자로 구성된다. 음어 음성을 인식하기 위해서는 7개의 모음음소와 12개의 자음음소(7개의 초성과 5개의 종성)가 필요하며, 앞에서 제시한 알고리즘에 따라 실험하였다. 초기 음성자판의 형태는 각각 자음 12개와 모음 7로 구성된 두 개의 음성자판으로 하였고, 계층적으로 생성되게 하였다. 전처리 기법으로는 FFT분석과 Cepstrum분석의 차이가 미세하며, 여기서는 FFT분석으로 다루겠다.

본 실험결과를 다루는 앞으로의 표에서는 음성자판에서의 각 음소들의 총 발생 빈도수, 대표 횟수(winning cell)를 기록할 때의 정규화 값(W_{ijk}')의 최대값, 최소값, 대표값, 자판 생성 기준값 및 실제 각 음성자판에서 음소들의 사용도를 나타내는 자판 활용도(구성 음소수/자판의 전체 크기)등의 결과를 알 수 있다. 또한, 부음성자판의 생성시에 자판 활용도값의 비교로 구성 음소들의 특징 벡터를 이루는 각 주파수군의 최적의 전처리 기법을 선택한다.

초기 음성자판의 형태에서 초기 자음자판을 레이블링한 결과는 표 1과 같다. 초기 자음자판의 자판 활용도가 0.479(276/576)로 나타났으며, 이때의 자판 생성 기준값은 0.213으로 나타남을 알 수 있다. 결국 초기 자음자판에서 제거되어 분리될 음소들은 ㄴ, ㄱ, ㄷ, ㅍ, (ㅂ)으로 부자음자판을 생성하고, 그외의 음소들은 본자음자판의 구성 음소들이 된다.(음소들중에서 괄호안의 음소는 종성을 나타낸다)

초기 음성자판의 형태에서 초기 모음자판을 레이블링한 결과는 표 2와 같다. 초기 모음자판의 자판 활용도가 0.734(144/196)로 나타났으며, 이때의 자판 생성 기준값(C)는 0.506으로 나타남을 알 수 있다. 초기 모음자판에서 제거되어 분리될 음소들은 하나도 없음을 알 수 있다. 결과적으로, 초기 모음자판의 경우는 자판 생성 기준값이 0.506이므로 새로운 부모음자판의 생성이 없는 자기조절식 음성자판의 형태가 된다.

부자음자판의 경우, 표 1에서 부자음자판을 생성하여 분리된 구성 음소들에 대하여 특징 벡터를 이루는 각 주파수군의 전처리 기법을 결정하고 레이블링한 결과는 표 3과 같다. 각 주파수의 저주파 채널군과 고주파 채널군에 대하여 FFT 분석과 Cepstrum 분석을 교환한 결과, 음성자판의 자판활용도는 그림 6에서 보는 바와 같이 Cepstrum/Cepstrum(저주파 채널군/고주파 채널군) 분석이 0.65로 가장 높고, 각 음소들의 총 빈도수와 대표값 또한 Cepstrum/Cepstrum 분석의 경우가 최적의 경우임을 알 수 있다. 결과적으로, 그림 3에 따라 자판 활용도값이 가장 높게 나타난 Cepstrum/Cepstrum 분석을 부자음자판에 적용하기로 한다.

표 1. 초기 자음자판의 레이블링 결과
Table 1. Labeling result of original consonant map

음 소	빈 도	최 대 값	최 소 값	대 표 값
ㄴ	13	0.4	0.05	0.13
ㅅ	34	0.9	0.05	0.32
ㅎ	21	0.75	0.1	0.26
ㅊ	56	0.9	0.05	0.56
(ㄹ)	17	0.6	0.15	0.31
(ㅁ)	43	0.9	0.15	0.41
(ㅅ)	32	0.95	0.05	0.52
(ㅇ)	17	0.65	0.05	0.32
ㄱ	9	0.4	0.05	0.20
ㄷ	10	0.2	0.05	0.11
ㅍ	14	0.3	0.05	0.12
(ㅂ)	8	0.2	0.05	0.14

표 2. 초기 모음자판의 레이블링 결과
Table 2. Labeling result of original vowel map

음 소	빈 도	최 대 값	최 소 값	대 표 값
ㅏ	17	0.95	0.1	0.61
ㅗ	18	0.95	0.05	0.65
ㅓ	18	0.95	0.05	0.64
ㅜ	23	0.95	0.05	0.78
ㅣ	13	0.8	0.1	0.54
ㅔ	27	0.95	0.1	0.68
ㅕ	28	0.95	0.05	0.62

선택된 전처리 기법에 따라 레이블링 하였을 때 자판 활용도는 0.65(65/100)로 나타났고, 자판 생성 기준값은 0.175로 나타났다. 이때 표 3에서 알 수 있듯이 음소 ㄴ만이 앞의 알고리즘의 조절 규칙에 따라

부자음자판에서의 음소의 대표값은 향상되었으나, 부자음자판의 구성 음소들사이의 상관관계에서 기여도가 적으므로 본자음자판으로 되돌려 보내진다.

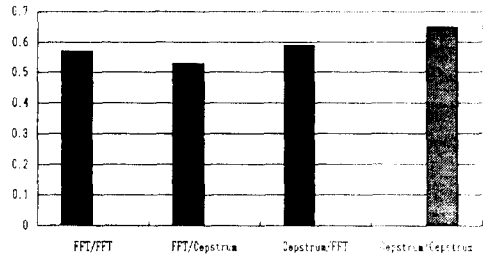


그림 3. 1차 부자음자판의 자판활용도값의 비교
Fig 3. Comparison Map-application value of sub-consonant map

표 3에서 본자음자판으로 되돌려 보내진 ㄴ음소를 제외하고 나머지 음소들로 Cepstrum/ Cepstrum 분석을 적용한 부자음자판을 2차 레이블링한 결과는 표 4와 같다. 자판 활용도는 0.734(47/64)로 나타났고, 자판 생성 기준값은 0.288로 나타났다. ㄴ음소가 되돌려 보내짐으로서 전체 구성 음소들의 상관관계가 높아짐을 알 수 있고, 자판 활용도값도 향상됨을 알 수 있다. 결국 자판 생성 기준값이 0.288이므로 더 이상의 새로운 부자음자판의 생성이 없는 자기조절식 음성자판의 형태가 된다. 또한, 이때의 결과로 2차 레이블링된 부자음자판의 구성음소들이 한국어 파열음의 형태라는 것을 알 수 있다. 즉, 언어학적 지식없이 학습한 기판의 분할이 최소한 언어학적 지식에 어긋나지 않음을 확인할 수 있다.

부자음자판에서 되돌려진 ㄴ음소를 포함하여 본자음자판을 레이블링한 결과는 표 5와 같다. 자판 활용

표 4. 부자음자판의 2차 레이블링 결과
Table 4. The second labeling result of sub-consonant map

음소	빈도	최대값	최소값	대표값
ㄱ	12	0.75	0.05	0.31
ㄷ	12	0.55	0.1	0.29
ㅍ	13	0.95	0.05	0.36
(ㅂ)	10	0.95	0.05	0.30

도는 0.543(139/256)이되고 표 5에서보면 초기 자음자판보다 안정된 자판의 형태 즉, 최적의 자기조절식 음성자판을 이룬다. 초기 자음자판과 대표값을 비교하였을 때(표 1과의 비교) 8개의 음소중 ㅎ, (ㅅ)이 약간 변이지고 나머지 6개의 음소들의 대표값은 향상됨을 알 수 있으며, 자판 활용도값 또한 향상됨을 알 수 있다. 또한, 초기 음성자판에서 부음성자판을 계층적으로 자동 생성해나가는 자기조절식 음성자판(SAPM)으로 형성되가는 과정의 각 음성자판의 자판 활용도값의 비교는 그림 4에 나타나 있다.

표 5. ㄴ음소가 되돌려진 본자음자판의 레이블링 결과
Table 5. Labeling result of "ㄴ" phoneme returned main-consonant map

음소	빈도	최대값	최소값	대표값
ㄴ	11	0.6	0.05	0.24
ㅅ	25	0.95	0.05	0.38
ㅎ	11	0.4	0.1	0.25
ㅊ	36	0.95	0.05	0.67
(ㄹ)	14	0.95	0.15	0.48
(ㅁ)	24	0.95	0.05	0.63
(ㅅ)	25	0.85	0.05	0.39
(ㅇ)	14	0.85	0.05	0.65

표 3. 각 전처리 기법의 부자음자판 1차 레이블링 결과
Table 3. The first labeling result of sub-consonant map using each pre-processing method

음소	FFT/FFT		FFT/Cepstrum		Cepstrum/FFT		Cepstrum/Cepstrum	
	빈도	대표값	빈도	대표값	빈도	대표값	빈도	대표값
ㄴ	11	0.14	10	0.14	11	0.17	10	0.15
ㄱ	12	0.26	12	0.23	12	0.25	14	0.29
ㄷ	10	0.17	8	0.16	13	0.25	14	0.21
ㅍ	13	0.24	13	0.27	12	0.22	15	0.31
(ㅂ)	11	0.24	10	0.25	11	0.24	12	0.22

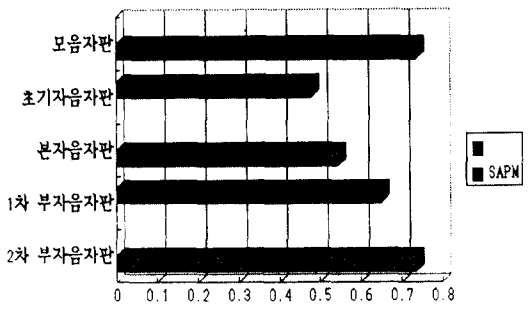


그림 4. 각 음성자판의 자판 활용도 비교
Fig 4. Comparison Map-application value of each phonotopic maps

2. 자기조절식 음성자판 구조의 실험 결과

제안된 계층적 자기조직화 분류기를 이용하여 자기조절식 음성자판을 형성하는 실험을 하였다. 이미 실험된 결과들은 음성자판들을 구성하는 음소들이 제시된 활성화 척도값과 자판 활용도값 그리고 계층적 자기조직화 분류 학습 알고리즘을 이용하여 계층적으로 자동 생성됨을 보았다. 이제 실험 음성 데이타로 구성된 자기조절식 음성자판의 형태를 보면 다음과 같다. 실험에 사용된 실제 음성자판들의 크기는 구성 음소수를 n이라 할때 $2n \times 2n$의 격자형 네트워크로 학습한 후, 각 2×2 크기의 윈도우에서 대표 음소를 정하는 등의 방식으로 $n \times n$의 자판으로 축소하였다. 각각의 원이나 괄호는 하나의 노드를 나타내며, 괄호 안에 레이블된 음소는 종성을 나타내고 원 안에 레이블된 음소는 초성을 나타낸다. 원 안의 대각선은 레이블이 되지 않았거나, 모호한 레이블이 된 경우를 나타낸다. 그림 5는 초기 자음자판이 레이블링되었을 때 각 구성 음소들의 실험결과를 보여준다. 그림 6은 초기 자음자판에서 계층적 자기조직화 분류과정을 거치고 난 후 자기조절식 음성자판이 되었을때 본자음자판과 부자음자판의 각 구성 음소들의 실험결과를 보여준다. 이 그림들에서 볼 수 있는 바와 같이 각 음성자판에서의 자판 활용도나 한국어 음소별 특질상 유사 계열의 음소 분류가 자기조절식 음성자판의 형태일때 더 좋아졌음을 알 수 있으며, 어떠한 한국어적 전문적인 지식 없이도 한국어 음소별 특징에 맞는 음성자판을 구성함으로 확장 음소에 대해서도 쉬운 분류가 가능해짐을 알 수 있다.

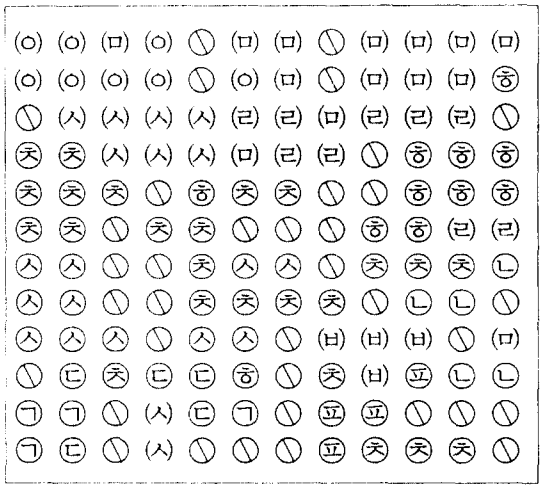
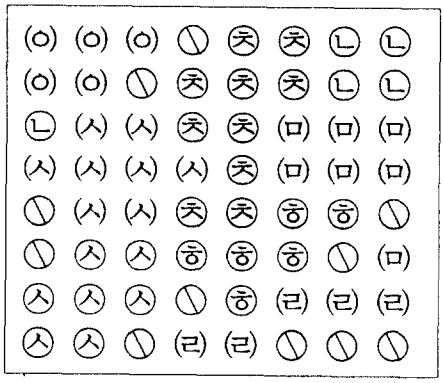


그림 5. 초기 자음자판의 레이블링
Fig 5. Labeling of original consonant map



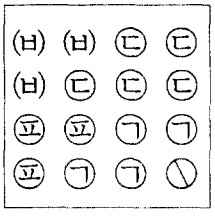


그림 6. 자기조절식 음성자판의 레이블링
(a) 본자음자판 (b) 부자음자판
Fig 6. Self-adaptive phonotopic map labeling
(a) main-consonant map (b) sub-consonant map

Ⅶ. 결 론

본 논문에서는 계층적 자기조직화 분류기(HSOC)를 제안하고 신경망 타자기에 응용하여 자기조절식 음성자판을 형성하는 새로운 학습 알고리즘을 제안하고 실험하였다. 또한, 부음성자판의 구성 음소들에 대하여 저주파 및 고주파 채널간의 전처리 기법을 교환하여 최적 기법을 사용할 수 있게 하였다. 제안된 HSOC는 한국어 음소 특질상 유사 음소들간의 복잡한 경계의 분류, 입력 데이타에 대한 적응성을 해결하기위해 다 계층 신경망의 구조로 제시하였다.

실험결과, 제안된 HSOC를 이용하여 신경망 음성타자기에서 다수의 음성자판을 자기조절식으로 생성해나가는 자기조절식 음성 자판을 구성하였고, 각 자판의 음성자판의 구성 음소들이 한국어 음소의 특성(예를 들면, 한국어 파열음 형태)에 맞게 분류가 되고 전체 네트워크의 구조와 크기가 점진적으로 동적 구성되었으며, 결국 입력으로 사용된 한국어 음성의 분류에 적합하게 전체 네트워크의 크기와 구조를 최적으로 자동결정하는 자기조절식 음성자판을 생성할 수 있었다.

본 연구방식은 인간이 미리 설정해놓은 모델보다 주어지는 데이타에 적합하게 자기조절 적응 구성능력이 있으며, 전문적인 언어학적 지식 없이도 한국어 음소별 특징에 맞는 음성자판을 구성함으로써 확장 음소에 대해서도 쉬운 분류가 가능하다.

앞으로 다른 전처리 기법을 고려한 기판의 평가방법 등을 개선한 음성자판을 생성하여 실제로 한국어 음성인식 시스템에 적용할 수 있을 것으로 판단된다.

참 고 문 헌

1. 이 아정, 이 기철, 변 영태, "음운타자기에 관한 한국어 음어 음성 인식", 한국정보과학회 가을 학술 발표 논문집 21(2), pp.467-470, 1994.
2. V.W. Zue, H.C. Leung, "Phonetic Classification Using Multilayer Perceptrons," *Proc. ICASSP*, pp. 525-528, 1990.
3. J. Lampien, "On Clustering Properties of Hierarchical Self-Organizing Maps," *Artificial Neural Networks 2*, pp.1219-1222, 1992.
4. P. Morasso, A. Pareto and V. Sanguineti, "SOC: A Self-Organizing Classifier", *Artificial Neural Networks 2*, pp.1223-1226, 1992.
5. J.A. Kangas, T.K. Kohonen and J.T. Laaksonen, "Variants of Self-Organizing Maps," *IEEE Trans. Neural Networks* 1(1), pp.93-99, Mar. 1990.
6. T.K. Kohonen, "The Neural Phonetic Typewriter," *IEEE Magazine*, pp.11-22, Mar. 1988.
7. T.K. Kohonen, *Self-Organization and Associative Memory*, Springer-Verlag, 3rd ed, 1989.
8. J.A. Freeman, *Neural-Networks*, CNS, 1991.
9. D.P. Morgan, C.L. Scofield, *Neural Networks and Speech Processing*, Kluwer Academic Publishers, 1991.
10. R.D. Strum and D.E. Kirk, *1st Principles of Discrete Systems and Digital Signal Processing*, Addison-Wesley, 1988.

정 담(Dam Chung) 정회원

1969년 2월 7일 서울 출생

1994년 : 순천향대학교 전산통계학과(학사)

1996년 : 홍익대학교 전자계산학과(석사)

1996년 1월~현재 : 삼성전자(주) 연구원

관심분야 : 신경 회로망, 음성 인식, 이동 통신, 개인 휴대 통신

이 기 철(Kee Cheol Lee) 정회원

1977년 : 서울대학교 전자공학과(학사)

1979년 : 한국과학원 전산학과(석사)

1987년 : Univ. of Wisconsin-Madison 전기 및 컴퓨터공학과(박사)

1980년~1982년 : 국방 과학 연구소 연구원

1989년~현재 : 홍익대학교 컴퓨터공학과 교수

관심분야 : 신경망, 제약만족 및 스케쥴링, 귀납적 학습, 시스템 소프트웨어

변 영 태(Young Tai Byun) 정회원

1977년 : 서울대학교 전기공학과 (학사)

1979년 : Indiana Univ. 전산학과 (석사)

1987년 : Univ. of Texas at Austin 전산학과(박사)

1977년~1979년 : 육군 복무

1979년~1982년 : D.E.C. in Korea 근무

1992년~현재 : 홍익대학교 컴퓨터공학과 교수

관심분야 : Knowledge Representation and Reasoning, Machine Learning, Neural Network, Autonomous Agents