

Cooperative Growing Hierarchical Recurrent Self Organizing Model for Phoneme Recognition

Case Study

Chiraz J*, Arous N, Ellouze N

(National School of Engineers of Tunis (ENIT)), Tunisia.

Abstract

In this paper, we propose a system of a tree evolutionary recurrent self-organizing models. Inherited from the Growing Hierarchical Self-Organizing Map GHSOM. The proposed GHSOM variants are characterized by a hierarchical model, composed of independent RSOMs (recurrent Self-Organizing Map). The case study of the proposed system is phoneme recognition in continuous speech and speaker independent context. GHSOM variants serve as tools for developing intelligent systems and proposed artificial intelligence application.

Keywords: Hierarchical Self-Organizing Map; Neural Network; Recurrent SOM; Speech Recognition; Unsupervised Learning; Cooperative System.

***Corresponding Author:**

Chiraz Jlassi,
(National School of Engineers of Tunis (ENIT)), BP 37, Belvédère 1002
Tunis, Tunisia.
E-Mail: Chiraz_jlassi@yahoo.fr

Received: March 11, 2015
Accepted: March 20, 2015
Published: March 23, 2015

Citation: Chiraz J, Arous N, Ellouze N (2015) Cooperative Growing Hierarchical Recurrent Self Organizing Model for Phoneme Recognition. *Int J Comput Neural Eng.* 02(1), 11-15.
doi: <http://dx.doi.org/10.19070/2572-7389-150003>

Copyright: Chiraz Jlassi© 2015. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Introduction

Among the large number of research publications discussing the SOM (Self-Organizing Map) [1, 2, 18, 19] different variants and extensions have been introduced. One of the SOM based models is the Growing Hierarchical Self-Organizing Map (GHSOM) [3-6]. The GHSOM is a neural architecture combining the advantages of two principal extensions of the self-organizing map, dynamic growth and hierarchical structure. Basically, this neural network model is composed of independent SOMs (many SOM), each of which is allowed to grow in size during the training process until a certain quality criterion regarding data representation is met. Consequently, the structure of this adaptive architecture automatically adapts itself according to the structure of the input space during the training process.

Speech production is a continuous and dynamic process. For a neural network to be dynamic it must be given a memory. So recurrent SOM (RSOM) [7] is similar to the original SOM except

that each neuron has an associated recursive differential equation. So the growing hierarchical recurrent self organizing map GH_RSOM [8, 9] use a hierarchical structure of multiple layers where each layer consists of a number of independent maps. Each unit of the map defines a difference vector which is used for selecting the best matching unit and also for adaptation of weights of the map. Weight update is similar to the SOM algorithm, except that weight vector vectors are moved towards recursive linear sum of past difference vectors and the current input vector.

The contribution of this work is to create a cooperative system composed of associated recurrent learning algorithm based on the GHSOM model. The goal of the system of GH-RSOMs is to create autonomous systems, the parts of which can learn from each other.

In addition to this, each GH-RSOM system participates in the decision in recognition phase.

The front-end preprocessor to the proposed competitive system is a matrix of real-valued 12-dimensional vectors of mel cepstrum coefficients. Each output unit is described by a general centroid vector and information relating to each phoneme class described by a mean vector, a label and an activation frequency [10].

In section 2, we explain the principles of the basic growing hierarchical self-organizing map algorithm and the variant of GHSOM named GH_RSOM. In section 3, we propose the new cooperative system of GH-RSOMs by detailing each of its competitive learning algorithms. In Section 4, we illustrate experimental results of the application of each isolated competitive model and associated GH_RSOMs in classification of phonemes of TIMIT speech corpus.

Basic GHSOM

The key idea of the growing hierarchical self organizing model

(GHSOM) is to use a hierarchical structure of multiple layers where each layer consists of a number of independent SOMs (Figure 1). Only one SOM is used at the first layer of the hierarchy. For every neuron in this map a SOM might be added to the next layer of the hierarchy. This principle is repeated with the third and any further layers of the GHSOM.

The GHSOM grows in two dimensions: horizontally (by increasing the size of each SOM) and hierarchically (by increasing the number of layers).

For horizontal growth, each SOM modifies itself in a systematic way very similar to the growing grid [11] so that each neuron does not represent too large an input space.

For hierarchical growth, the principle is to periodically check whether the lowest layer of SOMs have achieved sufficient coverage for the underlying input data. The basic steps of the horizontal growth and the hierarchical growth of the GHSOM are summarized in the algorithm below [12].

For growing in width, each SOM will attempt to modify its layout and increase its total number of neurons systematically so that each neuron is not covering too large an input space. The training proceeds as follows:

1. The weights of each neuron are initialized with random values.
2. The standard SOM training algorithm is applied.
3. The neuron with the largest deviation between its weight vector and the input vectors that represents is chosen as the error neuron.
4. A row or a column is inserted between the error neuron and the most dissimilar neighbour neuron in terms of input space.
5. Steps 2-4 are repeated until the mean quantization error (MQE) reaches a given threshold, a fraction of the average quantification error of neuron i , in the proceeding layer of the hierarchy.

As for deepening the hierarchy of the GHSOM, the general idea is to keep checking whether the lowest level of SOMs have achieved enough coverage for the underlying input data.

The details are as follows:

1. Check the average quantification error of each neuron to ensure it is above certain given threshold: it indicates the desired granularity level of a data representation as a fraction of the initial quantization error at layer 0
2. Assign a SOM layer to each neuron with an average quantification error greater than the given threshold, and train SOM with input vectors mapped to this neuron.

Cooperative System

In this section, we present a system of GH_RSOMs based on the association of unsupervised learning algorithms see (Figure 2).

The learning algorithms are:

- GH_RSOM based on a sequential learning.
- GH_seqopt_RSOM based on an optimized sequential learning.
- GH_DeS_RSOM with a conscience term.
- GH_Ad_RSOM with locally adapting neighborhood radii.

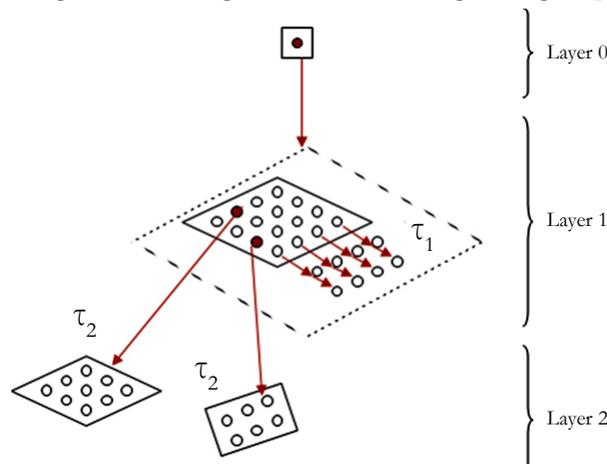
The different learning algorithms have the similar competition principle. So in this cooperative system each isolated subsystem can be considered as an autonomous system. But, each of them can learn from the other. The resulting of each subsystem is a growing hierarchical recurrent model.

Each subsystem during training phase, when a sample input vector is attributed to a BMU neuron, we save its corresponding vector, its label and updates its frequency activation, in order to enrich information [10]. By this way, each neuron of each map of the hierarchy is characterized by:

- A general centroid vector (GCV): determined by means of Kohonen update rule.
- Information relating to each phoneme class attributed to a neuron :
 - I. Mean vector (MV) of the phoneme class.
 - II. Label of the phoneme class.
 - III. Frequency activation of the phoneme class.

During recognition phase, a decision in each subsystem is operated in two steps. At a first step, for each test sample vector presented we search for the BMU among all general centroid vectors (GCV) of a map. Thereafter, inside the selected BMU neuron, we

Figure 1. Growing hierarchical self-organizing map.



search for the best mean vector (MV) of different classes of the selected neuron, in terms of minimal Euclidean distance see (Figure 3). For the selected best mean vector we retain its label and its frequency activation. Each subsystem presents to the decision system these informations (label and its frequency activation for example “ae(60)”). Decision rule maximises proposed solutions.

Reccurent GHSOM (GH_RSOM)

The recurrent SOM (RSOM) [7] is an extension to the Kohonen’s SOM that enables neurons to compete to represent temporal properties in the data. Therefore, the RSOM introduces leaky integrators (i.e. recurrent connectivity) into the neural output. A schematic picture of an RSOM is shown in (Figure 4).

The growing hierarchical recurrent self organizing map GH_RSOM is a variant of GHSOM which uses a hierarchical structure of multiple layers where each layer consists of a number of independent maps. Each neuron of the map defines a difference vector which is used for selecting the best matching neuron and also for adaptation of weights of the map. Weight update is similar to the SOM algorithm, except that weight vectors are moved towards recursive linear sum of past difference vectors and the current input vector. In the GH_RSOM each map of each layer is a recurrent SOM (RSOM) [8].

Optimized sequential learning (GH_seqopt_RSOM)

The sequential recurrent learning algorithm is modified in such a way that each unit has its own specific learning rate α_i and neighborhood radius [14].

GH_RSOM with a conscience term (GH_DeS_RSOM)

GH-DeS-RSOM. is composed of independent RSOMs (many recurrent SOM), each neuron of each map of the hierarchy is characterized by a conscience term and more than a prototype vector [15].

GH_RSOM with locally adapting neighborhood radii (GH_Ad_RSOM)

The proposed GH-Ad-RSOM is a hierarchical model, that composed of independent RSOMs (many RSOM), based on locally adapting neighborhood radii [16].

Experimental Results

In this work we have implemented four variants of the growing hierarchical recurrent SOM and the cooperative systems. We made experiments on voiced segment of continuous speech. The system is composed of three main components: first, signal filtering, vowels mel cepstrum coefficients production. The input space is composed by 16 ms frames of 12 mel cepstrum coefficients. 3 middle frames are selected for each phoneme. The second component is the learning module. The third component is the phoneme classification module.

Each output unit of the proposed GHSOM variants and the cooperative systems is described by a general centroid vector and information relating to each phoneme class described by a mean vector, a label and an activation frequency.

TIMIT corpus was used to evaluate the proposed systems in continuous speech and speaker independent context. TIMIT database contains a total of 6300 sentences, 10 sentences spoken by each of 630 speakers from 8 major dialect regions.

The data is recorded at a sample rate of 16 KHZ at 16 bits per sample.

In experiments, we used the New England dialect region (DR1) composed of 31 males and 18 females.

We have implemented the cited GH_RSOM variants and compared each isolated subsystem with the cooperative systems. For all experiments parameter r_1 which controls the actual growth process of these models is set to 0.7 and the parameter r_2 which controls the minimum granularity of data representation is set to 0.02. All maps are trained for 200 iterations and the memory coefficient is set to 0.35.

According to (Table 1) and (Figure 5), cooperative systems im-

Figure 2. Cooperative Systems.

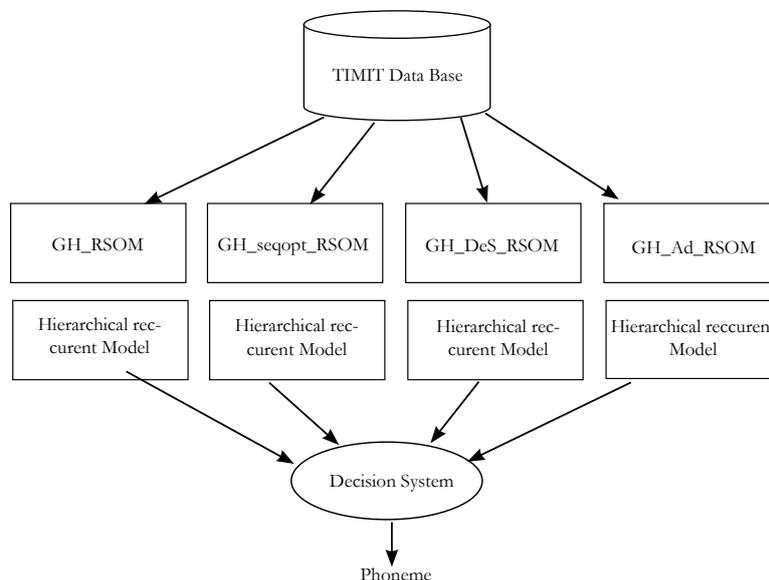


Figure 3. Recognition schema in each subsystem.

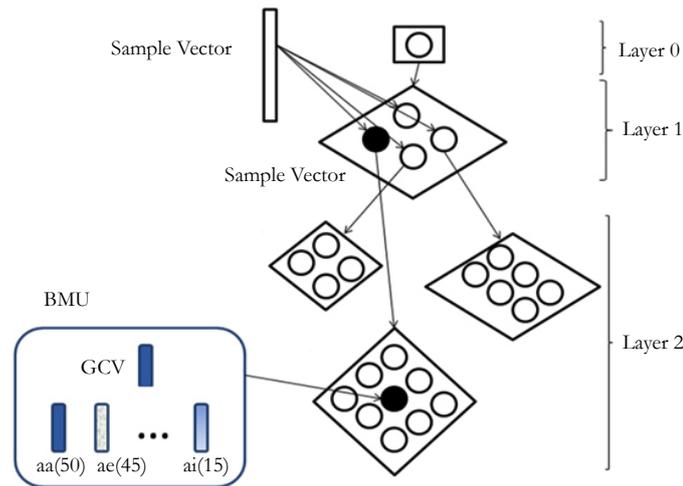


Figure 4. Recurrent self-organizing map (RSOM) (Each unit involves an exponentially weighted linear recurrent filter with the impulse response).

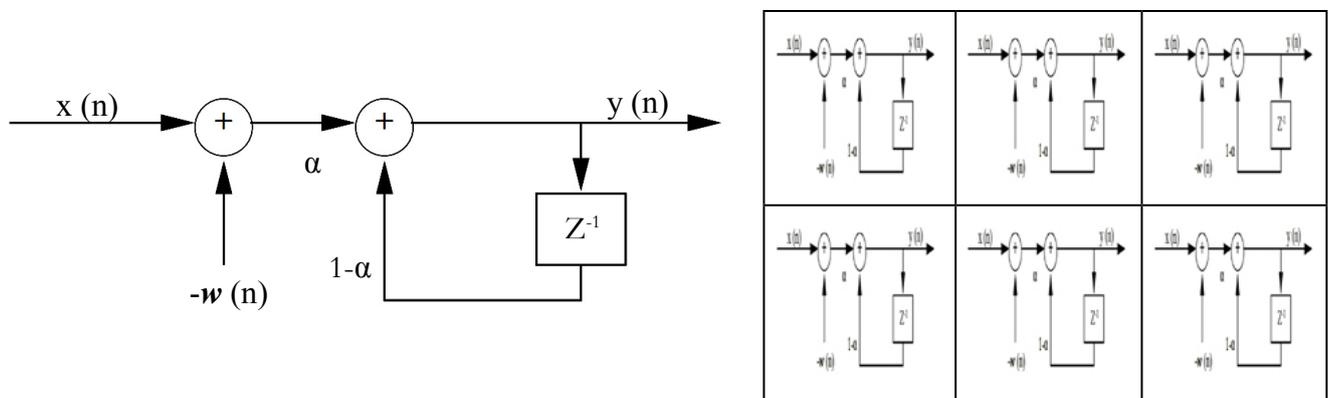
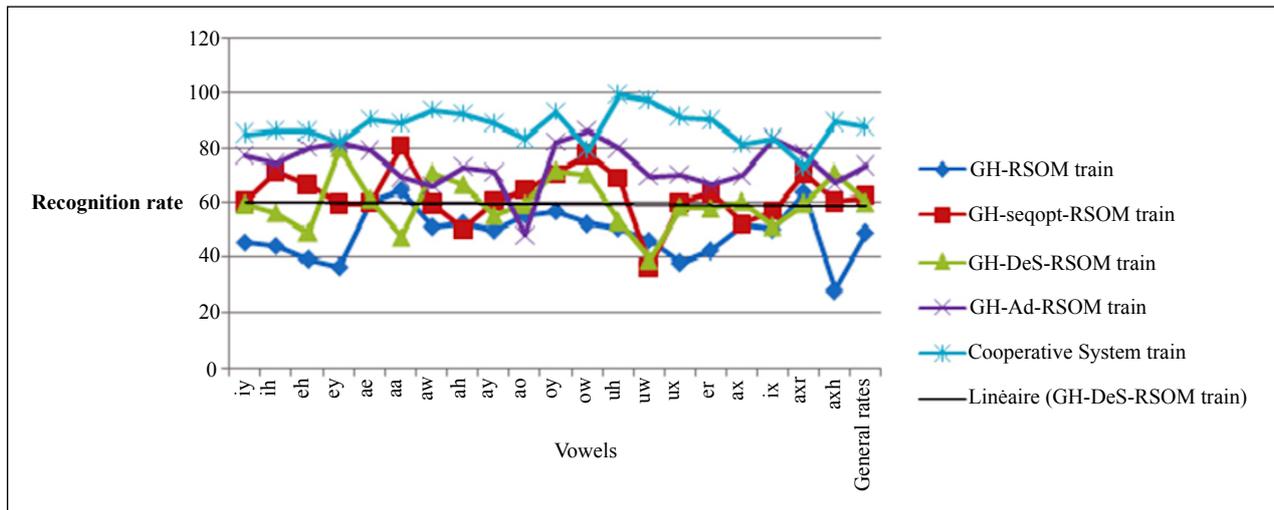


Table 1. Vowels recognition rates (training set and test set).

	GH-RSOM		GH-seqopt-RSOM		GH-DeS-RSOM		GH-Ad-RSOM		Cooperative systems	
	train	test	train	test	train	test	train	test	train	test
iy	45.48	30.21	60.01	40.28	59.38	40.22	77.45	60.19	84.95	73.13
ih	44.20	31.14	71.11	44.52	56.22	44.69	74.20	59.20	85.90	71.26
eh	39.50	26.38	66.23	52.30	49.30	39.72	80.08	66.38	85.77	68.60
ey	36.66	25.56	59.40	49.24	79.66	50.11	81.73	67.80	82.19	78.32
ae	60.28	40.20	59.66	38.52	61.20	50.44	79.20	56.66	90.33	74.18
aa	64.38	39.56	80.30	61.66	47.50	28.60	69.40	51.23	88.82	76.64
aw	51.20	37.11	59.28	33.48	70.74	59.20	66.30	49.19	93.63	69.44
ah	52.49	24.20	50.24	30.20	66.44	41.19	73.11	44.28	92.01	72.67
ay	49.89	29.50	60.44	29.50	55.32	27.40	71.29	52.30	88.55	73.28
ao	55.48	32.20	64.52	22.48	59.28	40.99	48.57	27.40	83.00	77.14
oy	56.78	20.12	70.28	29.60	71.55	46.71	82.19	56.80	92.98	68.23
ow	52.28	36.66	77.40	40.50	70.20	49.48	86.18	66.34	79.28	72.86
uh	50.78	20.14	68.71	41.28	53.40	26.50	79.88	50.19	99.02	76.60
uw	46.14	21.12	36.42	20.16	39.28	20.40	69.80	50.30	97.09	78.79
ux	38.20	20.16	59.88	46.34	58.40	47.20	70.28	51.80	91.11	77.50
er	42.58	18.42	63.29	52.20	57.89	39.40	66.47	42.20	89.84	72.45
ax	52.11	30.12	52.28	40.48	60.30	30.30	69.90	45.30	81.28	63.03
ix	50.22	31.67	56.77	49.30	51.44	31.29	83.38	56.80	83.32	66.43
axr	63.49	50.22	70.31	56.60	59.70	40.20	77.60	42.50	72.98	70.91
axh	28.14	19.38	60.12	41.38	70.66	55.19	67.20	37.80	89.23	80.23
General rates	49.01	29.20	62.33	41.00	59.89	40.46	73.71	51.73	87.56	71.69

Figure 5. Vowel Recognition Rates.



prove the recognition accuracy both in training and test set.

Conclusion

In this paper, we have presented three GH_RSOM variants, GH_RSOM with an optimized sequential learning, GH_RSOM based on a conscience term and GH_RSOM based on locally adapting neighborhood radii.

We have proposed also a system of multi_GH_RSOM based on the association of different GH_RSOM variants of unsupervised learning algorithms. The objective of such system is to create a cooperative system based on different competitive learning algorithms. The case study of such algorithms is phoneme recognition in continuous speech and speaker independent.

The aim of cooperative learning is to make different learning system collaborate, in order to reach at an agreement for a common dataset. The different learning system can produce different partitioning of the same dataset, finding a consensual classification from these results is often a hard task. the collaborative system tries to find the most relevant results For decision-making task.

The main results are as follows:

The proposed learning process based on multiple prototype vectors and information saved during training serves as a tool to reach better recognition accuracy.

GH-Ad-RSOM provides best recognition rates in comparison with other GH_RSOM variants. Cooperative system provides best recognition rates if it is proved by all subsystems.

As a future work, we suggest to study different learning variants of the recurrent GHSOM in domain application. For example hybridize GH_RSOM and genetic algorithm.

References

[1]. Kaski S, Kangas J, and Kohonen T (1998) Bibliography of self-organizing map (SOM) papers. *Neural Computing Surveys*. 1(3&4): 1-176.
 [2]. Oja M, Kaski S, and Kohonen T (2003) Bibliography of self-organizing map (SOM) papers. *Addendum Neural Computing Surveys* 3: 1-156.

[3]. Dittenbach M, Rauber A and Merkl D (2000) The growing hierarchical selforganizing map. Presented at the International Joint Conference on Neural Networks IJCNN.
 [4]. Dittenbach M, Rauber A and Merkl D (2001) Recent advances with the growing hierarchical self-organizing map. In *Advances in Self-Organizing Maps*.
 [5]. Dittenbach M, Rauber A and Merkl D (2002) Uncovering the hierarchical structure in data using the growing Hierarchical self-organizing map. *Neurocomputing* 48 (1-4): 199-216.
 [6]. Dittenbach M, Rauber A and Polzlbauer (2005) Investigation of alternative strategies and quality measures for controlling the growth process of the growing hierarchical self-organizing map. Presented at the International Joint Conference on Neural Networks IJCNN. 2954-2959.
 [7]. Varsta M, Heikkonen J, Millan. (1997) Context Learning with the Self-Organizing Map. Presented at Workshop on Self-Organizing Maps. 197-202.
 [8]. Jlassi C, Arous N and Ellouze N (2009) The growing hierarchical recurrent self-organizing map for phoneme recognition. ISCA tutorial and research workshop on Non-linear Speech Processing NOLISP. 184-190.
 [9]. Jlassi C, Arous N and Ellouze N (2010) The growing hierarchical recurrent self-organizing map for phoneme recognition. *Book Series : Lecture Notes in Computer Science Springer* 5933: 184-190.
 [10]. Arous N and Ellouze N (2003) Cooperative supervised and unsupervised learning algorithm for phoneme recognition in continuous speech and speaker-independent context. *Elsevier Science, Neurocomputing, Special Issue on Neural Pattern Recognition* 51: 225 - 235.
 [11]. Fritzke B (1995) Growing grid - a self-organizing network with constant neighborhood range and adaptation strength. *Neural Processing Letter* 2: 9-13.
 [12]. Tangsrapiroj S and M.H Samadzade (2006). Organizing and visualizing software repositories using the growing hierarchical self-organizing map. *Journal of Information Science and Engineering* 22: 283-295.
 [13]. Pampalk E, Widmer G., and Chan A (2004). A new approach to hierarchical clustering and structuring of data With self-organizing maps. *Intelligent Data Analysis Journal* 8(2): 131-149.
 [14]. Kangas J.A, Kohonen T, Laraksonen J. (1990) Variants of self organizing maps. *IEEE Transactions on Neural Networks* 1(1): 93-99.
 [15]. DeSieno D (1988) Adding a conscience to competitive learning. *IEEE International Conference on Neural Networks*: 117-124.
 [16]. Jlassi C, Arous N and Ellouze N (2010) Phoneme recognition by means of a growing hierarchical recurrent self-organizing model based on locally adapting neighborhood radii. *Cognitive computation journal, Springer* 2(3): 142-150.
 [17]. Kiviluoto K (1996) Topology preservation in self-organizing maps. *Proceeding of International Conference on Neural Networks (ICNN)*. 294-299.
 [18]. Benabdeslem K and Lebbah M (2007) Feature selection for Self Organizing Map. *International Conference on Information Technology Interface-ITI* . 45-50, Cavtat-Dubrovnik, Croatia
 [19]. Jin Xu, Guang Yang, Hong Man and Haibo He (2013) L1 Graph based on sparse coding for Feature Selection, *Lecture Notes in Computer Science, 7951: 594-601*.