Military Technical College Kobry El-Kobbah, Cairo, Egypt



6th International Conference on Electrical Engineering ICEENG 2008

Artificial Neural Network Application for Modeling of Teaching Reading Using Phonics Methodology (Mathematical Approach)

By

H. M. Hassan *

Saleh M. Al-Saleem **

Abstract:

Herein, Artificial Neural Network (ANN) Modeling is considered to mathematically formulate an interesting and rather challenging educational issue; namely, searching for optimality in an educational methodology for teaching children how to read. The adopted search approach is inspired by relevant artificial neural network modeling based on neuro-biological characterizations. That is rather than other classical approaches inspired by psychological and psycho-linguistics research directions.

dominant optimality of teaching reading phonically over other Fortunately, methodologies has been recently proven by a simulated but realistic model along with subsequent to an educational published results, field testing. Consequently, mathematical formulation of phonics methodology is a highly recommended research work to justify that optimality. Herein, mathematical formulation performed via comparative analogy with a naturally inspired artificial neural network (ANN) model. More precisely, that fulfilled on the basis of realistically simulated modeling of selforganized (unsupervised) learning paradigm originated from Hebb's learning rule. In other words, modeling of Hebbian rule essentially depends upon biological information processing to construct associative memory phenomenon after Pavlovian conditioning learning. Conclusively, presented mathematical formulation supported superiority as well as optimality of teaching reading using phonics methodology.

Keywords:

Biological Information processing, Artificial Neural Network Modeling, Educational Technology, Hebbian Learning, and psycho-learning experiments

- * Educational Technology Dept. at Faculty of Specified Education, Banha University, Egypt. Currently with Arab Open University. (Kingdom of Saudi Arabia Branch, IT Department)
- ** Arab Open University (Kingdom of Saudi Arabia Branch, IT Department)

1. Introduction:

Learning how to read is an essential step in the educational ladder. This is especially true for children during their primary school years. Thus, failure to achieve reading ability leads to nearly permanent learning disability during the following more advanced educational stages. For a long time, psycho-linguistics researchers as well as educationalists were trying effectively to find an optimal method for "how reading should be taught?" [1], [2]. During the last decade phonics method is replaced –at several U.S.A. schools - by other guided reading methods that are performed using literature-based activities [3]. Nevertheless, comparative evolutional analysis for both methodological approaches proved the superiority of phonics method [3]. It is worth mentioning here that learning by phonics is performed directly by concurrent association between the pronounced sound (phoneme) and its corresponding letter/word.

However researches in the field of psychology and linguistics are continuously performed to support obtained field results. Recently some evaluated field results tested the progressive outcome from teaching to read processes proved the optimality of phonic method adopted for teaching children how to read [1]. The National Institute of Health (NIH) in the US recently announced that when children are in elementary school, they may be learning "basic building blocks" of cognition and that after about 11 years of age, "children take these building blocks and use them." [4].

This paper is motivated by some published research work dealing with the relation between computer and education [5], [6]. Additionally this paper is well supported by other recently published papers adopting modelling and simulation of some psycho educational experiments [7, [8]. The objective of this paper is to justify and support the superiority and optimality of phonic approach over other teaching to read methods. In fulfillment of that objective, an elaborated mathematical formulation is introduced for a neuro-biologically based model. That model is built on the basis of some experimental work on animal learning psychological studies [9], [10] and is closely related to the learning paradigm known as "learning without a teacher"[11].

Additionally, the above presented view on teaching how to read methods could be characterized as one of pattern recognition and classification problem(s). This justifies adopting an ANN system for modelling biological observations detected during reading activities. A significant proportion of retained information (stored in human brain) is introduced in the form of patterns. Reading process is simply viewed as the pronunciation of any string of visualized letters as part of varied text patterns. The visual system must solve pattern recognition and classification problems in accordance with cognitive issues. In other words, the seen pattern should be transferred into its corresponding (previously stored) correlated auditory pattern [12].

Paper organization is as follows: At the following second section, a revision for detailed description of ANN generalized learning models will be presented. The basic concept of adopted model is described at the third section. Section four gives a generalized description of the model associated with mathematical equations. More detailed mathematical analysis and model formulations are introduced at the fifth section. Finally, sixth section presents conclusive remarks and discussions.

2. Revision of ANN learning model:

The block diagram shown below in Figure (1), presents an ANN model that realistically simulates a learning phenomenon. It combines both supervised and unsupervised learning paradigms.



Figure (1): Block diagram for an ANN model adopted for (adapted from [13])

Referring to above figure, error vector at any time instant (n) observed during learning processes is given by:

$$\overline{e}(n) = \overline{y}(n) - \overline{d}(n) \quad (1)$$

 $\overline{e}(n)$: Error correcting vector controlling learning adaptation process.

 $\overline{y}(n)$: Output signal of the model.

 $\overline{d}(n)$: Numeric value(s) of the desired /objective parameter of learning process (generally as a vector). Referring to above Figure (1), following equations are considered:

$$\begin{array}{ll} V_{k}(n) = X_{j}(n) \ W_{kj}^{T}(n) & (2) \\ y_{k}(n) = \phi(V_{k}(n)) = 1/(1 + e^{-\lambda v_{k}(n)}) & (3) \\ e_{k}(n) = |d_{k}(n) - y_{k}(n)| & (4) \\ W_{kj}(n+1) = W_{kj}(n) + \Delta W_{kj}(n) & (5) \end{array}$$

Where: X input vector, W weight vector, φ is the activation function, y is the output, e_k the error value, and d_k is the desired output. Noting that ΔW_{kj} (n) presents the dynamical change of weight vector value.

The above four equations are commonly applied for two different learners cognitive styles. They are namely: Field Dependent (FD) and Field Independent (FI) learning cognitive styles. Recently, both styles are realistically simulated considering supervised and unsupervised learning paradigms respectively, [14].

The dynamical change of weight vector value specifically for supervised paradigm is given by equation:

$$\Delta W_{kj}(n) = \eta \ e_k(n) \ x_j(n) \tag{6}$$

Where η is the learning rate value during learning process for both learning paradigms. However, for unsupervised paradigm, dynamical change of weight vector value is given by equation:

$$\Delta W_{kj}(n) = \eta y_k(n) x_j(n)$$
(7)

<u>3. Basic Model Concept:</u>

The reading process, by nature, consists of two basic and essential brain functions. Both functions are required to perform that process efficiently. The first function is to classify seen/heard patterns i.e. pattern classification. From neuronal networks point of view that function originates in the perception, and essentially requires supervisor's (teacher's) intervention for learning completion. Secondly, associative memory function which is originally based on classical conditioning motivated by Hebbian learning rule. It belongs to the principle of learning without a teacher (unsupervised). Referring to Figure (2) below, the illustrated teaching model is well qualified to perform simulation of above mentioned brain functions. Inputs to the neural network learning model at that Figure, are provided by environmental stimuli (unsupervised learning). The correction signal for the case of learning with a teacher is given by responses outputs of the model will be evaluated by either the environmental conditions (unsupervised learning) or by the teacher. Finally, the

teacher plays a role in improving the input data (stimulating the learning) by reducing noise and redundancy of the model input. That is according to the teacher's experience, he provides the model with clear data with maximum signal to noise ratio. However, that is not our case which is based upon unsupervised Hebbian autonomous learning, which equivalently called self organized learning.



Figure (2): Simplified suggested teaching reading model that interacts with environmental conditions, (adapted from [15])

4. Description of The Model:

Referring to the two figures shown below, the suggested models complies with that concept as the two inputs I1, I2 represent sound (heard) stimulus and visual (sight) stimulus respectively. The outputs O1, O2 represent pronouncing and image recognition processes respectively. In order to justify the superiority and optimality of phonic approach over other teaching to read methods, an elaborated mathematical representation is introduced for two different neuro-biologically based models. Any of the models needs to learn how to behave (to perform reading tasks). Somebody has to teach (for supervised learning) - not in our case - or rather for our learning process is carried out on the basis of former knowledge of environment problem (learning



Figure (3): Generalized model considering input stimuli and output responses, (adapted from [15])



Figure (4): The structure of the first model where reading process is expressed by conditioned response for seen letter/word (adapted from [7])

The model obeys the original Hebbian learning rule. The reading process is simulated at that model in analogues manner to the previous simulation of Pavlovian conditioned learning. The input stimuli to the model are considered as either conditioned or unconditioned. Visual and audible signals are considered interchangeably for training the model to get desired responses at the output. Moreover, the model obeys more elaborate mathematical analysis for Pavlovian learning process [7]. Above that, the model is modified following general Hebbian algorithm and correlation matrix memory [11].

The adopted model is designed to mirror simulation of previously measured performances of classical conditioning experiments. Model design concept is presented after the mathematical transformation of some biological hypotheses. In fact, these hypotheses are derived according to cognitive/ behavioural tasks observed during the experimental learning process. Generally, the output response signal varies as shown in the original Pavlov experimental work [9], where it is measured quantitatively in the exactness of pronouncing letter/word. In accordance with biology, the output response signal is

dependent upon the transfer properties of the output motor neuron stimulating pronunciation as unconditioned response (UCR) for heard phoneme (sound signal). However, this pronounced output is considered as conditioned response (CR) when input stimulus is given by only sight (seen letter/word). The structure of the model following originality of Hebbian learning rule; in its simplified form (with single neuronal output) as shown at Figure(3). Therein; A and C represent two sensory neurons (receptors), while B and D are presenting nervous subsystems developing output responses . By some details, that generalized simple structure drives an output response (pronunciation) that is as O_1 . However the other output response presented at Figure (3) as O_2 is obtained when input sound is considered as conditioned stimulus. Hence visual recognition as conditioned response of the heard letter/word is obtained as output O₂. In accordance with biology, the strength of response signal is dependent upon the transfer properties of the output motor neuron stimulating salivation gland. By referring to figure (4), it represents the classical conditioning learning process, where each of lettered circles A, B, and C presents a neuron cell body. The line connecting cell bodies represents an axon that terminates synaptic junctions. Both of two signals released out from sound and sight sensory neurons A and C are represented by y_1 and y_2 . The activation function for both neurons A and C are suggested to be as a fraction of signum function, rather than sigmoid, as suggested at [7]. Considering, some more simplification of any neuron cell arguments, the differential equation describing electrical neural activity has been suggested, as follows:

$$\frac{do_i}{dt} = \sum_{j=1}^n f(y_{ij}) - j(o_i)$$
(8)

where, y_{ij} represents the activity at the input (j) of neuron (i), $f(y_{ij})$ indicates the effect of input on membrane potential, $j(o_i)$ is nonlinear loss term combining leakage signals, saturation effects occurring at membrane in addition to the dead time till observing output activity signal. The steady state solution of the above differential equation (8), proved to be presented as transfer functions. Assuming, the linearity of synaptic control effect, the output response signal is given as:

$$O_i = \phi \left(\sum_{j=1}^n w_{ij} y_{ij} - \theta_i \right)$$
(9)

where, ϕ has two saturation limits, the function ϕ may be linear above a threshold and zero below or linear within a range but flat above θ_i is the threshold (offset) parameter, and w_{ij} synaptic weight coupling between two neuron (*i*) and (*j*). Specifically, the function (ϕ i) is recommended to be chosen as a ramp or sigmoid signal function [7]. However, the ramp function was used to represent the output response of the presented model, because of it's

EE162 - 8

mathematically similarity to sigmoid function. By referring to the weight dynamics described by the famous Hebb's learning law, the adaptation process for synaptic interconnections is given by the following modified equation:

$$\frac{d\omega_{ij}}{dt} = \eta z_i y_{ij} - a(o_i)\omega_{ij}$$
(10)

Where, the first right term corresponds to the unmodified learning (Hebb's law) and η is a positive constant. The second term represents active forgetting; a (o_i) is a scalar function of the output response (o_i). Referring to the model structure given at figure (4); the adaptation equation of the single neuronal model (between *i*, *j* nodes) is as follows:

$$w_{ij}^{\cdot} = -aw_{ij} + \eta y_i x_{ij} \tag{11}$$

activities. The k^{th} pair of patterns is represented by the key vector X_k and the memorized Y_k .Let the key vector X_k with m input space dimensionality, which represents implicitly both heard phoneme (sound signal) stimulus and its corresponding seen letter/word (visual signal) stimulus. Hence, this implicit vector should be decomposed into two smaller sub-vectors each with dimensionality less than (m). Assume that sound signal is represented (simulated) as vector with r dimensionality. Obviously the correlated visual signal is simulated as a vector with dimension (m-r). This means in practical application that diminishing any of the two smaller vectors implies non existence of either input stimuli. In other words input conditioned or unconditioned stimulus is detected by measuring input space dimensionality of vector X_k . Similarly, the memorized vector Y_k represents two unconditioned/conditioned responses to the input stimulus vector X_k. However the dimensionality of that memorized response vector differs from that for input vector. Consequently, considering that Y_k vector have 1 dimensionality; hence it decomposes into two smaller vectors as unconditioned and/or conditioned responses. Let Y_k vector with *l*-dimensionality implicitly include both output response signals i.e. when pronouncing signal vector have s-dimensionality the other recognizing process of seen letter/word is simulated as a vector with (l-s) dimensionality. As a consequence of above descriptions of both vectors X_k , Y_k the following equations illustrate well asassociated memorization process during and /or after performing completion of learning convergence (reading activities).

5.2 Associated Memorization Equations

Consider x'_k and x''_k are the two vectors simulating heard and seen input stimuli respectively. Similarly y'_k and y''_k are the two vectors simulating pronouncing and visual recognizing output responses respectively. The two expected unconditioned responses are described in matrix form as follows:

$$Y_{k}' = W(k) \cdot X_{k}'$$
, $k = 1, 2, 3, ..., q$ (12)

Where W(k) is a weight matrix determined solely by the input-output pair (X_k, Y_k)

$$y_{ki} = \sum_{j=1}^{r} w_{ij}(k) \cdot x_{kj} \qquad , i = 1, 2, ..., r$$
(13)

Where $W_{ij}(k)$, j = 1, 2, ..., r; are the synaptic weights of neuron *i* corresponding to the k^{th} pair of associated patterns of the input-output pair (X_k, Y_k) . We may express y_{ki} in the equivalent form

$$y_{ki} = \begin{bmatrix} w_{i1}(k), w_{i2}(k), \dots, w_{ir}(k) \end{bmatrix} \begin{bmatrix} x_{k1} \\ x_{k2} \\ \dots \\ x_{kr} \end{bmatrix}; i = 1, 2, \dots, s$$
(14)

Similarly for visual input stimulus $x_k^{"}$ and recognizing (of seen letter/ word) output response $y_k^{"}$

$$y_{ki} = \begin{bmatrix} w_{ir+1}(k), w_{ir+2}(k), \dots, w_{im-r}(k) \end{bmatrix} \begin{bmatrix} x_{kr+1} \\ x_{kr+2} \\ \dots \\ x_{km-r} \end{bmatrix}$$
(15)
$$i = s + 1, 2, 3, \dots, l$$

For conditioned response the input hearing stimulus x_k^{\dagger} results in recognizing of visual signal y_k^{\dagger} . However input seen letter/ word stimulus x_k^{\dagger} results in pronouncing that letter/ word as conditioned response vector y_k^{\dagger} which expresses the reading activity given by the equation

$$y'_{ki} = [w_{ir+1}(k), w_{ir+2}(k), ..., w_{im-r}(k)] \begin{bmatrix} x''_{kr+1} \\ x'_{kr+2} \\ \\ x''_{km-r} \end{bmatrix}$$
(16)
$$i = 1, 2, 3, ..., s$$

In a similar manner the other conditioned response for recognizing heard phoneme is described by the equation

$$y_{ki}^{"} = [w_{1}(k), w_{2}(k), ..., w_{r}(k)] \begin{bmatrix} \dot{x}_{kr+1} \\ \dot{x}_{kr+2} \\ \\ \dot{x}_{km-r} \end{bmatrix}; i = 1, 2, ..., s$$
(17)

As a result of the above equation the memory matrix that represents all q- pairs of pattern associations is given by m^*l memory correlation matrix as follows

 $M = \sum_{k=1}^{q} W(k)$, where W(k) weight matrix is defined by

$$W(k) = \begin{bmatrix} w_{11}(k) & w_{12}(k) & \dots & w_{1m}(k) \\ w_{21}(k) & w_{22}(k) & \dots & w_{11}(k) \\ \dots & \dots & \dots & \dots \\ w_{l1}(k) & w_{l2}(k) & \dots & w_{lm}(k) \end{bmatrix}$$
(18)

This weight matrix relating input stimulus vector with m-dimensionality x_k connected by synaptic with output response vector Y_k with *l*-dimensionality. The complete relation for input/ output relation is given by the following equation

$$\begin{bmatrix} y_{k1} \\ y_{k2} \\ \dots \\ y_{kl} \end{bmatrix} = \begin{bmatrix} w_{11}(k) & w_{12}(k) & \dots & w_{1m}(k) \\ w_{21}(k) & w_{22}(k) & \dots & w_{11}(k) \\ \dots & \dots & \dots & \dots \\ w_{l1}(k) & w_{l2}(k) & \dots & w_{lm}(k) \end{bmatrix} \begin{bmatrix} x_{k1} \\ x_{k2} \\ \dots \\ x_{km} \end{bmatrix}$$
(19)

It is worthy to note that, the above equation represents memory correlation matrix after learning convergence. Therefore, this matrix is given in other way as:

$$M = Y \cdot X^T \tag{20}$$

The above equation illustrates that all values of memory matrix M elements present synaptic weights relating key pattern X with memorized stored patterns Y. In other words, the relation between input patterns to the proposed model and that model's output patterns is tightly closed by the steady state values of the memory matrix M after reaching of learning convergence. Noting, that learning process obeys well the above presented ANN model performance illustrated at Figure (1).

6. Conclusions and Discussions:

However, there are known problems of associated memories such as pattern overlaps causing degraded storage and spurious retrievals [11], this paper considers mainly memories associated with classical conditioning learning process, [9], [10]. Moreover, this work has been accomplished following recent research approach concerned with relation between computer and educational technology, [13], [14], [17]. More specifically, the objective of presented work herein is to provide educationalists and psycho-linguistics researchers as well with mathematical formulation for a novel decisive solution concerned with the great debate issue for "how reading should be taught?"[1]. This objective is fulfilled exactly by adopting a neuro-biologically-based ANNs model supporting optimality and superiority of phonics method over other approaches for teaching children how to read.Additionally, that superiority is well supported by practical field results declared in U.S.A. , after application of teaching

reading by phonics methodology,[3].

Conclusively, adoption of ANNs modelling (biologically-based) seems to be a relevant tool to accomplish realistic simulation of such educational reading activities phenomenon. Additionally, effective support of our adopted phonics method is based upon mathematical analysis as well as biological evaluation and analysis of psychoeducational experimental studies. Moreover, correlation matrix memory is recursively updated according to successive provided pairs of stimuli until complete storage of all q pairs of patterns associations (referring to sub section 5.1 dealing with memory model). Finally, this work seems to open future research for more evaluation and elaborated studies which aim to achieve better optimised strategies concerned with other educational issues. This work proves previously suggested research direction when designing systems "The more biological based system models, the more optimality could be reached"[18]. As direct implication for future educational research issues: understanding learning processes carried out in ANNs models is recommended for increasing efficiency and effectiveness of some educational activities. As an example, the case of learning arithmetic using teaching method based on more biological based models. Finally, evaluation of some problems concerned with word recognition. may be carried out by above modeling approach, such as semantic paralexia observed via phonological reading route, and other articulatory system deficiency as Aphasia, [19], [20].

References:

- [1] Jeanne S. Chall 1996. : Learning to read, the great debate. Harcourt Brace.
- [2] Keith Rayner, et al., 2001: How Psychological Science Informs The Teaching Of Reading, Psychological Science In Public interest. Vol. 2, No. 2, 31-74
- [3] Keith Rayner, Barbara R. Foorman, Charles A. Perfetti, David Pesetsky, Mark S. Seiedenberg 2003: How Should Reading Be Taught?, Majallat Aloloom, volume 19, number 6/7 4-11.
- [4] Swaminathan,N 2007: Cognitive Ability Mostly Developed Before Adolescence, NIH Study Says, 2007. NIH announces preliminary findings from an effort to create a database that charts healthy brain growth and behavior. Scientific American letter, May 18.

- [5] Yu, Fancies T. 2002: Optical Neural Network & Education, Proceeding Sevens SPIE, Vol. 4588, 234-242.
- [6] Lee, F.L. 1998: A Computer Simulation of The Mathematics Learning Processes Using Harmony Theory, Beijing, China; Higher Education Press & Berlin: Springer-Verlag.
- [7] Hassan H. and Watany M, 2000: On Mathematical Analysis of Pavlovian Conditioning Learning Process using Artificial Neural Network Model , 10th Mediterranean Electro-technical Conf., Cyprus.
- [8] Hassan H. and Watany M., 2003: On Comparative Evaluation And Analogy For Pavlovian And Throndikian Psycho-Learning Experimental Processes Using Bioinformatics Modelling, AUEJ, 6,3, 424-432.
- [9] Pavlov, I.P 1927: Conditional Reflex, An Investigation of The Psychological Activity of the Cerebral Cortex, New York, Oxford University press.
- [10] Hebb, D.O. 1949: The organization of Behaviour, A Neuropsychological Theory, New York, Wiley.
- [11] Haykin S., 1999: Neural Networks, Englewood Cliffs, NJ: Prentice-Hall pp 50-60.
- [12] R. Beale and T. Jackson, 1990: Neural Computing, an introduction, Adam Hiler.
- [13] H.M. Hassan "Performance Evaluation of Artificial Neural Network Modeling Considering Learning / Training Convergence Time" published at International Conference of Electrical Engineering, M.T.C. Cairo Egypt, 24-26, Nov. 2004.
- [14] Hassan H.M 2005 :On Simulation of Adaptive Learner Control Considering Students' Cognitive Styles Using Artificial Neural Networks (ANNs), IEEE International Conference CIMCA 05, Veinna-Austuria, 28-31 Nov.
- [15] Hassan H.M.; et.al. Towards Evaluation of Phonics Method for Teaching of Reading Using Artificial Neural Networks (A Cognitive Modeling Approach)", published at IEEE Symposium on Signal Processing and Information Technology Seventh Symposium held in Cairo-Egypt 15-18 December 2007.
- [16] Freem n., A.J. 1994: Simulating Neural Networks with Mathematics, Addison-Wesley publishing company
- [17] Hassan, H.M. 2005: On Mathematical Analysis, and Evaluation of Phonics Method for Teaching of Reading Using Artificial Neural Network Models; published at SIMMOD, Jan.17-19; pp 254-262.
- [18] Caudil, M. 1992: The view from now", AI Export Magazine, pp.24-31.
- [19] Morton, J. 1979 :Word recognition, in Morton, J. and Marshall, J.C. (eds.), psycholinguistic Series, 2. Cambridge, Mass. MIT Press.
- [20] Temple, C. 1993: The Brain: An Introduction to Psychology of the Human Brain and Behaviour, Penguin Press England.

BREIF BIOGRAPHY

HASSAN M. HASSAN: is Associate Professor of Computer Science& Information Technology. Head of Educational Technology, (Computer Science Department) at College of Specified Education, Banha University, Banha - Egypt. He received his B.Sc. degree in Electrical Engineering from Military

Proceedings of the 6th ICEENG Conference, 27-29 May, 2008

Technical College (M.T.C.) Cairo-Egypt in 1970, at Communication Engineering. He worked as a Mandated Tutor / Lecturer at Department of Physics in M.T.C. during (1979-1990). In 1983, he received his M. Sc. degree from M.T.C. Cairo-Egypt, related to Microprocessors' Applications in Digital Communication. During (1983-1990), he chosen to perform some R&D missions concerned with Digitization of analogue systems. That is through his work as researcher at Technical Research Center / M.T.C. Cairo-Egypt. He held his Ph. D. from Ain Shams University Faculty of Engineering, Computer & Systems Department in summer 1996. Moreover, He attended successfully many of training courses concerned with Quality Assurance /control, and others related to development Staff members' capabilities at Banha University. He is a member of some Scientific Societies associated with Education and Communication as well. He was chosen as Chairman for some of scientific sessions, and referee to many papers for International conferences IEEE. His fields of interest are Artificial Neural Networks, Natural Inspired Computations, and their applications for modeling of Communication systems and for evaluation of learning processes /phenomena. Moreover, according to the German DFG organization scientific research program (during academic year 2005/2006) some interesting research work has been carried at Institute for Electronics, Signal Processing and Communications (IESK) Otto-von-Guericke-University Magdeburg that results in a set of recently published papers at well known international conferences.

Finally, He was the author / coauthor of about 35 published/accepted papers and technical articles at international specialized conferences and journals. That is during the time period from 1982 until 2007.

His E-mail address is hassan_sasa_2004@yahoo.com