Accepted Manuscript

Title: Age of Acquisition Effect: Evidence from Single Word Reading and Neural Networks

Author:

Ahmad Sohrabi [1,*]

[1]Department of Psychology, Faculty of Humanities and Social Sciences, University of Kurdistan, Sanandaj, Iran

To appear in: Basic and Clinical Neuroscience

Received date: 2014/07/19
Revised date: 2018/05/6
Accepted date: 2018/01/23

Please cite this article as:
Sohrabi, A. (In Press). Age of Acquisition Effect: Evidence from Single Word Reading and Neural Networks. *Basic and Clinical Neuroscience*. Just Accepted publication Jan. 23, 2018.

* Corresponding author: Ahmad Sohrabi, Department of Psychology, University of Kurdistan, Sanandaj, Iran, Postal Code: 66177-15175, Tel: +98-8733664600, E-mail: sohrabya@gmail.com

Abstract
Many studies have shown that words learned early in life are read better than words learned later and are less vulnerable to brain damage. In the first part of this research, 25 primary school students in the fifth grade read word groups learned initially during a previous grade. The words used in the experiments were 327 Farsi monosyllable words matched on other factors involved in Farsi word naming. The analysis of covariance (the consistency and frequency as covariates) on the data showed that words learned in earlier grades are read better than words learned later, showing the known Age of Acquisition (AoA) effect. In the second part of this study, a number of simulations were carried out to simulate AoA in word naming by a neural network model developed earlier based on connectionist approach. While previous studies have used random patterns, in this research words from primary school books were used. Like human data, words learned early by the model were read better than words learned later. However, there was a failure in replicating previous simulation of AoA in English reading by an algorithm called Quick prop here for Farsi. In addition, the model was lesioned by removing some hidden units to see its effect on word reading. As a result, words learned earlier were less vulnerable to damage compared to ones learned later. These effects are explained by considering the nature of learning in neural networks trained by error back-propagation.

1. Introduction

If the important variables in word recognition such as consistency, frequency and word length are being equal, it has been shown that still words learned early in life are read better than words learned later and are less vulnerable to brain damages. These effects have also been shown in other cognitive processes including object naming, face recognition, and spoken word recognition (Morrison & Ellis, 1995; Gerhand & Barry, 1998; Zevin and Seidenberg, 2004; Izura, Pérez, Agallou, Wright, Marín, Stadthagen-González and Ellis, 2011; Wilson, Ellis & Burani, 2012).

In several cognitive domains, early learning can result in a decrease in plasticity, limiting the ability to acquire new information. Phonological acquisition is a good example (Werker & Tees, 1984), i.e., learning the phonological structure of a language lessens the ability to learn new phonetic contrasts (e. g., in a second language).

Likewise, studies have shown that the ability to learn the morphology and syntax of a language drops steadily after approximately seven years of age (Flege, Yeni-Komshian, & Liu, 1999). However, other faculties such as lexical acquisition do not seem to be age-dependent that much (Markson & Bloom, 1997; McCandliss, Posner, & Givon, 1997).

Carroll and White (1973) were able to find a relationship between word learning age and later age processing speed. They showed that object naming latency has a high correlation with the age at which children learn the different object names. They further found, multiple regression analysis, that age of acquisition was the only significant independent variable for predicting the naming latency.

The better performance in the case of words learned in early childhood isn't due to more exposure (called frequency) in that time and later (called cumulative frequency) but has its own effect. Therefore, for example, words learned in early grades of primary school (e. g., grades 1 and 2) are recognized and read better than words learned later (e. g., grades 4 and 5). There are other important factors involved in word recognition such as frequency, consistency (Coltheart's N) and length of words (e.g., Colthert et al 1977; Seidenberg, 1985; in Farsi, Sohrabi, 1999). Words that are high frequent and/or consistent are better recognized than low frequent and/or inconsistent words.

Like many cognitive processes, word reading is simulated and explained well by connectionist model or artificial neural network (e. g., Seidenberg and, McClelland, 1989; Zorzi, et al, 1998, Sohrabi, 2001; Zevin and Seidenberg, 2004). Thus this research has two parts. One is about investigating word naming in human subjects and the other, simulating by connectionist models.

2. Materials and Methods

2.1. Part I: Investigating the age of acquisition effects by human data in primary school

In this part of the study, the aim was to show that there is an effect of AoA, controlling other important factors in naming words in primary school books in 5th grade students.

2.1.1. Materials

The items were 327 words from primary school books. All words were grouped based on their initial appearance in one of the five grades. In addition, two other important factors were also considered: consistency by Coltheart's N, frequency by objective measurement in primary school books as a good estimate of words that children of a small town in Iran are exposed to in primary school. All words were monosyllables.

2.1.2. Participants

Participants in this experiment were twenty five 5 graders near the end of the academic year. All participants were male with no problem in speech and vision. At the end of the experiment, playing a computer game and an ice-cream were offered to all participants.

2.1.3. Procedure

All words printed in an MS-Windows font with size 100 points each in a 3" by 6" card. The cards were put in random order. Subjects, sat in front of the experimenter, one by one, in a quiet room and all words presented to them after the following instruction:

"You will see some words from the primary school, each on a card. Read them aloud when they are shown one by one, as soon as possible."

The experiment took about 25 minutes for each subject. The reading errors were recorded by an assistant on a 3-level scale: Failure (3), failure with correction (2) and considerable delay (1).

2.2. Part II: Simulating AoA by a connectionist model

2.2.1. Methods

Neurocomputational modeling has beeen employed for simulating and explaining cognitive processes (e.g., McClelland and Rumlehart, 1986; Seidenberg and McClelland, 1989; Zorzi et al., 1998, Zevin and Seidenberg, 2004; Sohrabi and West, 2009; Ludvig, Sutton, and Kehoe, 2012) and lower and higher level neural functions (Daneshparvar and Daliri, 2012; Soltanzadeh and Daliri, 2014; Friston and Frith, 2015). However, the AoA had not been modeled until recent years (Ellis and Lambon-Ralph, 2000). Previously, McClelland et al. (1995) simulated the graded improvement of children's lexicon. They examined the outcomes of adding a novel concept (penguin) after the network had been trained. This was done using either focused or interleaved learning. With focused learning, the system is presented to new knowledge without interleaving with old knowledge, i.e., no further exposure to the earlier training set. Under this condition, information about penguins was learned rapidly but at a cost to pre-existing knowledge. In other words, the model underwent catastrophic interference. However, in the case of interleaved learning, i.e., the model being exposed to the old material alongside the new, then the new information was learned without cost to the old.

Ellis and Lambon-Ralph (2000) based on such findings studied the AoA effect on random patterns through some simulations. They showed that AoA effect is different from cumulative frequency and attributed it to inevitable consequence of losing flexibility in artificial neural networks as it is the case for matured subjects. This is because of error back-propagation algorithm that changes the connections weights at the early training more than later training. So, early training has more effect than late one. In this part of the research, using words from primary school books with their real frequency, some simulations were carried out to compare with human data.

2.2.2. Architecture of the model

A distributed connectionist model based on Seidenberg and McClelland (1989) was used which was adopted for Farsi language in a previous work (Sohrabi, 2001).

In such models neuron-like units are used for connection between input (letters) and output (phonemes).

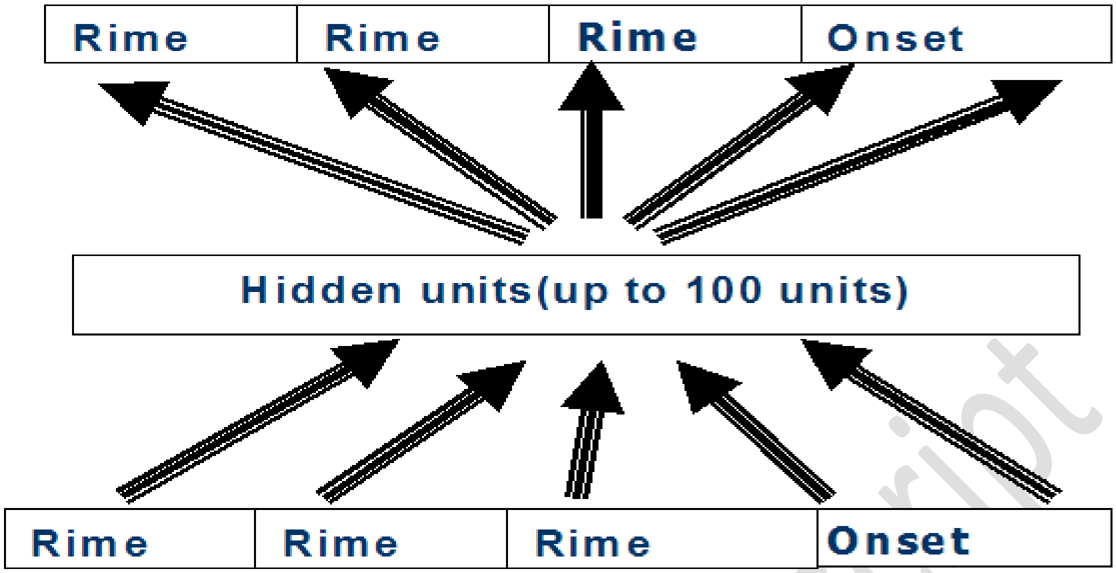

Figure 2. A distributed connectionist model based on Seidenberg and McClelland (1989) was used which was adopted for Farsi language in a previous work (Sohrabi, 2001).

In the model letters and phonemes are distributed among words, for each one a certain pattern of activation is involved. The input layer is 33 letters in Farsi language with Arabic script and its output layer is 28 phonemes (in contrast to English, Farsi has less phonemes than letters). For better representation of rimes that have important roles in reading Farsi (Sohrabi, 1999) like English (Coltheart et al, 1977), four slots of units were considered for input and output layers similar to Zorzi (1998) and Sohrabi (2001).
Additionally, because Farsi has deep (quasi-regular) orthography (Sohrabi, 1999), a hidden layer including 100 units was used. Thus a model with an input layer of 132 units and an output layer of 112 units was used in all simulations (see Figure 2). At the start of the training, connections were weighted initially by random values between +0.5 and -0.5. Then the model was trained by error back-propagation algorithm (Rumelhart et al, 1986). The learning rate was 0.05 in all simulations except for 1, 2 and 3 in which it was 0.01. The momentum value that speeds up the learning was 0.9 in all simulations but simulations 5, in which it was omitted to see its effect.

3. Results

3.1. Human Data

Mean errors for each word made by all subjects were analyzed by Analysis of Covariance (item analysis) as dependent factor.

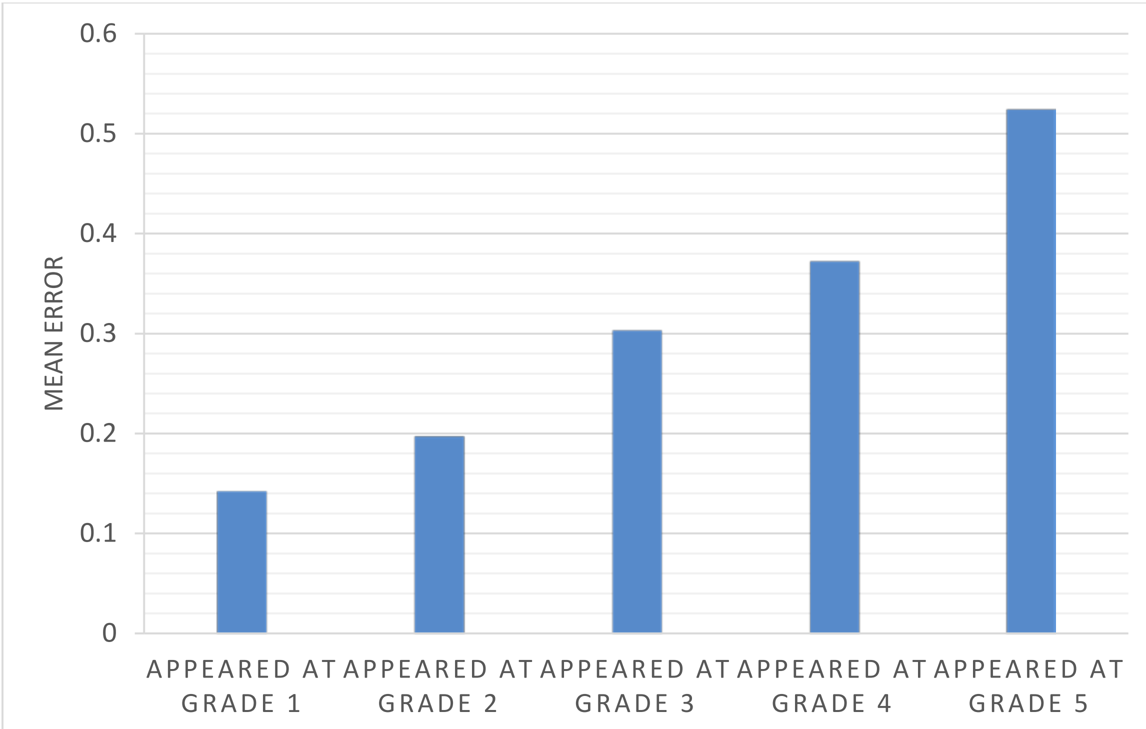

Figure 1. Mean error for word groups

Five grades at which words were initially learned made the five levels of fixed factor. Consistency and frequency were two covariates, as shown to be important factors in reading Farsi words by a multiple regression analysis (Sohrabi, 1999).The frequency as a covariate wasn't significant but included in analyses to control variance as shown in Tables 1 and 2 (also see Figure 1). The consistency covariate was significant, so AoA is mainly due to inconsistent words.

Table 1. Mean and Standard Deviation of Reading Errors in 5 Word Groups

Word Groups	N	Mean	*SD*
Appeared at Grade 5	78	0.524	0.5779
Appeared at Grade 4	49	0.3724	0.5393
Appeared at Grade 3	56	0.3036	0.4357

Appeared at Grade 2	62	0.1986	0.2913
Appeared at Grade 1	82	0.1441	0.3218
Total	327	0.3066	0.4649

As can be seen in Tables 1 and 2, word initially learned in early grades have less errors compared to those learned in later grades. The effect of word groups was significant [$F_{4, 320}$= 7.757; MSE = 0.17; $p < 0.001$].

Table 2. Analysis of Covariance on Reading Errors in 5 Word Groups

Source	SS	*df*	MS	F	Sig.
Frequency	0.123	1	0.123	0.696	0.405
Consistency	7.089	1	7.089	40.166	0.000
AoA	5.476	4	1.369	7.757	0.000
Error	56.476	320	0.176		

Though the differences between immediate grades were not significant (except for 4 and 5), other contrasts were significant. Additionally, there was a significant linear trend for AoA effect. Controlling other factors, the earlier the word, the less the error.

3.2. Modeling Data

3.2.1. Simulation 1

This simulation was intended to show that there is an effect of AoA even by controlling the frequency effect. First, the model was trained for 150 epochs on 82 early words and then for 150 epochs on 82 early words (from grade 1) as well as 87 late words (from grade 2). The late words were trained twice to be equal to early in frequency. The words were trained in a random order.

The result shows that early words have significantly less errors than late ones (see Figure 3).

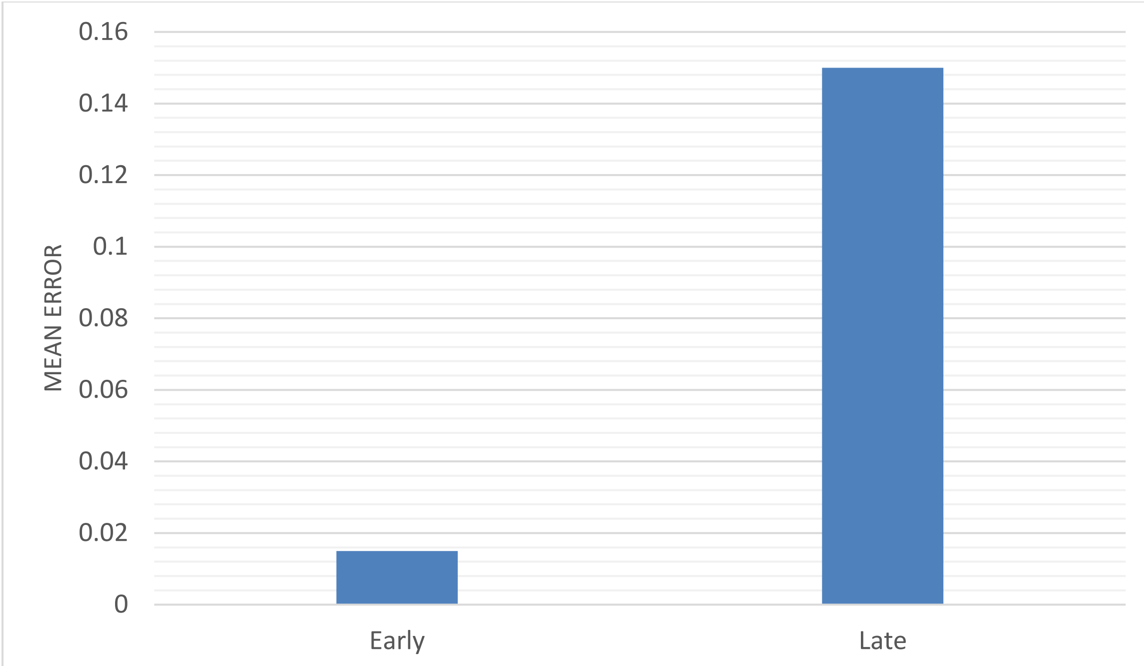

Figure 3. The errors for Early and Late after 300 epochs

3.2.2. Simulation 2

This simulation is the same as simulation 1. But here, training the mixed words groups continued twice as shown in Figure 4. The effect of AoA remains after several epochs and early group keeps its superiority, as shown in human being after some decades (Ellis et al., 1996).

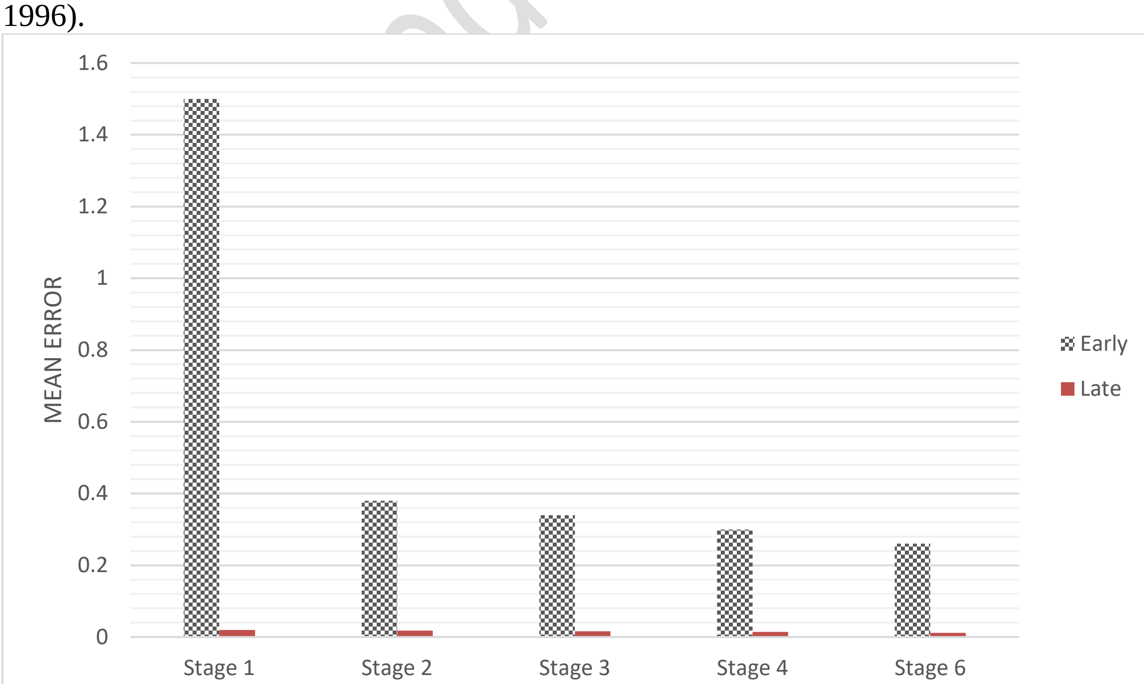

Figure 4. The errors for Early and Late after 600 epochs

3.2.3. Simulation 3

The aim of this simulation is to control the possible difference between the selected 2 groups. Thus the order to train the 2 groups reversed. Again, the gain for the early one was seen (see Figure 5).

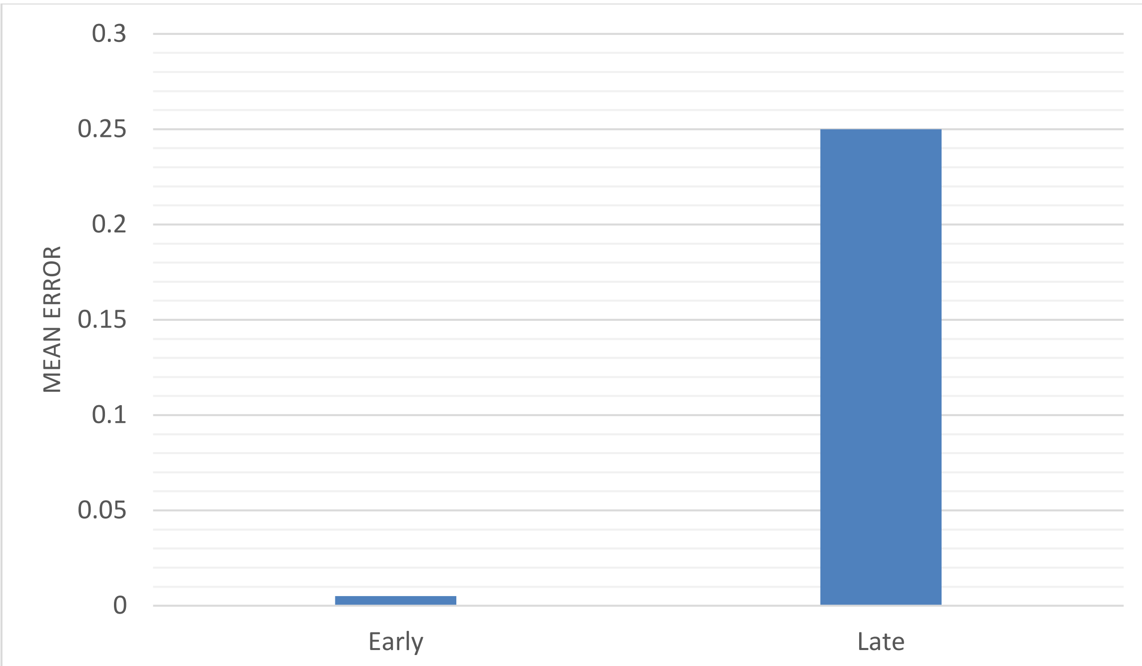

Figure 5. The errors for Early and Late after 300 epochs with reversed order.

3.2.4. Simulation 4

This simulation was like simulation 1 but the numbers of epochs in both groups were twice and the learning rate was 0.05.
As shown in Figure 6, the error of early was much less than late and it is independent of learning rate.

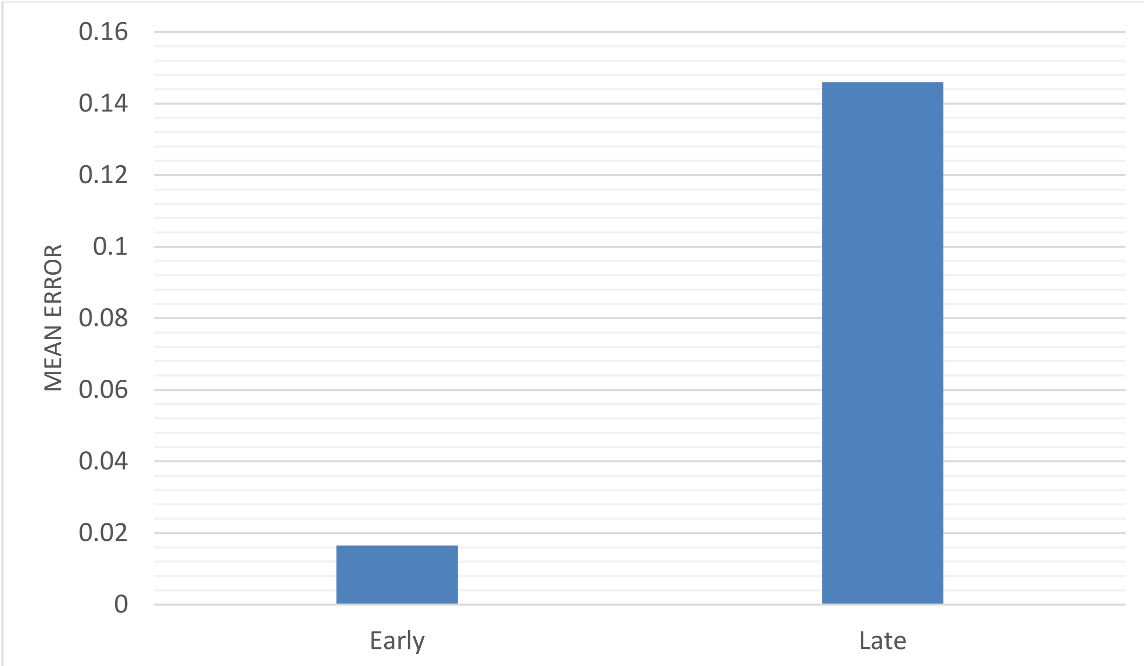

Figure 6. The errors for Early and Late after 300 epochs with learning rate 0.05

3.2.5. Simulation 5

Here the aim was to control the momentum, the value that speeds up the learning. Without momentum, after 3000 epochs (1500 early and 1500 mixed), AoA effect was observed like the other simulations as shown in Figure 7.

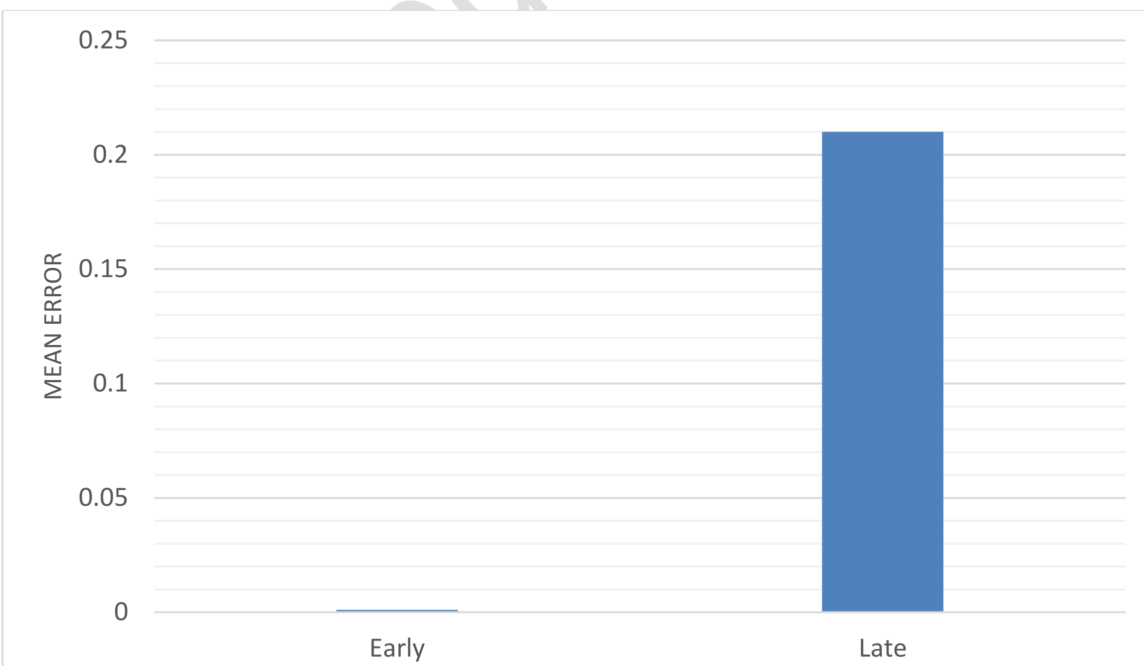

Figure 7. The errors for Early and Late after 3000 epochs

3.2.6. Simulation 6: Relation between AoA and brain damage

In this simulation the relation between brain damage and AoA (e. g., Ellis et al 1996) was simulated. The trained model in simulation 1 was lesioned at 3 levels of severity (5%, 10% and 20%). Two methods of lesioning were used (both have the same effects): setting weights to 0, and adding small number of noises to hidden neurons. In each one of the twenty samples, a random portion of neurons in the hidden layer was lesioned at each of the 3 levels of severity.

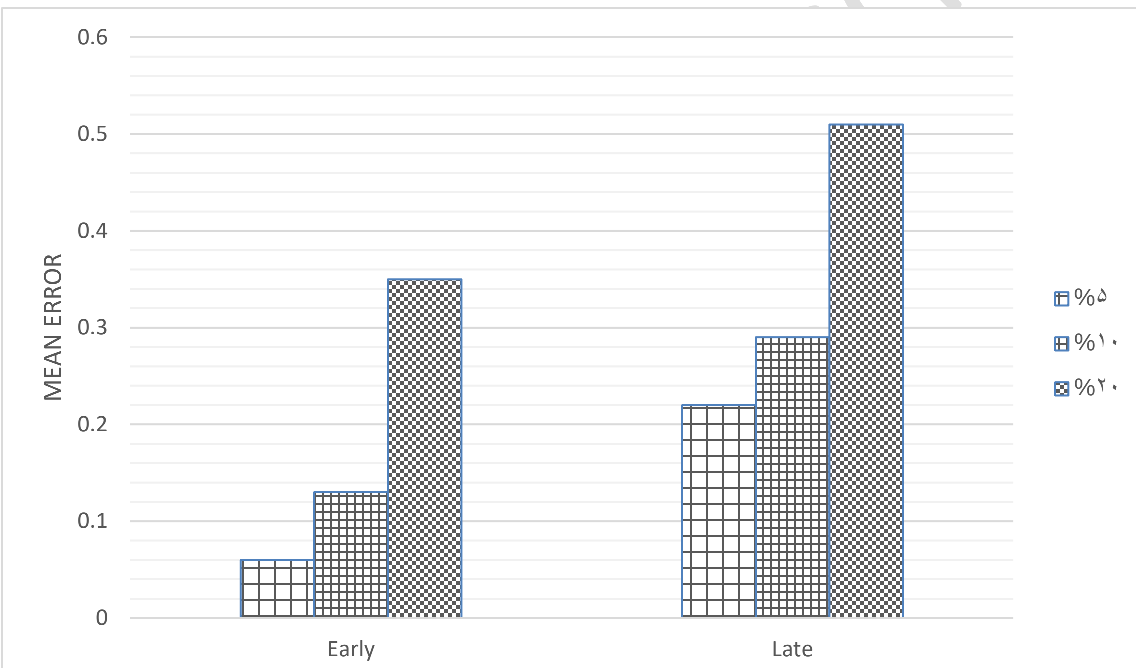

Figure 8. The effect of lesion on AoA at 3 levels of severity (5%, 10% and 20%).

The result shows that errors for late group were raised more than for early. And there was an increase in error as a function of severity of lesion. Thus as have been seen in the disease such as aphasia and dementia, the effect of damage in late experience is more than that of an early one (see Figure 8).

3.2.7. Simulation 7: The nature of AoA effect

As noted earlier, there is more change in start of learning by error back-propagation because of initial random weights and using sigmoid activation function. Activation is in its highest level when the inputs are around zero (thus at the beginning of learning and in the case of early items) and there will be little space to change as learning progresses (see Figure 9).

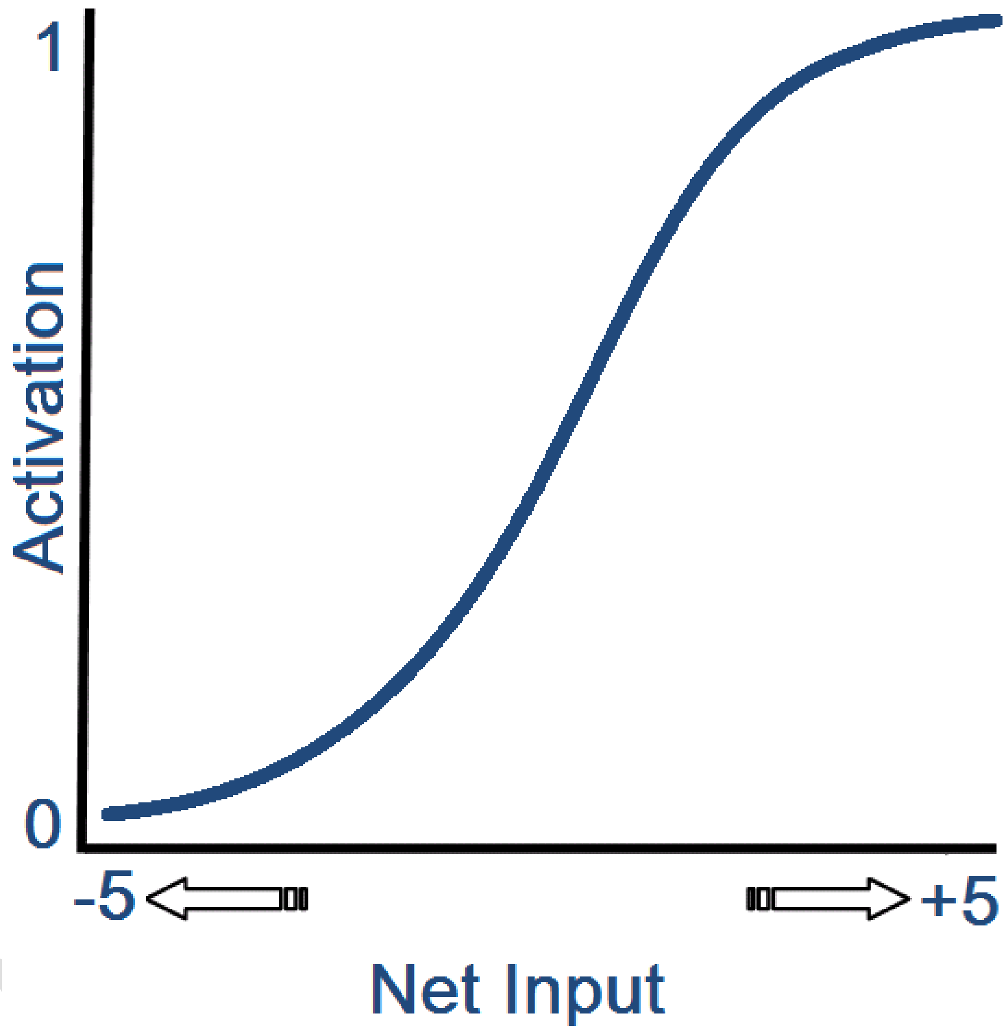

Figure 9. The logistic activation function with its slope is greatest when its input is around 0 as in start of training while decreases in both positive and negative directions as training proceeds.

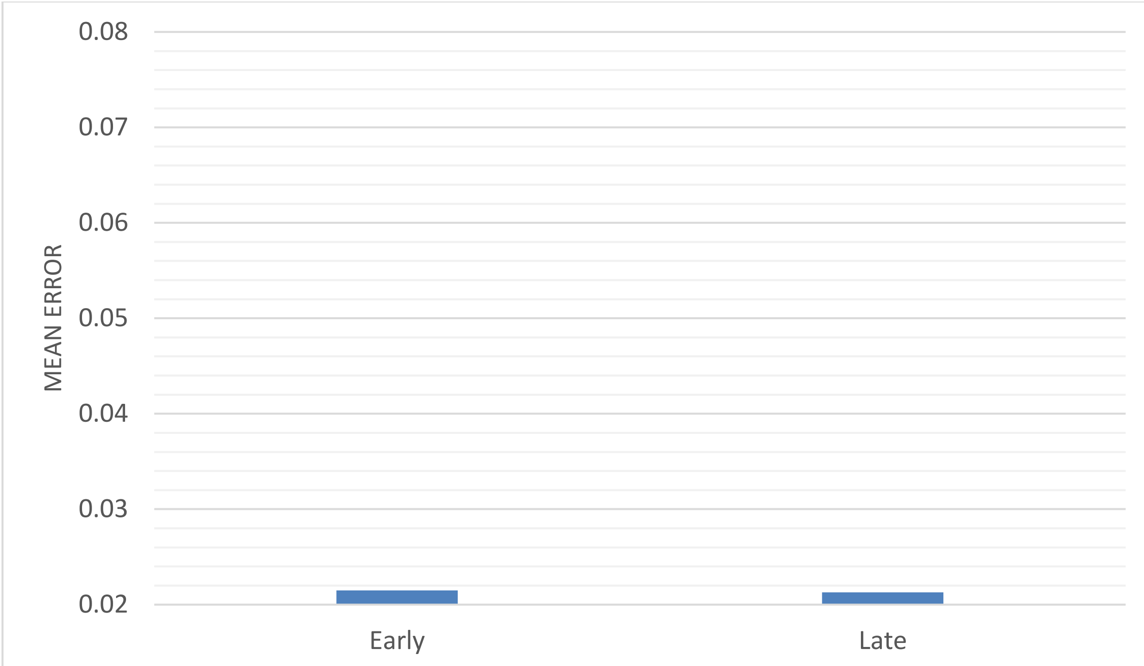

Figure 10. The effect of a modified algorithm, Quick prop, on AoA

This has been mentioned by Ellis and Rambon-Ralph (2000) but hasn't been practically shown. Here with another algorithm called Quick prop (Fahlman, 1987) the simulation 1 was replicated. Because this algorithm prevents activation function moving to its extreme value, it causes flexibility in network and eliminates AoA (see Figure 10). But Ellis and Lambon-Ralph have shown AoA even by using this algorithm and it is, presumably, because of using random patterns instead of real words.

4. Discussion

In the first part of this research, the effect of AoA in word reading, controlling other factors, was shown in 5th grade students. They read words initially learned in first grades better than ones learned in last grades. The consistency covariate was significant that is consistent with the assumption of Zevin and Seidenberg (2004). Thus AoA is mainly due to inconsistent words. Age of acquisition effects results from a loss of plasticity related to successful mastering in a task. This phenomenon can occur in several types of learning tasks including reading. It happens due to an interaction between AoA and consistency that was shown by covariance analysis in the current study.
When a new word is learned after an old one, it can catastrophically affect the learning of the first one if the first one is not being presented even for refreshing its learning. But if the first word is being presented alongside the second one, it keeps its superiority even after a long time.
As an explanation of AoA by connectionist models, it is worth mentioning that in these types of models, initially random values are assigned to the weights and output units take values of 1 or 0. The weight adjustments that occur during back-propagation with a logistic activation function are proportional to the unit activation (see also; Sohrabi, 2002). Thus,

these adjustments mostly occur with the activations in the middle of the logistic function (inputs are around 0), as happens when small, random weights are used for early initialization of the network. Therefore, the plasticity involved in learning the early-trained patterns are lost. This effect resembles inflexiblity of human brain to new learning such as infrequent and new words, despite its great flexibility during this age following neurofeedback (Rahmati et al., 2014) and education (Kranfick, et al., 2011).
When the model was lesioned by removing some hidden units to see its effect on word reading, words learned earlier were less vulnerable to damage compared to ones learned later. These effects can be explained by considering the nature of learning in neural networks trained by error back-propagation as mentioned above, i.e., early learning uses more resources and leave less for later learning, making it prone to be destroyed by damages.
Moreover, there was a failure in replicating previous simulation of AoA in English reading by an algorithm called Quick prop (Fahlman, 1987) here for Farsi, presumably due to the flexibility in the network that affects AoA. This happened in simulating AoA in Farsi, perhaps it is more regular than English. But Ellis and Lambon-Ralph have shown AoA even by using this algorithm and it is, presumably, because of using random patterns instead of real words. The problem with using random patterns also has been shown by Zevin and Seidenberg (2004). So, the relation between input and output are arbitrary instead of being quasi-regular as in the case of word naming. This gets even more important knowing that the AoA effect mainly occurs in irregular domains, as in less predictable Chinese characters (Chen, Zhou, Dunlap & Perfetti, 2007), less practiced reading aloud (Zevin & Seidenberg, 2004), and less regular stress in Italian word reading (Wilson, Ellis, & Burani, 2012).

Acknowledgements

I appreciate great helps from Reza Kormi-Noori during early stages of this work, and Shahin Fakhraei for her help on running the experiment. Also, I'd like to thank the principal, teachers, and students at Vahdat Elementary School, Saqqiz, Kurdistan, Iran, for their excellent helps in this study.

References

Chen, B. G., Zhou, H. X., Dunlap, S., & Perfetti, C. A. (2007). Age ofacquisition effects in reading Chinese: Evidence in favor of the arbitrary mapping hypothesis. *British Journal of Psychology, 98, 499–506.*

Coltheart, M., Davelaar, E., Jonasson, K., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), Attention & Performance VI (p. 535-555). Hillsdale, NJ: Erlbaum.

Coltheart, V., Laxon, V.J., & Keating, C. (1988). Effects of word imageability and age of acquisition on children's reading. *British Journal of Psychology, 79, 1-12.*

Daneshparvar, Z. and Daliri, M. R. (2012). A Simulation-Based Study of Dorsal Cochlear Nucleus Pyramidal Cell Firing Patterns. *Basic and Clinical Neuroscience, 3, 2, 22-31.*

Ellis, A. W., & Lambon Ralph, M. A. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26, 1103-1123.*

Ellis, A., & Morrison, C. (1998). Real age-of-acquisition effects in lexical retrieval *Journal of Experimental Psychology: Learning, Memory and Cognition, 24, 515-523.*

Fahlman, S. (1989). Fast learning variations on back-propagation: An empirical study. In M.C. Mozer, P. Smolensky,D. Touretsky, J.L. Elman, & A. Weigand (Eds), Connectionist Models: Proceedings of the 1988 Summer School.San Mateo: Morgan Kaufman.

Friston, K., & Frith, C. (2015). Active inference, communication and hermeneutics. *Cortex, 68, 129-143.*

Gerhand, S., & Barry, C. (1998). Word frequency effects in oral reading are not merely age-of-acquisition effects in disguise. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24, 267-283.*

Izura, C. Pérez, M. Agallou, E. Wright, V. Marín, J. Stadthagen,González, H. & Ellis, A. (2011). Age/order of acquisition effects and the cumulative learning of foreign words: A word training study. *Journal of Memory and Language, 64, 1, 32-58.*

Krafnick, A. J., Flowers, D. L., Napoliello, E. M., and Eden, G. F. (2011). Gray matter volume changes following reading intervention in dyslexic children. *NeuroImage, 57, 3, 733–741*.

Ludvig, E. A., Sutton, R. S., and Kehoe, E. J. (2012). Evaluating the TD model of classical conditioning. *Learning and Behaviour, 40, 305–319*.

Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature, 385, 813-815.*

McCandliss, B. D., Posner, M. I., & Givon, T. (1997). Brain plasticity in learning visual words. *Cognitive Psychology, 33, 88-110.*

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21, 116-153.*

Rahmati, N., Rostami, R., Zali, M. R., Nowicki, S., and Zare, J. (2014). The Effectiveness of Neurofeedback on Enhancing Cognitive Process Involved in Entrepreneurship Abilities among Primary school Students in District No.3 Tehran. *Basic and Clinical Neuroscience, 5, 4, 277-284.*

Rumelhart, D.E., & Todd, P.M. (1993). Learning and connectionist representations. In D.E. Meyer & S. Kornblum (Eds.), Attention and Performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience. Cambridge, MA: MIT Press.

Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986). Learning internal representations by error propagation. In D.E. Rumelhart, J.L. McClelland, and the PDP Research Group (Eds.), Parallel distributed processing: Explorations in the microstructure of cognition, Volume 1: Foundations. Cambridge, MA: MIT Press.

Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior, 23, 383-404.*

Seidenberg, M.S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96, 523-568.*

Sohrabi, A. (1999). The effects of frequency and consistency factors on visual word recognition among normal and mental retarded primary school students: A developmental cognitive approach, Unpublished dissertation for M.A. in Psychology, University of Tehran.
Sohrabi, A. (2001).A distributed developmental connectionist approach to word naming in normal, dyslexic and mental-retarded readers. *Paper presented at 1st International Conference on Cognitive Science, Tehran, Iran.*
Sohrabi, A. (2002). Cognitive deficits in schizophrenia: A connectionist approach. *Poster presented at Twelfth Annual Meeting of Canadian Society for Brain, Behavior and Cognitive Science, Vancouver, BC, Canada.*
Sohrabi, A., West, R.L. (2009). Positive and negative congruency effects in masked priming: a neuro-computational model based on representation, attention, and conflict. *Brain Research, 1289, 124–132.*
Soltanzadeh, M. J. and Daliri, M. R. (2014). Evaluation of Phase Locking and Cross Correlation Methods for Estimating the Time Lag between Brain Sites: A Simulation Approach. *Basic and Clinical Neuroscience, 5, 3, 205-211.*
Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development, 7, 1, 49-63.*
Wilson, M.A., Ellis, A.W., & Burani, C. (2012). Age-of-acquisition affects word naming in Italian only when stress is irregular, *Acta Psychologia, 139, 3, pp. 417-424.*
Zevin, J., & Seidenberg, M. (2004). Age-of-acquisition effects in reading aloud: Tests of cumulative frequency and frequency trajectory. *Memory and Cognition, 32, 31–38.*
Zorzi, M., Houghton, G., and Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual process model. *Journal of experimentalPsychology: Human Perception and Performance, 24, 1131-1161*.