7th International Conference on Computer and Knowledge Engineering (ICCKE 2017), October 26-27 2017, Ferdowsi University of Mashhad

Bilingualism advantage in handwritten character recognition:

A deep learning investigation on Persian and Latin scripts

Zahra Sadeghi
Department of ECE,
University of Tehran, Iran
Department of General Psychology,
University of Padova, Italy

Alberto Testolin
Department of General Psychology
and Padova Neuroscience Center,
University of Padova, Italy

Marco Zorzi
Department of General Psychology
and Padova Neuroscience Center,
University of Padova, Italy
IRCCS San Camillo Hospital
Foundation, Venice-Lido, Italy

***Abstract*— In this study, we investigated the effects of mastering multiple scripts in handwritten character recognition by means of computational simulations. In particular, we trained a set of deep neural networks on two different datasets of handwritten characters: the HODA dataset, which is a collection of images of handwritten Persian digits, and the MNIST dataset, which contains Latin handwritten digits. We simulated native language individuals (trained on a single dataset) as well as bilingual individuals (trained on both datasets), and compared their performance in a recognition task performed under different noisy conditions. Our results show the superior performance of bilingual networks in handwritten digit recognition in comparison to the monolingual networks, thereby suggesting that mastering multiple languages might facilitate knowledge transfer across similar domains.**

Keywords*— *bilingualism, handwritten character recognition, deep learning, hierarchical generative models, restricted Boltzmann machines, transfer learning

I. INTRODUCTION

Many empirical studies have uncovered the positive effects of knowing multiple languages in learning and cognition. For example, recent findings suggest that the volume of grey and white matter in frontal and temporal lobe regions of the cerebral cortex are greater in individuals with bilingual experience than monolinguals [1]. It has also been shown through a battery of experiments that learning two languages is advantageous for mental and linguistic abilities [2], and behavioral studies revealed that bilingualism greatly facilitates word learning [3].

The impact of bilingualism is also evident in visual word recognition. For example, Jared and Kroll (2001) investigated whether knowledge from one or two languages is activated in the process of naming words in English-French and French-English bilinguals, and found that spelling-to-sound correspondences from both languages can be activated simultaneously, even if one language is not relevant for the task. Other authors explored bilingual processing strategies to investigate the existence of a single, universal mechanism for parsing sentences in different languages, or of multiple subsystems that only partially overlap [5]. In this regard, the phonetic similarity of languages has shown to be a key parameter in the outcome of word learning [6]. Concretely, there are two different assumptions about word/character representation in brain. A number of researchers hold the hypothesis that word recognition in multilingual subjects is mediated by different areas of the cerebral cortex, each associated with processing of one specific language [7]. However, according to the opposite view which has been more prevalently accepted and more strongly supported by empirical evidence, information retrieval is non-selective and it is performed in an integrated or parallel form [8]–[10]. Ibrahimi and Eviatar suggested that, while only the left hemisphere is engaged in reading English, both left and right hemispheres are engaged in semitic language processing [11]. This result suggests a higher cognitive load in visual processing of semitic languages.

Another part of literature deals with the visual complexity of different scripts. For instance, the higher complexity of Arabic scripts is often related to the lower efficiency observed in letter recognition tasks [12] and orthographic processing [13]. However, it has been recently argued that biscriptal Arabic-English readers are more efficient in visual processing of complex letter shapes compared to monolinguals [14], thereby suggesting that learning multiple languages and scripts might facilitate knowledge transfer across perceptual domains.

In the present study we focused on Persian digits, which are very similar to Arabic numbers. The aim of our computational investigation is to provide a principled approach to study some implications of bilingualism within a connectionist framework. Given the astonishing complexity of human languages, it would be prohibitive to perform a comprehensive simulation spanning many different processing and representation levels, which would become even more prohibitive if we consider the presence of interacting languages within the same system. However, from a theoretical perspective even the study of simple character recognition might provide important insights. Indeed, written symbols often have a complex spatial structure,

978-1-5386-0804-3/17/$31.00 ©2017 European Union

which seems to be efficiently processed and represented using multiple levels of abstraction [15][16], where lower-levels visual features such as edges and grating (e.g., [17][18]) are successively combined in order to produce more abstract and invariant features, such as letter shapes [19]. On the one hand, it might be reasonable to assume that in bilingual individuals abstract symbols belonging to different languages are processed and represented in distinct regions of the visual cortex, which might facilitate word recognition since the constituting characters usually belong to a single script. However, at the same time it seems plausible that many neural resources are shared across different scripts/languages, which might have led to a better efficiency of the system thanks to the reuse of shared features. Indeed, the visual system might be highly optimized in order to exploit the common structure underlying most of the visual information, and the cultural evolution of writing systems might have selected written symbols that can be more easily represented by recycling preexisting visual features [20].

Bilingualism has been previously modeled by focusing on semantic and lexical aspects of language acquisition. Li and Farkas simulated Chinese-English bilingual processing using a computational model called DISLEX [21] which was based on a self-organizing map (SOM) trained with a Hebbian rule [22]. Sue and colleagues addressed the semantic similarities for creating a phrase-based translation system by developing a recursive neural network [23]. In the present paper, we investigated bilingual advantages in recognition of handwritten individual digits belonging to two different alphabets, Persian and Latin. Our computational approach is based on *deep learning*, but to differ from the most commonly used feed-forward architectures [24] we exploited stochastic, recurrent networks as building blocks for the deep architecture, which allow to learn hierarchical generative models in a completely unsupervised fashion [25], [26]. After learning a hierarchy of increasingly more complex visual features, each network can be tested on a supervised classification task by applying a simple, linear read-out module at the deepest level of representation. We trained separate deep networks on two different datasets, one composed by Persian handwritten digits [27] and the other composed by Latin handwritten digits [28]. Moreover, "bilingual" networks were obtained by jointly training on both datasets. After unsupervised learning, we first plotted the receptive fields of the hidden neurons belonging to different levels of the hierarchy, in order to analyze what type of visual features were extracted from different alphabets. We then asked each network to perform a recognition task, which was made more difficult by adding an increasing level of white Gaussian noise to the input patterns. The average classification accuracy was then compared, in order to assess the potential advantage of encoding the statistical distribution of both alphabets compared to the case where the generative model was learned on examples coming from a single dataset.

The paper is organized as follows. In Section 2, we present the different datasets used, the details related to the model's architectures and learning parameters, and the procedure adopted in our simulations. In Section 3, we present the results and discuss them. Section 4 concludes the paper with a general discussion and some possible future research directions.

II. METHOD

In the first phase of the study, we created monolingual networks which were trained on a single alphabet (in our case, either Persian or Latin). We then simulated bilingual networks, which were trained on digits selected from both scripts. In the second phase of the study, we measured the accuracy of the deep networks in the task of recognizing digits from both the alphabets, thereby comparing monolingual and bilingual abilities.

A. Datasets

Our experiments are carried out on HODA and MNIST digit datasets, which are both known as standard databases of handwritten digits and are developed for research purposes.

1) HODA datasest

The HODA dataset includes Persian handwritten digits which have been collected and made freely available [27]. Persian characters which are shown in TABLE I. are very similar in shape and form to Arabic numerals. HODA dataset consists of 60000 training images and 20000 test images The original images are in binary format and variable in size. Hence, for our simulation purposes all images were first rescaled to fit a 20x20 pixel box, which produced grey-valued patterns (i.e., real-valued). The obtained images were then padded by adding zero-value pixels in such a way to be fit in a 32x32 box with a black background.

1) MNIST dataset

The MNIST dataset contains 60000 training images and 10000 test images of handwriting Latin digits (TABLE II). The original images are gray-valued and centered to fit 28x28 pixel boxes [28], and were resized to 32x32 while preserving their aspect ratio in order to make it comparable with HODA dataset images of digits.

B. Network architecure and learning details

We used deep belief networks [29] composed by a stack of restricted Boltzmann machines [30]. The network architecture is represented in Fig. 1, and includes three hidden layers with, respectively, 500 units (first layer), 500 units (second layer), and 2000 units (third layer). This architecture is the same used in previous studies on digit recognition, and has shown to be robust to small variations in the size of the hidden layers [29]. Each RBM is trained in a completely unsupervised manner using the contrastive divergence algorithm [30]. The momentum coefficient is set to 0.5 for the first 5 epochs and then increased to 0.9. We exploited an efficient implementation on graphic processing units [31], which is based on a mini-batch learning scheme. Each mini-batch contained 100 patterns.

In principle, the RBMs constituting the building blocks of our deep network could be replaced by deterministic autoencoders, although it should be noted that these neural network architectures often give rise to different types of internal encoding [32].

978-1-5386-0804-3/17/$31.00 ©2017 European Union

TABLE I. CHARACTERISTICS OF THE HODA DATASET

Label	Spell	Character	#Train	#Test	Total Count
1	Sefr	۰	6000	2000	8000
2	Yek	۱	6000	2000	8000
3	Do	۲	6000	2000	8000
4	Se	۳	6000	2000	8000
5	Chahar	۴	6000	2000	8000
6	Panj	۵	6000	2000	8000
7	Shesh	۶	6000	2000	8000
8	Haft	۷	6000	2000	8000
9	Hasht	۸	6000	2000	8000
10	no	۹	6000	2000	8000
Total count			60000	20000	80000

TABLE II. CHARACTERISTICS OF THE MNIST DATASET

Label	Character	#Train	#Test	Total Count
1	0	6000	1000	7000
2	1	6000	1000	7000
3	2	6000	1000	7000
4	3	6000	1000	7000
5	4	6000	1000	7000
6	5	6000	1000	7000
7	6	6000	1000	7000
8	7	6000	1000	7000
9	8	6000	1000	7000
10	9	6000	1000	7000
Total count		60000	10000	70000

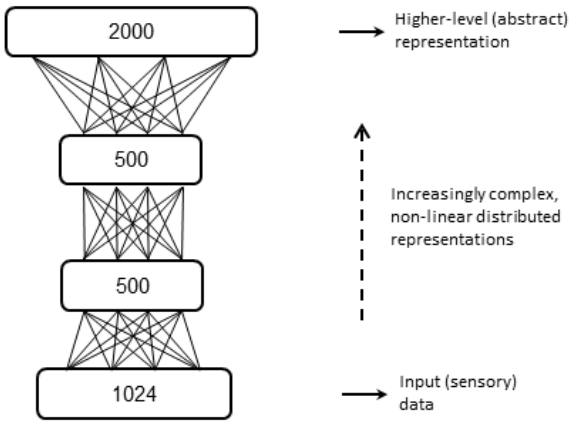


Fig. 1. Deep network architecture used in our simulations.

In order to evaluate the performance of the networks after unsupervised generative learning, we applied a linear classification module to the deepest (i.e., third) hidden layer. The supervised classification was implemented using a simple delta rule algorithm, with a learning rate of 0.0001. The accuracy of classification was obtained by computing the average rate of correct responses over all the test patterns.

C. Simulation procedure

We first evaluated the performance of each network trained separately on a single alphabet in presence of thirteen levels of visual noise. The noise levels are obtained by adding zero mean Gaussian noise with different levels of standard deviations from 0.01 to 0.4. Examples of noisy digits in each alphabet are illustrated in Fig. 2 and Fig. 3.

III. SIMULATIONS AND RESULTS

To assess the significances of the results, we fitted a logistic function to each of the ten curves obtained from ten trained networks. Fig. 4 illustrates the average accuracy of classification of digits of each dataset in the presence of noise. As can be noted from the graphs, the increase of noise level attenuated the classification performance, but it has similar effects on both networks which are separately trained on the two alphabets.

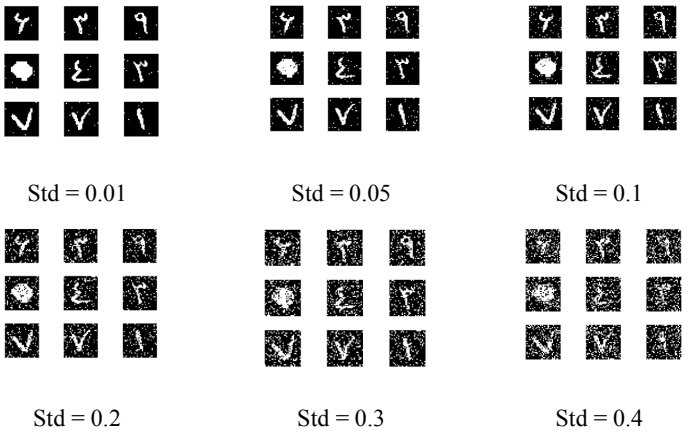


Fig. 2. Noisy digits sampled from HODA dataset (Std indicates the standard deviation of the Gaussian noise added).

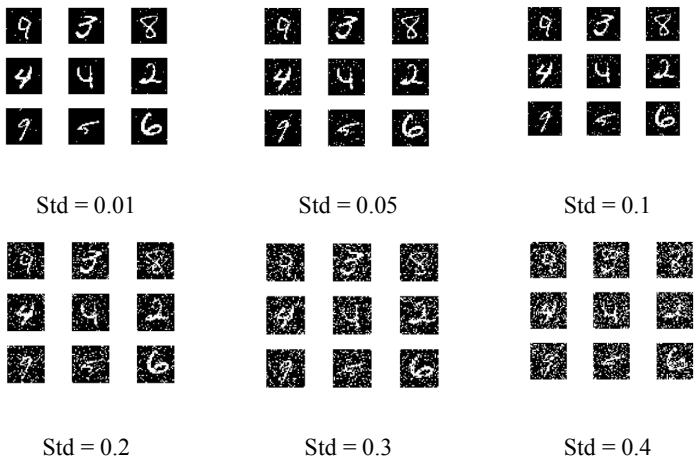


Fig. 3. Noisy digits sampled from MNIST dataset (Std indicates the standard deviation of the Gaussian noise added to the stimuli).

978-1-5386-0804-3/17/$31.00 ©2017 European Union

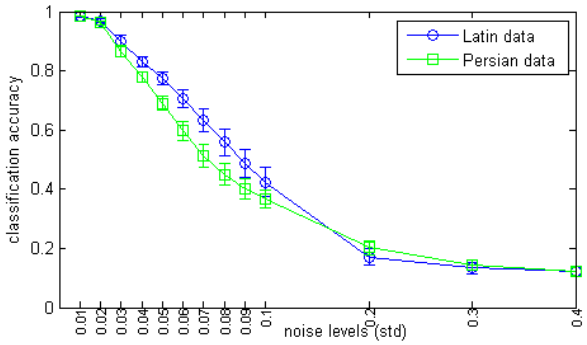
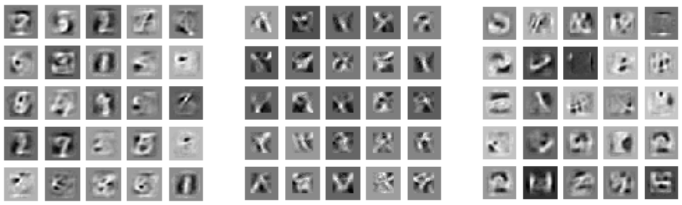


Fig. 4. Accuracy of monolingual networks in presence of different levels of noise.

Next, we simulated bilingual individuals and compared their performance with monolingual networks. Bilingual knowledge was constructed by simultaneously using the training sets of both scripts. This trained network was then employed to extract features for the test data from each alphabet separately. Examples of receptive fields of the third layer emerged after the training phase of each network are visualized in Fig. 5. It can be observed that, while the receptive fields of the output layer closely resemble the digit inputs in the monolingual cases, the receptive fields corresponding to the bilingual individuals contain more complicated patterns that seem to be constructed as a result of training on a combination of the digit sets from both alphabets. The learned features found in the receptive fields advocates the non-selective alphabets activation and puts forward the lexical bilingual pattern integration.

While no noticeable difference was found in comparing the performance of bilingual network on recognition of digits from each language dataset (Fig. 6), we observed a superior performance of the bilingual network over both the Persian monolingual (Fig. 7) and the Latin monolingual (Fig. 8) networks, across the whole range of noise levels. This was statistically confirmed by two sample t-tests on the slopes of the curves fitted to the data points. The corresponding t-test results are shown below each figure. A p-value smaller than 0.05 indicates that there is a significant difference between the two curves.



Latin monolingual network | Persian monolingual network | Bilingual network

Fig. 5. Visualization of a sample of receptive fields corresponding to the neurons in the third hidden layer.

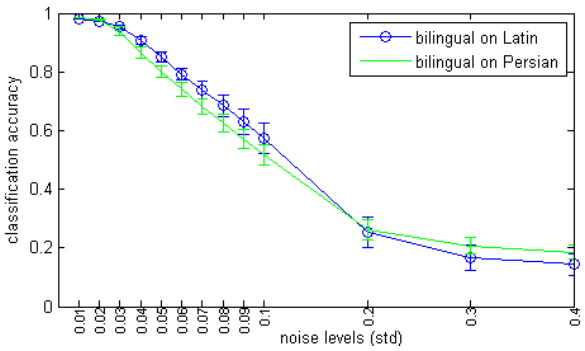
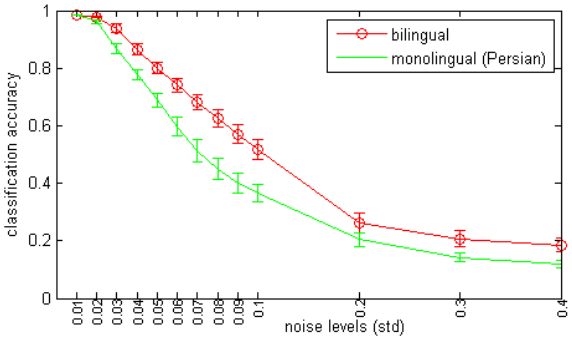
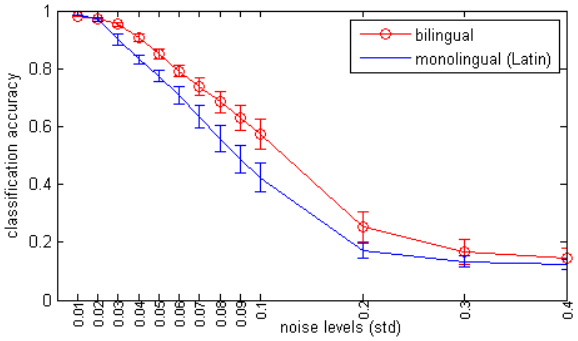


Fig. 6. Comparing performance of bilingual networks in each of the two separate scripts.



t-statistics	p-value	95 % confidence interval
-4.98	9.6856e-005	[-12.01 -4.89]

Fig. 7. Comparing Persian monolingual and bilingual networks in handwritten digit recognition.



t-statistics	p-value	95 % confidence interval
-7.44	6.8115e-007	[-14.18 -7.93]

Fig. 8. Comparing Latin monolingual and bilingual networks in handwritten digit recognition.

IV. General Discussion and Future Challenges

In this paper, we simulated visual symbol recognition in monolingual and bilingual individuals using deep neural networks. In all of our experiments, we used a three-layer deep belief network composed by a stack of restricted Boltzmann

978-1-5386-0804-3/17/$31.00 ©2017 European Union

machines. Bilingual networks were created by training the network on samples drawn from both alphabets. Performance of each network was examined by applying a linear readout on the deepest layer, which was implemented using the delta rule and whose aim was to classify the corresponding handwritten digits. We found that bilingual networks, which were equally exposed to different writing scripts, had a higher performance in digit recognition compared to the monolingual networks. Our results are in agreement with studies that support the positive impact and benefits of learning more than one language. In particular, we showed that learning the statistical features of two different alphabets increases the ability of recognition of handwritten digits even under noisy conditions.

One exciting avenue for future research would be to more systematically compare simulation results with empirical data collected on human observers. For example, the confusion errors of the model can be correlated with empirical confusion matrices, which are available both for Latin [33] and Arabic alphabets [14]. Moreover, it would be informative to more carefully investigate how learning over multiple datasets could facilitate knowledge transfer across similar perceptual domains [34], [35], for example by producing more abstract, high-level features that can be readily applied in different contexts.

V. REFERENCES

[1] R. K. Olsen *et al.*, "The effect of lifelong bilingualism on regional grey and white matter volume," *Brain Res.*, vol. 1612, pp. 128–139, 2015.

[2] S. Ben-Zeev, "The influence of bilingualism on cognitive strategy and cognitive development," *Child Dev.*, pp. 1009–1018, 1977.

[3] M. Kaushanskaya and V. Marian, "The bilingual advantage in novel word learning," *Psychon. Bull. Rev.*, vol. 16, n. 4, pp. 705–10, 2009.

[4] D. Jared and J. F. Kroll, "Do bilinguals activate phonological representations in one or both of their languages when naming words?," *J. Mem. Lang.*, vol. 44, no. 1, pp. 2–31, 2001.

[5] P. E. Dussias, "Sentence Parsing in Fluent Spanish-English Bilinguals.," 2001.

[6] M. Havy, C. Bouchon, and T. Nazzi, "Phonetic processing when learning words The case of bilingual infants," *Int. J. Behav. Dev.*, p. 165025415570646, 2015.

[7] L. D. Gerard and D. L. Scarborough, "Language-specific lexical access of homographs by bilinguals.," *J. Exp. Psychol. Learn. Mem. Cogn.*, vol. 15, no. 2, p. 305, 1989.

[8] C. Beauvillain and J. Grainger, "Accessing interlexical homographs: Some limitations of a language-selective access," *J. Mem. Lang.*, vol. 26, no. 6, pp. 658–672, 1987.

[9] A. M. B. De Groot, P. Delmaar, and S. J. Lupker, "The processing of interlexical homographs in translation recognition and lexical decision: Support for non-selective access to bilingual memory," *Q. J. Exp. Psychol. A.*, vol. 53, no. 2, pp. 397–428, 2000.

[10] R. Bijeljac-Babic, A. Biardeau, and J. Grainger, "Masked orthographic priming in bilingual word recognition," *Mem. Cognit.*, vol. 25, no. 4, pp. 447–457, 1997.

[11] R. Ibrahim and Z. Eviatar, "Language status and hemispheric involvement in reading: Evidence from trilingual Arabic speakers tested in Arabic, Hebrew, and English.," *Neuropsychology*, vol. 23, no. 2, p. 240, 2009.

[12] D. G. Pelli, C. W. Burns, B. Farell, and D. C. Moore, "Feature detection and letter identification.," *Vision Res.*, vol. 46, no. 28, pp. 4646–74, Dec. 2006.

[13] S. Tahan, T. Cline, and S. Messaoud-Galusi, "The relationship between language dominance and pre-reading skills in young bilingual children in Egypt," *Read. Writ.*, vol. 24, no. 9, pp. 1061–1087, 2011.

[14] R. W. Wiley, C. Wilson, and B. Rapp, "The Effects of Alphabet and Expertise on Letter Perception.," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 42, no. 8, pp. 1186–203, 2016.

[15] J. Grainger, A. Rey, and S. Dufau, "Letter perception: from pixels to pandemonium.," *Trends Cogn. Sci.*, vol. 12, n. 10, pp. 381–7, 2008.

[16] Z. Sadeghi, "Deep Learning and Developmental Learning: Emergence of Fine-to-Coarse Conceptual Categories at Layers of Deep Belief Network," *Perception*, vol. 45, no. 9, pp. 1036–1045, 2016.

[17] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol.*, vol. 160, no. 1, pp. 106–154, 1962.

[18] Z. Sadeghi, B. Nadjar Araabi, and M. Nili Ahmadabadi, "A Computational Approach towards Visual Object Recognition at Taxonomic Levels of Concepts," *Comput. Intell. Neurosci.*, 2015.

[19] S. Dehaene, L. Cohen, M. Sigman, and F. Vinckier, "The neural code for written words: a proposal.," *Trends Cogn. Sci.*, vol. 9, no. 7, pp. 335–41, Jul. 2005.

[20] S. Dehaene and L. Cohen, "Cultural recycling of cortical maps," *Neuron*, vol. 56, no. 2, pp. 384–398, Oct. 2007.

[21] R. Miikkulainen and S. Kiran, "Modeling the bilingual lexicon of an individual subject," *Adv. Self-Organizing Maps*, pp. 191–199, 2009.

[22] P. Li and I. Farkas, "3 A self-organizing connectionist model of bilingual processing," *Adv. Psychol.*, vol. 134, pp. 59–85, 2002.

[23] J. Su, B. Zhang, D. Xiong, R. Li, and J. Yin, "Convolution-Enhanced Bilingual Recursive Neural Network for Bilingual Semantic Modeling.," in *COLING*, 2016, pp. 3071–3081.

[24] Y. LeCun, Y. Bengio, and G. E. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[25] M. Zorzi, A. Testolin, and I. Stoianov, "Modeling language and cognition with deep unsupervised learning: a tutorial overview," *Front. Psychol.*, vol. 4, no. August, p. 515, 2013.

[26] A. Testolin and M. Zorzi, "Probabilistic Models and Generative Neural Networks: Towards an Unified Framework for Modeling Normal and Impaired Neurocognitive Functions," *Front. Comput. Neurosci.*, vol. 10, no. 73, Jul. 2016.

[27] H. Khosravi and E. Kabir, "Introducing a very large dataset of handwritten Farsi digits and a study on their varieties," *Pattern Recognit. Lett.*, vol. 28, no. 10, pp. 1133–1141, Jul. 2007.

[28] Y. LeCun and C. Cortes, "MNIST Optical Character Database at AT&T Research," *http://yann.lecun.com/exdb/mnist*, 1998.

[29] G. E. Hinton and R. Salakhutdinov, "Reducing the dimensionality

978-1-5386-0804-3/17/$31.00 ©2017 European Union

of data with neural networks.," *Science (80-.).*, vol. 313, no. 5786, pp. 504–7, Jul. 2006.

[30] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Comput.*, vol. 14, no. 8, pp. 1771–1800, 2002.

[31] A. Testolin, I. Stoianov, M. De Filippo De Grazia, and M. Zorzi, "Deep unsupervised learning on a desktop PC : A primer for cognitive scientists," *Front. Psychol.*, vol. 4, no. May, p. 251, 2013.

[32] A. Testolin, M. De Filippo De Grazia, and M. Zorzi, "The Role of Architectural and Learning Constraints in Neural Network Models: A Case Study on Visual Space Coding," *Front. Comput. Neurosci.*, vol. 11, no. March, pp. 1–17, 2017.

[33] I. C. Simpson, P. Mousikou, J. M. Montoya, and S. Defior, "A letter visual-similarity matrix for Latin-based alphabets," *Behav. Res. Methods*, pp. 431–439, 2012.

[34] Z. Sadeghi and A. Testolin, "Learning representation hierarchies by sharing visual features: A computational investigation of Persian character recognition with unsupervised deep learning," *Cogn. Process.*, vol. 14, pp. 1–12, 2017.

[35] A. Testolin, I. Stoianov, and M. Zorzi, "Letter perception emerges from unsupervised deep learning and recycling of natural image features.," *Nat. Hum. Behav.*, 2017.

978-1-5386-0804-3/17/$31.00 ©2017 European Union