Classifying the differences in gaze patterns of alphabetic and logographic L1 readers - A neural network approach

André Frank Krause[1 3], Kai Essig[1 3], Li-Ying Essig-Shih[2], and Thomas Schack[1 3]

[1] Faculty of Sport Science, Dept. Neurocognition & Action {andre_frank.krause, kai.essig, thomas.schack}@uni-bielefeld.de
[2] cultureblend IIT GmbH Universitätsstraße 25 essigshih@cultureblend.de
[3] Cognitive Interaction Technology, Center of Excellence
Bielefeld University, D-33615 Bielefeld, Germany

Abstract. Using plain, but large multi-layer perceptrons, temporal eye-tracking gaze patterns of alphabetic and logographic L1 readers were successfully classified. The Eye-tracking data was fed directly into the networks, with no need for pre-processing. Classification rates up to 92% were achieved using MLPs with 4 hidden units. By classifying the gaze patterns of interaction partners, artificial systems are able to act adaptively in a broad variety of application fields.

1 Introduction

With the technical progress and the minituarization of electronic devices, eye tracking became more and more popular in the recent years. The term eye tracking denotes the process of monitoring and recording the participants' gaze positions when they look at 2D or 3D stimuli. Researchers are interested in exact *gaze positions* and their temporal course, i.e., *spatial-temporal scan paths*. The analysis of eye movements yields valuable insights into the cognitive processes underlying information processing ("eyes are a window to the mind") [1] [2] - providing answers to questions like: " Which information are perceived as valuable?", "What is the temporal order of perceived objects?", "Where had the participants problems to understand important scene information (signified by a high number of fixations or long fixation durations)?". Eye tracking is of high interest in different research areas: Linguists are interested in the relations between the perception of spoken or written text and visual attention. Computer scientists are interested in the development of suitable human-computer interfaces for inutitive terminals or robotic systems.

Neural Networks have been developed with their main feature beeing their learning ability: They have the ability to learn a particular task autonomously from training data, whithout the need of explicit programming [3]. Neural nets have been applied to the field of eye tracking to reliably estimate the 3D gaze point from a participant's binocular eye-position data [4]. Neural nets have been used to analyze the diagnostic significance of dyslexis eye movements [5]. The authors measured eye movements of 52 school children (normal readers, mentally disabled readers and dyslexics) during

a reading task. Using a self-organizing map, the three groups formed individual clusters and could be clearly classified. [6] validate if the way a person reads influence the way they understand information and propose a novel method of detecting the level of engagement in reading based on a person's gaze-pattern. They organized some experiments in reading tasks of over thirty participants and the experimental outputs are classified with Artificial Neural Networks with approximately 80% accuracy. All these approaches have in common that the gaze data needs a high amount of pre-processing before it can serve as an input for the neural networks. Recent experiments revealed, that deep, big feedforward networks trained with standard online backpropagation can achive excellent classification results without any preprocessing of the input data [7] - making them a suitable tool for the flexible analysis of raw gaze-movement data. Here, we feed eye-tracking data, recoded in a reading experiment with German and Chinese natives, without any pre-processing into a feedforward network to classify the scan path as belonging to a Chinese or German reader. Eye-tracking combined with neural networks enable artificial systems to react adaptively in a broad variety of application fields.

2 Eye Movements in Reading

While reading, the visual information is projected on different areas of the retina, i.e., the foveal, parafoveal, and peripheral area. The fovea is the area of highest visual resolution in the retina. Because of the limited size of the fovea (2° of the visual field), only a small text area (around 8 letters) can be sharply analyzed during each fixation [8]. Figure 1 shows the perceptual span for the foveal and parafoveal areas during reading. Although humans are able to dissociate their attention from the foveal direction of gaze (Duchowsky, 2003), complex information processing tasks, like reading, require a strong coupling of gaze movements and attention [9]. During a fixation the eye remains directed at the same spot for a longer period (while reading: 200-400 ms), enabling visual information processing of the fixated region. Approximately 90% of the viewing time is spent on fixations [10].

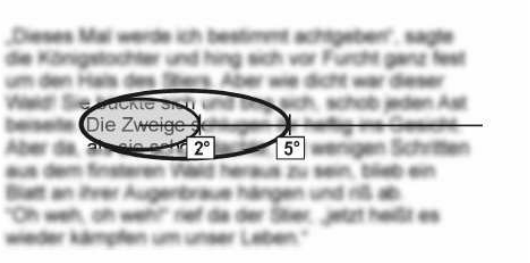

Fig. 1. The perceptual span for the foveal and parafoveal area during text reading.

Because only the fovea allows clear vision, the eye has to be moved so that different parts of the text can be processed in sequence. These movements are called *saccades*. Saccades are jerky eye movements from one fixation to another. During saccades the

pupil is accelerated up to 1000°/s, and an angular velocity of more than 300° per second is reached [11]. Saccade velocity and durations vary with the writing system, the individual fluency of the reader and the diffculty of the text.

Eye movement research with participants of different native languages revealed, that L1-Readers with an alphabetic writing system (e.g., English and German) have a slightly bigger perceptual span than those of a syllabographical and logographical writing system, such as Japanese or Chinese readers [12][13]. Furthermore, experienced readers have a larger perceptual span than beginners, i.e., the perceptual span can be widened by exercise [14]. Additionally, the perceptual span is larger when reading a simple compared to a complex text [15][13].

3 Data Collection

The eye movements of thirteen german native speakers (5 females and 8 males) and thirteen Chinese native speakers (9 females, 4 males), with an age between 26 and 37 years, were recorded while they read silently German text passages. The Chinese native speakers were required to have passed the DSH (German abbreviation for Deutsche Sprachprüfung für den Hochschulzugang), a language exam required to study at German higher education institutions and be at least enrolled for one year at a German high school. For the experiment, two literature text passages were used. Text A, is an excerpt from "'Der Richter und sein Henker"', Friedrich Dürrenmatt with 195 words (1027 letters). Text B is an excerpt from the novel "'Fabian. Die Geschichte eines Moralisten"', Erich Kästner with 201 words (1084 letters). The text was splitted in 4 sections and then presented in Times New Roman 16pt, double spaced, black colour on a gray background on a computer screen with 1024x768 pixel resolution (see Figure 2). For the experiment, an SMI Eyelink I eye tracker was used to record participants' eye movements. The collected eye-tracking data consists of (x,y) coordinates of all recorded fixations as well as the fixation durations in ms. The maximum length of scanpaths was 424 frames.

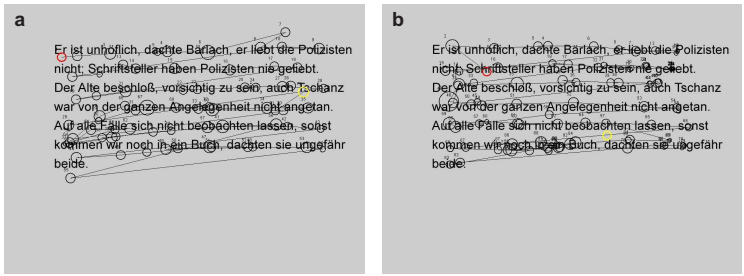

Fig. 2. Example scanpaths of a (a) chinese and (b) german reader.

4 Multilayer Perceptron Classification

To train a multilayer perceptron (MLP), the Matlab©2010b Neural Network Toolbox was used. The network had 424 input and two output neurons with a sigmoidal activation function (tanh). Each output neuron represents one class, with class 1 (logographic) coded as $\begin{pmatrix} 1 & 0 \end{pmatrix}$ and class 2 (logographic) as $\begin{pmatrix} 0 & 1 \end{pmatrix}$. The training data fed into the net had three data channels: the x- and the y-component of the scan path and the fixation duration (f). Before training, the temporal input data was scaled into the range $[-1, 1]$ and spatially unfolded into the input vector $\mathbf{u} = (x_0, y_0, f_0,\ x_1, y_1, f_1\ ....\ x_n, y_n, f_n)$. The default training algorithm for pattern classification used by the toolbox is Levenberg-Marquardt backpropagation. All toolbox options were kept at their default values, except the stopping criterion "'failed evaluation runs"'. The maximum allowed number of failed evaluation runs was increased to 50 (default = 7). This increased the classification performance slightly without impairing generalization. N-fold cross validation (N=10) was applied to estimate the classification performance.

Fig. 3a shows the performance of a single layer perceptron (N=100 training repetitions with different random network initialisations) compared to MLPs with hidden layers sizes from 1 to 10. The figure shows, that already a single layer perceptron can classify the data (median value 83%). The best network found had a classification performance of 88%. Adding a hidden layer improves performance (Fig. 3b). MLPs with 4 hidden units show a best performance of 92% and a median performance of 84%.

Adding more hidden layers did not improve the performance. 100 MLPs with two hidden layers and hidden layer sizes ranging from 1 to 15 were trained (N=100 repetitions of the training process with random initial network weights). Best mean value found was 84.3% and the best overall network had 92% classification performance. Fig. 4 shows mean and maximal performance for MLPs with two hidden layers.

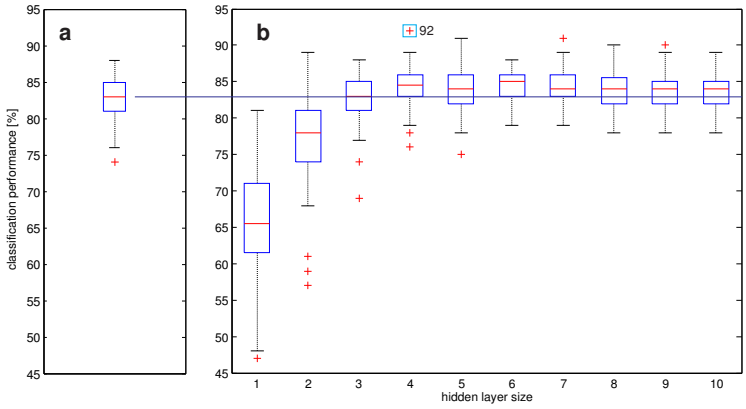

Fig. 3. Classification performance depending on the number of neurons in the hidden layer. Boxplots show n=100 repetitions of 10-fold crossvalidation runs over the training data. a) Performance of a single layer perceptron for comparison. b) Performance of perceptrons with one hidden layer.

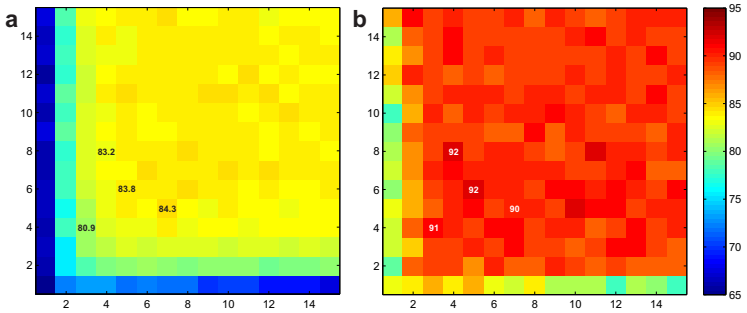

Fig. 4. a) Mean and b) maximum classification performance (colour coded) of MLPs with two hidden layers. N=100 training repetitions were performed. No combination of hidden layer sizes (1 to 15 units) resulted in an improvement over the single hidden layer MLPs.

4.1 Contribution of Data Components

To analyse the contribution of the several components of the eye tracking data, an MLP with four hidden units was trained (N=100 repetitions) seperately for each data channel. Fig. 5 shows that training the networks given each channel alone did not reach the performance of the networks that were fed with all three data channels. The figure also indicates that the vertical component of the scanpath contributes the most to classification success.

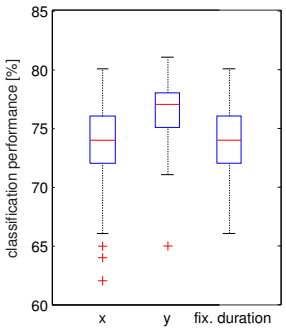

Fig. 5. Contribution of the eye-tracking data components. A MLP with 4 units in the hidden layer was used, N=100 training repetitions. The input to the network was either the x-, the y-component of the scan path or the fixation duration, only.

5 Conclusions

Given raw, non-preprocessed eye-tracking data, a simple feedforward network could be trained to achieve a high classification rate on an experimental data set regarding the reading performance of German versus Chinese L1-readers. The temporal eye-tracking data was spatially unfolded into a high number of network inputs.

A possible application scenario can be information terminals that are able to detect persons having difficulties reading the presented on-screen language and automatically offer a language selection menu. Also, a pre-evaluation of reading performance of aspirants for specialised professions (translators) are imaginable.

The process of recording data in empirical studies and test it in an adequate simulation will be important for further understanding of interaction scenarios. The question is, if the data which was measured in the real world, is also reflected in simulated behavior. One example would be humanoid robots. Will a robot perform like a grown up or more like a child, when fed with recorded data of adults or childrens? The automatic classification of the gaze behavior and human movement data will be of high importance to this field of research.

References

1. Essig, K.: Vision-Based Image Retrieval (VBIR) - A New Approach for Natural and Intuitive Image Retrieval. Saarbrücken: VDM Verlag. (2008)
2. Duchowski, A.: Eye Tracking Methodology: Theory and Practice. London, England: Springer (2003)
3. Haykin, S.: Neural Networks: A Comprehensive Foundation. Prentice Hall PTR Upper Saddle River, NJ, USA (1994)
4. Essig, K., Pomplun, M., Ritter, H.: A neural network for 3d gaze recording with binocular eye trackers. International Journal of Parallel, Emergent and Distributed Systems **21(2)** (2006) 79–95
5. Macaš, M., Lhotská, L., Novák, D.: Bio-inspired methods for analysis and classification of reading eye movements of dyslexic children. Technical report, University in Prague, Algarve, Portugal (October 3-5, 2005 2005)
6. Vo, T., Sumudu, U., Mendis, Gedeon, T.: Gaze pattern and reading comprehension. In: ICONIP'10 Proceedings of the 17th international conference on Neural information processing: models and applications. Volume Part II., Springer-Verlag, Heidelberg (2010)
7. Ciresan, D.C., Meier, U., Gambardella, L.M., Schmidhuber, J.: Deep big simple neural nets excel on handwritten digit recognition. CoRR (2010)
8. Just, M.A., Carpenter, P.A.: The psychology of reading and language comprehension. Boston: Allyn and Bacon (1987)
9. Essig-Shih, L.Y.: Effekte simultanen Hörens und Lesens auf das L2-Lesen: Eine Eyetracking-Studie im BereichDeutsch als Fremdsprache. PhD thesis, University of Bielefeld (2008)
10. Irwin, J.W.: Teaching reading comprehension process. Boston: Allyn & Bacon. (1991)
11. Rötting, M.: Parametersystematik der Augen- und Blickbewegungen für arbeitswissenschaftliche Untersuchungen. Aachen: Shaker. (2001)
12. Inhoff, A.W., Liu, W.: The perceptual span and oculomotor activity during the reading of chinese sentences. Journal of Experimental Psychology: Human Perception and Performance **24** (1998) 20–34
13. Rayner, K., Sereno, S.C.: Eye movements in reading. Psycholinguistic studies. In: Handbook of psycholinguistics. San Diego: Academic Press (1994) 57–81
14. Nerius, D., ed.: Deutsche Orthographie. 3. Aufl. Mannheim: Dudenverlag. (2000)
15. Rayner, K., Carroll, P.J.: Issues and problems in the use of eye movement data in study reading comprehension. In: New methods in reading comprehension research. Hillsdale: Erlbaum. (1984) 129–150