Neural Networks for Emotion Recognition Based on Eye Tracking Data

Claudio Aracena*, Sebastián Basterrech †, Václav Snášel†, and Juan Velásquez*
* Department of Industrial Engineering, University of Chile, Santiago, Chile
{caracena@ing.uchile.cl, jvelasqu@dii.uchile.cl}
† VŠB-Technical University of Ostrava, Ostrava, Czech Republic
{Sebastian.Basterrech.Tiscordio,Vaclav.Snasel}@vsb.cz

***Abstract*—We present an approach for emotion recognition using information of the pupil. In last years, the pupil variables have been used as an assessment of emotional arousal. In this article, we generate signals of pupil size and gaze position monitored during image viewing. The emotions are provoked by visual stimuli of colored images. Those images were taken from the International Affective Picture System which has been the reference for objective emotional assessment based on visual stimuli. For recognising the emotions we use the evolution of the eye tracking data during a window of time. The learning dataset is composed by the evolution of the pupil size and the gaze position, and labels associated to the emotional states. We study two kinds of learning tools based on Neural Networks. We obtain promising empirical results that show the potential of using temporal learning tools for emotion recognition.**

***Keywords*-Affective Computing; Computer-Human Interaction; Emotion Recognition; Neural Networks; Temporal Learning Problem**

I. Introduction

Emotion recognition is an interdisciplinary research area involving several fields that include Computer Sciences, Cognitive Science and Psychology. The study of emotional states is helpful for interpreting human actions, as well as for introducing human factors in artificial systems. Therefore, we can improve the relationship human-machine, for instance developing software that can, in some manner, understand human states and make actions according to them. Besides, the emotion recognition area is useful for generating knowledge in order to adapt in real-time devices producing appropriate responses according the subject emotions [1].

In the literature, we can find several types of input information for classifying emotional states, such as: face gesture, physiological responses, heart rate and brain signals.

We can get information about external expressions of emotions from face gesture, but these inputs are insufficient to get information from internal states. The pupil is strongly related with the nervous system, and it is highly sensitive to external stimuli. The pupil dilatation has been related with the cognitive processes, for example it has been associated with the emotionality and attention [2]. In addition, the behavior of the pupil presents well-identified limitations. The drawbacks related to the pupil measures are caused by its high sensitivity to the light, and the noise provoked by blinks and fast gaze movements.

In this article, we study the emotional states of a set of subjects using the behavior of their pupil size and gaze position along time. The goal of this article is generating a mapping between the pupil and gaze behavior and emotional states provoked by visual stimuli. As input information, we consider the pupil dilatation and the gaze position that are measured following standard protocols. As measure of emotional states we use an objective assessment developed by the *Center for the Study of Emotion and Attention* under the name of *International Affective Picture System (IAPS)* [3].

The pupil variables and the emotional states evolve in time, this means that the current pupil behaviour, gaze position and emotional state are affected by previous situations. As a consequence for generating a mapping between them we are using temporal learning tools. Specifically, we consider *Artificial Neural Networks (NNs)* that are commonly used for solving regression and classification tasks. In addition, we study the performance of a hybrid learning tool that consists of a binary decision tree wherein the nodes are NNs.

The article is organised as follows. Next section presents a background about the behaviour of the pupil and its relationship with the cognitive system. Section III introduces the experimental procedure used for the data collection, the preprocessing of the data, and an analysis of the dataset. The learning techniques are introduced in Section IV. Empirical results are presented in Section V. Final section contains some conclusions and a discussion about future research directions.

II. Emotionality and Pupil Behavior

For several years now, the *International Affective Picture System (IAPS)* has been the reference for objective emotional assessment based on visual stimuli. A large database of pictures was developed, the pictures represents a range of daily experiences such as: the household furniture to extreme encounters and a mutilated body. Each picture is rated according a subjective score given by a large group of people. This score is given by the subjects following their feelings of pleasure and arousal that each picture evokes during viewing. The familiarity of the human experiences is what evokes such an array of emotions. The pictures were classified according

978-1-4799-0652-9/13/$31.00 ©2013 IEEE

global indicators of these affecting scores. The IAPS provides a way to assess the emotions and it has been widely used in experiments studying emotion and attention. For more information about the experiment protocol used for generating the IAPS see [3].

The pupil size variation is controlled by the *Autonomic Nervous System (ANS)*, and it is recognised as a potential variable to measure the activities in the ANS. At the beginning of the 60s, have been presented several works studying the relationship between the emotion and the pupil dilatation [2], [4]–[6]. According to [6] the autonomic nervous system is sensitive to arousal emotions. The latest researches on pupilometry have confirmed that when the increasing of the pupil size has a linear relationship with the affective arousal. In [4] has been showed that when the emotions are provoked by unpleasant images the pupil tends to constrict its size. On the other hand, when the emotion is caused by pleasant pictures the pupil is dilated. A relationship between the pupil dilatation and cognitive load was analysed in [7], [8]. More evidence about pupil dilation and emotion is described in [6]. There exists also a relationship between pupil dilation and the stimuli affecting for seeing persons of opposite sex [9]. A relationship between the pupil dilation and sexual attraction was studied in [9].

There are several devices for measuring the pupil size, the most used technique is the analysis of the corneal and pupil reflex through thermographic camera recording. This camera measures the gaze reflex produced by an infrared light that is attached to the camera. As a consequence, it is an assessment technique less invasive that other devices used for measuring emotions such as: EEG signals, heart rate and skin conductance. The light enters the eye and begins the process of visual perception through the pupil. The diameter of that opening is determined by contractions of two opposing sets of muscles within the iris: the sphincter and pupillary dilator. The pupillary sphincter is under the control of the sympathetic nervous system and provokes the pupil constriction. The muscles in the parasympathetic nervous system control the pupil dilatation [10].

The accommodation reflex and the light are the initial relevant variables that impact in the pupil size variation. In addition of those variables, there are also visually insignificant fluctuations in the pupil dilatation, which are cognitively related. These movements appear to be attenuated reflections of changes in the brain activation system [10]. Figure 1 illustrates the normal evolution of the pupil size after a visual stimulus [11]. We can identify relevant pupil features in the curve: pupil contraction-dilation latency (A), pupil contraction (B) and pupil dilation (C). The pupil behaviour is highly dependent of the subject, it can vary among several normal subjects [11]. Although, it keeps a symmetry when both eyes are compared in the same subjects [11].

In spite of the advantages presented by thxe pupil size variable, the technique presents some well-identified limitations. The size variation is highly related with luminous reflex. For that reason luminous parameter has to be controlled. Additionally, the pupil size is most often affected by sources of noise, which are provoked by the blinks and fast gaze movements (also known in the literature as *saccades*). Therefore, in order to improve the classification efficiency of a machine learning tool, a pre-processing step should be performed in order of decreasing the noise impact.

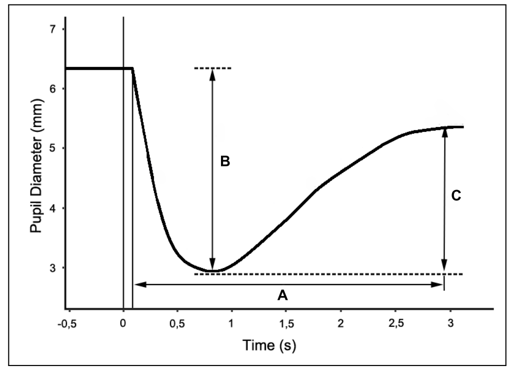

Figure 1. Normal pupil diameter behaviour after a visual stimulus. Figure extracted from [11].

III. Methodology

In this Section, we describe the protocol used for generating the dataset. Next, we present the technique used for pre-processing the data set before to consider it for using for training the learning models. The section ends with an analysis of the collected data set that is used for defining the learning tools in the next section.

A. Experimental Procedure for Collecting the Dataset

Participants: The data was collected during the experimental sessions with 4 subjects composed by 4 males with ages of 19, 24, 24 and 27 years olds. All subjects declared to have normal or corrected-to-normal vision and they did not have previous information of the purpose of the study. Each subject has provided written a participation consent approved by the *Ethical Committee of Faculty of Medicine of University of Chile*.

Materials and Design: We used a collection of 90 images that were randomly selected from the IAPS dataset [3]. The IAPS image is evaluated according the opinion of several subjects following the discreet scale 1, 2, ..., 9, then the mean of those opinions is the score used for qualifying the image. Therefore the IAPS score for each image is a continuous value in $[1, 9]$. We create a regular partition of this interval in three categories: negatives (C_1), neutral (C_2), and positives (C_3). The selection of images was done such that the collection has 30 images for each emotional status. All pictures were landscape in orientation (1024 x 768 pixels) and were displayed in a 24-bit color. Luminosity of each picture was measured by a photometer, and pictures were modified in order to get a similar mean luminosity (17 lux). The data was collected during sessions composed by trials that are defined as follows:

- A picture is randomly selected and it is displayed in the screen during 4 seconds. The pupil size is measured during this period.

- Then, in order to stabilise the pupil dilation a gray image is shown during a range of time of 2 and 4 seconds.

We use each picture twice, therefore the complete experiment process consists in 180 trials wherein each subject gaze 90 images. At the beginning of the experimental session for each subject an eye-tracking calibration is performed. In addition, a eye-tracking calibration is performed each time that 30 pictures are displayed to the subject. Figure 2 illustrates the protocol used during an experimental session, and Figure 3 shows a diagram corresponding to a specific experimental trial.

Experimental session						
Cal	Trial	$\cdots$	Trial 30	Cal	$\cdots$	Trial 180

Figure 2. Structure of an experimental session of each subject. The session starts with an eye-tracking calibration (Cal), then a sequence of 30 trials are performed, next an eye-tracking calibration is performed again. We repeat the that procedure until the trial 180 is performed.

Experimental trial	
Display IAPS image	Gray image
Measure (4 seconds)	2-4 seconds

Figure 3. Structure of an experimental trial for each subject. Each image was presented using random selection and displayed during 4 seconds. During this period the pupil dilatation was measured. Next, a gray image is displayed during a range of 2 till 4 seconds. We do not use the pupil size during the gray image is displayed.

Apparatus: The presentation of the picture was controlled by the software *Experiment Builder* made by *SR Research* [12]. Pictures were displayed on a $32''$ LG screen located in the experimental room, at a distance of 1 meter from where the subject was seated. Pupil size and gaze position were recorded using *EyeLink 1000* eye tracker made by *SR Research* [13], which consists in an infrared camera and an infrared light source plus a PC hardware for configuring and collecting eye-tracking data. The eye tracker recorded both eyes in a rate of 500 Hz during all the experiment. Subject's head was adjusted to a support that reduce head movements. *EyeLink 1000*'s infrared camera was located in a wood table in front of subject. The experimental room had no light during recording sessions.

B. Data Pre-processing

In order to extract the noise from the signals, we performed a pre-processing of the data. We start by defining the data notation. Let $X(t)$ be the original pupil size at time t before the data pre-processing, and let $Z(t)$ be the signal containing information about the pupil size after the pre-processing of $X(t)$. In addition, we use the point of gaze $(p_1(t), p_2(t))$ at each instant of time, where $p_1(t)$ is the gaze position on a horizontal axis and $p_2(t)$ is the gaze position on a vertical axis. Therefore, for each element $(X(t), p_1(t), p_2(t))$ there is the discreet variable $C_i(t)$, $i = 1, 2, 3$ that corresponds to a label given by the displayed IAPS image. The pre-processing procedure was as follows:

1) Blink extraction: When blinks are detected by the eye-tracking machine, the pupil size was measured as zero. To avoid this problem, we perform a linear interpolation between the last size measured before the blink and the first pupil size measure after the blink. Then, the pupil size during the blinking time was set to the value given by that interpolation function.
2) Saccade extraction: The saccades are fast movements of the eyes, for example, when the gaze changes quickly from one fixed point to another one. This phenomenon provokes in the pupil size signal some jumps, that we consider as noise. We apply the same interpolation technique that for removing the blinks. Thus, the beginning and ending points of a saccade are used for defining an interpolation function. The part of the signal that contains the saccade is completed with the value given by that interpolation.
3) High frequency extraction: We apply a low pass filter that only permits measuring frequencies lower than $2Hz$. That is in order to avoid artifacts in the signal.
4) Normalization: The eye-tracking instrument measures pupil size with an arbitrary metric. We apply the *z-score* normalization for the measured scores, which consists in the following. Given a period (Δt) wherein one IAPS picture is shown to the subject, at each time t the z-score $z(t)$ is computed as: $Z(t) = (X(t) - \mu(\Delta t))/\sigma(\Delta t)$, where $X(t)$ is the current measure, $\mu(\Delta t)$ is a mean value, and σ is a standard deviation. The values of $\mu(\Delta t)$ and $\sigma(\Delta t)$ are computed with the pupil size measured during an arbitrary Δt value.

C. Dataset Analysis

For analysing the data we follow a similar methodology that [2] and [6]. Figure 4 shows four graphs that depict the average pupil size with 95% confidence interval using Gaussian approximation by emotional state for four subjects [14]. Each graph corresponds to the evolution in time of the pupil size for a specific subject when an image is displayed. The red curve corresponds to the pupil size average when the stimuli was provoked by a negative image. The blue curve shows the average of size changes according time when positive images were displayed. The green curve shows the average of pupil size according time when a neutral image was displayed. We can see in Figure 4 that the behaviour of the pupil size differs from one subject to another one, although we can identify some common patterns. In most cases pupil size behavior is similar to the normal expected one shown in Figure 1. In the case of subject 2, the pupil size does not show a predominant pupil contraction, but it exhibits a large pupil dilation for the negative emotional state. In the three cases, the average of the pupil size when the images used were negatives is larger than the size when the images used were neutrals and positives. In all cases, positive and neutral pupil size have a similar behavior in both contraction and dilation.

Figure 5 shows the average of the normalized pupil size per emotional status. These results are different from ones

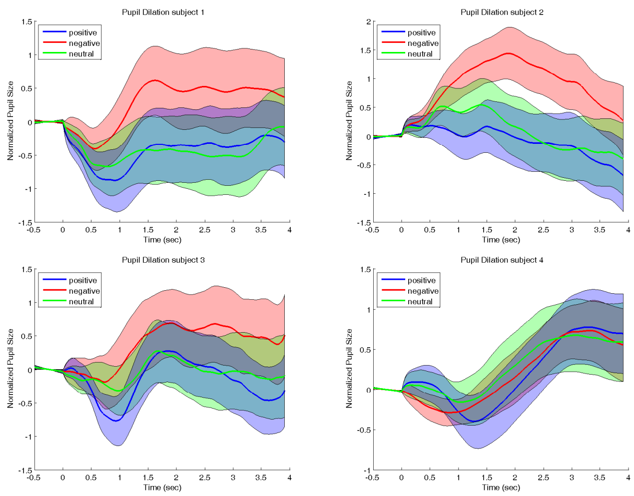

Figure 4. Average of the pupil size with a 95% confidence interval by emotional state for the four subjects.

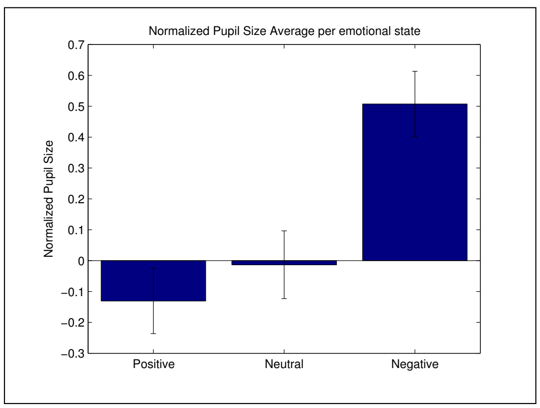

Figure 5. Averaged normalized pupil size (and s.e.m.) between 500 ms post-stimulus and stimulus offset of the different stimulus categories.

obtained by Bradley et al. [2] and Partala et al. [6]. In [2], the subjects were exposed to visual grayscale stimuli and pupillary changes were larger when viewing emotionally arousing pictures, regardless of whether these were pleasant or unpleasant. In [6], the subjects were exposed to auditory stimuli the results showed that the pupil size was significantly larger during both emotionally negative and positive stimuli than during neutral stimuli. Possible explanations for this phenomenon are different experiment settings between our study and the last two mentioned. In our study we used visual stimulus unlike Partala *et al*. where they used an auditory stimulus, and also we used pictures displayed in a 24-bit color unlike Bradley *et al*. who used pictures displayed in a 16-bit grayscale.

IV. Machine Learning Description

The goal of supervised learning consists in predicting the value of outcome measures based on a number of input measures [14]. In this article, the aim is finding a relationship between the gaze position and the pupil size with the emotional states. We assess the emotional states according the type of image used for stimulating the subjects. Therefore, we have that the outcome measure can have three values: negative, neutral and positive. The section follows with a description of the learning tools studied in this article.

A. Neural Networks

For generating the machine learning model, we consider the evolution of the input information in time. The learning model should be able to learn the geometrical characteristics of the patterns as well as their serial order. We use *Neural Networks (NNs)*, for an overview about this learning family see [14], [15]. In the case of working with NNs, the most common approach consists in representing the sequential order of the input patterns using the input neurons. This means that an input pattern is defined containing the information in an arbitrary range of time (time window). The input layer has dimensionality equal to the input pattern. The first element in the input pattern represents the first temporal event in the time window, the second element represents the second temporal event, and and so forth.

More formal, let q be the number of explanatory variables collected at each time, and let the output exogenous variable be an unidimensional value. We define a time window having s units of time. The lagged input information given by the explanatory variables is collected in the vector

$$\mathbf{u}(t) = (u_1(t-1), \ldots, u_1(t-s), \ldots, u_q(t-s)).$$

The output at each time is $y(t)$. In our approach we do not use delayed exogenous variables as inputs to the system (this means at time t we do not use the results of $y(t-k)$, $k > 1$). In our problem, we have $q = 3$ and an arbitrary value for s, then the input is composed by

$$\begin{aligned}\mathbf{u}(t) =& (Z(t-1), \ldots, Z(t-s), p_1(t-1), \ldots, p_1(t-s), \\ & p_2(t-1), \ldots, p_2(t-s)),\end{aligned} \tag{1}$$

where Z is the pupil size and (p_1, p_2) is the gaze point in 2D after the pre-processing of the data. Therefore, the learning set is defined as $\mathcal{L} = \{(\mathbf{u}(t), y(t)), t = 1, \ldots, T\}$. The standard approach of using feedforward NNs for solving this problem consists in using a network with $3s$ input neurons, arbitrary number of hidden neurons and one output neuron. The goal consists in inferring a mapping $\varphi(\cdot)$ given by a NN, in order to predict y, such that certain error function $E(\varphi(\mathbf{u}, \mathcal{L}), y)$ is minimized.

B. Decision Tree with Neural Networks

In the graphics of Figure 4 we can see that the pupil behaviour for negative images is clearly different that the behaviour for positive and neutral images. In particular, for the subjects 1, 2 and 3 this difference seems to be very clear. Although for the subject 4 the pupil size presents a "similar" evolution for the three types of images. As a consequence, we also experiment with the following automatic learning tool. We define a method for taking binary decisions when the image is negative and other cases. Next step, we train a new learning tool for classifying only between positive and neutral images.

Let $\mathcal{L}_1$ be the learning set when the output class is binary and corresponds to negative images, and neural and positive images. Let $\mathcal{L}_2$ be the learning set when the output class is binary and corresponds to neutral images and positive images. We independently train a NN (NN_1) using $\mathcal{L}_1$ and we train another NN (NN_2) using the set $\mathcal{L}_2$. Therefore, we have a decision tree where the decisions on the nodes are taken by using NN machines. Figure 6 illustrates the approach of using a binary decision tree with NNs as node decisions.

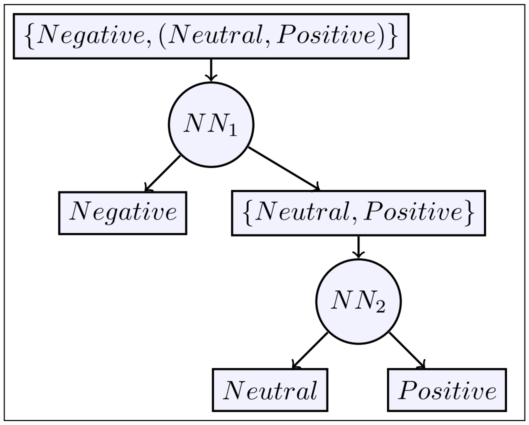

Figure 6. Binary decision tree.

V. Experimental Results

This section contains a description of the learning dataset, the setting of the model parameters, the experimental results. The section ends with a discussion about our results.

A. Generation of the Learning Dataset

The generation of the learning dataset was as follows. We have a long array of pupil size and gaze position values for each subject. These values were obtained with a frequency rate of 500 Hz. As the pupil behavior presents low frequency rates and for improving algorithms execution time, we apply a data subsampling from 500 Hz to 100 Hz. We do not consider measures corresponding to grey images, as well as we do not consider the pupil size during the subject is scoring the image. After some testing, we define sliding windows as follows. For generating the input pattern, we take the last 10 consecutive measures, then we take measures with a distance of 10 until 200, so at each time the input $\mathbf{u}(t)$ is composed by information taken in the following time windows:

$$\{t-1, t-2, t-3, \ldots, t-10, t-20, t-30, \ldots, t-200\}.$$

The total number of samples of the learning set is 271,678 (around 68,000 samples by subject).

B. Parameter setting

We run feed-forward neural network implemented in Matlab. We consider feedforward topologies, and we use the Matlab toolbox called *DeepLearnToolbox* [16]. As usual, we generate a partition with the 80% of the samples for training the network and the rest samples are used for testing the model. Feedforward Neural Network was configured with 3 layers (input, hidden and output layer). We evaluate the performance with 4 different topologies. The input layer is composed with 114 neurons and the output layer has 3 neurons. The hidden layer has 40, 60, 80 and 100 units. The learning algorithm used for training the parameters was back-propagation. We use $\tanh$ function as activation function of the neurons. The maximum number of iterations was set to 1000 epochs. The weight connections were random initialised into the range $[-4\sqrt{\frac{6}{n_i+n_h}}, 4\sqrt{\frac{6}{n_i+n_h}}]$, where n_i and n_h are the number of input and hidden neurons, respectivly. In order to avoid overfitting, we have a regularisation factor set to 1x10^{-4}.

C. Results

Table I presents the accuracy reached by the two learning models. We present the accuracy showing the percentage of well-classified testing samples. The first column refers to the learning dataset, the last one shows the accuracy reached by the decision tree model. The other columns of Table I show the accuracy reached by the NNs with different number of hidden units. The first four rows present results when we learn the model using learning dataset according to each subject. Last row presents the results when the models were trained using the whole learning dataset. The decision tree was built with NNs with 100 hidden nodes when the model is training using the dataset for each subject. When we use the whole dataset for training the decision tree, we use networks with 60 hidden units.

According to Table I we can see that the pupil behaviour is dependent of the subject. These results are coincident with the research presented in [11]. The accuracy of the learning tools adjusted for each subject outperforms the accuracy obtained when all subjects are used in the learning set. This characteristic happens for both: the NNs and the decision tree. From Table I we can also see that the model performs better when the neural network has 100 hidden units (only for the subject 3 the best topology was for a network with 80 hidden units). Therefore, we generate the decision tree using NNs with 100 hidden units in the tree nodes. Figure 7 presents the accuracy using a visualisation of bars. The vertical axis corresponds to the percentage of well-classified testing samples. The horizontal axis presents blocks of bars, each block corresponds to the learning dataset used for adjusting the model. From the figure we can see that the accuracy reached by the decision tree is higher than the reached one by the NNs. Therefore, it is better to classify the emotions using binary decisions. A first step consists in identifying negative emotions, the next step consists in discriminating the positive emotions. Despite that we are using only the evolution of the pupil size and the gaze position for recognising three classes of emotions, Table I presents promising results. In addition, if we consider the task of classifying between negative emotions and other types (accuracy of model NN_1), the results are still better. A decision tree with 100 hidden units reached the following results: 76.0$\%$ (subject 1), 82.8$\%$ (subject 2), 80.5.3$\%$ (subject 3), and 80.4% (subject 4). The model reached a 71.7% of accuracy when we consider the whole dataset.

Table I
PERCENTAGE OF TESTING SAMPLES WELL-CLASSIFIED BY THE MODEL. THE RESULTS WERE REACHED WITH DIFFERENT NUMBER OF HIDDEN NEURONS (40,60,80 AND 100) AND WITH THE DECISION TREE.

Subjects	NN				Tree
	40	60	80	100	
1	64.2	65.6	68.4	**69.8**	**74.5**
2	62.3	66.9	64.9	**68.5**	**73.2**
3	63.1	63.8	**68.6**	67.9	**62.3**
4	57.4	66.7	67.5	**70.6**	**78.4**
All subjects	46.6	**50.1**	49.9	47.5	**53.6**

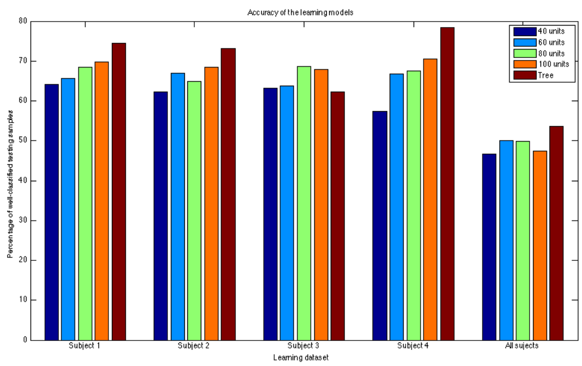

Figure 7. Visualization of the accuracy reached by the learning tools. Each block of bars corresponds to different learning dataset. From left to right the bars shows the accuracy reached by NNs with 40, 60, 80 and 100 hidden units and the decision tree.

VI. CONCLUSIONS AND FUTURE WORK

This article presents an approach for emotion recognition using learning tools. We study the states associated to human emotions as a natural phenomena that evolve in time. For recognising the emotions we use the evolution of the eye tracking data during a window of time. It has been shown that the pupil is strongly related with the emotional arousal and autonomic activation. The emotions are provoked by visual stimuli of colored images. Those images were taken from the International Affective Picture System which has been the reference for objective emotional assessment based on visual stimuli. Each picture has associated an emotional state, which are labeled as: negative, neutral and positive. We generate a learning dataset where the input information is given by the evolution of the pupil size and the gaze position, and the outputs are labels associated to the emotional states. We study two kinds of learning tools based on Neural Networks. We demonstrate that we can find a learning machine for estimating the emotions (with a *reasonable* accuracy) of the subjects when they are viewing color images. The additional interest of our approach, is that we are using only two measures (gaze position and pupil size) that are cheaper, less complex and less invasive of measuring and analizing than other signals, such as MEG or EEG that are normally used in studies related to emotional tasks.

In the near future, we plan to evaluate our approach with a larger number of subjects and other temporal learning tools. In addition, we are interested in incorporating other input features that can gives more information to the predictors. Besides, our approach can be applied for other types of stimuli, for instance auditory stimuli.

ACKNOWLEDGMENT

This article has been elaborated in the framework of the project *New creative teams in priorities of scientific research*, reg. no. CZ.1.07/2.3.00/30.0055, supported by Operational Programme Education for Competitiveness and co-financed by the European Social Fund and the state budget of the Czech Republic and supported by the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070), funded by the European Regional Development Fund and the national budget of the Czech Republic via the Research and Development for Innovations Operational Programme, and by the Project SP2015/105 *DPDM - Database of Performance and Dependability Models* of the Student Grand System, VŠB - Technical University of Ostrava.

REFERENCES

[1] Z. Zeng, M. Pantic, G.I. Roisman, and T.S. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(1):39–58, Jan 2009.

[2] Margaret M. Bradley, Laura Miccoli, Miguel A. Escrig, and Peter J. Lang. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45(4):602–607, 2008.

[3] M.M. Bradley P.J. Lang and B.N. Cuthbert. International affective picture system (iaps): Affective ratings of pictures and instruction manual. technical report a-8., 2008. University of Florida, Gainesville, FL.

[4] Eckhard H. Hess and James M. Polt. Pupil size as related to interest value of visual stimuli. *Science*, 132:349–350, August 1960.

[5] R. W. Picard. *Affective Computing*. M.I.T. Press, Cambridge, MA, 1995.

[6] Timo Partala and Veikko Surakka. Pupil size variation as an indication of affective processing. *International Journal of Human Computer Studies*, 59(1-2):185–198, July 2003.

[7] J. Hyönä, J. Tommola, and A.M. Alaja. Pupil dilation as a measure of processing load in simultaneous interpreting and other language tasks. *Quarterly Journal of Experimental Psychology*, pages 598–612, 1995.

[8] D. Kahneman and J. Beatty. Pupil diameter and load on memory. *Science*, 154, 1966.

[9] E.H. Hess and S.B. Petrovich. Pupillary behavior in communication. *Nonverbal Behavior and Communication*, 1987.

[10] John T. Cacioppo, Louis G. Tassinary, and Gary Berntson, editors. *Handbook of Psychophysiology*. Cambridge University Press, March 2007.

[11] Oliver Bergamin, M.Bridget Zimmerman, and Randy H Kardon. Pupil light reflex in normal and diseased eyes: Diagnosis of visual dysfunction using waveform partitioning. *Ophthalmology*, 110(1):106 – 114, 2003.

[12] SR Research. Experiment builder. Available at: http://www.sr-research.com/eb.html, date of access: April 14, 2015.

[13] SR Research. Eyelink. Available at: http://www.sr-research.com/eyelinkII.html, date of access: April 14, 2015.

[14] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of Statistical Learning*. Spring series in statistics. Springer-Verlag, New York, USA, 2001.

[15] Bing Cheng and D. M. Titterington. Neural Networks: A Review from a Statistical Perspective. *Statistical Science*, 9, February 1994.

[16] R. B. Palm. Prediction as a candidate for learning deep hierarchical models of data. Master's thesis, Technical University of Denmark, DTU Informatics, E-mail: reception@imm.dtu.dk, Asmussens Alle, Building 305, DK-2800 Kgs. Lyngby, Denmark, 2012.