YOUMING ZHANG

Biometric Verification of a Subject Based on Data Mining of Saccade Eye Movement Signals

ACADEMIC DISSERTATION
To be presented, with the permission of
the Board of the School of Information Sciences of the University of Tampere,
for public discussion in the Auditorium Pinni B 1096,
Kanslerinrinne 1, Tampere, on February 21st, 2014, at 12 o'clock.

UNIVERSITY OF TAMPERE

ACADEMIC DISSERTATION
University of Tampere, School of Information Sciences
Finland

Cover design by
Mikko Reinikka

Acta Universitatis Tamperensis 1899
ISBN 978-951-44-9355-3 (print)
ISSN-L 1455-1616
ISSN 1455-1616

Acta Electronica Universitatis Tamperensis 1381
ISBN 978-951-44-9356-0 (pdf)
ISSN 1456-954X
http://tampub.uta.fi

Suomen Yliopistopaino Oy – Juvenes Print
Tampere 2014

Abstract

This thesis focuses on biometric verification of subjects based on saccadic eye movements. Verification corresponds to two-class classification to recognize an authenticated user and to classify other subjects as impostors. Compared with other biometric signals or data, the possible advantages of eye movements can be as follows: harder to imitate, easier processing and faster computation. The thesis describes a procedure to use variables of saccade eye movements recorded. It analyses the variabilities between electro-oculography (EOG) and video-oculography (VOG) signals: i.e. eye movements were recorded with skin electrodes or with two special video cameras. When a signal was recorded with a low-frequency video camera device simulating a web camera, the sampling frequency of signals was enhanced using interpolation. The techniques of signal processing and statistics were also applied to analysis. In order to evaluate biometric accuracy, the test procedures for true positive rate (TPR) and true negative rate (TNR) were designed separately. Many classification methods were explored for verification performance, including both modified simple methods such as *k*-nearest neighbour searching and advanced methods such as neural networks and support vector machines. Approaches and other details in the verification procedure were improved through multiple tests and comparisons of the verification accuracies. Optimal parameters and settings of the classification methods used were found. With more and more saccades and subjects collected into training sets, a high TNR accuracy was gained, which was close to 95% at its best. It showed that, using saccade eye movements, it was possible to distinguish between an authenticated user and impostors. On the other hand, after multiple recordings of subjects, the high accuracy of TPR – close to 90% – also confirmed that an authenticated user can be recognized notwithstanding the variability of variable values of saccades between different sessions. Finally, better results given by signals with a relatively high sampling frequency of 250 Hz were obtained, and this could allow user verification based on eye movements to be applied in practice, along with the development of eye movement video cameras in future.

Keywords: Biometric verification, eye movements, saccade, signal analysis, classification, machine learning, data mining

Acknowledgements

I would like to express my appreciation to the individuals and organizations that supported me in my doctoral studies and research over the past four years. I could not have completed the work without their help.

Firstly, I would like to express my gratitude to my supervisor, Professor Martti Juhola, for offering me the opportunity to study at the University of Tampere School of Information Sciences, for his excellent guidance in my research and for the help he gave an international doctoral student with daily life in Finland. I was always able to obtain suggestions and help from him when I had problems.

I am also grateful to my colleagues in our research group, the Data Analysis Research Group (DARG), especially Henry Joutsijoki, PhD, and Jyrki Rasku, PhD. They gave many suggestions to help me with the problems of how to apply classification methods and record the signals. On the subject of living and studying in Finland, Yevhen Zolotavkin, PhD, also helped me a great deal.

The financial support mostly came from the Tampere Doctoral Programme in Information Science and Engineering (TISE) and from the Centre for International Mobility (CIMO), which enabled me to focus on my doctoral research. The signal recording devices in my research were provided by Professor Ilmari Pyykkö of the Department of Otorhinolaryngology at the Tampere University Hospital and Professor Kari-Jouko Räihä of the Visual Interaction Research Group (VIRG). I am very thankful to their support, and especially for the help provided by Oleg Spakov, PhD.

Professor Pekka Loula of the Tampere University of Technology in Pori and Professor Juha Röning of University of Oulu reviewed this thesis, and I am grateful to them for their valuable comments and suggestions.

Finally, I want to thank my family for their continual encouragement and support.

Youming Zhang

Table of contents

List of Abbreviations

Abbreviation	Description
1-D	One-Dimensional
ACC	Accuracy
AUC	Area Under the Curve
BP	Back-Propagation
ECG	Electrocardiography
EEG	Electroencephalography
EER	Equal Error Rate
EOG	Electro-Oculography
FAR	False Acceptance Rate
FNR	False Negative Rate
FPR	False Positive Rate
FRR	False Rejection Rate
KNN	K-Nearest Neighbour Searching
LDA	Linear Discriminant Analysis
LED	Light Emitting Diode
LMA	Levenberg-Marquardt Algorithm
LogDA	Logistic Discriminant Analysis
MDA	Multiple Discriminant Analysis
MLP	Multilayer Perceptron
NN	Neural Network
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
PCHIP	Piecewise Cubic Hermite Interpolating Polynomial
RBF	Radial Basis Function
ROC	Receiver Operating Characteristic
SVM	Support Vector Machine
TNR	True Negative Rate
TPR	True Positive Rate
VOG	Video-Oculography

List of Original Publications

I. Y. Zhang, J. Rasku and M. Juhola. Biometric verification of subjects using saccade eye movements. *International Journal of Biometrics*, Vol. 4, No. 4, 317-337, 2012.

II. M. Juhola, Y. Zhang and J. Rasku. Biometric verification of a subject through eye movements. *Computers in Biology and Medicine*, 43, 42-50, 2013.

III. Y. Zhang and M. Juhola. Biometric verification of a user based on eye movements. Accepted to *International Journal of Cognitive Biometrics*, 2013.

IV. Y. Zhang and M. Juhola. On Biometric Verification of a User by Means of Eye Movement Data Mining. In *Proceedings of the Second International Conference on Advances in Information Mining and Management (IMMM2012)*, 85-90, Venice, Italy, October 2012 (Best Paper award).

V. Y. Zhang, J. Laurikkala and M. Juhola. Biometric verification of a subject with eye movements, with special reference to temporal variability in saccades between a subject's measurements. Accepted to *International Journal of Biometrics*, 2013.

VI. Y. Zhang and M. Juhola. On applying signals of saccade eye movements for biometric verification of a subject. In *Proceedings of the 8th International Conference on Advances in Mass Data Analysis of Images and Signals in Medicine, Biotechnology, Chemistry and Food Industry (MDA 2013)*, 78-92, New York, USA, July 2013. (The next-best paper in the competition for the Best Paper award.)

Chapter 1

Introduction

Identification and verification based on biometric information [1,2,3,4] has been researched and developed for the past 20 years. Some of relevant methods are widely and deeply applied in many fields in society. Fingerprints [5,6,7,8], face images [9,10,11,12] and signatures [13] are well known by the public. For example, fingerprints are used for criminal identification and instead of passwords in many areas. Face images have also been used in access control systems. Handwritten signatures as an authentication method have been used for more than a thousand years of human history. Retina [14], iris [15,16,17] and palm print scanning [18], as well as voice [19] and electrocardiography (ECG) signals [20,21,22,23] are also increasingly being studied.

Biometrics can be divided up in various ways. According to character [24,25,26], they could be separated into physiological biometrics (fingerprints, face or iris images) and behavioural biometrics (voice or signatures). They could also be categorized by signal complexity as one-dimensional (1-D) or 2-D.

Identification [10,26,27,28] is a kind of classification that allows recognition of a given person from among a large group of members. Compared with identification of *n*-class classification, ***verification*** can be considered a simpler situation: it is only two-class classification in that, for example, the correct user of a computer has to be recognized and other possible subjects are simply determined as non-users or impostors (more discussion in Section 3.1).

With technological advances, the biometric identification and verification of a subject has become more and more reliable and convenient, but some drawbacks and disadvantages must still be noted. The first one is artificial imitation. A face image can be replaced by a photo; there are also many devices that can surreptitiously record a person's voice. It is not difficult to steal the fingerprints of a person, which could be left anywhere. A large data set of signals, such as images, is another problem for biometric identification and verification. For example, a complex image often contains a great number of features, which make for a huge data set, complicated computation and time-consuming processing. Moreover, measurement conditions should be considered as well. The iris and retina need an advanced and complicated device to be measured. Face images [28] are sensitive to changing factors such as illumination, glasses and hairstyle. A voice may be easy to obtain, but such a recording needs a background without noise and may also be affected by other circumstance factors. Therefore, one possible option is to try to find a novel signal. It should include less data for which the corresponding procedure of verification and computation should become simpler and faster. One-dimensional signals, electroencephalography (EEG) [29,30] and ECG have been researched for biometrics, but the number of publications is clearly far smaller than that for image data.

Eye movements [31,32,33,34,35,36,37] are an interesting and potential behavioural biometric objective for the purpose of classifying subjects. Just like EEG and ECG, eye movements are one-dimensional biometric signals. Their data amount is smaller than those of images, but they can contain enough information for the verification of subjects. For example, the latency of saccades is such a feature type of eye movements [38,39,40,41,42] (Figure 1.1). It represents the time that person needs to respond to a stimulation. Researchers investigated infants between four and eight months old to reveal the latency values of their saccades and attempted to find out how they became shorter just during these four months [43]. Moreover, as behavioural biometrics, eye movements are more difficult to copy or steal than such images as fingerprints. Since human beings cannot control some bodily responses consciously [44], it is also difficult to imitate the eye movements of other subjects. Although eye movements have to be measured by video cameras of high quality at present, with the technology developing, such video cameras will probably become cheaper and measurements will be simpler in future.

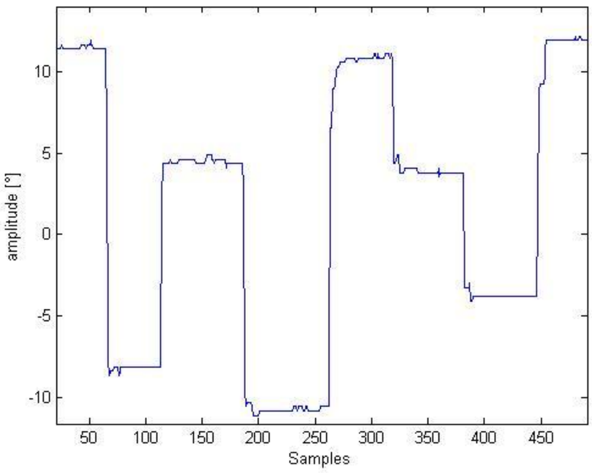


Figure 1.1. Several saccades of a 30 Hz signal.

Eye movements have been researched and investigated for decades in many different fields such as medicine, psychology, and other related areas. For example, latency values were used to distinguish between children and adults, and the results showed how latencies of children were longer than those of adults [45]. Some variables of eye movements in time domain are used to detect patients from the healthy. Over the past several years, they have also been studied for use in a human-computer interface [46,47]. A human can make a simple command to a computer or device by eye movements, for example to choose an icon on the screen.

Although eye movements are infrequently researched for the verification of subjects, some achievements have been made in using them for identification purposes. The following are a few first cases. In one study [44], researchers measured the blind spot on a subject's retina and, combined with latency, classified a user for authentication. After the calibration of data and finding the subject's blind spot, a target was displayed at a random position on the authentication screen which could be within or outside the blind spot. After the user followed the target displayed, saccades were produced and the system measured their latencies. A user was authenticated if the saccades occurred at 100 to 500 ms when the target displayed outside his/her blind spot or saccades were not induced when target was within the blind spot. It was supposed that subjects' blind spots are located at different places on their retinas.

In other research [48,49], a mathematical model of the oculomotor system based on oculomotor plant characteristics was used for biometric authentication. The model simulated the six muscles rotating the eyeball when it moves, and parameters of the model were used as input to classify the subjects. Researchers tested the false acceptance rate (FAR: the probability that a biometric system will incorrectly grant a non-user or impostor access) and false rejection rates (FRR, the probability that the biometric system will wrongly reject an access attempt by an authenticated user) with nearest-neighbour searching and C4.5 decision trees, and the results [49] were better than those of their previous work [48]. Moreover, at an early stage of eye movement classification [50,51], variables were extracted from a series of signals measurements, each measurement consisting of six integer values: stimulation position (sx, sy) and what left and right eye looked at (lx, ly, rx, ry). They were then studied by computing cepstrum signals, and the subjects were classified using naïve Bayes decision, nearest-neighbour searching, decision trees and support vector machines. Finally, pupil size, gaze velocity and the distance between a subject's eyes were also used in user recognition [52]. When measuring the signals, fast Fourier transform and principal component analysis were utilized for preprocessing. Nearest-neighbour searching was used for the classification between subjects. However, the distances between the eyes of subjects based on image analysis was the only input that was efficient in this study, and other variables of eye movements were applied in a minor role. In addition, graph matching based on eye movement was also used for verification [53]. However, these techniques are not real verification based on eye movements.

There are several types of eye movements [54], e.g. saccade, smooth pursuit movements and nystagmus. Saccades are the most popular and easiest for biometric verification or identification of subjects because their waveforms are the simplest of all, they are easy to stimulate and they are the fastest eye movements: in other words, obviously shorter signals can be used. Based on various measurement ways used here, eye movement signals can be classified as electro-oculographical (EOG) [38,39,55] and video-oculographical (VOG) [34]. The two signals look similar, but the difference is the recording approach. The former uses skin electrodes attached to the corners of the eyes of a subject to record signals on the basis of potential difference between the retina and cornea of an eye. The latter uses two small video cameras to obtain eye movements. Compared with an EOG signal, a VOG measurement is much simpler and faster to make. Due to the equipment limitation, the frequency of VOG is not often as high as that of EOG, but the less noisy and more practical VOG method has become the main method of measuring eye movements at present. Due to the regular development of electrical devices, high-frequency cameras have already been obtained, but they are still expensive. For these two reasons, this thesis mainly includes research on and considers how to make a biometric verification of a user on the basis of VOG saccade eye movements. The verification of subjects using EOG signals plays a secondary role as a comparison with VOG signals.

The research questions and hypothesis are as follows. The thesis will attempt to explain and test them in the following chapters.

- Can an individual be separated from others by saccade eye movements based on single recordings?
- Can an authenticated or right user be distinguished from others by saccade eye movements based on multiple recordings?
- As to saccade eye movements, is the verification effect better when it is based on EOG or VOG?
- How is it possible to process signals measured with a low-frequency device, corresponding to a normal web camera.
- Which classification methods, parameters and models suit verification and how are their results evaluated?
- The hypothesis was that problems associated with the above questions are possible to solve and the verification task given can successfully be performed.

The outline of this thesis consists of the following parts: Chapter 2 provides general information on the eye movement signals used, which include EOG and VOG figures, saccade structure and variables (features). How to make eye movement measurements, signal processing, selecting variables and computation of variable values are also given. Chapter 3 describes the statistical information, relevant application and classification procedure. Results of individual publications are surveyed in Chapter 4, and development and progress of the research as a whole is also presented. Chapter 5 concentrates on discussion and a conclusion.

Chapter 2

Signal processing and analysis

2.1 Recording of eye movements

2.1.1 EOG and VOG recordings

Two types of eye movement data sources were used: EOG and VOG. The data set of EOG signals was from earlier research [34,56], and its sampling frequency was as high as 400 Hz. It was old and quite noisy [39,40,42], but the higher sampling frequency compared to VOG signals was interesting in obtaining enough information about eye movements. As new data sets, VOG signals were more important and practical, and thus more relevant in user verification, but their disadvantage was the low measurement frequency: for example, the main type of VOG signals in this thesis consisted of only 30 Hz. On the other hand, this was not merely a poor detail, since the use of this low sampling frequency could be understood to simulate the situation of moderate eye movement video cameras that may appear in computers and other devices in the near future. For example, in a smart phone like the Samsung® SIII, the front camera is used to check whether its user is looking at the screen and, if so, its screen saver is not switched on.

In fact, few studies utilized both horizontal and vertical eye movements to select different verification or identification variables [36,57]. It can bring more variables, indeed, compared to the use of horizontal eye movements only, but the procedure of extracting variable values is slightly more complicated. In addition, because recordings should be as short as possible to make a verification test fast, this means shortage of data, and whether this would be a weakness or not for verification using several variables was not studied in [36,57]. Therefore, to make measurements and variable selection as simple and practical as possible, only horizontal eye movements were measured in this research, instead of two directions. On the other hand, vertical eye movements are sensitive to eye blinks in EOG, and wide vertical angles cannot be recorded as exactly as wide horizontal eye movements, so horizontal eye movements are more suitable for EOG measurement than vertical.

For EOG measurements, skin electrodes were used to record potential differences when eye movements are induced. The signals had been recorded monocularly at the same time from both eyes, with one skin electrode close to each outer eye corner and a ground electrode on the forehead. The sampling frequency of signals was 400 Hz, which was amplified to the scale of ± 10 V, converted with a 13-bit analogue-digital converter and filtered digitally with a low-pass filter with a 70 Hz cutoff. The measurement system applied the constant amplitude stimulations of 60° at the beginning and end of each signal to adjust the calibration.

In a VOG test, there is one video camera for each eye recording eye movements when the position of a pupil changes. There is a built-in image processing system in the VOG system to find the pupil of an eye in order to compute eye movements on the basis of the position of the pupil. The sampling frequency was 30 Hz, but in the most recent paper also 250 Hz was used. In fact, the former VOG system required no separate calibration [58] (except when the system was installed for the very first time) as long as the pupil of each eye could be caught in images. Since in VOG there were two video cameras, one for each eye, two horizontal signals were recorded at every measurement. The one with higher quality, which contained less possible noise or artefacts such as eye blinks, was selected from among two. The amplitude accuracy of both measuring techniques was better than 1°.

Generally speaking, EOG is often noisier than VOG, because the former can contain abundant noise, such as that arising from facial muscles due to talking, smiling, frowning or gasping. Therefore, a subject was not allowed to do these things during measurements. VOG measurements are much more ‘user-friendly’ in these aspects, since they do not involve these problems, and a subject only needs to keep his or her eyes open, be alert and be responsive to stimulations shown. However, intensive facial expression can also cause trouble with these measurements. Thus, subjects were asked to concentrate on the measurement under relaxed conditions. Moreover, in order to avoid the light reflection of the camera in calibration (with the device of 250 Hz sampling frequency) and to help the subjects concentrate on a measurement, the recording laboratory room at the cellar level equipped with high-quality shields against electrical distortion, noise and tremble was utilized for all VOG measurements.

Figure 2.1. A subject was making a measurement of eye movements with the VOG system (Visual Eyes®, Micromedical Technologies, UK) with an image resolution of 320×240 and a sampling frequency (frames per second) of 30 Hz.

Figure 2.1 shows the procedure for measuring VOG eye movements. With the gaze, the subject followed a red dot (Light Emitting Diodes, or LEDs) in the black bar in front of him. When the measurement started, one LED was turned on and, after a short interval, it was switched off and another switched on immediately. This stimulation series was continued till the end of a measurement that was exactly the same for all subjects. It looked like the red dot 'jumped' from one location to another on the horizontal axis. The locations of the black bar and chair were fixed and the distance between the eyes and bar was also kept constant, which means the angle from an eye to two sides of the bar was fixed and available for calculating. Actually, the distance of an eye to the bar cannot be exactly the same for every individual in a practical measurement, but the negative influence of a slight alteration is minor, because alteration in visual angles between dots would change very little. To avoid a situation that subjects would attempt to predict the next stimulation movement of the light dot (LED), the directions of the light dot 'jumping' (to left or right) were designed to be random from the viewpoint of a spectator. Although intervals between switching on and off were about two seconds, they were still unexpected for a spectator, for the sake of the irregular variability. On the other hand, in order to obtain more data, one measurement series had to be repeated by each subject several times. Thus, a stimulation movement also had to be varying so that a subject could not learn or remember it. However, the same stimulation series (with fixed amplitudes and intervals) was shown to each subject so that their responses could be compared to each other. It was also important that any valid saccades would be reliable responses for the stimulation movements arranged. This type of saccade stimulation has been used in medical investigations for decades as a standard-like convention [32,38,59]. Furthermore, it was important that each subject showed several responses to similar stimulations in data analysis to collect sufficient data for classification decisions. This was important because machine learning algorithms were then able to learn variable values of individuals from the data.

The type of stimulation movement employed was as simple as possible to guarantee that it was easy to follow and did not cause fatigue, even if it was repeated consecutively many times for a subject. Additionally, to distinguish a subject, as simple stimulation as possible is a reasonable approach. Otherwise more repetitions of data collecting after complicated stimulations would be needed. A test base of a complicated stimulation would be easier at risk of failing, but this simple one normally yielded no failure.

Utilization of the above-described stimulation (red light dot) stems from the idea that subjects could use eye movements to log in to their computer, private system, account, webpage etc. Without any password, the computer or system could recognize its legal user or reject impostors. The idea was also that there would be a light dot on the screen: the user would stare at it and then follow the dot jumping horizontally after short intervals. The whole procedure is quite similar to the measurement procedure described above. Both stimulation amplitude (lengths of the light dot jumps) and intervals between jumps were varied. In addition, most stimulation amplitudes should be large enough, such as 40-60°, because the large amplitudes would probably guarantee that variability occurred in saccade features between subjects [34,60]**.** Large amplitudes of saccades were necessary for verification, but smaller ones could be interspersed with the large ones in the series of stimulation movements so that the angle and direction (left or right) changed unexpectedly to make it random-like for a spectator.

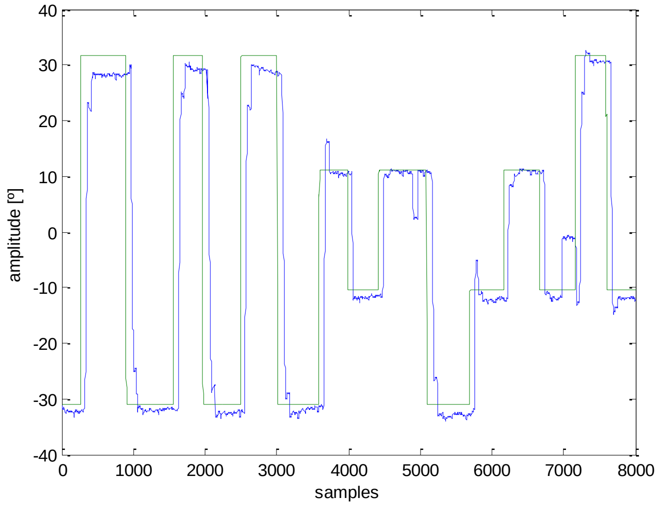


Figure 2.2. A 20-second EOG signal sampled at 400 Hz. The smooth step signal is its stimulation signal recorded at the same time.

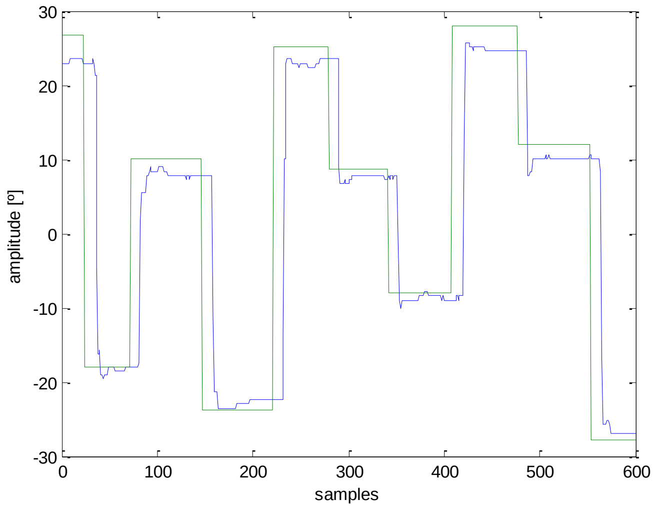


Figure 2.3. A 20-second VOG signal sampled at 30 Hz with its stimulation signal.

Figures 2.2 and 2.3 show the segments of an EOG and VOG signal and the corresponding stimulation signals (blue for eye movements and green for stimulation in colours). Since the sampling frequencies of the EOG and VOG signals were 400 Hz and 30 Hz, the corresponding numbers of the 20-second samples were 8000 and 600. The vertical axis is amplitude (angle) in the figures, where value 0 means the position is in the middle of the bar. The right side is defined as positive and left as negative. For EOG signals, a measurement lasted for 80 seconds and included 12 or more large saccades. However, for the VOG signals, there were only four large saccades in one measurement 64 seconds in duration. Thus, subjects had to repeat the tests a few times. Since the frequency of VOG signals was so low, only 30 Hz, it was 'artificially' increased with interpolation (to be presented in Section 2.2) before data analysis and calculation of variable values.

The EOG signals were recorded at a university hospital, and a physician checked all voluntary subjects for the ability to do the test without impediment. The distance between the target of a computer-controlled light dot and a subject was fixed at 1.40 m. Since subjects had to wear a 'mask' with the video camera system, spectacles could not be used in the VOG measurements. Whether a subject was able to see the light dot accurately enough was checked in advance in order to avoid potential problems such as severe myopia. In addition, the constant distance between the target of the LED bar and a subject was 0.74 m, shorter than for EOG.

2.1.2 Higher sampling frequency for VOG recordings

Since verification results when using VOG (30 Hz, but interpolated before classification) were worse than those with EOG in the first and second studies of the thesis and the low sampling frequency was considered as the main reason for the worse result, another recording system of VOG eye movements was also used in the latter part of this thesis.

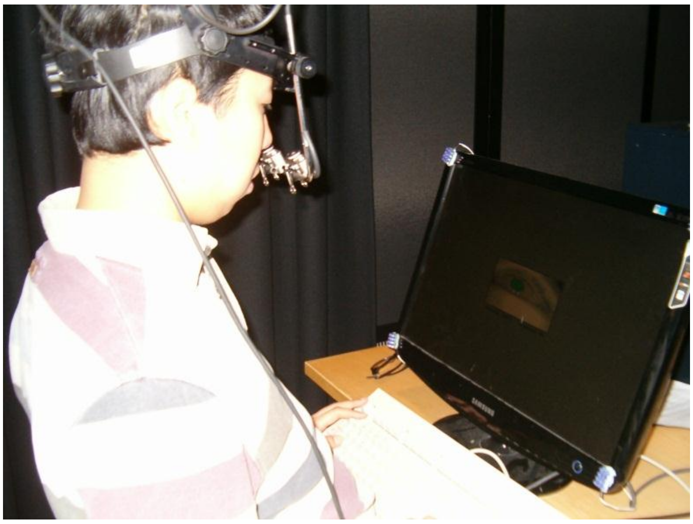


Figure 2.4. A subject performing calibration in the VOG system called EyeLink® (SMI, Berlin, Germany), the sampling frequency of which is 250 Hz; the maximum angle is ±30 °in the horizontal direction and accuracy is 0.1 °for pupil locations.

For the other VOG system (Figure 2.4), the technique of eye movement cameras (based on pupil measurements and computing differences in pupil positions in successive video images), the way to use the video cameras (attached to the headband and one camera for each eye) and the measurement procedure (light dot and following its seemingly random jumping) are quite similar to the preceding VOG system. However, there were certain differences:

1. The 250 Hz sampling frequency was still not as high as that of EOG signals, but it was found to be enough for user verification and, finally, interpolation was then found to be unnecessary.
2. The purpose of the stimulation series employed and subject verification with saccade eye movements in this thesis was to simulate whether this technique could be used in PC computers in future. However, the width of the LED bar in the preceding system is too wide to display the same stimulation on a computer screen. In order to induce large enough angles, the distance of a subject’s eyes to the screen of the computer was a constant 45 cm, closer than before.
3. Before starting measurements, every subject performed a calibration by alternately looking at eight points on the computer screen (the centre, the middle of each of the four sides, and each of the four corners) to calibrate fixations.
4. The duration of a measurement series was still about 60 seconds, but the average intervals of stimulations was reduced from approximately 2 seconds to 1 second, which means that the number of stimulation movements (and saccades) increased from 30 to 54. After excluding small stimulation angles used, clearly less than 51 º, 30 stimulations remained for the verification task. Although the efficiency of data collecting improved a great deal, the measurement series was still repeated a few times for each subject. When larger data sets of each subject were now obtained compared to the preceding arrangement, more reliable and convincing accuracy results were expected.

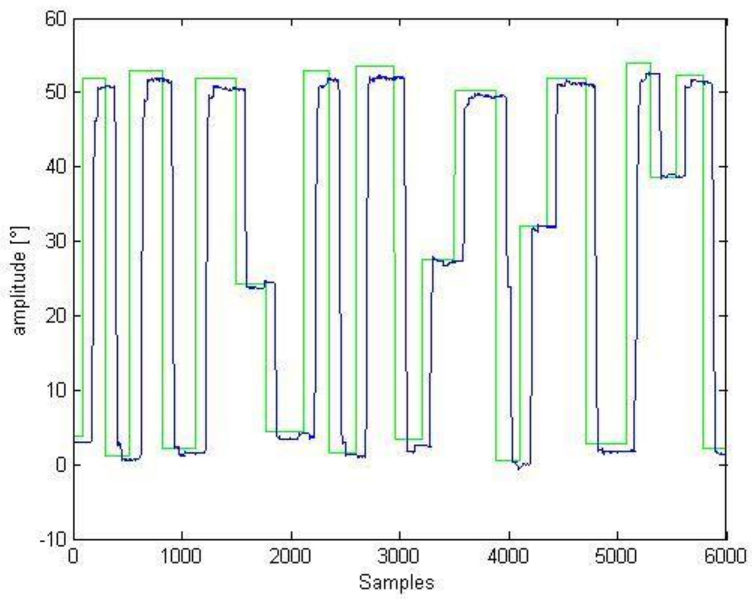


Figure 2.5. A 24-second segment of 250 Hz VOG stimulation and saccades.

The only difference (in addition to the data) between Figure 2.5 and Figure 2.3 is the values on the vertical axis, because the initial left point on the screen was set to be

0 in this system. No negative influence for variable computation was observed, since all variables related to amplitudes were calculated based on difference values, not absolute values.

2.2 Signal processing

An eye movement signal is a simple biometric one-dimensional time series, and only large-angle saccades were used for user verification so that all methods of signal processing were only applied in time domain. This was needed since the saccade variables applied were dependent on time and thus, in a way, functions of time. Furthermore, small differences between individuals' saccades might not appear between them if such as power spectra of saccades computed with fast Fourier transform were compared. However, this could be an opportunity for study in future. The use of time domain variables was, however, supported by the fact that, in medicine and psychology, these time domain variables are known to be affected by disease and age and vary slightly between individuals.

In a way, it was difficult to compute reliable variable values from the VOG signals from the first studies, because the sampling frequency was 30 Hz and there were only six-seven samples in one large-amplitude saccade. Therefore, the low sampling frequency had to be increased in an artificial way. Interpolation [61] is a method of constructing new data points within a range of a discrete set of known data points. The simplest interpolation technique is nearest-neighbour interpolation, but it would 'destroy' the saccade (eliminating features in Figure 2.6) and make it look like a stimulation movement. Thus, it could not be used for eye movement signals. Linear interpolation was another simple choice, but it was too straightforward and was not well-suited for non-linear signals. For the curvature of saccades, it was best to also test the other alternatives, which were a cubic spline curve and piecewise cubic Hermite interpolating polynomial (PCHIP) [62].

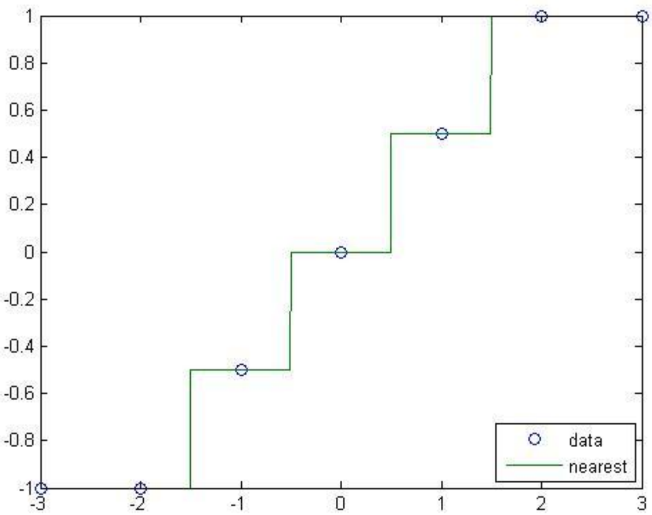


Figure 2.6. Nearest-neighbour interpolation corrupts a simulated eye movement signal.

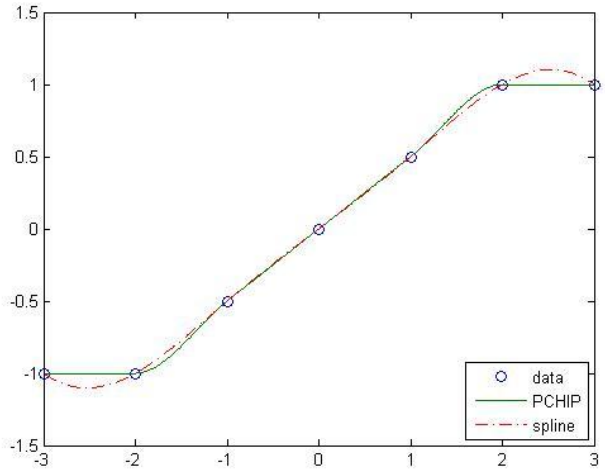


Figure 2.7. Comparison of spline and PCHIP interpolations used in a simulated eye movement.

The waveforms of spline and PCHIP are quite similar, since spline interpolation produced almost the same results as PCHIP did. However, spline chose the slopes differently and the second derivative of the interpolating function is continuous (that of PCHIP is probably not). Therefore, spline produced a smoother result and, if the data consisted of values of a smooth function, spline might bring a more accurate result. On the other hand, PCHIP had no overshoots and less oscillation if the data were not smooth. Figure 2.7 and one experiment in a publication (Chapter 4) show that the spline interpolation effect is close to that of PCHIP.

Interpolation also benefits signal recordings, however. In recording 250 Hz VOG signals, some rare missing samples and erroneous samples (about 0.05% of all) occurred in saccades. Interpolation to estimate new samples from the two sides' samples of a beginning and end was used to insert them or to replace erroneous samples.

Interpolation is a good way to increase frequency and benefits the computation of variables, but it also brings some negative effect. For example, it would enhance the velocity values and then increase the maximum velocity [39,63]. The maximum velocity of a VOG saccade from a healthy person is approximately 700-800 °/s for large amplitudes above 50 °, but it might reach 1000 °/s after interpolation used. The more the frequency of interpolation is increased, the higher the velocity expected. Fortunately, the problem was not very serious for subject verification, because the key point of this is whether variability between subjects is obvious enough to distinguish them, but not whether the values of variables are accurate enough in some medical field and not to assess possible disease.

Besides interpolation, a median filter [42] was used in signal processing: it dampened some high-frequency noise before interpolation and made saccades smoother.

2.3 Saccades and analysis of their variables

2.3.1 Saccade structure and variables

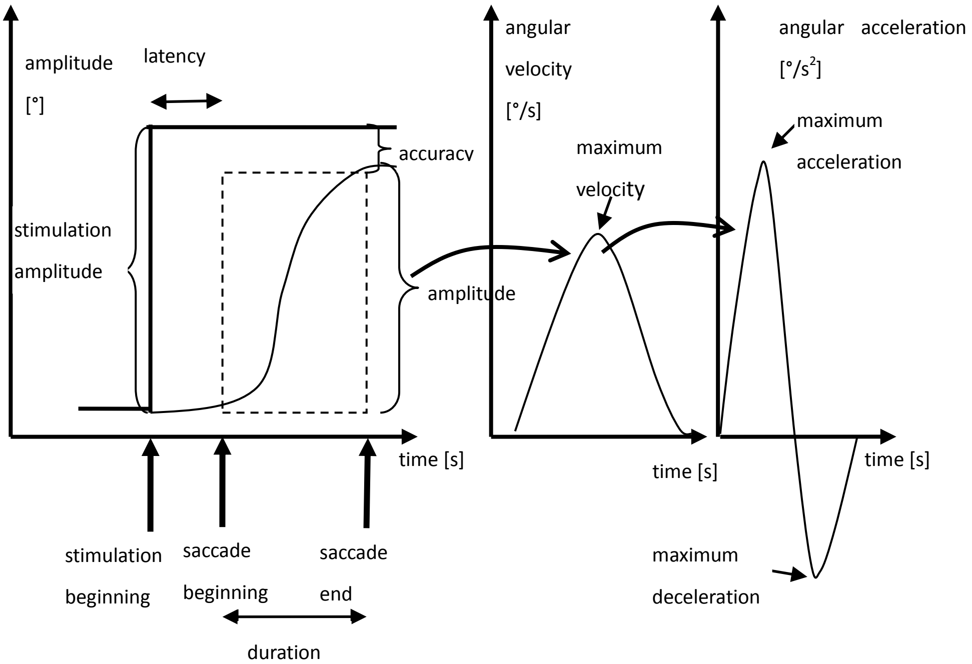


Figure 2.8. A schema for a stimulation angle and corresponding saccade, including the variables used: amplitude, accuracy, latency, maximum angular velocity, maximum angular acceleration, maximum angular deceleration and duration.

From each complete saccade (Figure 2.8), the following variables were selected for the verification task:

- Latency is the time it takes for a subject to react to a stimulation. In other words, it is the temporal difference between the beginning of a stimulation and the saccade. However, in principle, subjects might move their eyes prior to a stimulation, or a latency value might be smaller than a minimum reaction time of a human caused by lack of concentration, fatigue, carelessness or other factors. A latency limit of 0.12s, being physiologically reasonable, was set for saccades to be rejected. In addition, the threshold of a saccade beginning was set to be 10 °/s at first and 50 °/s later. Latency might be one of the most important variables in eye movement verification. Some studies even only used it as the main feature to classify subjects [44]. It also plays a significant role in this thesis.
- Amplitude is the angle that the eye moves. It is a simple variable, but sometimes a saccade is followed by another saccade or even two small partial saccades called corrective saccades because the brain is correcting the direction of gaze, since some subjects occasionally cannot catch the stimulation, particularly those with large amplitudes, in one step. Corrective saccades are rather infrequent occurrences, and the largest (first) saccade or angle would be selected to

determine an amplitude value in such a case.

- Maximum or peak velocity in a saccade velocity curve is physiologically interesting, since there may be clear differences between, for example, healthy people and patients with vertiginous diseases [34,56]. The velocity curve is equal to the first derivative of the corresponding saccade. (Of course, it is approximate since digital signals are discrete.) Increasing the sampling frequency of a saccade signal, e.g. with interpolation, would increase the maximum velocity somewhat; it also would be affected by other methodological factors such as recording methods and devices, analogue and digital filters, and especially their cut-off frequencies and sampling frequencies.
- Accuracy is the difference of the angles between the ends of a stimulation movement and its response, the saccade. Its computation is not simply a matter of using the absolute values of the differences of two angles. Depending on the directions of stimulation and eye movements, an accuracy value was defined to be positive if a saccade amplitude exceeded the angle of its stimulation; otherwise, it was negative. Negative values occurred more frequently than positive in practice.
- Maximum acceleration and deceleration are the greatest changes in velocity increasing and decreasing. Thus, these are given from the second derivative of a saccade in a signal segment. For most saccades, the absolute value of the maximum acceleration is somewhat larger than that of the maximum deceleration, which means the velocity goes up rapidly and drops slightly less rapidly again. Therefore, the velocity threshold of saccade ends was set to be 30 °/s, smaller than that of 50 °/s mostly used for beginnings.
- Duration means how long it takes to complete a saccade. In other words, it is the difference between a saccade's beginning and end in terms of time. Since the classification effect of duration did not seem to be very good on the basis of variable analysis and statistics, duration was only used in a few publications (in Chapter 4) in this research.
- Besides the above variables, others could be selected for verification in future, especially with eye movement camera systems of higher sampling frequencies. For example, mean velocity during a saccade and the time of maximum velocity from the beginning of a saccade are worth studying. However, mean velocity comes from the ratio of amplitude to duration, so it would be unnecessary if amplitude and duration are used at same time.

Since a saccade is a one-dimensional signal, all computation of variables was based on difference values of positions of samples. Basically, all other variables depended more or less on amplitude, for example, the greater the amplitude, the greater maximum velocity. Variable values of EOG saccades were obtained from earlier research [34,56], but this thesis only presents how to calculate variable values for VOG signals. The first method used to compute velocity (approximation of the first derivative) was the formula for a two-point central difference differentiator [39,40,63]:

$$\mathrm{v}(l) = \frac{x(l+m) - x(l-m)}{2Tm}, l = m+1, \ldots, N-m$$

where x is an eye movement signal, v is a velocity signal to be computed, $2m$ is a window length, time interval T is equal to $1/f$ (f is sampling frequency) and N is the number of signal samples.

The other method used for velocity computation was the slope of linear regression [41]**:**

$$\mathbf{y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}$$

where **y**, **X**, **ß** and **ε** are the corresponding matrix or vectors of response, regressor, slope and error.

Ordinary least squares (OLS) or linear least squares [64] is a simple and common method of estimating the unknown parameter, **ß** in this case.

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

The two-point central difference differentiator and linear regression look like a low-pass filter. The longer the window length is, the stronger the effect of frequency cutoff or the lower the cutoff frequency. Since the maximum velocity (and the velocity of one sample) is influenced by computational methods [40], their values are not exactly the same. Based on classification results, the former method was applied more frequently than the latter. Nevertheless, differences between their results were slight.

In addition, the variable values of a subject varied depending on measurements of different recording devices [65].

2.3.2 Analysis of variable values

In order to predict the separation capability of six variables (features), a ratio of inter-individual to intra-individual variability was used for evaluating the usefulness of variables. The formula [66] was based on the calculation of mean and standard deviations

$$r_j = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\bar{u}_{ij} - \bar{a}_j\right)^2}}{\frac{1}{n}\sum_{i=1}^{n}\sqrt{\frac{1}{p_i}\sum_{k=1}^{p_i}\left(u_{kj} - \bar{u}_{ij}\right)^2}}$$

where j denotes a variable, n is equal to the number of subjects, $\bar{u}_{ij}$ is equal to the mean of variable j of subject i, $\bar{a}_j$ is the mean of variable j for all subjects, u_{kj} the value of variable j of saccade k for subject i and p_i is the number of the saccades for subject

i.

In this formula, the word inter-individual means the difference between saccades of individual subjects. Thus, a relatively high value would be good for verification of subjects. On the contrary, intra-individual means the variability within one subject and a relatively lower value is expected. Therefore, a ratio higher than 1 of a variable may indicate that it is capable of distinguishing subjects.

At a later stage of this thesis, the formula was modified as follows:

$$r_j = \frac{\sqrt{\frac{1}{\sum_{i=1}^{n} p_i} \sum_{i=1}^{n} \sum_{k=1}^{p_i} \left(u_{kj} - \bar{a}_j\right)^2}}{\frac{1}{n} \sum_{i=1}^{n} \sqrt{\frac{1}{p_i} \sum_{k=1}^{p_i} \left(u_{kj} - \bar{u}_{ij}\right)^2}}$$

The difference was in inter-individual computation, which was calculated with the standard deviation of means of subjects in the former and with the standard deviation of all saccades in the latter. The latter method was seen to be more reasonable of use standard deviation.

2.4 Data reduction and visualization

In biometric verification, many variables are usually extracted: for example 15 distances and 6 amplitude variables in one ECG study [67]. However, some variables are relational or dependent on each other, which means that no one variable could provide independent information for verification. Moreover, increasing the number of variables (data dimensions) would perhaps improve verification efficiency, although some problems are also caused such as more complicated analysis and slower computation. However, above a certain point of accuracy, additional variables do not improve the result any more or even have a negative influence on verification. Therefore, one approach to resolving the problem is to reduce the excessive dimensionality by combining variables. For this, Principal Component Analysis (PCA) [68] and Multiple Discriminant Analysis (MDA) [68] are well known methods which project high dimensional data onto a lower dimensional space.

2.4.1 Principal component analysis

Principal component analysis (PCA) finds a projection which can best represent data according to the least-square error. Firstly, in a set of n d-dimensional samples, it is not difficult to prove that the minimum squared error criterion function can be resolved by the mean $\mathbf{m}$ of samples. However, the mean is too simple to represent the data. A more interesting approach is to use a one-dimensional representation by projecting the data onto a line. The line can be written as

$$\mathbf{x}=\mathbf{m}+a\mathbf{e}$$

where **e** is the direction of the line and a corresponds to the distance of any sample **x** from the mean **m**.

Sample $\mathbf{x}_k$ can be represented by $\mathbf{m}+a_k\mathbf{e}$ and minimum $\boldsymbol{J}$ can be found by the optimal set of coefficients a_k

$$\boldsymbol{J}(a_1, a_2, \dots, a_n, \mathbf{e}) = \sum_{k=1}^{n} \|(\mathbf{m} + a_k\mathbf{e}) - \mathbf{x}_k\|^2$$

$$= \sum_{k=1}^{n} a_k^2\|\mathbf{e}\|^2 - 2\sum_{k=1}^{n} a_k\mathbf{e}^t(\mathbf{x}_k - \mathbf{m}) + \sum_{k=1}^{n} \|\mathbf{x}_k - \mathbf{m}\|^2$$

Since (Euclidean norm) $\|\mathbf{e}\| = 1$, setting the derivative equal to zero to obtain the minimum, we obtain

$$a_k = \mathbf{e}^t(\mathbf{x}_k - \mathbf{m})$$

That is to say, the minimum solution can be obtained geometrically as long as one projects $\mathbf{x}_k$ onto the line of direction **e** (equals a_k), which passes the mean of the samples (Figure 2.9).

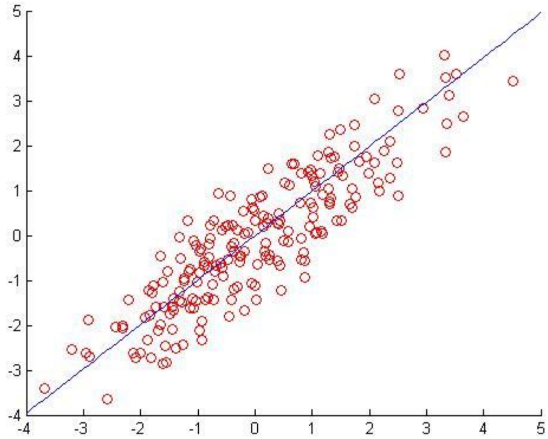


Figure 2.9. The samples (red circles) are projected onto the optimal line which best represents them.

In order to find the optimal direction **e** of the projected line, a scatter matrix **S** should be defined as

$$\mathbf{S} = \sum_{k=1}^{n} (\mathbf{x}_k - \mathbf{m})(\mathbf{x}_k - \mathbf{m})^t$$

The squared error criterion function could be written using a_k

$$\boldsymbol{J}(\mathbf{e}) = -\sum_{k=1}^{n} \mathbf{e}^t(\mathbf{x}_k - \mathbf{m})(\mathbf{x}_k - \mathbf{m})^t\mathbf{e} + \sum_{k=1}^{n} \|\mathbf{x}_k - \mathbf{m}\|^2 = -\mathbf{e}^t\mathbf{S}\mathbf{e} + \sum_{k=1}^{n} \|\mathbf{x}_k - \mathbf{m}\|^2$$

It is clear that the maximum of $\mathbf{e}^t\mathbf{S}\mathbf{e}$ would lead to the minimum $\boldsymbol{J}$. Using the method of Lagrange multipliers [69] to maximize $\mathbf{e}^t\mathbf{S}\mathbf{e}$ (λ as a multiplier) and derivative of **e** to make the gradient equal to zero, we can see that **e** must be an eigenvector of scatter matrix **S**.

$$\mathbf{S}\mathbf{e}=\lambda\mathbf{e}$$

Since $\mathbf{e}^t\mathbf{S}\mathbf{e} = \lambda\mathbf{e}^t\mathbf{e} = \lambda$, the conclusion is that the eigenvector corresponding to the greatest eigenvalue of the scatter matrix is the best direction of the line whose projection has the best least-square error.

According to the result, the formulas can be extended from the one-dimensional projection to d'-dimensional ($d' < d$) projection by original d-dimensional data.

$$\mathbf{x} = \mathbf{m} + \sum_{i=1}^{d'} a_i\mathbf{e}_i$$

$$\boldsymbol{J} = \sum_{k=1}^{n} \left\| \left(\mathbf{m} + \sum_{i=1}^{d'} a_{ki}\mathbf{e}_i \right) - \mathbf{x}_k \right\|^2$$

The minimum criterion function can be obtained when the vectors $\mathbf{e}_1,\ldots,\mathbf{e}_{d'}$ are the d' eigenvectors corresponding to the greatest eigenvalues of the scatter matrix.

2.4.2 Fisher linear discriminant and multiple discriminant analysis

PCA is a good approach to finding the components to represent data, but it might have no effect on classification discriminant. In other words, PCA finds a direction to represent data and the projecting line of best distinguishing data is, in another direction, used in discriminant analysis (Figure 2.10).

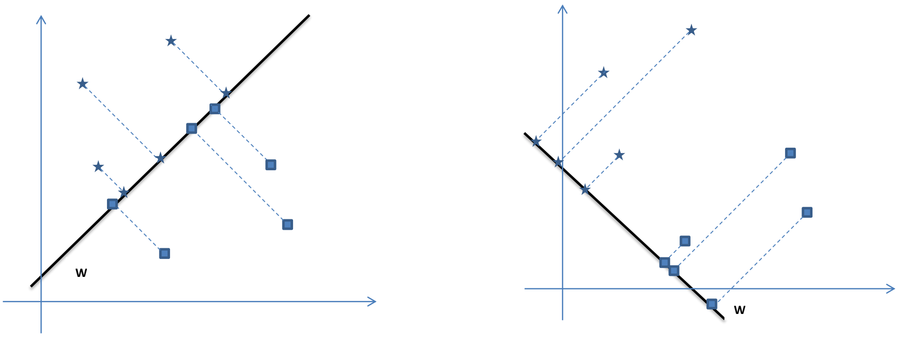


Figure 2.10. The same data set of two classes projects two lines in different directions **w**. The projected samples are mixed in the figure on the left and well separated in the figure on the right.

Let suppose that there are a set of n d-dimensional samples $\mathbf{x}_1,\ldots,\mathbf{x}_n$ that belong to two classes, C_1 and C_2, and these samples are projected onto the line as

$$y=\mathbf{w}^t\mathbf{x}$$

where **w** is the direction of the line.

It is simple to obtain the mean $\vec{m}_i$ of projection data from class i

$$\mathbf{m}_i = \frac{1}{n_i}\sum_{\mathbf{x}\in C_i}\mathbf{x}$$

$$\vec{m}_i = \frac{1}{n_i}\sum_{y\in C_i} y = \frac{1}{n_i}\sum_{\mathbf{x}\in C_i}\mathbf{w}^t\mathbf{x} = \mathbf{w}^t\mathbf{m}_i$$

where n_i and $\mathbf{m}_i$ are the number and mean of original data corresponding to the class.

The distance between the means of the projected samples of two classes is

$$|\vec{m}_1 - \vec{m}_2| = |\mathbf{w}^t(\mathbf{m}_1 - \mathbf{m}_2)|$$

Besides between-class scatter, for a good efficiency in distinguishing classes, the within-class scatter based on standard deviation should also be calculated:

$$\vec{s}_i^{\,2} = \sum_{y\in C_i}(y - \vec{m}_i)^2$$

where $\vec{s}_1^{\,2} + \vec{s}_2^{\,2}$ is within-class scatter.

The criterion function $\boldsymbol{J}$ of the Fisher linear discriminant [70] is

$$\boldsymbol{J} = \frac{|\vec{m}_1 - \vec{m}_2|^2}{\vec{s}_1^2 + \vec{s}_2^2} = \frac{|\mathbf{w}^t(\mathbf{m}_1 - \mathbf{m}_2)|^2}{\vec{s}_1^2 + \vec{s}_2^2}$$

Here $\mathbf{w}$ which maximizes $\boldsymbol{J}$ is the best direction that separates the two projected classes.

Similarly to PCA, scatter matrix $\mathbf{S}_i$ is given as

$$\mathbf{S}_i = \sum_{\mathbf{x} \in C_i} (\mathbf{x} - \mathbf{m}_i)\,(\mathbf{x} - \mathbf{m}_i)^t$$

Thus, the above equations can be rewritten as

$$\vec{s}_i^2 = \sum_{\mathbf{x} \in C_i} (\mathbf{w}^t\mathbf{x} - \mathbf{w}^t\mathbf{m}_i)^2 = \mathbf{w}^t\mathbf{S}_i\mathbf{w}$$

$$\vec{s}_1^2 + \vec{s}_2^2 = \mathbf{w}^t\mathbf{S}_W\mathbf{w}$$

where $\mathbf{S}_W = \mathbf{S}_1 + \mathbf{S}_2$.

$$|\vec{m}_1 - \vec{m}_2|^2 = |\mathbf{w}^t(\mathbf{m}_1 - \mathbf{m}_2)|^2 = \mathbf{w}^t\mathbf{S}_B\mathbf{w}$$

where $\mathbf{S}_B = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^t$. Therefore, the criterion function $\boldsymbol{J}$ can be rewritten as

$$\boldsymbol{J} = \frac{\mathbf{w}^t\mathbf{S}_B\mathbf{w}}{\mathbf{w}^t\mathbf{S}_W\mathbf{w}}$$

It is well known that maximum $\boldsymbol{J}$ satisfies

$$\mathbf{S}_W^{-1}\mathbf{S}_B\mathbf{w} = \lambda\mathbf{w}$$

So $\mathbf{w}$ equals the eigenvector of $\mathbf{S}_W^{-1}\mathbf{S}_B$.

However, since $\mathbf{S}_B\boldsymbol{w}$ is always in the direction of $\mathbf{m}_1 - \mathbf{m}_2$, it is not necessary to compute the eigenvalue and eigenvector. Thus, it can be shown that the maximum $\boldsymbol{J}$ occurs when

$$\mathbf{w} = \mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2)$$

In fact, when the samples from two classes are satisfied as normally distributed with the equal covariance, the best solution is written as

$$\mathbf{w} = \mathbf{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

where $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ are the mean and covariance of the samples and the equation is also equivalent to linear discriminant analysis (LDA) [71].

Multiple discriminant analysis (MDA) extends from projecting data of two classes onto one-dimensional space to projecting c class data onto at most (c-1)-dimensional space, which applies the same theory as discriminant analysis. (The details are not presented in this thesis.) The columns of the optimal $\mathbf{W}$ (the matrix of $\mathbf{w}_i$, d by (c-1)) maximizing the criterion function are still the eigenvectors of $\mathbf{S}_W^{-1}\mathbf{S}_B$ corresponding to the largest eigenvalues

$$\mathbf{S}_W^{-1}\mathbf{S}_B\mathbf{w}_i = \lambda_i\mathbf{w}_i$$

Neither PCA nor MDA are classification methods and merely benefit the separation efficiency by reducing or combining excessive variables at the pre-processing stage. However, since at most seven variables only were extracted in this thesis, the main purpose of applying them was data analysis and visualization (in Section 3.2.4).

Chapter 3

Verification procedure and methods

3.1 Identification and verification

In biometric classification field, there are two types of tasks: identification and verification. The former is defined as distinguishing a particular individual from among a group of subjects. The latter is a simpler task, which is defined as recognition of a real subject and determining that other possible subjects are non-users or impostors. If identification is like the question of 'What's your name?', verification is the question 'Are you XXX?' whose answer is 'Yes' or 'No'. In other words, identification is *n*-class classification and verification is binary or two-class classification.

Identification is certainly more complicated than verification. It has been used in many areas in society such as criminal identification and access control systems (multi-user). However, an advanced method almost always means more complicated and more time-consuming computation. Actually, identification is not necessary for the situation of only one authenticated user in daily life: for example, a private computer or other devices. Moreover, when people log in to any account or system (online or offline), the password is a kind of verification after inputting one's username. Compared with identification, verification is a simpler, obviously faster and easier computational task. The use of these terms sometimes varies. Some biometric research into what was called identification was actually verification [36,53]. The classification task of the present thesis is verification.

Table 3.1. Classification accuracy for identification and verification in case of *n* subjects.

Accuracy	*Identification*	*Verification*
95%	Excellent	Excellent
50%	Fair (if *n* is large)	No effect (random guess)
5%	Poor	The same as 95% if its opposite option is taken

The meaning of classification accuracy is different in identification and verification (Table 3.1). In identification, the higher the accuracy value, the better efficiency in classification. However, if an accuracy value is close to 50% (no matter whether it is higher or lower) in verification, the classification is poorer. On the other hand, if the value is closer to 100%, the result is much better. Since verification is like two-class classification, choosing the opposite class could have the same effect as what is subtracted from 100%: if the accuracy is less than 50%, for example 5%, then it equals 95%. That is why it is impossible for the average or whole accuracy of verification to be lower than 50% and all points in a receiver operating characteristic (ROC) curve are above the 50% line (Section 3.2).

3.2 Statistic information and its application in biometric verification

3.2.1 Errors of two types

In statistics theory, there are two error types [72]: Type I errors and Type II errors. The former is the incorrect rejection of a true null hypothesis, and the latter is the failure to reject a false null hypothesis (Table 3.2). For example, if the null hypothesis is pregnant, a pregnant lady classified as non-pregnant is a Type I error. In contrast, testing a non-pregnant lady as pregnant is a Type II error.

Table 3.2. Type I and II errors.

	Null hypothesis is true	Null hypothesis is false
Reject null hypothesis	Type I error	Correct
Fail to reject null hypothesis	Correct	Type II error

Type I errors are also called false positives (FPs) or false acceptances and the corresponding correct outcomes are true negatives (TNs). Another name for a Type II error is a false negative (FN) or false rejection, and the corresponding opposite is a true positive (TP). However, the variability of the above terms is merely in their names, not their inherent meaning, connection or computation. The corresponding meanings in the eye movement verification of the present thesis are in Table 3.3.

Table 3.3. FP, TN, FN and TP meanings.

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Tested saccade comes from</td></tr>
<tr><td>An authenticated user</td><td>An impostor</td></tr>
<tr><td rowspan="2">Classified as</td><td>Authenticated user</td><td>True positive</td><td>False positive</td></tr>
<tr><td>Impostor</td><td>False negative</td><td>True negative</td></tr>
</table>

Some relative accuracy rates can also be calculated as follows.
True positive rate (TPR):

$$TPR = \frac{TP}{P} \times 100\% = \frac{TP}{TP + FN} \times 100\%$$

True negative rate (TNR):

$$TNR = \frac{TN}{N} \times 100\% = \frac{TN}{TN + FP} \times 100\%$$

False positive rate (FPR):

$$FPR = \frac{FP}{N} \times 100\% = 100\% - TNR$$

False negative rate (FNR):

$$FNR = \frac{FN}{P} \times 100\% = 100\% - TPR$$

Total accuracy (ACC):

$$ACC = \frac{TP + TN}{P + N} \times 100\%$$

where *P*, *N*, *TP*, *FN*, *TN* and *FP* are the corresponding numbers of positive (number of saccades of an authenticated subject), negative (number of saccades of impostors), true positive, false negative, true negative and false positive cases.

In fact, the above terms are variously named within biometrics, machine learning, statistics or other fields, but they are the same ones indeed. For example, in the publications of this thesis, the term ‘accuracy’ was used to mean correctly classified subjects either being the authenticated user or being impostors. These test options were designated Conditions 1 and 2, which were used because the two tests were run separately, not as usual, because of the small data sets at the beginning of the research. This approach enabled the use of the scarce data more extensively for tests than the conventional way of running them at the same time. They could also have been called true positive or true negative rates, or the terms in the following section could have been used. Notwithstanding the terms used, the aim of binary classification is always the same as in verification: to recognize who is the authenticated user of a device or authenticated subject in general and who is not. 'Accuracy' was preferred in the thesis, since the interest was not only aimed at using saccades for biometric purposes, but also how to perform the classification task needed as efficiently as possible.

3.2.2 Equal error rate

In binary classification, a group value of the false positive rate – which could also be called the false acceptance rate (FAR) – and false negative rate or the false rejection rate (FRR) can be obtained according to a specific threshold. With this threshold changed, the group values also vary. Usually speaking, when one error rate increases, the other comes down (Figure 3.1 [26]): more specifically, if FRR(*t*) is an increasing function and FAR(*t*) is a decreasing function with *t* growing. Threshold *t* can have a wide scale, which can be presented in many ways. One error rate only cannot represent good accuracy, because the other rate might be high. When the error rates are equal, i.e. in that specific situation, the common value is referred to as the equal error rate (EER). The value indicates that the proportion of false acceptances is equal to the proportion of false rejections. The lower the equal error rate value, the higher the accuracy of the classification system.

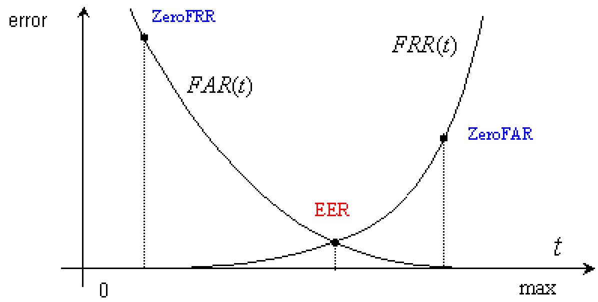


Figure 3.1. EER denotes the system error when FRR=FAR. ZeroFAR denotes FRR when FAR=0,f and ZeroFRR denotes FAR when FRR=0.

Generally, FPR and FNR depend on the threshold *t*. The most common *t* is the ratio of samples from the authenticated subject and non-user subjects in a training set. If the ratio of authenticated samples is higher, a tested sample is easily classified as authenticated user. It is certain that FPR is high and FNR is low. The threshold *t* can be defined 'large' in that case. In the opposite case, if there are more samples of non-user subjects in a training set, it is highly possible to judge the tested sample to be that of an impostor. Low FPR and high FNR are the result when the threshold is small. In addition, other factors also can explain and represent threshold *t*, such as classification method parameters, risk indexes and threshold in judgment.

In biometric verification research, equal error rate is often calculated for convincing accuracy. However, an error of one type is frequently more serious than an error of the other in practice or for an application. For example, most people may believe that the error of authenticated user rejection (FNR) is more tolerable than accepting an impostor (FPR) in an access control system. That is to say, the threshold setting is increased to make access more difficult for impostors, although some authorized people may find it is also more difficult to gain access. Authenticated users can try to access again if they are rejected, but it could cause danger or losses if an impostor enters. Therefore, they would prefer System 1 to System 2 (Table 3.4), whose EERs are both 15%. It is certain that few people have the opposite opinion (that authenticated user rejection is worse than an impostor granted access). In any case, the key point is setting the threshold to various desired security levels depending on different situations.

Table 3.4. Comparison of two access control systems.

	System 1	*System 2*
FPR	10%	20%
FNR	20%	10%
EER	15%	

3.2.3 Receiver operating characteristic

A receiver operating characteristic (ROC) curve [73] is another way to present the accuracy of biometric classification based on two types of errors.

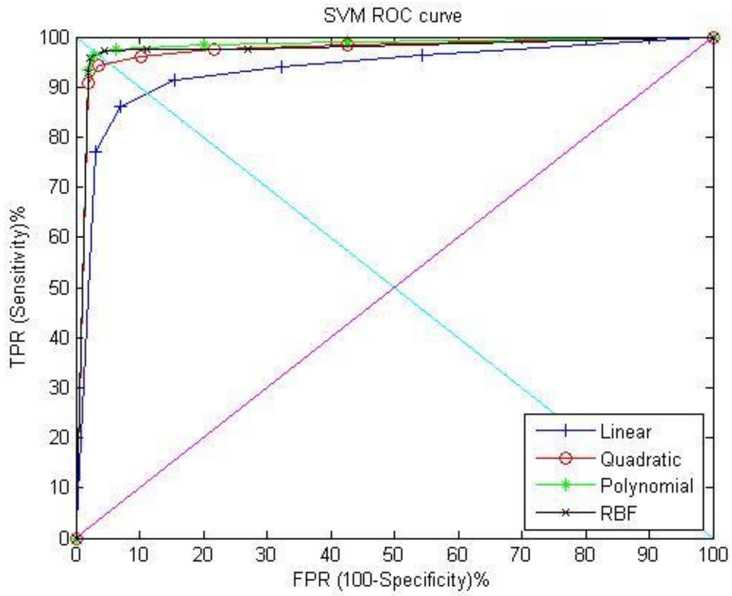


Figure 3.2. Exemplar ROC curves concerning classification of support vector machine (SVM) with four kernel functions (RBF for radial basis function). The horizontal axis of the ROC corresponds to FPR and vertical to TPR. The terms 'sensitivity' and 'specificity' (true negative rate) are sometimes also used. The line of (0,0) to (100,100) is the 50% line and (0,100) to (100,0) is the EER line.

Each ROC curve in Figure 3.2 represents an individual classification method or individual parameter of a method which consists of some points. Each point is the accuracy (a group of two errors) of the corresponding methods or parameters based on one threshold value. The closer the point is to the top left, the higher the accuracy is. In contrast, the accuracy is low if the point is near the 50% line. The intersection point of each curve and the EER line is the corresponding EER point. Although the EER point might not be the point closest to the top left of the curve, it is still considered the best accuracy for the corresponding method or parameter.

There is another way to calculate accuracy: the area under the curve (AUC). The larger the AUC, the better the accuracy. Compared with the EER from an individual threshold or situation, it is a kind of average accuracy of a classification method or parameter and presents efficiency from a macroscopic viewpoint.

3.2.4 Two test conditions

According to the theory of two types of errors, two test conditions were designed in the user verification based on saccades. The purpose of Condition 1 is checking whether the system can recognize an authenticated or right user. In other words, Condition 1 calculated the true positive rate (if succeeded) and false negative rate (if failed).

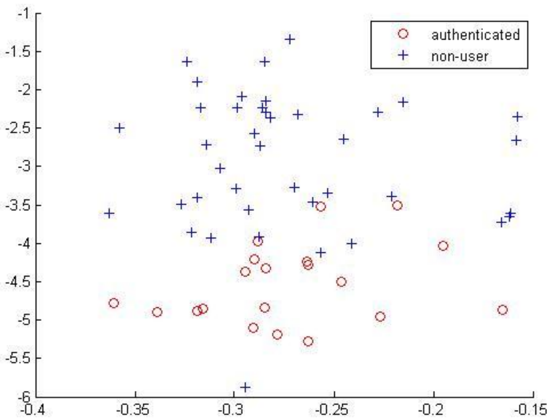


Figure 3.3. Condition 1 as presented after data dimension reduction. The original six-dimensional data (latency, amplitude, accuracy, maximum velocity, maximum acceleration and deceleration) was reduced to two dimensions by MDA. It shows how the majority of the saccades of the authenticated user were separated from those of non-user subjects (blue cross). Thus it is probable that, in this two-dimensional space, most saccades of the authenticated user here would be classified as coming from the authenticated subject (red circle).

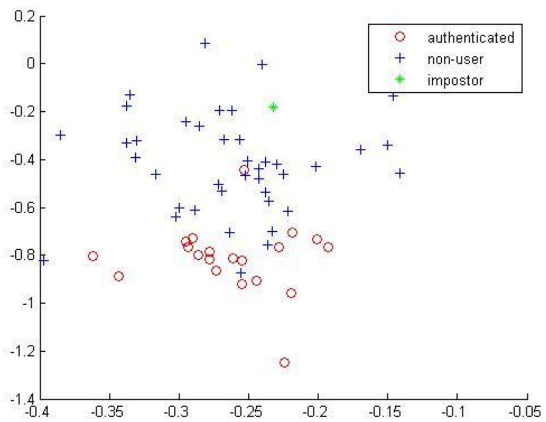


Figure 3.4. For Condition 2, the impostor saccade (green star) was far away from the saccades of the authenticated user and succeeded in being classified as a non-user subject. The method of dimension reduction was the same as for Condition 1 by MDA in the preceding figure.

By contrast, there are three types of subjects in a set in Condition 2: authenticated users, non-users and impostors. An impostor's saccade should be rejected after comparing it with those of the authenticated subject and classified as a non-user subject, although it comes from the third subject. It is usually probable that it resembles more saccades of non-user subjects than those of the authenticated when the former are more dispersed in the space, because they originate from several non-user subjects. Condition 2 computes the true negative rate and false positive rate.

Note that verification is not based on a single saccade but on a set of test saccades. As a result, even though a few of them resulted in false classifications, if a majority correspondingly resulted in correct decisions, this would be promising for classification.

3.3 Verification procedure

Firstly, two statistics terms used in machine learning are presented: *k*-fold cross-validation [74] and leave-one-out [74]. In *k*-fold cross-validation, the original data set is randomly partitioned into *k* subsets of equal size or as equal as possible. Of the *k* subsets, a single subset is retained as the validation data to test a model, and the remaining ($k-1$) subsets are used as training data. The cross-validation process is then repeated *k* times (the folds), with each of the *k* subsets used exactly once as validation data. The *k* results from the folds can then be averaged (or otherwise combined) to produce a single estimate. The advantage of this method over repeated random sub-sampling is that all subsets are used for both training and validation, and each sample (or case or instance) is used for validation exactly once. Ten-fold cross-validation is commonly used, but in general $k>1$ remains an unfixed integer parameter.

Leave-one-out is a special case of *k*-fold cross-validation which makes *k* equal to the number of samples in all the data, i.e. only a single sample from the data is used for testing and the remaining n-1 for training. The procedure is repeated *n* times so that each sample is used once as the validation data. The size of each training set is maximized, which makes, in a way, the best starting point for training a model.

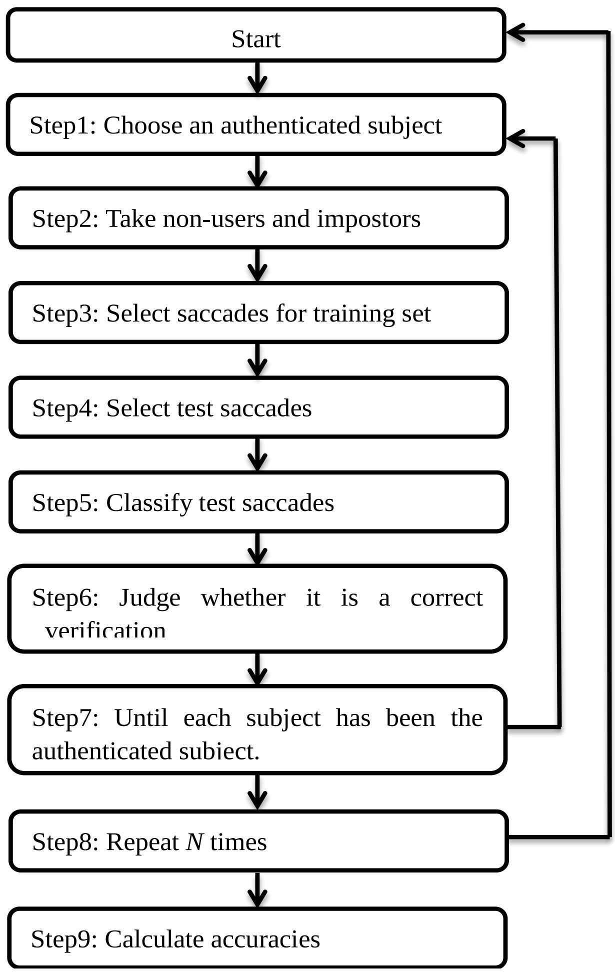


Figure 3.5. The schema of the biometric verification procedure.

For Step 1, choose a subject to be the authenticated one.

For Step 2, take the rest n-1 subjects as non-user subjects in a test of Condition 1; take a part of the remaining subjects as non-users and the others as impostors in a test of Condition 2. The ratio of the numbers of non-user subjects and impostors in Condition 2 depended on the number of all subjects and some other factors. It was usually subjects randomly selected half-and-half.

For Step 3, select some saccades of the authenticated user and non-user subjects for a training set. The selection of saccades of the authenticated subject or user depended on various cases: for example, one saccade was taken to be tested and the rest were picked for a training set (leave-one-out) if the number of saccades of each subject was small, or selected by various measurements series or sessions (k-fold cross-validation) when the number of saccades was large. For example, saccades from one series or session of measurements were put into a test set and the rest of the sessions into the corresponding training set. The ratio of the saccades of the authenticated user to those of non-user subjects in a training set was also different depending on various situations. The saccades of non-user subjects were randomly taken and on average from every non-user subject, so it also affected how subjects were divided into non-user and impostor subjects in Condition 2. In fact, in order to keep the same ratio and above distribution in selection, sometimes only part of the subjects were used as non-users and impostors and a few others were abandoned.

For Step 4, select the rest of the saccades of the authenticated user for the test set in Condition 1, and pick some saccades of impostors for the test set in Condition 2. The selection of the saccades of impostors also depended on various cases: sometimes one impostor was chosen randomly and sometimes it was an average picked from all the impostors.

For Step 5, classify all test saccades using a classification method.

For Step 6, sum up the numbers of correct and incorrect test results and judge whether it is a correct verification or not. If a tested saccade was classified as being from the authenticated user, it was a correct result. Otherwise, it was incorrect for Condition 1. By contrast, the test was correct if it was classified as an impostor for Condition 2. The judgment of correct verification was obtained if the number of correct test results was larger than or equal to 50% of all tests. However, the threshold of 50% was not used in all situations. For example, the points on the curve in Figure 3.2 came from different thresholds, or the *k*-nearest-neighbour classification (in Publication 1 of Chapter 4) used a different approach.

For Step 7, the authenticated subject was selected one by one from the group of all subjects in a data set until each subject in the data set was selected once to be the authenticated subject.

For Step 8, since the non-user subjects, impostors and saccades from both were selected randomly, N iterations instead of 1 only made the accuracy more reliable and convincing.

For Step 9, according to the numbers of subjects and iterations, the number of all verification tests could be acquired. TPR and TNR were the numbers of correct verification results for Conditions 1 and 2 divided by the total.

In fact, the above procedure was just a basic or general template. As my thesis research developed, more and more saccades and subjects were measured, so the specific verification procedures at different stages varied slightly. For example, training sets of Conditions 1 and 2 are not the same in this template, since subjects were scarce in the early publications. When the number of subjects increased, the training sets for the two conditions became equal.

Chapter 4

Publications, their themes and results

This chapter consists of six articles presenting the development and progress of the research, such as data sets from small to large; the classification methods applied, from simple to complicated; the classification procedure of two test conditions, from the two various at the beginning to the unified form at the end; and the computation of saccade variables.

Publication I included large-amplitude saccades of VOG signals sampled at 30 Hz. In addition, EOG signals sampled at 400 Hz from 30 subjects were obtained, each subject with 12 saccades. The former were measured at the beginning of the present research, but the latter had been measured earlier in connection with other studies. However, their horizontal stimulations and other details were well comparable to those of the new VOG signals. Estimated signals at 1000 Hz from 30 subjects of the VOG signals, each of which also included 12 saccades, were computed by increasing from 30 Hz by interpolation. Verification of subjects based on saccades was explored, and the results of EOG and VOG signals were compared.

In Publication II, the data sets of both EOG and VOG signals were extended to include 40 subjects, but still 12 large amplitude saccades from each subject. In the case of the EOG signals, 19 subjects were healthy young people, and 21 were mostly middle-aged, otoneurological patients (with vertigo and balance problems) from the Ear, Nose and Throat Clinic at the Helsinki University Central Hospital. The various methods of calculating variables were compared. Depending on the ratio of inter-individual to intra-individual variability, variables were evaluated to distinguish individual subjects or separate the healthy from the patients. Verification was studied partly with different methods from those in Publication I.

Publication III abandoned the old EOG signals and only focused on VOG signals. It dealt with the effect of a priori probability for verification, and the optimal ratio of saccades of the authenticated subject to those of non-users was found. A so-called balance ratio by copying cases of the minority class to reach approximately equal sizes between two classes was used to pursue the EER. LogDA was also applied in this publication.

Publication IV described how to apply advanced classification methods for eye movement verification. It compared the verification performance of multilayer perceptron and radial basis function neural networks and that of support vector machines with various parameter values. Finally, the number of the subjects measured with the VOG system reached 132.

Publication V concentrated on multiple measurement sessions for each of the subjects: in other words, repeated measurement sessions. It was studied whether an authenticated user could be verified by saccades between different sessions. Data sets from two groups were tested. One group of 22 subjects was recorded five times in over two months. The accuracy results obtained were relatively good, but results from the other group with the small data set measured with a very long interval of approximately 16 months were not satisfactory as expected.

Publication VI started using VOG signals with a relatively high sampling frequency (250 Hz) and another eye movement recording system. The efficiency of biometric verification results of these saccade recordings was improved, and each subject produced 120 saccades per session. Condition 1 and Condition 2 used the same training set. The verification accuracy with each method increased, obviously due to the higher sampling frequency than the 30 Hz used in the earlier VOG recordings.

4.1 Publication I: The first tests and comparison of EOG and VOG data verification

This is the first article of the thesis that mainly considers whether saccades could be used for subject or user verification and compares the performance between EOG and VOG signals. Since it is the beginning stage of the research, only eye movements of 30 subjects were recorded as VOG signals, and 30 subjects with EOG signals came from an earlier study. In these data sets, every subject with EOG or VOG data had 12 saccades of large amplitudes used for verification tests.

Based on how many subjects were used, the classification was divided into two parts. In Part I, data from 19 subjects was used. In addition to one subject as an authenticated user, 18 subjects were selected randomly from a remaining population as non-user subjects for tests of Condition 1. For tests of Condition 2, 18 subjects were also taken randomly and divided into half non-users and half impostors. Part II was simpler: all 30 subjects were used, and the ratio of subjects from the subject categories (2 or 3) was 1:29 for tests of Condition 1 and 1:15:14 for tests of Condition 2. It should be noted that the training set arrangements in Conditions 1 and 2 were different regardless of which part, since the number of subjects was so small concerning machine learning techniques in general. In addition, there were only 12 saccades from each subject. Thus, leave-one-out cross-validation was used in training and in tests of Conditions 1 and 2 since it is appropriate for small data sets.

The classification methods applied in this publication were linear discriminant analysis, quadratic discriminant analysis, naïve Bayesian rule, *k*-nearest-neighbour searching (KNN) and *k*-means clustering. The judgment of three previous methods (Step 7 in the classification procedure) was usually a majority vote: if the sum of correct test results was larger than that of incorrect test results, a classification decision was correct; otherwise, it was incorrect. However, Step 7 of two last methods was slightly complicated.

KNN [68] might be the simplest approach in the classification field:

- Calculate the distance between a test sample and all examples in the training set.
- Select *k* examples closest to the test sample.
- Assign the test sample to the most frequent class among its *k* nearest neighbours.

However, it was very sensitive to the ratio of samples of each class in a training set (the ratio effect will be discussed more in Section 4.3). That is to say, the TPR was quite low and the TNR is extremely high in this case. Therefore, instead of majority votes, the KNN judgments were modified to

$$\frac{x}{ka} > \frac{a-1}{a-1+bc}$$

where x is the sum of correctly classified saccades of the correct or authenticated subject in all tests, a is the number saccades of the authenticated subject (12 in this case), b is the number of selected saccades from incorrect or non-user subjects, c is the number of non-users (9 and 15 in Parts I and II) and k is the number nearest neighbours searched for. In other words, for tests of Condition 1, there is a correct verification if the ratio of correct neighbour and all neighbours is larger than the ratio of the samples of the saccades of the authenticated subject or user to all samples in a training set. Otherwise, it is incorrect. It is clear that the opposite applies to tests of Condition 2.

The *k*-means clustering method [68] is usually used in unsupervised classification, and the label of a tested sample cannot then be obtained directly. The judgment of verification was modified in this case as whether or not the number of saccades of the authenticated user is the majority in the cluster (or clusters) in which tested saccades were gathered. If it is, the verification is correct for Condition 1 or incorrect for Condition 2.

Because KNN and *k*-means clustering are methods based on distance calculation, normalization [75], which transforms the values measured from different scales into a common scale (here [0,1] as usual or [0,100]), was also used:

$$\mathrm{y} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

where x is an original variable value, y is the scaled value, and x_{max} and x_{min} are the maximum and minimum of a variable on an original scale.

Moreover, the methods of calculating distance also affect results, so performances of four distance calculation methods – Euclidean, city block, cosine and correlation [76] – in *k*-means clustering were compared.

Besides verification tests with the EOG data, there were two alternatives for the variables applied to the VOG data. One, V4, also used the EOG data variables: amplitude, accuracy, latency and maximum velocity. In addition to the V4 variables, the other alternative, V7, included duration, maximum acceleration and maximum deceleration.

The results obtained showed:

1. Many groups with high average accuracies of tests of Conditions 1 and 2 (for example, 97% for EOG with *k*-means clustering and 92% for EOG with naïve Bayesian) indicated that saccades can be used in subject verification.
2. Although interpolation increases the sampling frequency by estimating, the general performance based on the low frequency of the VOG data was still worse than that produced by the high frequency of the EOG data. For example, the average accuracy in Part II was 94% for EOG by linear discriminant, but 84% for V4 and 83% for V7.
3. Applying normalization with KNN and *k*-means clustering is reasonable, but the effect was minor.
4. The differences between distance measures for *k*-means clustering were not obvious.
5. The additional three variables of V7 compared to V4 benefited verification only slightly.

4.2 Publication II: Computation and comparison of variables

It is known that a classification result may depend greatly on the variables used and their values. Different variable values may be obtained with different techniques to compute them. Thus, methods for selecting variables and computing their values are rather important in signal processing and machine learning. In Chapter 2, two methods (two-point central difference differentiator and slope of linear regression) were described fro approximating derivation of signals. They also acted like low-pass filters. Still, the maximum velocity values of the two-point central differentiator were rather similar to those of the slope of linear regression.

The ratio of inter-individual to intra-individual variability (introduced in Chapter 2 above) is an elementary approach to examining the separation ability of a variable. The result showed that, with the increasing window length of each method, the separating performance (ratio value) of some variables improved, but then decreased. These ratios were smaller than 1.0, but the classification results presented were good for the VOG data. This means that, for these ratios, it is difficult to give some lower bound to predict how they may affect separating performance between classes in classification. The differences between the results from the two methods were slight (Table 4.1). Since the results of two-point central difference differentiator were a little better for the amplitude variable than the slope of linear regression results, the former were used in all later tests in this and later publications.

Table 4.1. Ratio of inter-individual to intra-individual variability for the variables of amplitude and accuracy for the VOG signal saccades of 40 subjects. Signal differentiation (computing velocity and acceleration/deceleration) had been computed using two-point central difference differentiator or slope of linear regression.

Window length (samples)	*Amplitude*		*Accuracy*	
	Two-point	Slope	Two-point	slope
3	0.71	0.71	0.62	0.62
5	0.77	0.73	0.64	0.64
7	0.77	0.73	0.64	0.64
9	0.67	0.71	0.65	0.65

In addition, data from more subjects with EOG and VOG data were used in this publication than in Publication I. There were 40 subjects for each of the two types. For the EOG signals, the subjects were separated into two groups: 19 healthy subjects and 21 patients. It was difficult to separate a patient out using accuracy and amplitude values of saccades, which was also denoted by the corresponding ratio values (0.49 and 0.45). On the other hand, patients usually reacted to the stimulation slower than the healthy subjects, so the ratio value of latency was relatively large (1.44). That is to say, latency was a significant variable in separating the healthy subjects from the patients. This was the only publication of the thesis in which patient data was used. The motive behind using patient data was to see whether verification was possible with somewhat different EOG signals compared to the EOG signals of the healthy subjects. Since the best classification results of the EOG data reached 90%, this patient material obviously did not worsen them. Perhaps it might even indicate that patients suffering from vertigo and other otoneurological symptoms [34] could, after all, perform user verification tests with their saccades.

4.3 Publication III: A priori probability

A priori probability [77] is the proportion of each class in a training set (hereby it is the ratio of samples [saccades] of the authenticated user to those of non-users in a training set), which plays a significant role in classification. It always affects or may even directly decide a classification result. The principle can be presented with the following formula for some methods, for example the Bayesian rule:

$$P(w_j|\mathbf{x}) = \frac{P(\mathbf{x}|w_j)P(w_j)}{P(\mathbf{x})}$$

where $P(w_j)$ is the a priori probability of class w_j, $P(\mathbf{x}|w_j)$ is the conditional probability of $\mathbf{x}$ with condition w_j, $P(\mathbf{x})$ is the evidence and $P(w_j|\mathbf{x})$ is the a posterior probability. A sample to be tested is determined to be or classified as w_i if $P(w_i|\mathbf{x}) > P(w_j|\mathbf{x})$, otherwise w_j.

From the formula, it is easy to see that a higher a priori probability leads to a higher posterior probability. That is to say, the tested samples are, with a high probability, classified according to the class label that corresponds to the higher a priori probability.

Although some other methods affected by a priori probability are hardly derived from the formula, a classification result can also be presented corresponding to such methods. See the results of logistic discriminant analysis (or logistic regression, LogDA) in Table 4.2.

Table 4.2. Means and standard deviations of LogDA accuracy of 68 VOG subjects for five different ratios of saccades of the authenticated user to those of non-users in a training set. The accuracies of Condition 1 and Condition 2 became lower and higher, respectively, step by step as the number of non-users' saccades (the latter number of ratios) increased in a training set.

Accuracy (%)	*Ratio*				
	19:20	19:40	19:60	19:120	19:180
Condition 1	98.2 ±1.8	32.8 ±3.4	10.3 ±2.1	0.2 ±0.5	0 ±
Condition 2	86 ±3	100 ±	100 ±	100 ±	100 ±

One problem is how to find an optimal a priori probability to get the best accuracy for both Conditions 1 and 2 at the same time. If only pursuing the EER, it can simply obtain the answer: an a priori probability of 50% or a ratio of m:m, where m is equal to the number of samples (saccades) in a class in a training set. However, another problem was that each subject did not have enough saccades (12 or 20 at that time) since large-amplitude saccades were scarce. If the a priori probability of the authenticated user was 50%, only a few subjects could be taken to a training set from non-users. This situation can easily bring good or bad with it randomly, possibly causing the standard deviation in tests repeated several times to be quite high and the classification accuracy to be unconvincing.

Balancing the ratio of two classes in a training set was a good solution which increased the saccades of the authenticated user in a training set by repeatedly extracting saccades of the authenticated user to make the ratio of saccades from the authenticated user and from non-users to reach or be close to m:m. This way, more subjects could be taken into a training set while making the final classification accuracy correspond more closely to the EER.

Table 4.3. Means and standard deviations of LogDA classification accuracy of 68 VOG subjects for five different ratios after using the balanced ratio.

Accuracy (%)	*Ratio*				
	19:20	38:40	57:60	114:120	171:180
Condition 1	98.2 ±1.8	98.5 ±1.8	97.9 ±1.2	98.4 ±1.8	97.9 ±1.2
Condition 2	86 ±3	90.4 ±3.7	92.8 ±3.4	96.2 ±2.8	96.3 ±2.2

In addition to LogDA (Table 4.3), the effect of a balanced ratio was also great for some other methods which are sensitive to a priori probability. For example, for the naïve Bayesian method, accuracy was improved from 21% in Condition 1 and 100% in Condition 2 (a ratio of 19:180) to 88% in Condition 1 and 76% in Condition 2 (ratio of 171:180). However, it did not resolve the problem for all the methods, e.g. decision trees, whose accuracy was only enhanced by 0.3% (1.3% to 1.6%) in Condition 1 and with no effect in Condition 2 compared with the ratio of 19:180. Obviously, with decision trees, one cannot take advantage of 'copying' existent samples in a training set.

On the other hand, for the methods which were heavily influenced by a priori probability, the performance of the balanced ratio was not apparent, but it still more or less benefited accuracy. For example, after the balancing ratio, the accuracy of Condition 1 increased 8% and Condition 2 decreased 3% for linear discriminant analysis originally with the ratio of 19:120. Therefore, balanced ratio was used for all methods in later research.

4.4 Publication IV: Applying advanced classification methods

With the verification performance achieved in the preceding research, saccades were found to be valid for the verification of subjects. However, with saccades from more subjects and more saccades recorded from each of them, the classification accuracy of some simple methods began to decrease: for example, the average accuracies of Conditions 1 and 2 for LogDA (the best method in the preceding study) declined to 82% (132 subjects) from 97% (68 subjects). Therefore, it was necessary to apply some advanced and complicated classification methods to test biometric verification using saccadic eye movements. Neural networks [78,79] and support vector machines [68,80] were used, which are often effective and practical methods in classification. Since the purpose of this thesis is not algorithmic theory research and the two above-mentioned methods are mature and complicated, the main parts of the methods are merely generally presented.

4.4.1 Support vector machine

To simplify, when a training data set cannot be linearly separated in an original variable space, a support vector machine projects the data into a higher dimensional space and then finds an optimal hyperplane to separate the classes. This method is well suited to binary classification problems.

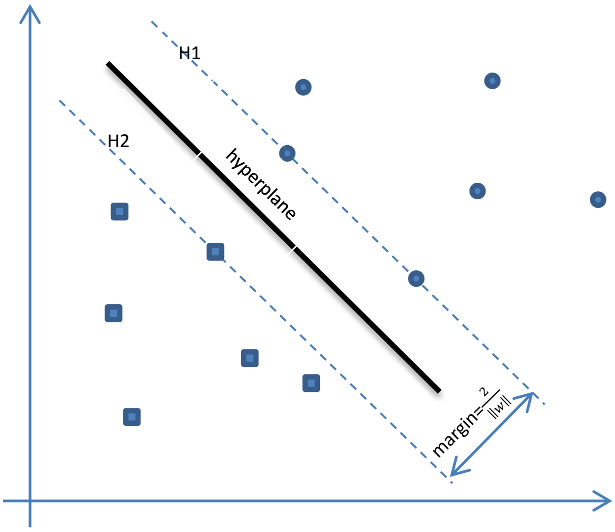


Figure 4.1. An example of an optimal hyperplane and margin linearly separated in a two-dimensional space, where the margin comes from distances of subjects to the hyperplane.

A hyperplane can be defined as

$$\mathrm{g}(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b$$

where $\mathbf{w}$ is a weight vector, b is a bias and the symbols $\langle \cdot, \cdot \rangle$ denote the inner product of two vectors.

When $g(\mathbf{x}_i)>0$, $\mathbf{x}_i$ belongs to Class 1; otherwise it belongs to Class 2. The value or label (y) of a class can then be defined as 1 or -1. Thus, the samples satisfy inequality

$$y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) > 0$$

According to the Figure 4.1, the margin is $\frac{2}{\|\mathbf{w}\|}$ [81]. It is easy to conclude that a larger margin leads to a smaller error rate. Thus, the task is to find the largest margin of the hyperplane, in other words, to minimize $\|\mathbf{w}\|$ [82]. Simultaneously, $\mathbf{w}$ and b are rescaled so that the samples closest to the hyperplane (the points on lines H1 and H2) satisfy

$$|y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b)| = 1$$

So all samples also satisfy

$$y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1 \ or \ y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) - 1 \geq 0$$

Thus, the equation of the problem can be modified as

$$\min \quad \left\{\frac{1}{2}\|\mathbf{w}\|^2\right\}$$

subject to $\quad y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) - 1 \geq 0 \quad i = 1, \dots, n$

where n is the number of training samples and, in order to simplify the computation, $\frac{1}{2}\|\mathrm{w}\|^2$ replaces $\|\mathrm{w}\|$.

To find the solution of **w**, Lagrangian formulation [83] must be used and is constructed as

$$\boldsymbol{J}(\mathbf{w}, b, \alpha) = \frac{\|\mathbf{w}\|^2}{2} - \sum_{i=1}^{n} \alpha_i \left[y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) - 1\right]$$

where α_i is the Lagrange multiplier of the corresponding sample.

Let us minimize ***J***() with respect to the weight vector **w** and maximize it with respect to the undetermined multipliers $\alpha_i \geq 0$. When using Kuhn-Tucker [83,84] construction, the optimization can be reformulated as maximizing

$$\boldsymbol{Q}(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle$$

subject to $\quad \sum_{i=1}^{n} \alpha_i y_i = 0 \ and \ \alpha_i \geq 0, \ i = 1, \dots, n.$

The solution to **w** is

$$\mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i$$

In fact, only the multipliers which are larger than zero (the corresponding samples are on the lines H1 and H2) affect the weight vector **w** and they are called 'support vectors'. The hyperplane is

$$\mathrm{g}(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i y_i \langle \mathbf{x}_i, \mathbf{x} \rangle + b$$

Now, there are still two problems: how to handle the nonlinearly separated samples and how to map the data to a higher dimensional space.

Actually, the most samples can be linearly separated, but a few cannot in some cases. If a hyperplane is always modified according to these samples, the procedure becomes slower and more complicated. Two parameters are needed [85]. One is slack variables e_i measuring deviation of an example from the ideal condition of example separability. Hence, training samples can be presented as

$$y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) > 1 - e_i \ \ or \ \ y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) - 1 + e_i > 0$$

Another parameter is the cost factor, C, a positive value showing how serious an incorrectly classified sample is. If C is too small, the error rate might increase. On the other hand, if C is larger, generalization of a model will be weak. A suitable C is necessary for a model. Cost calculation is typically

$$\mathrm{C}\sum_{i=1}^{n} e_i$$

The Least Squares Support Vector Machine (LS-SVM) [86], which was used here, computes Cost as follows

$$\frac{C}{2}\sum_{i=1}^{n} {e_i}^2$$

Also, C is not a constant under various conditions: for example, some special samples are crucial and cannot be separated incorrectly and other samples are tolerable. Also, differences between a priori probabilities of each class may be too large.

Secondly, it is known that computation in a high-dimensional space is more complex than that in a low-dimensional space (inner product, for example). Therefore, it needs a function, *K*, to simplify it:

$$K(\mathbf{a}, \mathbf{b}) = \langle \varphi(\mathbf{a}), \varphi(\mathbf{b}) \rangle$$

where $\mathbf{a},\mathbf{b} \in \mathrm{X}$, φ is the projection from the input space X to space Y (lower to higher), and *K* is called a kernel function [81].

Moreover, the kernel function is valid only if it satisfies the condition of Mercer's theorem [81,84]. The kinds of kernel functions used in this thesis were

- Polynomial: $K(\mathbf{x}, \mathbf{w}) = (\langle \mathbf{x}, \mathbf{w} \rangle + 1)^d$, where *d* is the order of the kernel function, and it is linear and quadratic when *d* is 1 and 2.
- Radial basis or Gaussian function: $K(\mathbf{x}, \mathbf{w}) = e^{-\frac{\|\mathbf{x}-\mathbf{w}\|^2}{2\delta^2}}$, where the δ is the standard deviation of **x**.

The optimization problem is as follows:

$$\min \qquad \left\{\frac{1}{2}\ \|\mathbf{w}\|^2 + \frac{C}{2}\sum_{i=1}^{n} {e_i}^2\right\}$$

subject to

$$\sum_{i=1}^{n} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}_j) + b - 1 + e_i \geq 0$$

Using the same calculation procedure as mentioned above, a hyperplane can be obtained:

$$\mathrm{g}(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b$$

SVM was one of the methods with the best accuracies for verification, especially with the kernel of radial basis function (84% for Condition 1 and 92% for Condition 2 for 132 subjects), and the computation was also faster than that of neural networks.

4.4.2 Neural networks

A neural network (NN, or sometimes called an artificial neural network, ANN) is a mathematical model inspired by biological neural networks. NN is a ‘star’ level method in machine learning, which can be used for pattern classification, predictive modelling, adaptive control, etc., since it is systematic and complex. Only two types of NN were used: multilayer perceptron (MLP) and radial basis function (RBF) [78].

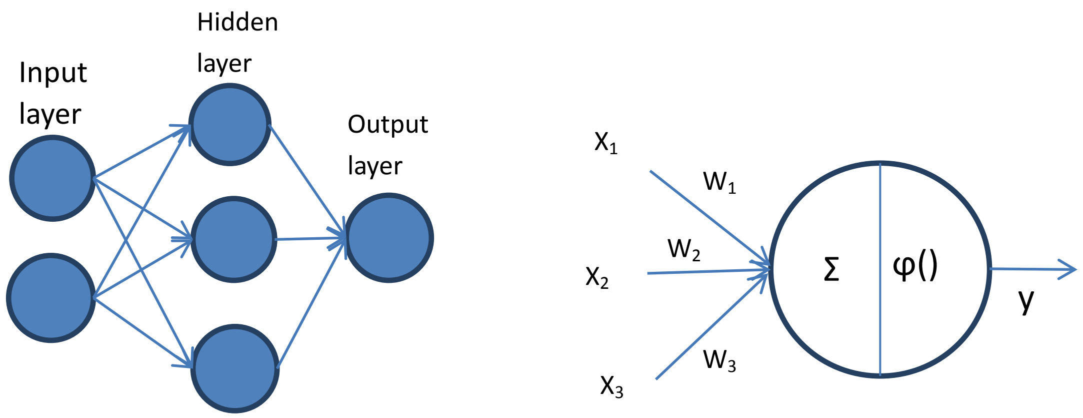


Figure 4.2. On the left, an example of a simple neural network consisting of input, hidden and output layers. On the right, an example of a simple neuron consisting of input $\mathbf{x}=(x_1,x_2,x_3)$, output $\mathbf{y}$, weight vector $\mathbf{w}=(w_1,w_2,w_3)$ and activation function φ.

Compared with a single layer perceptron, MLP consists of multiple layers of nodes. Each node that is not on the input layer is a node (neuron) with a nonlinear activation function, which is usually a logistic function such as

$$f(x) = \frac{1}{1 + exp(-ax)}$$

where a is the slope of the function.

Nonlinearity is very important. Otherwise, the relation of input to output can be obtained with a single layer perceptron. Moreover, MLP uses a supervised learning technique (for example) called back-propagation (BP) [87] to train a network. As the key part of MLP, BP can be divided into two steps:

- Step 1: Use the forward and backward propagation to get the deltas of all output and hidden nodes, where the delta is the error of a desired response and real output.
- Step 2: Calculate the gradient of the weight based on deltas and the input of the corresponding nodes, then update the weights of nodes (add or subtract a ratio of gradient).

The ratio called ‘learning rate’ is a significant parameter that controls the effect and speed of training. This thesis applied several kinds of training algorithms for MLP. Training procedures of most algorithms are relatively fast. Note also that the amount of the data used in the tests was, after all, highly restricted. However, the tests were repeated several times according to the leave-one-out principle. The performance of Levenberg-Marquardt algorithm (LMA) [79] was the best: its average accuracy was 8%-20% higher than those of other training algorithms.

RBF is another popular NN type that was also one of the methods to produce the best classification accuracies. The differences between MLP and RBF are [78]:

1. RBF usually has a single hidden layer, but MLP can have more.
2. Computation in the MLP nodes, which are located in the hidden or output layer, uses the same procedure (although not necessarily the same activation function). The RBF nodes in the hidden or output layer vary, depending on the function aspect and one the model used.
3. The hidden layer of RBF is nonlinear (always transform to a higher space) and the output layer is linear. Compared with this, both hidden and output layers are nonlinear in this thesis.
4. There are various ways to calculate an activation function of a hidden layer node. In RBF, a function is calculated using the Euclidean distance between the input $\mathbf{x}$ and the centre node $\mathbf{c}_i$, e.g.
$$f(\mathbf{x}) = \sum_{i=1}^{n} exp(-\frac{\|\mathbf{x}-\mathbf{c}_i\|^2}{2\delta^2})$$
whereas, in MLP, an inner product of the corresponding input and weight vectors is computed, e.g.
$$f(\mathbf{x}) = \sum_{i=1}^{n} \varphi(\langle \mathbf{w}_i, \mathbf{x} \rangle + b_i)$$
5. Both use a nonlinear transform from input to output, and RBF constructs local approximations, but MLP pursues global approximations. See some example results of MLP and RBF networks in Table 4.4

Table 4.4. Some accuracies given by MLP and RBF networks with some appropriate parameter values.

Accuracy (%)	*MLP*		*RBF*	
	one output node	*two output nodes*	*Goal=0.08*	*Goal=0.1*
Condition 1	78.41±2.5	79.55±2.6	83.36±2.6	88.53±1.8
Condition 2	80.45±3.2	82.65±2.5	88.86±3.9	88.86±1.9

Although NN could generate good accuracies in the test runs made, the computation time was a problem in principle, for example when the size of a training set was large or the goal of model (expected error) was set too small. Setting optimal parameter values can more or less resolve the problem. Besides reducing computation time, it also can improve the classification accuracy. However, the parameter setting (for example, kernel function selection in SVM, how many hidden layers and how many nodes in a hidden layer in MLP, and the maximum learning step and learning rate) was another problem while applying these advanced methods, as it prevented a smart or systematic approach to find the optimal parameter values, which depend on different cases indeed. Therefore, multiple tests are the only solution at present. This computation time constraint only pertains to the training phase of a neural network. Thus, subject to a practical verification situation, training would be required while updating a model with new or extended training data.

4.5 Publication V: Multiple recordings

The preceding section included the two tasks of subject verification: whether subjects can be separated (to reject an impostor) and whether a certain subject can be recognized (accepted as the authenticated user). The earlier result showed that each subject can be distinguished with good accuracy in Condition 2. However, the previous high accuracy of Condition 1 was not able to completely show that each subject could be recognized almost perfectly. All the saccades of each subject were recorded in one session only, although a session included a few successive measurements. Thus, whether the possible variabilities of saccades of each subject between sessions affected results or not was unknown.

For some types of biometric data other than eye movements, the value of a feature or variable is usually fixed – for example, the distance between the eyes or the width of nose in face recognition – no matter how many times they are measured. Instead, the variable values of a single subject are recorded remembering variability for the behavioural biometric signal, although they are within a certain range. The core of subject verification concerns classifying various variable values of a subject. In our earlier recordings, the saccades of each subject came from one session, which contained a series of three to five measurements and lasted for, at most, 20 minutes altogether. Thus, a high accuracy of Condition 1 merely verifies that, despite the possible variability in an authenticated subject's saccades within one session, the subject can be recognized. The question arose how performance would be in multiple sessions. Obviously, thus far, only one earlier biometric eye movement study, in addition to the present research, included such temporal tests, but only some of the subjects were measured twice, at an interval of one week [36].

An early study [88] proposed that the average latency could be different over a random period of three days. Subjects' physiological properties may also be affected by many outer factors such as fatigue, alcohol, drugs, disease and age [89]. Because of practical restrictions, it was not possible to research the independent effect of each of the above-mentioned factors for saccade variables. For example, it is difficult to define the degree of fatigue precisely, and investigations into the response to alcohol are very complex to perform and require a great deal of time and a physician to supervise the tests and guarantee the safety of subjects. In reality, such tests can be conducted with a small number of subjects only. Therefore, the present thesis had to ignore the influence of those factors and focus on the variability of saccadic variable values over a period of time. Still, in Publication II, some of the data were obtained from otoneurological patients who suffered from vertigo and balance problems. Nevertheless, every subject taking part in a verification test was required to abstain from alcohol and medication for 24 hours before a measurement session. A subject not satisfying this requirement would normally have been excluded, but there was one subject on continuous medication.

For this publication, the two following data sets were measured:

Group I consisted of 22 subjects, and each subject contributed 200 saccades over five separate days. The intervals between the five measurement days were different: 7 days, 26±7 days, 36±10 days and 1 day. This way, both varying and fixed intervals were included. Every day was comprised of two sessions: one in the morning and the other in the afternoon, with five hours between them. Each session included a series of five measurements, each with four saccades of the large amplitude.

Group II consisted of 12 subjects, and each of them made 32 saccades. Of these 32, 12 were from the first session and 20 from the second, with an interval between the sessions of approximately 16 months.

The objective with Group I was to study a rather long period with regard to user verification compared with everyday life, because, for example, computer logins are usually performed almost daily. Because the results in the medical literature were rarely obtained in the relevant area of saccades and have varied perhaps since the 1970s, with variable values of saccades of a subject either varying or not varying statistically significantly over the course of time – such as between morning and afternoon or a longer time – it was interesting to investigate whether such possible variability might affect user verification results.

Results obtained showed that a subject can be verified fairly well within periods of Group I. For example, RBF network gave a classification accuracy of 86% for Condition 1 (85% for Condition 2) for Group I or 88% for Condition 1 (86% for Condition 2) by using SVM, but good accuracy based on Group II could not be expected when the period was very long. However, the size of the data set in Group II was very small because it was, for practical reasons, difficult to find subjects to take part in both measurement sessions. Therefore, the research done with Group II only hints at the conclusion drawn. Deeper and more extensive research is needed.

4.6 Publication VI: Subject verification with a relatively high sampling frequency of saccade signals

The VOG signals used in the previous publications were recorded with a low sampling frequency of 30 Hz. Thus, the other eye movement camera system with a higher sampling frequency of 250 Hz was utilized to record VOG signals tested in this publication. According to Section 2.1, processing these signals with the higher sampling frequency was performed quite similarly to those previously, but more saccades were collected in order to increase training sets for machine learning algorithms. There were 55 subjects in the data set of this publication, and each subject contributed a total of 120 saccades with four measurements per session. There was a difference in the verification procedure between this publication and the earlier ones. Since the size of the data set (both the number of subjects and the number of saccades from each subject) was larger than before, the training sets of Conditions 1 and 2 were the same. In other words, except for one subject as the authenticated user, the remaining subjects were separated into either non-users or impostors. Both Conditions 1 and 2 used the same training models for verification tests. On the other hand, with experience developing, only the classification methods which had the best accuracies in the previous publications of the research were included; some basic methods were abandoned in this publication. Moreover, how to obtain the optimal parameter values of some advanced methods was still being studied.

Table 4.5. Comparison of the best accuracies of low and high sampling frequencies for VOG signal verification according to LogDA, MLP and RBF networks, and SVM.

Accuracy (%)		*LogDA*	*MLP*	*RBF*	*SVM*
30 Hz VOG	Condition 1	86	92	88	87
	Condition 2	78	84	88	78
250 Hz VOG	Condition 1	91	95	96	96
	Condition 2	89	95	95	95

From Table 4.5, it is easy to see the considerable improvement in accuracy from 30 Hz to 250 Hz signals. Actually, on average, 96% accuracy is the best and virtually optimal result in the entire thesis. For the sake of the recording situations, two data sets varied as to the number of subjects and their saccades, but the results still indicated that the sampling frequency plays a significance role in user verification with saccades.

In addition, interpolation was also used in this publication, but the effect was not as crucial as before. Sometimes it had a slightly negative influence on the verification results. Therefore, it can be concluded that interpolation, as an estimation approach here, is unnecessary for relatively high sampling frequencies, at least with the present 250 Hz data.

Chapter 5

Conclusions

The thesis considered the biometric verification of subjects using saccadic eye movement signals, and it consisted of six publications which specifically presented the entire research procedure, including signal recording, signal processing and analysis, extraction and computation of variables, classification using machine learning methods, and evaluation of results obtained. The final aim of the research is attempting to use saccadic eye movements as a new form of 'password' applied to personal devices such as a computer or possibly even a mobile phone, just as other biometric signals (fingerprint, face image, iris image, etc.) are used. One of the main factors restricting development of this technique is the specifications of the eye movement video cameras currently available, but there is no doubt that the quality of these devices will improve and their prices will decrease in the future. In principle, any device and system with such a camera system could apply this new biometric technique.

As the simplest eye movement with regard to their waveform structures, saccades were used in the research presented here. In fact, the saccades used were only measured in the horizontal direction. Compared with other studies in which saccade data were collected from two-dimensional space (vertical and horizontal), processing and analysis may have been slightly easier. There were two kinds of saccade signal types used: EOG and VOG signals. The technology for EOG signals is quite old, but its advantage was the relatively high sampling frequency: 400 Hz. In addition, recording EOG signals was not as convenient for subjects as recording VOG signals was, due to the use of skin electrodes, and such recordings were easily affected by noise. Thus, EOG signals were not the main data source: they were simply used for comparison with VOG signals for use in biometric verification. The results of Publication I indicated that the sampling frequency of signals is one of the key points for verification. Although EOG signals were commonly noisier than VOG signals, the 400 Hz sampling frequency benefited the accuracy better than that of VOG signals, whose sampling frequency was only 30 Hz but interpolated to produce a high 'artificial' sampling frequency. This conclusion was again verified when new VOG signals with the sampling frequency of 250 Hz were measured for Publication VI. Using the same classification methods and their same parameter settings, the classification accuracies for the 250 Hz VOG signals were 5%-7% higher on average than for the interpolated 30 Hz signals.

Interpolation was used to increase the sampling frequency, which aids the accurate extraction of variable values and calculation to a certain extent in signal processing. However, as an estimation technique, interpolation does not essentially improve the signal quality, for example, it is difficult to ‘repair’ some disappeared features (e.g. spike [90,91]) caused by a low sampling frequency very well. Moreover, interpolation was not suitable for all situations. Especially, when the sampling frequency was originally high enough such as 250 Hz, the influence might be even slightly negative on classification accuracies. Therefore, interpolation of signals over 250 Hz was quite unnecessary. One final conclusion drawn is that in future a more scarce interpolation for signals of 30 Hz would be sufficient and used, e.g. up to 250 Hz or no more than 400 Hz, based on our experience with these three sampling frequencies so far and earlier results that the information content of saccades of large amplitudes is at most 70 Hz [38,39]. Just like interpolation, good computational methods for variables can benefit classification accuracies, but they cannot change results fundamentally. On the other hand, as regards both these factors, they alter the original values of more or less every variable, which is not obviously reasonable for medical investigations, for example. However, their effects are positive in classification for biometric verification purposes as long as subjects can be separated easier and better.

A priori probability, or the ratio of saccades of an authenticated user to those of non-users in a training set, is another important factor that impacts verification results. The tested samples were easily classified in the higher a priori probability class, which was discussed in Publication III. Adjusting the a priori probability of saccades of an authenticated user and those of other subjects (non-users) is an effective way to obtain the EER of results. When it is difficult to switch the a priori probability, for example when the number of saccades is not sufficient for a good result in machine learning, balancing the above-mentioned ratio is a good solution. Saccades of the smallest class were copied to adjust the ratio for a training set and make accuracy more closely to correspond to the EER. Although balancing this ratio does not fit in some classifiers or performing some of them is not apparent (linear discriminant analysis), it still improves the accuracy of some other methods considerably (LogDA). On the other hand, it should be noticed that sometimes the EER cannot represent the optimal accuracy of a method, especially in various situations. Firstly, because the two errors (FAR and FRR) involved in the EER do not increase or decrease linearly (Figure 3.1), the EER may not be the smallest average error possible. On an ROC curve, it might not be the closest point to the top left, either. Secondly, the accuracy of a method from a macroscopic point of view is not completely dependent on an EER point, but decided based on the area under the ROC curve (Figure 3.2). Ultimately, since the one error is frequently more serious than the other in practice, the EER is obviously not optimal in a specific application.

The characters of the classification methods applied were from the simple to the complicated. Decision mechanisms of some methods, such as *k*-nearest neighbour searching, were modified based on different situations. The results showed that the performance of simple methods could suffice at an early stage of research when numbers of subjects and saccades were small, but their accuracies began to decline when the size of a training set was extended. Although complicated methods could improve accuracy, they also brought problems. One of the problems was setting suitable parameter values. Typically, the more complicated the method, the more parameters need to be set. Since there is no perfect solution to find the optimal parameter values, multiple tests is the only reasonable approach. For example, a matrix of parameter combinations was usually used for finding the best accuracy in our research. In addition, time consuming as another problem should be considered. For example, both too small goal (output error) in multilayer perceptron networks and too large cost parameter value in support vector machines would make the computation of these methods slow. However, sometimes it has to be done to obtain a satisfying result. Hence, how to balance between classification accuracy and execution time is important in biometric verification which is supposed to be run fast in practice.

Multiple recordings of each subject were discussed in Publication V. In fact, from the accuracies obtained, we can conclude that using saccadic eye movements to verify an authenticated user or subject from multiple sessions (88% for Condition 1 with 14 subjects in Publication V) is more difficult than separating subjects (96% for Condition 2 with 68 subjects in Publication III), because the values of the variables of saccades varied. Along with a longer time, the range of variability would obviously be extended due to rather unpredictable and unknown factors. In other words, verification was more difficult when the interval between recording sessions was longer. For the present, aside from the present research, few studies based on the classification of eye movements have focused on multiple recordings, so it will be one direction our future research could take. In terms of the length of an interval, several combinations of multiple recordings are good to explore, such as intervals of four weeks for several recording days, morning and afternoon recordings on several successive days, and several recordings for intervals even longer than four weeks.

In this thesis, classification concerns binary classes, which is also used by most studies that consider the verification or identification of subjects based on eye movements. Compared with identification of n-class (n>2) classification, verification is simpler. However, verification can completely satisfy the requirements of all personal devices and most security systems that have one authenticated user only. Especially concerning safety, a model of a high accuracy has to be the primary aim. Identification can be considered as the combination of n verifications, and accuracy would be probably lower than that of verification, since with verification error is amplified for n subjects. Identification was initially attempted, but the accuracy was inferior to those of the verification tests performed, especially when the number of subjects was large. Regardless of the accuracy of identification using eye movements, as an advanced classification model, it may still be another direction for future research to go in. At any rate, subject to verification, promising results were obtained.

Personal Contributions

All publications in this thesis were mostly the work of the author of the thesis and Martti Juhola (hereafter referred to as MJ). The EOG signals in Publications I and II came from MJ's earlier research. The VOG signals in all publications were recorded by the author, with the help of Jyrki Rasku measuring some of the VOG signals in Publications I and V. All publications were written jointly by the author and MJ. Jorma Laurikkala performed the statistical tests in Publication V and wrote a minor part of its text. The author of the thesis designed the test setup, performed all the tests, implemented the classification models and evaluated the results in all publications. MJ provided guidance and supervised the entire research project.

Bibliography

[1] R. Luis-Garcia, C. Alerola-Lopeza, O. Aghzooutb, and J. Ruiz-Alzola, "Biometric identification systems," *Signal Processing*, vol. 83, pp. 2539-2557, 2003.

[2] E. Danielyan, "The lures of biometrics," *The Internet Protocol Journal*, vol. 7, no. 1, 2004.

[3] V.M. Mane and D.V. Jadhav, "Review of multimodal biometric: applications, challenges and research areas," *International Journal of Biometrics and Bioinformatics*, vol. 3, no. 5, pp. 90-95, 2009.

[4] A.K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology* , vol. 14, no. 1, 2004.

[5] S. Chang et al., "Artifical finger detection by spectrum analysis," *International Journal of Biometrics*, vol. 3, no. 4, pp. 376-389, 2011.

[6] A. Jain, L. Hong, and R. Bolle, "On-line fingerprint verification," *IEEE Transactions on Pattern Analysis and Machine Interlligence*, vol. 19, no. 4, pp. 302-314, 1997.

[7] C. Kant and R. Nath, "Reducing process-time for fingerprint identification system," *International Journal of Biometrics and Bioinformatics*, vol. 3, no. 1, pp. 1-9, 2009.

[8] X. Chen, J. Tian, and X. Yang, "A new algorithm for distored fingerprints matching based on normalized fuzzy similiarity measure," *IEEE Transactions on Image Processing*, vol. 15, no. 3, pp. 767-776, 2009.

[9] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 3, pp. 210-227, 2009.

[10] R. Chellappa, P. Sinha, and P.J. Phillips, "Face recognition by computers and human," *IEEE Computer*, pp. 46-55, Feb. 2010.

[11] G. Kukharev, A. Tujaka, and P. Forczmanski, "Face recognition using two-dimensional CCA and PLS," *International Journal of Biometrics*, vol. 3, no. 4, pp. 300-321, 2011.

[12] A. Adler and M.E. Schuckers, "Comparing human and automatic face recognition performance," *IEEE Transactions on System, Man, and Cybernetics - Part B: Cybernetics*, vol. 37, no. 5, 2007.

[13] M. Erbilek and M. Fairhurst, "Framework for managing ageing effect in signature biometrics," *IET Biometrics*, vol. 1, no. 2, pp. 136-147, 2012.

[14] W. Barkhoda, F. Akhlaqian, M.D. Amiri, and M.S. Nouroozzadeh, "Retina identification based on the pattern of blood vessels using fuzzy logic," *EURASIP Journal on Advances in Signal Processing*, 2011.

[15] M. Negin et al., "An iris biometric system for public and personal use," *IEEE Computer*, vol. 33, no. 2, pp. 70-75, 2000.

[16] Z. Sun, Y. Wang, T. Tan, and J. Cui, "Improving iris recognition accuracy via cascaded classifiers," *IEEE Transactions on System, Man, and Cybernetics - Part C: Applications and Reviews*, vol. 35, no. 3, pp. 435-441, 2005.

[17] M.A.M. Abdullah, F.H.A. Alu-Dulaimi, W. Al-Nuaimy, and A. Al-Ataby, "Efficient small template iris recognition system using wavelet transform," *International Journal of Biometrics and Bioinformatics*, vol. 5, no. 1, pp. 16-27, 2011.

[18] M.V.N.K. Prasad, "A novel technique for palmprint classification and authentication," *International Journal of Biometrics*, vol. 2, no. 1, pp. 87-108, 2010.

[19] A. Kounoudes, V. Kekatos, and S. Mavromoustakos, "Voice biometric authentication for enhancing internet service security," *Information and Communication Technologies ICTTA '06*, vol. 2, 2006.

[20] S.A. Israel, J.M. Irvine, A. Cheng, M.D. Wiederhold, and B.K. Wiederhold, "ECG to identify individuals," *Pattern Recognition*, vol. 38, pp. 133-142, 2005.

[21] S. Chantaf, A. Nait-Ali, P. Karasinski, and M. Khalil, "ECG modelling using wavelet networks: application to biometrics," *International Journal of Biometrics*, vol. 2, no. 3, pp. 236-249, 2010.

[22] F. Sufi and I. Khalil, "Faster person identification using compressed ECG in time critical wireless telecardiology applications," *Journal of Networks Computer Applications*, vol. 34, no. 1, pp. 282-293, 2011.

[23] G. Wubbeler, M. Stavridis, D. Kreiseler, R.D. Bousseljot, and C. Elster, "Verification of humans using the electrocardiogram," *Pattern Recognition Letters*, vol. 28, no. 10, pp. 1172-1175, 2007.

[24] R.V. Yampolskiy and V. Govindaraju, "Behavioural biometrics: a survey and classification," *International Journal of Biometrics*, vol. 1, no. 1, pp. 81-113, 2008.

[25] A.K. Jain, P. Flynn, and A.A. Ross, *Handbook of Biometrics*.: Springer, 2008.

[26] Introduction to Biometric Systems. [Online]. http://www.biometrika.it/eng/wp_biointro.html

[27] A.K. Jain, S. Prabhakar, L. Hong, and S. Pankanti, "Filterbank-Based Fingerprint Matching," *IEEE Transactions on Image Processing*, vol. 9, no. 5, pp. 846-859, 2000.

[28] L. Torres, "Is there any hope for face recogition," in *Proceedings of 5th International Workshop on Image Analysis for Multimedia Interactive Service*, Lisboa, Portugal, 2004.

[29] R. Palaniappan and D.P. Mandic, "EEG Based Biometric Framework for Automatic Identity Verification," *The Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology*, vol. 49, pp. 243-250, 2007.

[30] R.B. Paranjape, J. Mahovsky, L. Benedicenti, and Z. Koles, "The electroencephalogram as a biometric," in *Proceedings of Electrical and Computer Engineering*, Toronto, 2001, pp. 1363-1366.

[31] C.H. Morimoto and M.M.R. Mimica, "Eye gaze tracking technique for interactive applications," *Computer Vision and Image Understanding*, vol. 98, pp. 4-24, 2005.

[32] H. Aalto, M. Juhola, and I. Pyykkö, "Visual stimulator using laser beam for eye-movement studies," *Medical & Biological Engineering & Computing*, vol. 27, no. 1989, pp. 332-335, 1989.

[33] M. Juhola, T. Hirvonen, H. Aalto, and I. Pyykkö, "Computational problems in the analysis of eye movement signals in the determination of vestibulo-ocular reflex," *Computers in*

Biomedical Research, vol. 30, pp. 49-60, 1997.

[34] M. Juhola, H. Aalto, and T. Hirvonen, "Using result of eye movement signal analysis in the neural network recogition of otoneurological patients," *Computer Methods and Programs in Biomedicine*, vol. 86, pp. 216-226, 2007.

[35] L. Autio, M. Juhola, and J. Laurikkala, "On the neural network classification of medical data and an endeavour to balance non-uniform data sets with artificial data extension," *Computers in Biology and Medicine*, vol. 37, pp. 388-397, 2007.

[36] C. Holland and O.V. Komogortsev, "Biometric identification via eye movement scanpaths in reading," in *Proceedings of International Joint Conference on Biometrics(IJCB)*, Washington, DC, USA, 2011, pp. 1-8.

[37] M. Juhola, "On sytactic pattern recognition of eye movement signals," *Research Advances in Biomedical Engineering*, vol. 1, 2000.

[38] A.T. Bahill, A. Brockenbrough, and B.T. Troost, "Variability and development of a normative data base for saccadic eye movements," *Investigative Ophthalmology & Visual Science*, vol. 21, no. 1, pp. 116-125, 1981.

[39] M. Juhola, V. Jäntti, and I. Pyykkö, "Effect of sampling frequencies on the velocity of slow and fast phases of nystagmus," *Biological Cybernetics*, vol. 53, pp. 67-72, 1985.

[40] M. Juhola, "The effect of digital lowpass filters on the maximum veloctiy of saccade eye movements," *Computers in Biology and Medicine*, vol. 16, no. 5, pp. 361-370, 1986.

[41] M. Juhola, "Comparison of regression methods in approximation of the maximum velocity of saccadic eye movement," *Automedica*, vol. 12, pp. 165-173, 1990.

[42] M. Juhola, "Median filtering is appropriate to signals of saccadic eye movements," *Computers in Biology and Medicine*, vol. 21, no. 1/2, pp. 43-49, 1991.

[43] C.A. Nelson and P.F. Collins, "Neural and behavioral correlates of visual recognition memory in 4- and 8-month-old infants," *Brain and Cognition*, vol. 19, no. 1, pp. 105-121, 1992.

[44] M. Nishigaki and D. Arai, "A user authentication based on human reflexes using blind spot and saccade response," *International of Journal of Biometrics*, vol. 1, no. 2, pp. 173-189, 2008.

[45] Q. Yang, M.P. Bucci, and Z. Kapoula, "The latency of saccades," *Journal of Neurophysiology*, vol. 43, no. 9, pp. 2939-2949, 2002.

[46] A. Hyrskykari, "Eyes in ettentive interfaces: Experiences from creating iDict, a gaze-aware reading aid," University of Tampere, Doctoral thesis 2006.

[47] P. Majaranta, "Text entry by eye gaze," University of Tampere, Doctoral Thesis 2009.

[48] O.V. Komogortsev, S. Jayarathna, C.R. Aragon, and M. Mahmoud, "Biometric identification via an oculomotor plant mathematical model," in *Proceedings of the Symposium on Eye Tracking Research & Applications*, New York, NY, USA, 2010, pp. 57-60.

[49] O.V. Komogortsev, "Biometric authentication via oculomotor plant characteristics," in *Proceedings of 5th IAPR International Conference on Biometrics*, New Delhi, USA, 2012, pp. 413-420.

[50] A. Kapczyński, P. Kasprowski, and P. Kuźniacki, "Modern access control based on eye movement analysis and keystrok dynamics," in *Proceedings of International Multiconference*

on Computer Science and Information Technology, Wisla, Poland, 2006, pp. 477-483.

[51] P. Kasprowski and J. Ober, "Eye movements in biometrics," in *Biometric Authentication: International Workshop, Lecture Notes in Computer Science*, vol. 3087, Prague, Czech Republic, 2004, pp. 248-258.

[52] R. Bednarik, T. Kinnunen, A. Mihaila, and P. Franti, "Eye-Movements as a Biometric," in *Proceedings of 14th Scandinavian Conference*, vol. 3540, Joensuu, Finland, 2005, pp. 19-22.

[53] I. Rigas, G. Economou, and S. Fotopoulos, "Biometric identificaiton based on eye movement and graph matching thchniques," *Pattern Recognition Letters*, vol. 33, pp. 786-792, 2012.

[54] Applied Vision Research Unit Institute of Behavioral Sciences. [Online]. http://lecerveau.mcgill.ca/flash/capsules/articles_pdf/type_eye_movements.pdf

[55] R. Andersson, M. Nyström, and K. Holmqvist, "Sampling frequency and eye-tracking measures: how speed affects durations, latencies, and more," *Journal of Eye Movement Research*, vol. 3, no. 3, pp. 1-12, 2010.

[56] M. Juhola, T. Tossavainen, and H. Aalto, "Influence of lossy compression on eye movement signals," *Computers in Biology and Medicine*, vol. 34, pp. 221-239, 2004.

[57] Z. Liang, F. Tan, and Z. Chi, "Video-based biometric identification using eye tracking technique," in *Proceedings of IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC)*, Hong Kong, 2012, pp. 728-733.

[58] D.A. Skoog, F.J. Holler, and S.R. Crouch, *Principles of Instrumental Analysis*, International Student ed.: Brooks/Cole, 2007.

[59] A. Kaminiarz, K. Königs, and F. Bremmer, "The main sequence of human optokinetic afternystagmus (OKAN))," *Journal of Neurophysiology*, vol. 101, no. 6, pp. 2889-2897, 2009.

[60] N. G. Henriksson, I. Pyykkö, L. Schalén, and C. Wennmo, "Velocity pattern of rapid eye movements," *Acta Otolaryngologica*, vol. 89, pp. 504-512, 1980.

[61] R.E. Crochiere and L.R. Rabiner, *Multirate Digital Signal Processing*.: Prentice Hall, 1983.

[62] MathWorks. Documentation Center. [Online]. http://www.mathworks.se/help/matlab/ref/interp1.html

[63] M. Juhola and I. Pyykkö, "Effect of sampling frequencies on the velocity of slow and fast phase of nystagmus," *International Journal of Bio-Medical Computing*, vol. 20, pp. 253-263, 1987.

[64] T.L. Lai, H. Robbins, and C.Z. Wei, "Strong consistency of least squares estimates in multiple regression," in *Proceedings of the National Academy of Sciences of the United States of America*, vol. 75(7), 1978, pp. 3034–3036.

[65] J.B.J. Smeet and I.T.C. Hooge, "Nature of variablility in saccades," *Journal of Neurophysiology*, vol. 90, no. 1, pp. 12-20, 2002.

[66] Y. Y. Gu, Y. Zhang, and Y. T. Zhang, "A novel biometric approach in human verification by photoplethysmographic signals," in *Proceedings of the 4th Annual IEEE Conference on Information Technology Application in Biomedicine*, UK, 2003, pp. 13-14.

[67] Y. Wang, K.N. Plataniotis, and D. Hatzinakos, "Integrating analytic and appearance attributes for human identification from ECG signals," in *Proceedings of Biometrics Symposium*, Baltimore, MD, 2006, pp. 1-6.

[68] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*, second edition ed.: Wiley-Interscience, 2001.

[69] D.P. Bertsekas, *Nonlinear Programming*, Second ed.: Athena Scientific, 1999.

[70] R.A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, no. 2, pp. 179-188, 1936.

[71] W.N. Venables and B.D. Ripley, *Modern Applied Statistics with S*, 4th ed.: Springer, 2002.

[72] D.J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, 5th ed.: CRC Press, 2011.

[73] J. A. Swets, *Signal Detection Theory and Roc Analysis in Psychology and Diagnostics: Collected Papers*.: Lawrence Erlbaum Associates, 1996.

[74] G. Seymou, *Predictive Inference*.: Chapman and Hall, 1993.

[75] Y. Dodge, *The Oxford Dictionary of Statistical Terms*., 2003.

[76] MathWorks. Documentation Center. [Online]. http://www.mathworks.se/help/stats/kmeans.html

[77] A.M. Mood, F.A. Graybill, and D.C. Boes, *Introduction to the Theory of Statistics*, 3rd ed.: McGraw-Hill, 1974.

[78] S. Haykin, *Neural Networks, A Comprehensive Foundation*, 2nd ed.: Prentice Hall, 1999.

[79] H. Demuth, M. Beale, and M. Hagan, *Neural Network ToolboxTM 6 User's Guide*.: The MathWorks, 2009.

[80] H. Joutsijoki, "Variations on a Theme, The Classification on Benthic Macroinvertebrates," University of Tampere, Doctoral thesis 2012.

[81] N. Christiani and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, UK: Cambridge University Press, 2000.

[82] B.E. Boser, I.M. Guyon, and V.N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the 5th Annual Workshop on Computational Learning Theory (COLT '92)*, New York, NY, USA, 1992, pp. 144-152.

[83] C.J.C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121-167, 1998.

[84] B. Schölkopf and A.J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, and Beyond*. USA: MIT Press, 2002.

[85] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273-297, 1995.

[86] J.A.K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, pp. 293-300, 1999.

[87] R. Rojas, *Neural Networks - A Systematic Introduction*.: Springer-Verlag, 1996.

[88] L. Schalén et al., "Intra-individual variation in oculomotor performance in man," *Acta Otolaryngologica*, vol. Supplementum.406, pp. 212-217, 1984.

[89] A. Lanitis, "A survey of the effects of aging on biometric identity verification," *International Journal of Biometrics*, vol. 2, no. 1, pp. 34-52, 2010.

[90] M. Juhola, "Detection of the spike artifact of saccade eye movements with a digital filter,"

Automedica, vol. 9, pp. 313-319, 1987.

[91] M. Juhola et al., "An identification technique for the spike artefact of saccadic eye movements," *Biological Cybernetics*, vol. 57, pp. 415-420, 1987.

Publication I

Biometric verification of subjects using saccade eye movements

Youming Zhang, Jyrki Rasku and Martti Juhola

Biometric verification of subjects using saccade eye movements

Youming Zhang, Jyrki Rasku
and Martti Juhola*

School of Information Sciences,
33014 University of Tampere, Finland
E-mail: Youming.Zhang@uta.fi
E-mail: Jyrki.Rasku@uta.fi
E-mail: Martti.Juhola@cs.uta.fi
*Corresponding author

Abstract: Biometric verification of subjects as users of computers or other devices has mainly based on fingerprints, face, iris or other images. We developed biometric verification using eye movements to be measured with eye movement videocameras. We measured saccades using the same stimulation for each subject. Our data included signals recorded in two manners: electro-oculographically from 30 subjects and with a videocamera system from additional 30 subjects. Verification tests were run with *k*-means clustering, linear and quadratic discriminant analysis, Naïve Bayes rule and *k* nearest neighbour searching. The highest accuracies were obtained with *k*-means clustering, discriminant analysis and Naïve Bayes rule, up to 90% and even close to 100% at their best.

Keywords: biometric verification; eye movements; saccades; signal analysis; classification; data analysis.

Reference to this paper should be in the following form: Zhang, Y., Rasku, J. and Juhola, M. (2012) 'Biometric verification of subjects using saccade eye movements', *Int. J. Biometrics*, Vol. 4, No. 4, pp.317–337.

Biographical notes: Youming Zhang, MSc is a Postgraduate of Computer Science. He took his BSc Degree in Xiamen University, China, 2004, and MSc Degree in University of Science and Technology Beijing, China, 2007. He studies biometric verification and classification.

Jyrki Rasku, PhD is a Researcher. He took his MSc and PhD Degrees in the University of Tampere, 2003 and 2009. Nowadays he works with the issues concerning the use of force platform in the clinical applications of human balance.

Martti Juhola, PhD is a Professor at the University of Tampere, Finland. His research interests include biomedical signal analysis, biometrics, pattern recognition and data analysis. He was Assistant, Researcher and Lecturer at the University of Turku, 1980–1992 and professor in the University of Kuopio, Finland, 1992–1997.

1 Introduction

Fingerprints and face images are perhaps the most usual biometric means to verify a subject. Numerous computational techniques have been developed for these biometric images, e.g., Chang et al. (2011), Chellappa et al. (2010), Danielyan (2004), Jain et al. (1997), Kant and Nath (2009), Kukharev et al. (2011), Mane and Jadhav (2009), Rani et al. (2008), Shih and Liu (2011), in which the recognition process started from preprocessing and image analysis. Other biometric images have also been studied, for example, iris images (Abdullah et al., 2011; Arivazhagan et al., 2009; Danielyan, 2004; Dey and Samanta, 2010), palmprints (Prasad, 2010) or other images recorded from subjects. In addition, these alternatives have been combined to produce multimodal processes (Mane and Jadhav, 2009) and perhaps to improve verification.

Verification of a user or subject is generally seen as a situation in which the actual user of a computer has to be determined and other possible subjects should be determined as non-users or imposters (Bednarik et al., 2005; Chellappa et al., 2010). Identification is usually seen as a more extensive computational task, in which any individual can be identified and distinguished from others in a group of subjects. We can see the former as a binary classification problem and the latter as a multiclass classification problem. In the present research we describe a novel technique to utilise saccade eye movements for verification purposes, as a simulation to verify an actual user of a computer or some device including a measuring component for eye movements.

Our motivation to develop a verification technique applying eye movements arose from our earlier, long-term research in the field of otoneurological eye movement studies, e.g., Aalto et al. (1989), Juhola et al. (1985, 1997, 2007), Juhola, (1986). Of course, one reason was the technical development over the last 15 years of new videocamera systems to facilitate eye movement studies for various purposes (Morimoto and Mimica, 2005). In addition, we noticed how the values of a few essential features computed from eye movements varied fairly clearly between individuals (Juhola et al., 2007) which formed a sound basis for an objective to exploit eye movements in the process of verifying subjects. As the research of eye movements for human-computer interaction is currently very active, we may assume that in the future such systems can be used to aid interaction with computers in addition to a mouse and keyboard by registering the targets of the user's gaze on a computer screen. Maybe such videocamera systems will be like the webcameras of today, cheap and easy to use. Therefore a verification procedure based on eye movements would be a timely and expedient property for a computer system including eye movement cameras.

Saccades are probably the simplest eye movements (see Figures 1 and 2) to detect with signal analysis (Bahill et al., 1981; Baloh et al., 1976; Juhola et al., 1985, 2007, Juhola, 1986). They are also the fastest eye movements, in fact the fastest movements of any performed by a human being. They are very easy to stimulate. Most of the eye movements performed in daily life are saccades while moving the gaze from one target to another. These properties naturally give additional motivation to design a verification procedure based on saccades and not, for instance, on other eye movement types such as smooth pursuit movements. Using saccades we can deal with short signals of no longer than one to a few minutes being long enough for verification, since they can include tens of saccades. Compared to images this is an advantage because of the decidedly smaller quantities of data, which may reduce the computation times required for verification and simplify the recognition process as such. When eye movement signals are one-dimensional signals, these include much less data than in images.

To the best of our knowledge, one-dimensional physiological signals except voice have so far only seldom been investigated for the purpose of biometric verification or identification (Deshpande and Holambe, 2011). Still, voice is obviously a difficult area, not only because of recognition difficulties of voice signals as such, but since there can be so many disturbing factors such as surrounding noise from several sources, for instance, other speakers and traffic. As to other one-dimensional signals, studies (Chantaf et al., 2010; Israel et al., 2005; Sufi and Khalil, 2011) seem to include virtually only Electrocardiographic (ECG) signals, obviously being the most explored signal type in biomedical engineering. Since any accurate recording of ECG signals always requires a fixed contact to some parts of the limbs or body, its use for rapid verification is complicated. Furthermore, a subject should be at rest, for instance, should not have exerted before a measurement is taken. Otherwise, the intravariation between the ECG signals of an individual might be considerable. Thus these ideas have been perhaps at their best for special purposes, e.g., identifying a patient within a hospital, where ECG signals are recorded from time to time for medical investigations and follow-up. On the other hand, the advantage here is that ECG signal analysis has been studied very extensively for several decades and there are effective computational techniques available in that field.

Eye movements have very rarely been studied for user verification purposes. Recently there have been four attempts to utilise eye movements for verification or identification. In one of these (Nishigaki and Arai, 2008), they detected the blind spot on a subject's retina. If an object of a subject's gaze was displayed at a position outside the blind spot in the visual field of the right user or subject, he or she saw it. In other words, the right subject moved the gaze to it while performing an eye movement. Another subject whose blind spot was very slightly different from that of the correct one should not have seen it, obviously making no saccades during the following one second recorded. The technique seemed to be complicated as every subject had to lean against a chin rest. In addition, there was a possibility that a subject made extraneous eye movements during this 1 s; he might have moved his eyes although did not see the actual object. It is inherent for everyone to constantly shift the gaze while looking at the surroundings – this has perhaps been very important in the distant past in our biological development to survive when human beings were both prey and predators, and, e.g., in traffic at present. During scientific tests extraneous eye movements may be forbidden, but not in natural behaviour expected in the routine use of computers. The investigation included no machine learning algorithm, which was our crucial idea in order to facilitate distinguishing between the right user and others and to adapt to the possible slow intraindividual alteration of a subject's saccades in the course of time.

Secondly, eye movements were studied (Kapczyński et al., 2006; Kasprowski and Ober, 2004) by computing the cepstrum of a signal and by classifying results according to naïve Bayes decision, nearest neighbour searching, decision trees and support vector machines. Thirdly, pupil sizes, gaze velocity and distance between eyes were used (Bednarik et al., 2005) for the biometric objective. Here fast Fourier transform and principal component analysis were computed for eye movement signals. Nearest neighbour searching was applied to the data tested according to the leave-one-out manner. Nevertheless, this technique was chiefly based on using a distance between eyes (images) and eye movements were in a minor role. In any case, their results proposed discriminatory information between subjects in eye movements. Fourthly, a mathematical model of the oculomotor system was used for verification (Komogortsev et al., 2010) focusing on saccade trajectories. The parameter vectors of the model were used as input for the classification to distinguish subjects. Verification was executed by applying nearest neighbour and C4.5 tree classifications.

Figure 1 This includes a 20 s Electro-oculographic (EOG) signal and its stimulation signal, which is the more regular and smoother of these. The stimulation signal precedes the EOG signal, because the subject has followed the stimulation light dot by his gaze, except concerning an extraneous small saccade on the right (starting approximately at sample 7000). Such a saccade was not used as an acceptable case because it was no response to any actual stimulation movement (see online version for colours)

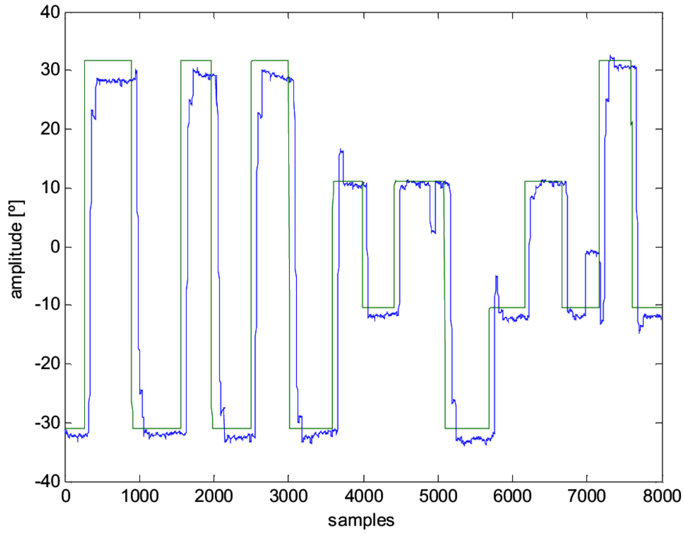
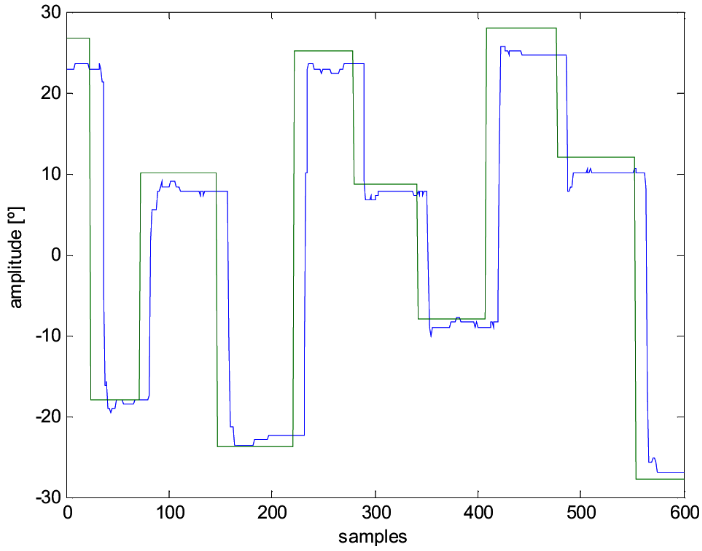


Figure 2 A VOG signal of 20 s and the corresponding smooth stimulation signal (see online version for colours)



2 Recording eye movements

We applied two data sources. The more important data was measured with a videocamera or Video-oculogram (VOG) system (Visual Eyes, Micromedical Technologies, UK). However, since its sampling frequency (frame rate per second) was low, 30 Hz, we used another data set recorded earlier (Juhola et al., 2007). The same stimulation procedure was applied to both, so their results are comparable. The advantage of the latter Electro-oculographic (EOG) data set was that its sampling frequency was as high as 400 Hz. We knew that this might be a critical issue given the earlier research (Andersson et al., 2010; Bahill et al., 1981; Juhola et al., 1985) since, of course, the higher sampling frequency enables gathering more information on eye movement signals.

In VOG there is a videocamera for each eye registering horizontal and vertical eye movements according to positional changes of the pupils in images. In EOG skin electrodes are placed close to the eye corners to register potential differences changing along with the eye movements. To make the stimulation as simple and practical as possible we applied horizontal eye movements only. In addition, EOG is better for horizontal than vertical eye movements, since the latter are sensitive to eye blinks and so wide vertical angles cannot be recorded as accurately as horizontal movements. EOG is typically noisier than VOG, because the former may include abundant noise, such as that originating from facial muscles because of talking, smiling, frowning, gasping etc. Therefore, a subject is advised not to do these during tests. VOG is much more 'user-friendly' in many respects, since it excludes the described problems provided that a subject remembers to keep her eyes open.

Signals such as those in Figures 1 and 2 were measured with the EOG and VOG techniques. The videocamera system worn by an author is seen in Figure 3. With his gaze, he followed a light dot (LEDs) in the black bar in front of him. The light dot was altered rapidly to another place in the bar (actually one LED was switched off and another switched on) so that the angle formed by them in the direction of the spectator seemed to be random from the spectator's viewpoint. Such angles were constant when the distance of the eyes from the bar was constant. However, any slight alteration in this distance would have had only a negligible effect. In addition, varying the time intervals between jumps of the light dot made the stimulation movements random-like for the spectator, although they formed a fixed series of stimulation movements shown for each subject. Such a series was complicated enough so that it could not be learnt although it was repeated several times for an individual. It was important to avoid any proactive saccades that would not have been authentic responses to the stimulation movements arranged. This type of saccade stimulations has been applied to medical investigations for decades as a standard convention, for instance, Aalto et al. (1989), Bahill et al. (1981), Baloh et al. (1976), Kaminiarz et al. (2009). On the other hand, for data analysis it was important that there were several responses to similar stimulations from each subject so that a machine learning algorithm was able to learn the feature values of individuals from the data.

The stimulations employed were used as if in the initialisation of a subject's computer session, which he or she begins by logging into the computer. The idea was not to write a password, but that the computer would recognise its legal user by recording the user's eye movements during the initialisation of the computer system. The purpose was that the computer would present the same stimulation series of light dot jumps on its screen. The user was due to look at the dot jumping approximately once in two seconds for a minute or so. Both stimulation amplitude (lengths of jumps of the light dot) and time intervals between jumps were varied, and most amplitudes should be large enough, such as 40–60°. The large amplitudes guaranteed that variation occurred for saccade features between subjects (Henriksson et al., 1980; Juhola et al., 2007). The large saccade amplitudes were used for verification, but occasionally a smaller one could be interspersed so that the angle was changed surprisingly in the stimulation series to make it random-like for a spectator.

Figure 3 A subject was following the target with his gaze on the bar in front of him. The small red LED was the target light, which jumped abruptly from one place to another along the bar (see online version for colours)

Using both EOG and VOG we utilised data measured from two disjoint sets of individuals each including 30 people. The EOG signals, the duration of which was 80 s, consisted of 12 or more large saccades. Since the VOG signals of duration 64 s included only four large saccades (above 40º), three such segments were measured from each subject. Since the sampling frequency of EOG was 400 Hz and that of VOG 30 Hz only, the VOG signals were linearly interpolated to raise its (artificial) frequency (13 times 30 Hz) up as close as possible to that of EOG in order to enable comparisons between the two techniques and to make VOG 'more accurate' as regards saccade features. The effect of increasing sampling frequencies on saccade features, particularly maximum velocity, was presented earlier (Andersson et al., 2010; Bahill et al., 1981; Juhola et al., 1985). Interpolation, of course, is not the same as an original measurement using a higher sampling frequency, but it can be used as an estimate.

The EOG signals had been recorded monocularly at the same time from both eyes with two skin electrodes and a ground electrode on the forehead. The signals were recorded at 400 Hz, amplified to a scale of ± 10 V, converted with an analog-digital converter of 13 bits and filtered digitally with a lowpass filter of 70 Hz cutoff. Calibration was accomplished with the signals themselves by employing the constant amplitude stimulations of 60º at the beginning and end of each signal. The VOG system included a built-in image processing system to find the pupil of an eye in order to compute eye movements on the basis of the positions of the pupil. The sampling frequency was 30 Hz interpolated up to 390 Hz. The system required no separate calibration (except when the system was installed for the very first time). Since in VOG there were two videocameras, one for each eye, two horizontal signals were received at every measurement. The better one, with less possible noise or artefacts such as eye blinks, was chosen from these two. The amplitude accuracy of both measuring techniques was 1º or better.

The EOG signals had been recorded at a university hospital, and a physician had checked all the voluntary subjects for being able to do the test without any impediments. Spectacles could be used since skin electrodes attached to the corners of the eyes had been used. There had been approximately as many females as males among the 30 subjects and their approximate mean age had been 45 years. The distance between the target of a computer-controlled light dot and a subject had been 1.40 m. The VOG signals were measured from a younger population of 20 males and 10 females, whose mean age was 29 ± 10 years. Since spectacles could not be used in the VOG measurements, the ability of all subjects to see accurately enough was checked first to avoid possible problems such as severe myopia. Associated with the age, two subjects only had presbyopia. In addition, the distance between the target of the bright LED light and a subject was 0.74 m for VOG measurements, shorter than for EOG. There were two different groups: the former (EOG) with ages from young to old and with both sexes equally, and the latter (VOG) as a fairly homogeneous age group of mainly young males. It was hard to find clear indications from the physiological literature showing whether a subject's sex might have any effects on saccades. We have not observed anything like this in our several earlier eye movement studies. Obviously, age can have effects. Therefore, it was interesting to have two quite different groups.

3 Signal analysis and forming data for verification

The EOG eye movement signals were considered according to the method presented, e.g., in Juhola (1986) and Juhola et al. (2007). The VOG eye movement signals, being usually less noisy than EOG, were processed with conventional, straightforward signal analysis methods. The objective in both was to identify saccades from them, i.e., the beginning and end of every saccade as accurately and correctly as possible so that features could be computed from the saccades detected. The principle in both techniques was to approximate the first derivative, which equals the angular velocity of eye movements. Detecting clear, rapid changes in this reveals saccade beginnings and ends. A threshold criterion of 10 s was used for velocity. In addition to this, stimulation signals had to be considered so that we knew at which time each stimulation movement (a jump of the light dot) had started. This was an easy task, because stimulation signals are noiseless and very regular, as seen in Figures 1 and 2.

The EOG data included 12–35 large saccades from each subject. The VOG data consisted of exactly 12 large saccades from a subject. After the detection of saccades the features of latency, amplitude, accuracy and maximum velocity (Figure 4) were computed from every acceptable saccade found from a signal. Latency or reaction time is the time between the beginning of a saccade and its stimulation. An accuracy value is equal to the difference of the amplitudes (angles) of a stimulation movement and its response. A saccade amplitude is more frequently less than its stimulation amplitude, but sometimes also greater. Finally, the maximum of the velocity curve was computed (Figure 4). For the EOG and VOG signals the means and standard deviations of the features are given in Table 1. The negative accuracy denotes smaller saccade amplitudes than stimulation amplitudes. Thus these fairly large standard deviations denoted opportunities to distinguish subjects from each other. The differences of the means between the techniques came from the different subjects and the different measurement techniques.

Table 1 Means and standard deviations of features in the EOG and VOG data sets and their ratios between interindividual variation and intraindividual variation

Data set	*Amplitude* (º)	*Accuracy* (º)	*Latency* (s)	*Maximum velocity* (º/s)	*Duration* (s)	*Maximum acceleration* (º/s^2)	*Maximum deceleration* (º/s^2)
Means and standard deviations							
EOG	53 ± 11	–7 ± 11	0.231 ± 0.110	631 ± 121			
VOG	47 ± 11	2 ± 8	0.216 ± 0.058	965 ± 280	0.182 ± 0.055	42980 ± 23667	40464 ± 24757
Ratios r_j of interindividual and intraindividual variations							
EOG	0.81	0.85	1.39	1.33			
VOG	0.97	0.68	0.40	0.71	0.36	0.71	0.75

The features described above are commonly used in medical and physiological tests, because changing in these can reveal peculiarity of a human being's physiology. Further, others are sometimes also computed. We still computed the duration, maximum angular acceleration and maximum angular deceleration (Table 1) of the saccades of the VOG data in order to see whether these could improve the verification results of our main data. The duration is equal to the time difference between the beginning and end of a saccade. The acceleration curve is the approximated second derivative during a saccade (Figure 4). The latter part of this curve consists of deceleration (in the opposite direction in Figure 4). The maxima of both parts form two additional physiologically meaningful features.

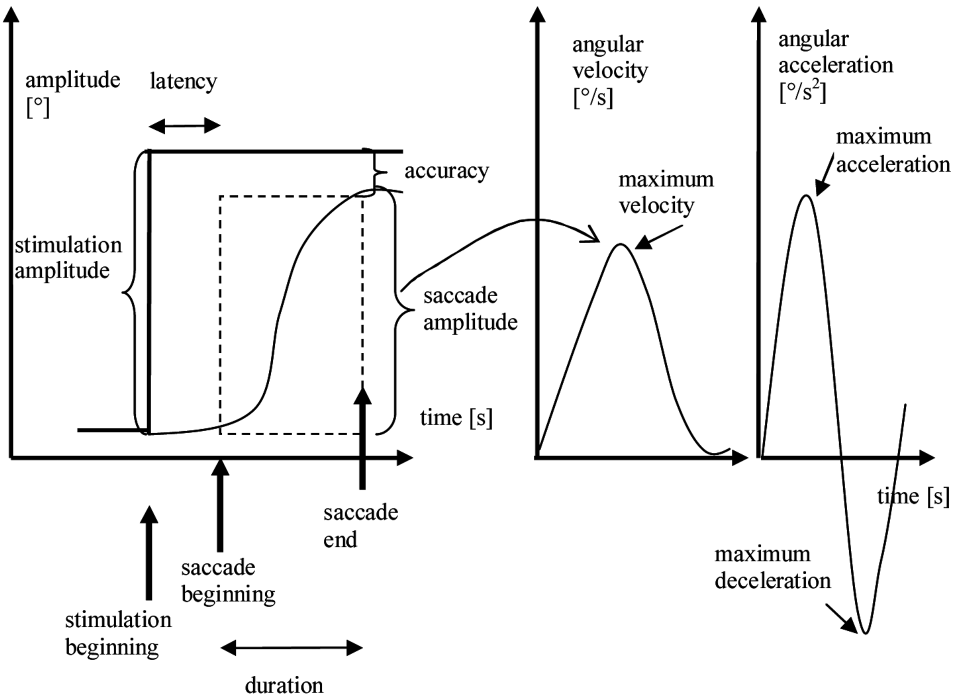
To further explore the separation ability of the features we calculated ratios of interindividual and intraindividual variations in the following (Gu et al., 2003). Here j denotes a feature, n is equal to the number of subjects, $\bar{u}_{ij}$ is equal to the mean of feature j of subject i, $\bar{e}_j$ the mean of feature j for all subjects, u_{kj} the value of feature j of saccade k for subject i and p_i the number of the saccades for subject i. The higher the ratio, the better the distinguishing property of a feature is met:

$$r_j = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\bar{u}_{ij} - \bar{e}_j)^2}}{\frac{1}{n}\sum_{i=1}^{n}\sqrt{\frac{1}{p_i}\sum_{k=1}^{p_i}(u_{kj} - \bar{u}_{ij})^2}}.$$

The results r_j in Table 1 indicated that the features of accuracy, latency and maximum velocities were better to distinguish in EOG than VOG because of their greater values in EOG. Thus these predicted that EOG saccades could be verified better. Perhaps the originally low sampling frequency of VOG also affected this, even after the interpolation, so that the ratios of VOG were less than for EOG, apart from the amplitudes. Nonetheless, we cannot draw any firm conclusions about this, since the two data sets were entirely disjoint, not only measured with the different techniques, but also from different subjects.

We restricted ourselves to the preceding time domain variables, only excluding possible frequency domain variables. This choice was based on the extensive use of these time domain variables in such areas of medicine as physiology, ophthalmology, otoneurology and neurophysiology and medical informatics since the 1960s (Bahill et al., 1981; Baloh et al., 1976; Boghen et al., 1974; Bollen et al., 1993; Henriksson et al., 1980; Juhola et al., 1985, 1997, 2007; Pyykkö et al., 1984; Robinson, 1964; Schmidt et al., 1979; Thomas and O'Beirne, 1967). We therefore knew that the features selected can express various physiological phenomena. As regards verification with eye movements, cepstrum was applied (Kasprowski and Ober, 2006), and fast Fourier transform (spectrum) and principal component analysis were used (Bednarik et al., 2005). Nevertheless, the significance of eye movements was minor in the latter, since the verification computation was chiefly on the basis of the image analysis subject to the distance of eyes and pupil diameters. Naturally, the use of frequency domain is worth studying although not included in the present research.

Figure 4 An ideal saccade curve on the left from which seven features can be computed: amplitude, accuracy, latency, duration, maximum angular velocity, maximum angular acceleration and maximum angular deceleration. All are physiological features used in medical, psychological etc. investigations



4 Verification tests

Two test conditions were applied to simulate the verification of a user on the basis of saccade eye movements. For the first test condition we needed two classes: saccades of the right user and those of others called non-users. For the second test condition we needed a third group of subjects, excluding the right user and non-users used for a training set. The third group then formed a test set of imposters.

Condition 1:

For ii=1,..,t **do** {main loop is repeated for the sake of random selections}
  For i=1,..,n **do** {n equals the number of subjects in the whole set}
    For l=1,..,a **do** {a equals the number of saccades of subject i}
      To form a training set:
      Take other a-1 saccades of a user i other than the lth **and select randomly** b saccades from each of c subjects j (non-user), j=1,…,c, $j\neq i$, $c\leq n$-1.
      To form a test set:
      Take a user's lth saccade (excluded in training) to be the test saccade.
      Test (classify with method x) and **check** whether either correct or incorrect classification was found.
    End
    The majority of either correct or incorrect classifications decides the verification result for subject i.
  End
  Compute the numbers of correct and incorrect verifications.
End
Compute the means of correct and incorrect verifications for t iterations.

Condition 2:

For ii=1,..,t **do** {main loop is repeated for the sake of random selections}
  For i=1,..,n **do** {n equals the number of subjects in the whole set}
    To form a training set:
    Take a saccades of user i **and select randomly** 2b saccades from each of approximately d=c/2 subjects j (non-user) **randomly selected**, j=1,…,d, $j\neq i$, $c\leq n$-1.
    For k=d+1,…,c, $k\neq i$, $c\leq n$-1 **do** {these subjects k are imposters}
      To form a test set:
      Randomly select an imposter's saccade from kth subject to be the test saccade.
      Test (classify with method x) and **check** whether either correct or incorrect classification was found.
    End
    The majority of either correct or incorrect classifications decides the verification result for subject i.
  End
  Compute the numbers of correct and incorrect verifications.
End
Compute the means of correct and incorrect verifications for t iterations.

First, the right user was due to be verified as such in the first condition. The former pseudocode described how a training set and its corresponding test set of one saccade were built for the classification of n subjects. Since the leave-one-out testing method was used, one saccade at a time formed a test set and all other saccades of the same subject (the right user) were a part of a training set jointly with some saccades randomly taken from other subjects (non-users).

For the second condition, we had to divide subjects excluding the right user into non-users and imposters, each of these two groups being approximately equal parts of n–1 subjects. Test saccades were taken from the group of imposters.

The ratio between the number of the saccades of the right user and that of non-users could have been selected in numerous ways, but it was reasonable to set more saccades in the latter, which should represent a clearly larger area in the feature space. We determined two different selections to form these ratios as follows.

For the first selection and for the first condition there, the saccades of every subject as the right user were taken and these were tested against one saccade ($b = 1$) from other $c = 18$ subjects as non-users (not the right user) randomly chosen from 29 subjects. Alternately each of $n = 30$ subjects was in the role of the right user. For the second test condition the saccades of each subject (right user) were taken against $2b = 2$ saccades times some other $d = 9$ subjects (non-users). Additional 9 random subjects were used as imposters due to be verified as such (not the right user). Since there were at least 12 (large amplitude) saccades from each subject, we varied this selection when there were more (only in EOG). Thus, $a = 12$ saccades were taken for the right user. In this way, there were a user's 11 saccades versus non-users' 18 saccades in a training set of the first condition and a user's 12 saccades versus non-users' 18 saccades in that of the second condition.

For the second selection and for its first condition there were again $a = 12$ saccades for the right user. Then $b = 1$ saccade was taken randomly from each of $c = 29$ non-users. In the second condition $2b = 2$ saccades were taken randomly from each of $c = 15$ non-users. Here the saccades of $d = 14$ imposters were naturally used merely for testing, not for training, since in reality they would not have been known in advance. When 10 more or less different training sets had been built, we could run $t = 10$ test rounds for 30 subjects using both EOG and VOG data, i.e., 60 individuals in total. The results were then computed for 300 test series for every classification setup. Thus, there were a user's 11 saccades versus non-users' 29 saccades in a training set of the first condition and a user's 12 saccades versus non-users' 30 saccades in that of the second condition.

Because the number of saccades was rather small, we ran leave-one-out tests for both data sets as described. This is appropriate for small data sets. A test result was checked as to whether it was correct: in the first test condition a saccade of the right user denoting this individual and in the second test condition a saccade of an imposter denoting non-users' saccades. Our verification problem was a binary classification task for both conditions.

If an entire guess had been made for classification in the first condition of the second selection, it would have been incorrect, since the a priori probability of incorrect classification was 29/40, greater than 0.5. Instead, that of the correct classification was 11/40 in every training set. Therefore, no pure guess would have helped here, but a machine learning algorithm really had to learn the features from a data set. Thinking of the situation more abstractly, we can understand that the binary classification task contained a feature space of the current features and values, in which every right user consisted of a minor part and the corresponding non-users the rest, a major part of the feature space used. Imposters were probably within the volume of the feature space, but their feature values were not known in advance as for those of a training set. On the average, imposters ought to resemble more the non-users of a training set than the right user. More similar cases ought to be present among non-users, because non-users predominated in a far larger part of the feature space volume used than that of a single correct user.

We ran our classifications using k-means clustering, k nearest neighbour searching, linear and quadratic discriminant analysis and naïve Bayes rule. These methods were chosen since they can be trained even with relatively small training sets. They can cope with situations where a class distribution between two classes is rather imbalanced, for instance, 10% and 90% of training cases. (Although we did not test so biased distributions this time, they are in our future plans.) For example, multilayer perceptron networks might be unsuitable due to the reasons presented (Siermala and Juhola, 2006; Autio et al., 2007). Computational time complexities were not crucial here, since the numbers of input data were relatively small, probably not more than a few hundred training cases and fewer test cases. Naturally, there are other classification methods that could be as effective as these tested. For example, support vector machines could be such since they are designed especially for binary classifications, but we shall address other classification methods in our future research.

For clustering we also tested different distance measures and feature values either normalised into interval [0,1] or without normalisation. As seen in the previous means of the features (in Section 3), their scales varied considerably. Thus normalisation might have affected something in machine learning. In addition, to compare EOG and VOG results we ran VOG tests with the basic four features. Furthermore, we tested the VOG data set with all seven features as described above.

5 Test results

As mentioned, 10·30 random test series were executed in the manner of leave-one-out among a set of 30 subjects in both EOG and VOG data. There were two selections for the sizes of the groups of non-users (9 or 15 subjects) and imposters (9 or 14 subjects) and two test conditions for these: correct user verification and imposter verification. All the computation was executed with Matlab R2010a™ (MathWorks Inc., USA). The results are described in the following, first for the first selection and then slightly more concisely for the second selection.

For the first selection we performed tests by using *k*-means clustering either without or with feature value normalisation. We tested four distance measures: Euclidean and city block (Manhattan) in Table 2, and cosine and correlation distance measures in Table 3. (To limit the number of results presented we did not give standard deviations, which were mostly small, a few percent or less.) The numbers of clusters were tested from 2 to 6. Greater numbers of clusters were not applied since there were only 29 (or 30 for the second condition) cases altogether in a training set in our binary classification. We found that greater numbers of clusters would also have started to yield empty clusters. Understandably, this was due to the small number of training cases. For the VOG data, there were two alternatives of the features applied. V4 included amplitude, accuracy, latency and maximum velocity. In addition to these, V7 comprised duration, maximum acceleration and maximum deceleration. The results are given as accuracies in percentages, in other words, how many classifications were correct related to all cases tested. If false rejection rates are desired (Type I error or false negative rate), these are formed by decreasing an accuracy value from 100% in the first condition. Correspondingly, false acceptance rates (Type II error or false positive rate) can be calculated in the second condition.

Looking at the best accuracies in Tables 2 and 3 we found that the results of the EOG data set were better than those of the VOG data set for the condition 1. Instead, for condition 2 there were no such differences. The best accuracies of condition 1 were typically obtained with 5 or 6 clusters. Their differences were small between all clusters for condition 2, except occasionally in 2–4 clusters of EOG. Subject to the best VOG results, condition 2 was better classified than condition 1, but between the best EOG results no differences could be seen.

Table 2 Selection 1: Clustering results in percentages of *k*-means (*k* equal to 2, …, 6) without and with feature value normalisation to [0,1] for Euclidean and city block distance measures. EOG denotes the results of the EOG data set, V4 the VOG data set with four features and V7 with seven features. The best value or values of every column are given in bold face and their mean is *B*

	With normalisation						*Without normalisation*					
	Euclidean distance measure											
	Condition 1			*Condition 2*			*Condition 1*			*Condition 2*		
k	*EOG*	*V4*	*V7*	*EOG*	*V4*	*V7*	*EOG*	*V4*	*V7*	*EOG*	*V4*	*V7*
2	58 ± 6	32 ± 5	25 ± 7	93 ± 5	**97 ± 2**	**98 ± 2**	69 ± 5	15 ± 4	16 ± 5	87 ± 6	**99 ± 2**	98 ± 3
3	88 ± 5	56 ± 8	48 ± 5	88 ± 6	95 ± 4	96 ± 2	84 ± 6	42 ± 8	51 ± 8	96 ± 3	93 ± 5	97 ± 4
4	94 ± 5	69 ± 7	63 ± 7	93 ± 6	96 ± 3	**98 ± 2**	89 ± 5	62 ± 8	63 ± 8	95 ± 4	96 ± 3	96 ± 3
5	**97 ± 3**	79 ± 3	79 ± 6	**98 ± 3**	**97 ± 4**	**98 ± 4**	89 ± 5	75 ± 9	80 ± 6	**98 ± 2**	96 ± 3	**99 ± 2**
6	**97 ± 3**	**83 ± 5**	**83 ± 5**	97 ± 2	**97 ± 2**	**98 ± 2**	**97 ± 2**	**85 ± 11**	84 ± 7	96 ± 4	97 ± 3	98 ± 2
B	*93*						*94*					
	City block distance measure											
	Condition 1			*Condition 2*			*Condition 1*			*Condition 2*		
	EOG	*V4*	*V7*	*EOG*	*V4*	*V7*	*EOG*	*V4*	*V7*	*EOG*	*V4*	*V7*
2	62 ± 9	34 ± 7	25 ± 4	96 ± 3	97 ± 4	**99 ± 2**	68 ± 7	28 ± 7	30 ± 8	89 ± 4	96 ± 3	96 ± 3
3	86 ± 8	55 ± 9	55 ± 7	94 ± 4	**98 ± 2**	98 ± 3	84 ± 6	52 ± 7	49 ± 5	96 ± 3	97 ± 3	98 ± 2
4	96 ± 4	66 ± 7	63 ± 6	96 ± 4	97 ± 3	97 ± 3	90 ± 3	67 ± 5	61 ± 7	95 ± 4	95 ± 4	97 ± 3
5	96 ± 4	83 ± 6	78 ± 5	96 ± 2	**98 ± 2**	**99 ± 3**	**97 ± 4**	78 ± 4	73 ± 8	97 ± 4	*98 ± 3*	98 ± 2
6	**97 ± 2**	**86 ± 4**	**84 ± 7**	**98 ± 2**	**98 ± 2**	98 ± 3	**97 ± 3**	**89 ± 4**	**87 ± 6**	**99 ± 2**	97 ± 5	**99 ± 1**
B	*94*						*95*					

Table 3 Selection 1: Clustering results in percentages of *k*-means (*k* equal to 2, …, 6) without and with feature value normalisation to [0,1] for cosine and correlation distance measures. EOG denotes the results of the EOG data set, V4 the VOG data set with four features and V7 with seven features. The best value or values of every selection (column) are given in bold face and their mean is *B*

	With normalisation						*Without normalisation*					
	Cosine distance measure											
	Condition 1			*Condition 2*			*Condition 1*			*Condition 2*		
k	*EOG*	*V4*	*V7*	*EOG*	*V4*	*V7*	*EOG*	*V4*	*V7*	*EOG*	*V4*	*V7*
2	53 ± 6	24 ± 6	25 ± 7	93 ± 2	**99 ± 2**	**98 ± 2**	55 ± 6	843 ± 9	28 ± 4	91 ± 5	94 ± 3	96 ± 4
3	79 ± 4	42 ± 7	41 ± 9	87 ± 6	97 ± 4	*98 ± 3*	82 ± 6	61 ± 9	40 ± 6	88 ± 7	95 ± 4	97 ± 2
4	91 ± 3	60 ± 6	65 ± 8	93 ± 4	97 ± 2	96 ± 1	93 ± 3	72 ± 8	56 ± 7	95 ± 3	95 ± 6	**99 ± 2**
5	**96 ± 4**	73 ± 7	79 ± 7	93 ± 2	98 ± 3	**98 ± 2**	95 ± 3	83 ± 3	66 ± 9	**97 ± 4**	96 ± 5	**99 ± 2**
6	**96 ± 3**	**83 ± 6**	**84 ± 5**	**97 ± 2**	97 ± 3	**98 ± 4**	**98 ± 2**	**88 ± 4**	**76 ± 4**	**97 ± 3**	**97 ± 3**	98 ± 2
B	*93*						*93*					

Table 3 Selection 1: Clustering results in percentages of k-means (k equal to 2, …, 6) without and with feature value normalisation to [0,1] for cosine and correlation distance measures. EOG denotes the results of the EOG data set, V4 the VOG data set with four features and V7 with seven features. The best value or values of every selection (column) are given in bold face and their mean is B (continued)

	Correlation distance measure											
	Condition 1			*Condition 2*			*Condition 1*			*Condition 2*		
	EOG	*V4*	*V7*	*EOG*	*V4*	*V7*	*EOG*	*V4*	*V7*	*EOG*	*V4*	*V7*
2	49 ± 9	20 ± 5	18 ± 4	93 ± 6	98 ± 2	**98 ± 2**	60 ± 5	42 ± 9	25 ± 7	88 ± 5	94 ± 4	95 ± 4
3	76 ± 7	43 ± 9	43 ± 6	90 ± 3	97 ± 3	97 ± 3	87 ± 6	57 ± 5	40 ± 7	88 ± 5	96 ± 3	97 ± 3
4	92 ± 4	59 ± 7	60 ± 9	93 ± 4	96 ± 3	**98 ± 3**	92 ± 5	69 ± 8	52 ± 4	93 ± 4	97 ± 3	98 ± 2
5	93 ± 5	71 ± 7	76 ± 7	**96 ± 4**	97 ± 3	97 ± 3	95 ± 4	78 ± 7	62 ± 8	95 ± 4	98 ± 2	98 ± 4
6	**96 ± 3**	**83 ± 3**	**80 ± 7**	**96 ± 3**	**99 ± 1**	**98 ± 2**	**99 ± 2**	**89 ± 6**	**78 ± 7**	**96 ± 3**	**99 ± 2**	**99 ± 2**
B				*92*						*93*		

We also computed means B of the best accuracies of the columns to roughly estimate possible differences between distance measures and with or without normalisation. Whether the normalisation of the features was applied revealed no differences. For the results within single distance measures, the situations varied slightly, but generally there were no differences between their best values. In most of all cluster numbers there were none, but occasionally differences greater than 5% appeared between the use of V4 and V7 for 2–4 clusters of condition 1 in the VOG data set. Considering still the means of the best values and comparing the four distance measures with each other we noticed that there were virtually no differences between them.

Next we ran tests using k nearest neighbour searching, linear and quadratic discrimination analysis, and naïve Bayes rule. All tests were implemented similarly to that mentioned above for clustering. Nonetheless, we did not normalise feature values except in k nearest neighbour searching. Since there were k (>1) nearest neihgbours involved in every classification instead of 1 compared to all other classification methods, we did not use directly majority vote. The verification procedures in Section 4 were modified to indicate a correct verification in condition 1 provided that

$$\frac{x}{ka} > \frac{a-1}{a-1+bc},$$

where a, b and c were defined in Section 4 and k is the number nearest neighbours and x equals the number of correctly classified saccades of subject i. Here the left side was compared to the a priori probability of a correct verification. For condition 2 the opposite operator (≤) was employed since correct verification decisions then corresponded to matching with non-users' saccades more frequently than with those of a right user. The results are presented in Tables 4 and 5. The Euclidean distance measure was applied to these tests.

While running k nearest neighbour searching its maximum was 11, since no more than 12 saccades were used for a right user, in other words, for the smaller class.

According to Table 4, the tests of condition 1 were classified better than for condition 2. According to Table 5, linear discriminant analysis generated the best results for condition 1. Instead, quadratic discriminant analysis was best in condition 2.

Table 4 Selection 1: Results in percentages for *k* nearest neighbour searching (*k* equal to 1, 3, 5, 7, 9, or 11). EOG denotes the results of the EOG data set, V4 the VOG data set with four features and V7 with seven features. The best value of each column is given in bold face

	Condition 1			*Condition 2*		
k	*EOG*	*V4*	*V7*	*EOG*	*V4*	*V7*
1	73 ± 8	82 ± 6	82 ± 6	58 ± 9	**74 ± 8**	63 ± 5
3	82 ± 6	84 ± 5	89 ± 5	62 ± 7	73 ± 5	67 ± 7
5	86 ± 6	88 ± 3	87 ± 5	**66 ± 7**	71 ± 7	**68 ± 9**
7	**87 ± 6**	**89 ± 4**	**91 ± 4**	59 ± 6	65 ± 9	64 ± 8
9	85 ± 4	88 ± 2	89 ± 2	51 ± 7	58 ± 4	57 ± 3
11	82 ± 3	87 ± 1	87 ± 2	60 ± 6	51 ± 6	53 ± 8

Table 5 Selection 1: Results in percentages for linear and quadratic discriminant analysis and naïve Bayes rule. EOG denotes the results of the EOG data set, V4 the VOG data set with four features and V7 with seven features. The best value of each column is given in bold face

	Condition 1			*Condition 2*		
Method	*EOG*	*V4*	*V7*	*EOG*	*V4*	*V7*
Linear discriminant	**99 ± 1**	84 ± 5	**82 ± 4**	78 ± 4	70 ± 6	76 ± 9
Quadratic discriminant	96 ± 2	**86 ± 5**	37 ± 7	85 ± 8	**83 ± 5**	**92 ± 4**
Naïve Bayes rule	97 ± 3	78 ± 3	80 ± 4	**87 ± 5**	**83 ± 7**	80 ± 7

We still computed tests for the second selection mentioned above, which incorporated more non-users and more saccades of non-users in training sets than in the first selection. On the basis of the a priori probabilities of its two classes, the right user and non-users, condition 1 could become more difficult to verify and vice versa for condition 2.

We ran *k*-means clustering tests similar to those shown in Tables 2 and 3. Nevertheless, since the results obtained were quite similar between the four distance measures, Table 6 only includes results for the Euclidean measure. They indicated how the increase of non-users' saccades in training sets significantly decreased accuracies in condition 1. On the other hand, those of condition 2 increased virtually up to 100%. The magnitudes of the changes in condition 1 were surprising, although changes were indeed expected. For condition 2 the changes were small, because the accuracies were already close to 100% in Table 2 and the a priori probabilities in selection 2 favoured condition 2.

Table 6 Selection 2: Clustering results in percentages of *k*-means (*k* equal to 2, …, 6) without and with feature value normalisation to [0,1] for the Euclidean distance measures. EOG denotes the results of the EOG data set, V4 the VOG data set with four features and V7 with seven features. The best value or values of every column are given in bold face and their mean is *B*

k	*With normalisation*						*Without normalisation*					
	Euclidean distance measure											
	Condition 1			*Condition 2*			*Condition 1*			*Condition 2*		
	EOG	*V4*	*V7*	*EOG*	*V4*	*V7*	*EOG*	*V4*	*V7*	*EOG*	*V4*	*V7*
2	28	12	10	97	**100**	**100**	21 ± 8	6 ± 2	6 ± 3	99 ± 2	**100**	**100**
3	42	18	13	99	**100**	**100**	38 ± 4	9 ± 2	14 ± 5	**100**	**100**	**100**
4	60	26	20	98	99	**100**	50 ± 8	19 ± 5	21 ± 5	99 ± 1	**100**	**100**

Table 6 Selection 2: Clustering results in percentages of *k*-means (*k* equal to 2, …, 6) without and with feature value normalisation to [0,1] for the Euclidean distance measures. EOG denotes the results of the EOG data set, V4 the VOG data set with four features and V7 with seven features. The best value or values of every column are given in bold face and their mean is *B* (continued)

k	*With normalisation*						*Without normalisation*					
	Euclidean distance measure											
	Condition 1			*Condition 2*			*Condition 1*			*Condition 2*		
	EOG	*V4*	*V7*	*EOG*	*V4*	*V7*	*EOG*	*V4*	*V7*	*EOG*	*V4*	*V7*
5	72	35	26	99	99	98	56 ± 7	28 ± 5	22 ± 7	**100**	**100**	**100**
6	**78**	**46**	**34**	**100**	99	99	**63 ± 9**	**36 ± 8**	**34 ± 9**	**100**	**100**	**100**
B	*76*						*72*					

Finally, we tested nearest neighbour searching (Table 7) and the other three classification methods (Table 8). Compared to the results in Table 4, the method of nearest neighbour searching gave slightly better results for condition 2, as expected, but only a few percent poorer for *k* equal to 1 in condition 1. Linear and quadratic discriminant analysis and Bayes rule altered the best results of condition 2 from Table 5 to Table 8.

Table 7 Selection 2: Results in percentages for *k* nearest neighbour searching (*k* equal to 1, 3, 5, 7, 9, or 11). EOG denotes the results of the EOG data set, V4 the VOG data set with four features and V7 with seven features. The best value of every column is given in Bold face

k	*Condition 1*			*Condition 2*		
	EOG	*V4*	*V7*	*EOG*	*V4*	*V7*
1	69 ± 5	79 ± 4	80 ± 6	**70 ± 8**	**80 ± 5**	75 ± 5
3	83 ± 6	88 ± 2	87 ± 5	66 ± 12	72 ± 8	**76 ± 6**
5	86 ± 5	86 ± 6	87 ± 4	62 ± 9	69 ± 11	75 ± 5
7	86 ± 5	89 ± 3	89 ± 3	61 ± 8	67 ± 5	73 ± 6
9	**90 ± 4**	**90 ± 4**	**91 ± 6**	58 ± 8	69 ± 7	67 ± 8
11	87 ± 5	88 ± 3	89 ± 2	54 ± 5	60 ± 7	60 ± 7

Table 8 Selection 2: Results in percentages for linear and quadratic discriminant analysis and naïve Bayes rule. EOG denotes the results of the EOG data set, V4 the VOG data set with four features and V7 with seven features. The best value of each column is given in bold face

Method	*Condition 1*			*Condition 2*		
	EOG	*V4*	*V7*	*EOG*	*V4*	*V7*
Linear discriminant	89 ± 6	**87 ± 4**	**84 ± 3**	89 ± 6	87 ± 4	87 ± 3
Quadratic discriminant	**97 ± 2**	85 ± 3	25 ± 2	93 ± 4	89 ± 4	**98 ± 2**
Naïve Bayes rule	90 ± 2	65 ± 3	69 ± 3	**98 ± 3**	**97 ± 2**	92 ± 4

6 Conclusion and discussion

In the following we draw conclusions on the results obtained. On the basis of Tables 2-8 nearest neighbours produced poorer results for condition 1 compared to those of other methods. Further, selection 2 was more successful for condition 2 than selection 1 according to Tables 4 and 7. Although the results for condition 2 could be improved from selection 1 to selection 2, the results for condition 1 did not drop. Unlike with the other methods, the results of *k*-means clustering for condition 1 were greatly impaired along with this change, where clustering favoured the majority class of non-users. Instead, linear and quadratic discriminant analysis and naïve Bayes rule were fairly intolerant of it in condition 1, but could improve results in condition 2. Neither normalisation nor choice of distance measure seemed to affect the results in clustering.

Computing with or without normalisation did not lead to differences in these data sets, but since the scales of the seven features applied are very different, it is reasonable to return to this issue later in the future research after having collected larger VOG data sets. Viz., latency and duration are roughly in [0.05,0.5], amplitude in [10,70], accuracy in [–40,30], maximum velocity in [100,1100] and maximum acceleration and deceleration in [10000,100000]. The current VOG data was our preliminary data set. In the VOG data the differences between the results of either four or seven features varied and were mostly small, a few percent. Thus both could be applied.

The results introduced could not be easily compared with the results of the verification tests presented for fingerprints and face images, among others, since these test situations and methods were very different. However, looking at classification accuracy values only, our results turned out well. It was possible to verify a right subject (condition 1) up to 90% and even close to 100% with the EOG data and also to detect an imposter as such at its best for the current data. For those other eye movement or related results (Bednarik et al., 2005; Kapczyński et al., 2006; Kasprowski and Ober, 2004), they obtained various results for subject identification. For 9 subjects they obtained average false acceptance rates of 1.4-17.5% and average false rejection rates of 12.6-35.6% depending on a classification method (Kasprowski and Ober, 2004), for 47 subjects average false acceptance rates of 4.8% and average false rejection rates of 9.4% (Kapczyński et al., 2006), and for 12 subjects 90% accuracy based mostly on distance between eyes (not actually on eye movements) (Bednarik et al., 2005). Nearest neighbour searching yielded false acceptance rates of 5.4 % and false rejection rates of 56.6 %, but C4.5 trees gave poor false acceptance and good false rejection rates (Komogortsev et al., 2010). Altogether, they recorded 68 subjects, but only 41 subjects passed criteria set for the analysis. Our highest accuracies were better than those of the few other studies published so far.

Although EOG recordings are not relevant in the planned routine use of eye movements for the verification of users because of the skin electrodes needed, they were useful in the present research to predict how the results might have been better while waiting for more effective videocamera systems in the future regarding their sampling frequency (frame rate per second). As was seen, the results obtained with EOG were sometimes (for condition 1 in Tables 2, 3 and 6) slightly better than those with VOG. A probable cause is the higher sampling frequency of EOG applied, 400 Hz, compared to the low one for the VOG data, 30 Hz only. There are VOG systems with higher frequencies up to at least 500 Hz, but they are expensive. After all, we also showed here that it is possible to verify a user with a low frequency camera, which is a beneficial property when considering the use of eye movement for verification.

As one-dimensional signals eye movements can be fairly easily measured and rapidly analysed in the theoretical time complexity sense compared to image data. Eye movements can also be measured in difficult circumstances such as in dim light. The stimulation can be run within one minute, which is enough to include 30–40 saccades, perhaps only in 30–45 s.

What could be possible problems concerning user verification based on eye movements? Falsification is out of the question here since it is virtually impossible to imitate some one else's eye movements. Modern videoacameras can function well in difficult circumstances regarding illumination and temperature. An interesting issue is ageing (Lanitis, 2010) for most biometric techniques. Saccades may become slower with age, which would decrease, e.g., maximum velocity and latencies could become longer. However, the meaning of such possible phenomena is negligible in the current context of user verification, because this can always be implemented so that the verification system is adaptive, where after each acceptable login the training data buffer of the users' saccade features would be updated with a new item, leaving out the oldest one. A few dozen items would be sufficient in such a data buffer. Thus the period from which the content of the buffer is collected would be short, perhaps a few weeks. Moreover, computers, mobile phones etc. are seldom used for more than five years. A more drastic effect on eye movements might be caused by some disease affecting eye movements (Henriksson et al., 1980; Juhola et al., 1997, 2007; Pyykkö et al., 1984). These, however, are very infrequent. The adaptation property of the verification system would then be very useful.

A problem could be a possible variability in individuals' saccade feature values. If a subject's saccades vary too much at short intervals, say during days, this may cause difficulties in distinguishing his or her saccades from those of others. However, such studies have been reported showing no significant differences between different measurement times. For instance, no statistically significant differences had been obtained when average maximum velocities of 58 healthy subjects were computed within an interval of two weeks (Bollen et al., 1993). Nevertheless, we are going to study this matter in the future.

In the future we shall collect measurements from more subjects and develop our technique on the basis of the research introduced. We believe that eye movements could be used for verification when eye movement videocamera systems are used like webcameras at the moment. The encouraging results of the verification experiments presented support these objectives well. There are still several other classification methods worth testing. Logistic discriminant analysis has sometimes been effective. Like support vector machines they are designed for binary classification in particular. Neural networks such as multilayer perceptron networks, learning vector quantisation networks, self-organising maps (Kohonen networks), and radial basis function networks are possible, but neural networks frequently require a large amount of training data. Thus their use might be complicated. Decision trees may cope well with small amounts of data and imbalanced class distributions.

Acknowledgements

We are grateful to Prof. Ilmari Pyykkö of the Department of Otorhinolaryngology, Tampere University Hospital, Timo Hirvonen, MD and Heikki Aalto, PhD of the Department of Otorhinolaryngology, Helsinki University Central Hospital, Finland, for medical advice on saccades and aid in recording signals.

References

Aalto, H., Juhola, M. and Pyykkö, I. (1989) 'Visual stimulator using laser beam for eye-movement studies', *Medical & Biological Engineering & Computing*, Vol. 27, pp.332–335.

Abdullah, M.A.M., Alu-Dulaimi, F.H.A., Al-Nuaimy, W. and Al-Ataby, A. (2011) 'Efficient small template iris recognition system using wavelet transform', *International Journal of Biometrics and Bioinformatics*, Vol. 5, No. 1, pp.16–27.

Andersson, R., Nyström, M. and Holmqvist, K. (2010) 'Sampling frequency and eye-tracking measures: how speed affects durations, latencies, and more', *Journal of Eye Movement Research*, Vol. 3, No. 3, pp.1–12.

Arivazhagan, S., Ganesan, L. and Srividya, T. (2009) 'Iris recognition using multi-resolution transforms', *International Journal of Biometrics*, Vol. 1, No. 3, pp.254–267.

Autio, L., Juhola, M. and Laurikkala, J. (2007) 'On the neural network classification of medical data and an endeavour to balance non-uniform data sets with artificial data extension', *Computers in Biology and Medicine*, Vol. 37, pp.388–397.

Bahill, A.T., Brockenbrough, A. and Troost, B.T. (1981) 'Variability and development of a normative data base for saccadic eye movements', *Investigative Ophthalmology & Visual Science*, Vol. 21, No. 1, pp.116–125.

Baloh, R.W., Kumley, W.E. and Honrubia, V. (1976) 'Algorithm for analyses of saccadic eye movements using a digital computer', *Aviation, Space, and Environmental Medicine*, Vol. 47, No. 5, pp.523–527.

Bednarik, R., Kinnunen, T., Mihaila, A. and Fränti, P. (2005) 'Eye-movements as a biometric', *14th Scandinavian Conference on Image Analysis*, *Lecture Notes in Computer Science*, Springer-Verlag, Vol. 3540, pp.780–789.

Boghen, D., Troost, B.T., Daroff, R.B., Dell'Osso, L.F. and Birkett, J.E. (1974) 'Velocity characteristics of normal human saccades', *Investigative Ophthalmology*, Vol. 13, No. 8, pp.619–622.

Bollen, E., Bax, J., van Dijk, J.G., Koning, M., Bos, J.E., Kramer, C.G.S. and van der Velde, E.A. (1993) 'Variability of the main sequence', *Investigative Ophthalmology & Visual Science*, Vol. 34, No. 13, pp.3700–3704.

Chang, S., Secker, J., Xiao, Q., Reid, B., Bergeron, A. and Almuhtadi, W. (2011) 'Artifical finger detection by spectrum analysis', *International Journal of Biometrics*, Vol. 3, No. 4, pp.376–389.

Chantaf, S., Nait-Ali, A., Karasinski, P. and Khalil, M. (2010) 'ECG modelling using wavelet networks: application to biometrics', *International Journal of Biometrics*, Vol. 2, No. 3, pp.236–249.

Chellappa, R., Sinha, P. and Phillips, P.J. (2010) 'Face recognition by computers and human', *IEEE Computer*, February, pp.46–55.

Danielyan, E. (2004) 'The lures of biometrics', *The Internet Protocol Journal*, Vol. 7, No. 1, http://www.cisco.com/web/about/ac123/ac147/archived_issues/ipj_7-1/lures_of_biometrics.html

Deshpande, M. and Holambe, R.S. (2011) 'Robust speaker identification in the presence of car noise', *International Journal of Biometrics*, Vol. 3, No. 3, pp.189–205.

Dey, S. and Samanta, D. (2010) 'Fast and accurate personal identification based on iris biometric', *International Journal of Biometrics*, Vol. 2, No. 3, pp.250–281.

Gu, Y.Y., Zhang, Y. and Zhang, Y.T. (2003) 'A novel biometric approach in human verification by photoplethysmographic signals', *Proceedings of the 4th Annual IEEE Conference on Information Technology Applications in Biomedicine*, UK, pp.13–14.

Henriksson, N.G., Pyykkö, I., Schalén, L. and Wennmo, C. (1980) 'Velocity patterns of rapid eye movements', *Acta Otolaryngologica*, Vol. 89, pp.504–512.

Israel, S.A., Irvine, J.M., Cheng, A., Wiederhold, M.D. and Wiederhold, B.K. (2005) 'ECG to identify individuals', *Pattern Recognition*, Vol. 38, pp.133–142.

Jain, A., Hong, L. and Bolle, R. (1997) 'On-line fingerprint verification', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 4, pp.302–314.

Juhola, M. (1986) 'A syntactic method for analysis of saccadic eye movements', *Pattern Recognition*, Vol. 19, pp.353–359.

Juhola, M., Aalto, H. and Hirvonen, T. (2007) 'Using results of eye movement signal analysis in the neural network recognition of otoneurological patients', *Computer Methods and Programs in Biomedicine*, Vol. 86, pp.216–226.

Juhola, M., Hirvonen, T., Aalto, H. and Pyykkö, I. (1997) 'Computational problems in the analysis of eye movement signals in the determination of vestibulo-ocular reflex', *Computers in Biomedical Research*, Vol. 30, pp.49–60.

Juhola, M., Jäntti, V. and Pyykkö, I. (1985) 'Effect of sampling frequencies on the maximum velocity of saccadic eye movements', *Biological Cybernetics*, Vol. 53, pp.67–72.

Kaminiarz, A., Königs, K. and Bremmer, F. (2009) 'The main sequence of human optokinetic afternystagmus (OKAN)', *Journal of Neurophysiology*, Vol. 101, No. 6, pp.2889–2897.

Kant, C. and Nath, R. (2009) 'Reducing process-time for fingerprint identification system', *International Journal of Biometric and Bioinformatics*, Vol. 3, No. 1, pp.1–9.

Kapczyński, A., Kasprowski, P. and Kuźniacki, P. (2006) 'Modern access control based on eye movement analysis and keystroke dynamics', *International Multiconference on Computer Science and Information Technology*, Wisla, Poland, pp.477–483.

Kasprowski, P. and Ober, J. (2004) 'Eye movements in biometrics', *Biometric Authentication: International Workshop*, *Lecture Notes in Computer Science*, Springer-Verlag, Prague, Czech Republic, Vol. 3087, pp.248–258.

Komogortsev, O.V., Jayarathna, S., Aragon, C.R. and Mahmoud, M. (2010) 'Biometric identification via an oculomotor plant mathematical model', *Proceedings of the 2010 Symposium on Eye Tracking Research & Applications*, Austin, Texas, pp.57–60.

Kukharev, G., Tujaka, A. and Forczmanski, P. (2011) 'Face recognition using two-dimensional CCA and PLS', *International Journal of Biometrics*, Vol. 3, No. 4, pp.300–321.

Lanitis, A. (2010) 'A survey of the effects of aging on biometric identity verification', *International Journal of Biometrics*, Vol. 2, No. 1, pp.34–52.

Mane, V.M. and Jadhav, D.V. (2009) 'Review of multimodal biometrics: applications, challenges and research areas', *International Journal of Biometrics and Bioinformatics*, Vol. 3, No. 5, pp.90–95.

Morimoto, C.H. and Mimica, M.M.R. (2005) 'Eye gaze tracking techniques for interactive applications', *Computer Vision and Image Understanding*, Vol. 98, pp.4–24.

Nishigaki, M. and Arai, D. (2008) 'A user authentication based on human reflexes using blind spot and saccades response', *International Journal of Biometrics*, Vol. 1, No. 2, pp.173–189.

Prasad, M.V.N.K. (2010) 'A novel technique for palmprint classification and authentication', *International Journal of Biometrics*, Vol. 2, No. 1, pp.87–108.

Pyykkö, I., Dahlen, A-I., Schalén, L. and Hindfelt, B. (1984) 'Eye movements in patients with speech dyspraxia', *Acta Otolaryngologica*, Vol. 98, pp.481–489.

Rani, J.S., Devaraj, D. and Sukanesh, R (2008) 'Robust face recognition using wavelet transform and autoassociative neural network', *International Journal of Biometrics*, Vol. 1, No. 2, pp.231–252.

Robinson, D.A. (1964) 'The mechanics of human saccadic eye movement', *Journal of Physiology*, Vol. 174, pp.245–264.

Schmidt, D., Abel, L.A., Dell'Osso, L.F. and Daroff, R.B. (1979) 'Saccadic velocity characteristics: intrisinc variability and fatigue', *Aviation, Space, and Environmental Medicine*, Vol. 50, No. 4, pp.393–395.

Shih, P. and Liu, C. (2011) 'An effective colour feature extraction method using evolutionary computation for face recognition', *International Journal of Biometrics*, Vol. 3, No. 3, pp.206–227.

Siermala, M. and Juhola, M. (2006) 'Techniques for biased data distributions and variable classification with neural networks applied to otoneurological data', *Computer Methods and Programs in Biomedicine*, Vol. 81, pp.128–136.

Sufi, F. and Khalil, I. (2011) 'Faster person identification using compressed ECG in time critical wireless telecardiology applications', *Journal of Network Computer Applications*, Vol. 34, No. 1, pp.282–293.

Thomas, E.L. and O'Beirne, H. (1967) 'Curvature in the saccadic movement', *Archives of Ophthalmology*, Vol. 77, pp.105–109.

Publication II

Biometric verification of a subject through eye movements

Martti Juhola, Youming Zhang and Jyrki Rasku

Computers in Biology and Medicine 43 (2013) 42–50

Contents lists available at SciVerse ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Biometric verification of a subject through eye movements

Martti Juhola*, Youming Zhang, Jyrki Rasku

Computer Science, School of Information Sciences, 33014 University of Tampere, Finland

ARTICLE INFO

Article history:
Received 23 March 2012
Accepted 26 October 2012

Keywords:
Biometric user verification
Classification
Machine learning
Eye movements
Saccades

ABSTRACT

Matching digital fingerprint, face or iris images, biometric verification of persons has advanced. Notwithstanding the progress, this is no easy computational task because of great numbers of complicated data. Since the 1990s, eye movements previously only applied to various tests of medicine and psychology are also studied for the purpose of computer interfaces. Such a short one-dimensional measurement signal contains less data than images and may therefore be simpler and faster to recognize. Using saccadic eye movements we developed a computational verification method to reliably distinguish a legitimate person or a subject in general from others. We tested features extracted from signals recorded from saccade eye movements. We used saccades of 19 healthy subjects and 21 otoneurological patients recorded with electro-oculography and additional 40 healthy subjects recorded with a videocamera system. Verification tests produced high accuracies.

1. Introduction

Fingerprints [1,3], face [2–4] or iris images [1,2] have been intensively studied and developed for the biometric verification and identification of subjects. Further, human palms [5] have been studied for the biometric verification and identification. Multimodal verification has also been researched, for example by applying fingerprint, face and hand geometry data [2].

We define verification as a condition according to [4,6] where, for example, the right user of a computer is due to be recognized and any other person to be recognized as an impostor. Actually, the subject could be a certain person in general such as a patient in a hospital, but since the test is made with a computer we call him or her as a user here. Identification [4] can be understood as a more complicated situation, where any person given has to be recognized from among a very large group of members. In the present study we introduce a method which applies a subject's rapid eye movements called saccades to the biometric verification.

Fingerprint, face, iris, palm and possibly other images are two-dimensional data or even include more than two dimensions, which contain great numbers of data. Our aim was to apply smaller data quantities pursuing faster and, above all, simpler signal analysis and pattern recognition tasks for verification. Using simpler recognition objects than, for instance, faces in images, computational complexity can be reduced. An advantage could be faster computation. Computation of eye movement signals could also be done with lesser powerful processors as in mobile phones. A disadvantage might be the loss of wider degree of variables in the sense of more versatile data types.

Even after intensive research on the use of fingerprints, face and iris images, these objects still consist of difficulties for the sake of their complicated data types subject to the automatic recognition to be performed by a computer. For example, face images of an individual may considerably vary according to acquisition conditions [4] the pose of the face with respect to the camera, illumination, facial expressions, wearing glasses, sunglasses or a hat, aging, or changing hairstyle.

Since short one-dimensional signals can include far less data than two-dimensional images, the former could be easier and faster to be analyzed. Obviously, very few attempts only have been made to utilize one-dimensional signals for the biometric verification, except voice [2]. These studies have considered almost merely ECG signals [7–12]. These approaches have been methodological with respect to their signal analysis techniques, but less considering their applications, for instance, with the idea to use identification while transferring ECG signals between a patient's mobile phone and a hospital [12]. ECG signals of good quality require measurements with skin electrodes attached to the chest. The detection of several precise features, e.g., durations of parts of QRS complexes, requires that electrodes are attached to the certain locations of the body. The requirements directly restrict the biometric verification of a subject to special applications such as signal transferring of patient ECG data.

The biometrics verification techniques described above are called physiological. The other alternative is to apply behavioral biometrics [13] where verification is frequently on the basis of

* Corresponding author. Tel.: +358 40 1901716.
E-mail addresses: Martti.Juhola@cs.uta.fi (M. Juhola), Youming.Zhang@uta.fi (Y. Zhang), Jyrki.Rasku@uta.fi (J. Rasku).

0010-4825/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compbiomed.2012.10.005

muscle control such as keystrokes, gait or signature. An important characteristic of biometrics then depends on time or duration between the beginning and end to do something specifically determined that is used to verify a subject. Behavioral biometrics researchers attempt to quantify behavioral traits exhibited by subjects in order to separate different feature values between these [13]. For example, gait feature fusion was presented in which verification was computed with hidden Markov models [14]. Mouse dynamics is also studied for biometrics [15]. Gaze based verification [16] and visual attention research [17] are related to our present study, but the interest in these areas usually covers such issues as at what location or at what object a subject looks in the visual field.

A novel alternative for the verification task is to apply eye movements, saccades particularly. These are very fast eye movements that can be measured with eye movement cameras. Eye movements are generated by six eye muscles of each eye. The eye muscles are functioning constantly even when the eyes do not move keeping the eye directed at a fixed target. If the gaze is due to move to a new target, the ‘‘control signal’’ from the brain to the eye muscles changes to move the eyes to a new position and to keep them there [18,19]. The visually controlled ocular stabilizing systems of man produce saccade, nystagmus and smooth pursuit eye movements. The saccade system responds to an error in the direction of the gaze with respect to the position of a target by starting a fast movement called saccade, to correct this retinal position error and brings the visual target to the fovea in the shortest possible time. Smooth pursuit eye movements require much longer stimulation periods than saccades. Nystagmus is a reflexive movement occurring sometimes spontaneously. From these, saccades are probably the easiest and most practical eye movement type to be applied to straightforward eye movement tests. Once started the trajectory and angular velocity of saccades cannot be voluntarily altered [20]. Therefore, these are suitable for biometric use.

One of few previous studies touching on biometric authentication and eyes applied measurements of the blind spot [21]: if an object is displayed at a position outside the blind spot in the visual field of a legitimate subject, he sees it and moves his gaze to it, in other words, performs a saccade to look at it. Another subject whose blind spot is slightly different should not see it, making probably no saccade during the following 1 s recorded. The technique seemed to be complicated, because a user rested against a chin rest. No machine learning ability (such as nearest neighbor searching in our later presentation) was applied, and no accurate saccade features were calculated for the verification purpose [21]. These are the main points of our approach so that the system can be ‘‘taught’’ according to a legitimate subject.

Eye movements for biometrics were studied and concisely described in [22,23] by computing cepstrum for a signal and by classifying its output values with nearest neighbor searching, naïve Bayesian rule, decision trees and support vector machines.

The third study [6] on eye movements for the biometric purpose applied pupil sizes, gaze velocity and distance between eyes and proposed to use eye movements as an additional biometric to be integrated with other biometric means. Either fast Fourier transform or principal component analysis or both were run for eye movement signals and then performed 3-nearest neighbor searching with the leave-one-out crossvalidation. After all, the identification was mostly based on the distance between eyes as the best biometric feature found. As the authors wrote [6], this is no feature of eye movements, although was effective in the identification. In summary, their results showed that there is discriminatory information in eye movements.

The fourth approach involved a model of the oculomotor system used for subject verification by means of saccade trajectories [24]. Verification was accomplished with nearest neighbor searching and C4.5 decision trees.

We developed a method for the verification of a subject, who uses some device including a computer, an instrument to measure eye movements and applicable software to control this instrument and detect eye movements and ultimately to verify whether he or she is a legitimate subject.

When we have studied eye movement signals for long [25–32] to develop signal analysis, pattern recognition, data mining and classification methods for medical purposes, especially otoneurological balance investigations, we have used physiologically important features (variables or attributes) there. Recently, we observed how such feature values frequently varied considerably between individuals [31]. Because these features directly express a subject's reactions to a visual stimulus, originate from his or her natural behavior and are fairly straightforward to be computed, we used them in the present research. Saccades are easy to stimulate and natural (voluntary) while reading or looking at the surroundings all the time and even during the REM (rapid eye movements) phase of sleep. Using saccades only parts of eye movements signals are necessary to include in the actual recognition process, since after the recognition of saccades signal segments between them can be left out, which further decreases data used for verification reducing processing.

2. Methodology

2.1. Measurements and saccade features

Our eye movement experiments simulated a condition where a person sits down at a computer and the computer system has to verify him or her to be or not to be the legitimate, authenticated subject. The system consisted of a device (Fig. 1) able to detect a subject's eye movements, saccades (Fig. 2), and a program that computed such physiologically or medically interesting features from saccades as amplitude [°], latency [s], accuracy [°] and maximum angular velocity [°/s] (Fig. 3). Latency is the time difference between the beginnings of the stimulus movement and response, saccade, which is a voluntary, rapid eye movement. Accuracy is equal to the difference of the amplitudes of the stimulation and saccade. To compute the maximum angular velocity, the first derivative was approximated by differentiating an eye movement signal numerically and searching for the maximum velocity during the eye movement. Its location is

Fig. 1. The subject followed, by his gaze, the light spot (in the LED bar) horizontally jumping abruptly from one location to another after varying intervals.

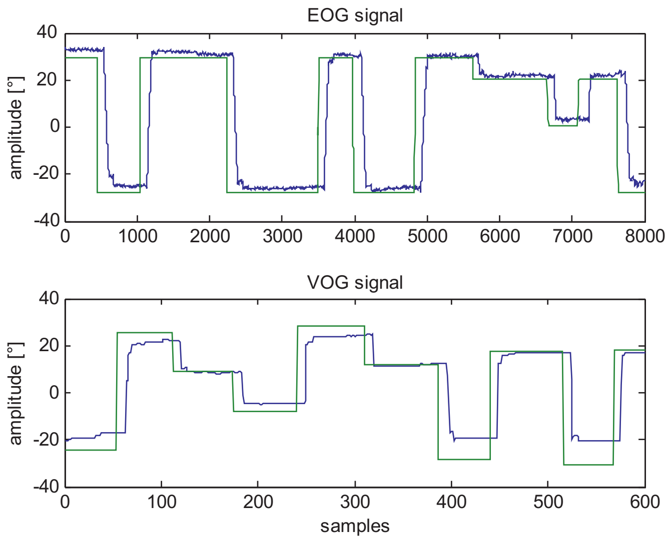


Fig. 2. The upper pairs are 20 s parts from an electro-oculographic (EOG) signal and its stimulation signal. The lower pair is from those of a video-oculography (VOG) measurement. The smooth step-like signals are the stimulations followed by the responses, saccades, after short latencies. The movements upward are horizontal eye or stimulation movements from left to right and those downward from right to left.

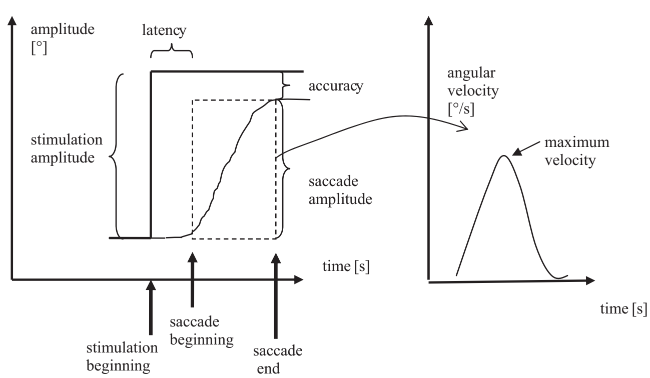


Fig. 3. The left part of the figure depicts an ideal saccade as a response to its step-like, very fast stimulation movement. The right part presents the angular velocity (approximated first derivative) curve of the saccade. Saccade features computed from the signals and commonly employed in medical investigations of eye movements are: latency as time difference between the beginnings of the stimulation movement and saccade, saccade amplitude, accuracy as difference between the amplitudes of the stimulation and saccade, and maximum velocity.

approximately in the middle of the duration of a saccade. We selected these features, because they and some other are generally employed for diagnostic purposes in medical and physiological eye movement studies in order to reveal exceptional values affected by diseases and disorders. We took these four particularly after having observed how clearly they varied between individuals [31].

2.2. Stimulation

In our simulation the task of a subject was to follow, by the gaze, a small stimulation target corresponding to a moving target on the computer screen for approximately one minute so that eye movements could be recorded during that session. This was an easy task and it is of inherent curiosity to follow a moving target in the visual field by the gaze.

We used a stimulation target that jumped horizontally in the visual field of a subject and was a switched-on LED in the eye movement stimulation device as in Fig. 1. By switching off one LED and another on in the bar in Fig. 1 the system moved the stimulation light abruptly from one location to another. Intervals between jumps and also amplitudes were randomly varied slightly so that a subject could not anticipate movements of the light. We applied this stimulation type [28] commonly used for decades in medical eye movement investigations since it was very easy for a subject and rapid to perform. It was important that a spectator could not anticipate a stimulation movement, because the anticipation would mean that he or she would not respond to an actual stimulation, but would make somewhat random eye movement. Naturally, the motivation for this stimulation type was also that we knew that subjects' responding eye movements, saccades, differed more or less between individuals [31]. For classification methods based on machine learning it is important that there are as many as possible eye movements (or data in general) stimulated in the similar way to collect a large enough training set so that a classifier could be trained to reliably separate subjects. On the other hand, a measurement should be as short as possible so that subjects did not become fatigued, a test was easy to make and verification results could be computed fast. According to our experience [25–32] and numerous medical sources, e.g., [33], a suitable duration was from 30 s to a few minutes to be sufficiently long for the signal analysis purpose, but not too long to cause fatigue or tiredness.

Our experiments simulated a situation, where a subject was due to look at a small, horizontally jumping target and his or her eye movements were recorded for the verification purpose. Every subject was seated in chair at a fixed location and with the same distance from the stimulation device. Although the heights of the subjects naturally varied, the stimulation angles between their eyes and stimulation light dots were always identical for all subjects. At first, the test was carefully explained and demonstrated to everyone to avoid misunderstandings how to do it. In addition to a testee and the researcher, nobody else attended in the laboratory. The objective was then that the computer program is able to recognize or verify whether the person is a legitimate subject or not.

Although horizontal, vertical and even torsional (the rotary axis approximately the same as the visual axis) eye movements can be measured, horizontal stimulation movements were only used, because we could then restrict ourselves to one one-dimensional signal, producing simpler and less data, faster to analyze than two one-dimensional signals like horizontal and vertical eye movements, or two- or three-dimensional signals such as images. Our objective was to form as straightforward verification task as possible, since this approach enabled short measurements. Since a training set of saccades was inevitably small because of short measurements, it was important to use only a few features to enable successful test circumstances for a classifier. If there had been several features designed to be used for verification, there should obviously have been a greater number of training cases, more data or longer signals for a classifier. On the other hand, there is few features used with eye movements and the currently employed are typically used, for example, in otoneurology [31] and other medical subspecialties.

2.3. Eye movement recording and tests

Since the 1960s eye movements are measured electro-oculographically (EOG) by setting skin electrodes on the corners of a subject's eyes. Such signals can be recorded with as high sampling frequency as the analog–digital converter used enables. In our previous studies [25,29,30], we found that 400 Hz is sufficiently high to measure accurately such short and sensitive features as latency and maximum velocity for the medical purposes. First, we used such electro-oculographical signals (Fig. 2) stimulated as mentioned above and recorded at 400 Hz and lowpass filtered below about 100 Hz. The information content of saccades is known to be below 50–70 Hz [25,34]. Since EOG utilizes skin electrodes, this is not practical for the biometric purpose. However, we were interested in comparing its results to those of the other technique with a lower sampling frequency, because we knew that the sampling frequency of 400 Hz was high enough for the recognition of saccades, the fastest movements of a human.

The modern way to measure eye movements is to employ two small videocameras, one for each eye, to follow the pupils of a subject's eyes. Such systems have recently been started to use for clinical eye movement research and also to develop human–computer interfaces [35]. Signals (Fig. 2) given by this two-dimensional (horizontal and vertical) video-oculography system (Visual Eyes, Micromedical Technologies, UK) can be typically measured with a low sampling frequency, in this case with 30 Hz (image resolution 320×240). This is no topmost sampling frequency for eye movements, since there are systems applying [6] 50, 60, 120 Hz or even more like 250 Hz [23] and also including the third, torsional dimension. Nevertheless, our interest was to employ a moderate sampling frequency so that it would be closer to webcamera-like eye movement videocameras probable in future computers.

It is rather obvious than potential that in the future human–computer interfaces will be extended to other tools in addition to a keyboard and mouse. Eye movement cameras can be used to follow targets of a subject's gaze on a computer screen. Further, we were interested in experimenting with the videocameras of the low resolution and sampling frequency to study whether it is possible to record eye movements accurately enough so that their sensitive features could be extracted, notwithstanding the low resolution and sampling frequency, from eye movement signals for the biometric verification of subjects. Since 30 Hz was far lower than 400 Hz used for EOG signals, we linearly interpolated eye movement signals given by the present video-oculograhpy (VOG) system. Twelve samples were interpolated between each pair of consecutive samples in a VOG signal because the rounded ratio of 400 and 30 Hz is 13. Virtually the same sampling interval, 2.5 ms, was then the time interval for latency (Fig. 3). The amplitude resolution was 0.5–1.0° for eye movement signals calibrated (built-in in the systems).

During all tests every subject was asked not to move the head. The duration of the EOG signals were 80 s, and that of the VOG signals 64 s. Since the latter included fewer stimulation movements of large amplitudes (over 40°) than the former, three VOG signals were recorded successively from every subject. The signals (20 s parts in Fig. 2) included stimulation movements following each other after 1–3 s. More or less randomly varying intervals between stimulations were important so that no subject could learn to guess when a stimulation movement appeared. Similarly, stimulation amplitudes (Fig. 3) were randomly changed. Not only maximum amplitudes, which were 60° for EOG and 48° for VOG signals, were stimulated, but also smaller to achieve a randomly varying stimulation sequence from a subject's viewpoint. Therefore, a subject could not learn such a stimulus sequence although it was repeated. The larger the amplitudes, the greater differences could be obtained for feature values computed from saccades [31,36–38]. Variation between subjects enabled separation between these in the recognition. If there had been no randomness in time intervals and amplitudes of stimulation movements, a subject could have attempted to anticipate some of stimulation movements. No anticipation was allowed, because anticipated saccades are no actual responses for stimulations. No saccade, with the latency less than 0.120 s, was incorporated into the data sets. This value was seen as the lower limit [31,36] for physiological reasons. Visual information is transferred via eyes into the brain to detect a stimulation and then information back as a "command" to move eyes. Thus a saccade appearing before its actual stimulation or before the limit of 0.120 s after the stimulation would have been rejected as anticipation.

Although smaller amplitudes (just used for variation to avoid anticipation) than either 60° for EOG or 48° for VOG signals were also stimulated, such small amplitudes were excluded in the actual biometric verification, since apparently the differences of the feature values of saccades with smaller amplitudes would be minor between subjects. Viz., feature values depend on stimulation amplitudes [25–27,29,36–38]. For instance, large saccade amplitudes yield greater maximum velocities than those of small amplitudes, since a saccade needs time to accelerate. Measurements of all test subjects included at least 12 large saccades (60° for EOG or 48° for VOG). Employing both measuring ways eye movements of 40 subjects were measured, 80 subjects altogether. All our subjects measured with VOG were healthy. Among those measured with EOG there were 19 healthy subjects and 21 otoneurological patients.

2.4. Signal analysis and final data for verification

The EOG eye movement signals were analyzed according to a method presented earlier, among others [31]. The VOG signals containing less noise than the EOG signals were analyzed with conventional signal analysis techniques. The main approach of both was to compute their angular velocity signals by approximating the first derivatives of the original position signals and from velocity signals to detect the locations where the velocity values rapidly increased or decreased corresponding to the beginnings or ends of saccades. A threshold velocity value of 10°/s was employed here. The beginnings of the stimulation movements were also detected being simple, because their intervals were known and their step-like signals were very regular (Fig. 2).

The EOG data consisted of $40 \times 12 = 480$ saccades from 19 voluntary, healthy subjects and from 21 otoneurological patients with the average age of around 50 years and roughly equally of both sexes. The VOG data also contained 12 saccades from each of 40 voluntary, healthy subjects (14 females and 26 males), whose average age was 29 ± 9 years. We tested different groups to include versatile data. After the recognition of every valid saccade its amplitude, accuracy, latency and maximum velocity were computed. The means and standard deviations of the amplitude, latency, accuracy and maximum velocity features were $52 \pm 11°$, 0.249 ± 0.137 s, $-5 \pm 9°$ (the negative accuracy corresponded to undershooting, i.e., smaller response amplitudes than stimulations) and $621 \pm 121°$/s in the EOG data and $47 \pm 11°$, 0.215 ± 0.055 s, $2 \pm 8°$ and $966 \pm 296°$/s in the VOG data. The greater maximum velocity of the latter is a consequence of the different recording systems.

Velocity values (approximated first derivative) of VOG data were computed with the formula of two-point central difference differentiator (m equal to 3) where x is an eye movement signal, v is a velocity signal to be computed, time interval T is equal to $1/f$ (f sampling frequency) and N is the number of signal samples:

$$v(l) = \frac{x(l+m) - x(l-m)}{2Tm}, l = m+1,..,N-m \quad (1)$$

There exist various manners to approximate derivatives. For instance, another is to compute linear regression in a sliding window through a signal and apply slope values given by its formula (optimal in the sense of least squares sums). To clarify these alternatives we also computed maximum velocity values on the basis of the two-point central difference differentiator with m equal to 1–4, when the window length was $2m+1 = 3$, 5, 7

Table 1
Means and standard deviation of maximum velocity [°/s] values computed with different formulas for 40 VOG signals.

Formula	Window length			
	3	5	7	9
Two-point central differentiator	987 ± 307	979 ± 303	966 ± 296	947 ± 286
Slope of linear regression	987 ± 307	981 ± 304	971 ± 299	959 ± 292

Table 2
Ratios between interindividual and intraindividual variations for the features of amplitude, latency, duration and maximum velocity: for the VOG signals results given for all differentiators and for the EOG signals results given separately for healthy subjects and patients.

Method and data set		Feature			
VOG	Window length	Amplitude	Latency	Accuracy	Maximum velocity
Two-point central differ. differentiator	3	0.71	0.44	0.62	0.68
	5	0.77	0.44	0.64	0.68
	7	0.77	0.44	0.64	0.68
	9	0.67	0.44	0.65	0.68
Slopes of linear regression	3	0.71	0.44	0.62	0.68
	5	0.73	0.44	0.64	0.68
	7	0.73	0.44	0.64	0.68
	9	0.71	0.44	0.65	0.68
EOG: Slopes of linear regression					
19 healthy	6	0.75	0.95	0.75	1.35
21 patients	6	0.45	0.98	0.49	1.36
40 subjects	6	0.71	1.44	0.58	1.44

and 9 samples, and with slopes of linear regression within the window length of 3, 5, 7 and 9 samples. Results obtained for the VOG signals are given in Table 1. Their differences were small. Thus we chose the alternative mentioned above (window length of 7 samples). Note how these differentiators function as lowpass filters: the longer the filter window, the slightly smaller maximum velocities [25,39].

In principle the approximation formula of velocity values also affected values of latency, duration and amplitude values slightly. However, for the VOG signals such differences were very small because the signals were interpolated from 30 to 390 Hz which made the interpolated signals virtually noiseless. Since no noise would affect the locations of beginnings and ends of saccades, effects of the velocity computation on latency, duration and amplitude values were negligible. Nevertheless, we applied those different formulas together with the following concept.

To examine the separation ability of the features we calculated ratios of interindividual and intraindividual variations in the following [40] where j denotes a feature, n equals the number of subjects, $\bar{u}_{ij}$ equals the mean of feature j of subject i, $\bar{a}_j$ is the mean of feature j of all subjects, u_{kj} is the value of feature j of saccade k of subject i and p_i is the number of the saccades of subject i. The greater the ratio, the better discriminating feature was found:

$$r_j = \frac{\sqrt{1/n \sum_{i=1}^{n} (\bar{u}_{ij} - \bar{a}_j)^2}}{(1/n) \sum_{i=1}^{n} \sqrt{1/p_i \sum_{k=1}^{p_i} (u_{kj} - \bar{u}_{ij})^2}} \tag{2}$$

Results given by the preceding formula are shown in Table 2 for all four features and separately for EOG and VOG signals. Moreover, results of EOG healthy subjects and patients are also shown. For the EOG signals velocities had earlier been computed with the slopes of linear regression using the window length of 6 samples. (At the beginning the EOG signals were also filtered with a median filter of length 11 samples [29,31], because these signals sampled at 400 Hz were relatively noisy compared to the VOG signals.) All result values were small indicating relatively large intraindividual variation. The ratios for latency and maximum velocity were greater for the EOG saccades than for the VOG saccades. Therefore, the EOG saccades would probably enable better classification results. Presumably this came from the higher original sampling frequency of the EOG signals. Nonetheless, we cannot be sure about this since the subjects of the two sets as well as their saccades were not the same. Hereafter we consider the signals of all EOG subjects together as one data set.

3. Recognition tests for the verification of subjects

Our objective was to design a technique which could be used to verify a legitimate subject, but at the same time be able to reject any other subjects as impostors. This approach can be divided into two test conditions. For the former we describe in the following how a legitimate user could be recognized. For the latter we present a test procedure how all other subjects could be rejected as impostors. In future the term called "non-user" corresponds to other subjects than a legitimate user. Some of non-users' saccades were employed to be one part of a training set, whereas the other part was based on saccades of a legitimate user. Because our data sets were measured from two rather limited groups of 40 subjects, we employed leave-one-out testing procedure that is suitable for small data sets. As normal in machine learning, the main idea was to vary randomly the contents of training sets in order to enable statistically reasonable results for tests. See Appendix 1.

In the preceding ways we built two test conditions to simulate that a computer user had to be verified through saccades and an impostor was not verified to be a legitimate user. These were two opposite conditions relevant to any biometric verification manner. We constructed two alternatives to vary a ratio between the numbers of saccades of a legitimate user and non-users in a training set. For alternative A we took $a=11$ saccades of a legitimate user (see variables from Appendix 1) and one saccade from each of $c=18$ non-users for condition 1. For condition 2 we then took $a=11$ saccades of a legitimate user and $2b=18$ saccades from $b=9$ non-users and tested with two saccades from each of 9 impostors. Note that impostors had to be disjoint from non-users, in other words, impostors' saccades could not exist in a training set since they had to be unknown to the verification system. For alternative B we increased the share of non-users in training sets. For condition 1 we now took $a=11$ saccades of each subject (as a legitimate user) and one saccade from each of $c=38$ non-users. For condition 2 we took $a=11$ saccades of a legitimate user and $2b=38$ saccades from $b=19$ non-users. We tested with one saccade from each of 20 impostors.

We repeated every test series $p=10$ times. Subject to both EOG and VOG data, for alternative A we obtained $10 \times 40 \times 12 = 4800$ tests for condition 1 and $10 \times 40 \times 18 = 7200$ tests for condition 2. For alternative B we then obtained again 4800 tests for condition 1 and $10 \times 40 \times 20 = 8000$ tests for condition 2.

For verification we applied k-nearest neighbor searching, linear and quadratic discriminant analysis and naïve Bayesian classification, where a saccade represented by its four features (amplitude, accuracy, latency and maximum velocity; Fig. 3) was compared to saccades of a training set, either from a legitimate user or other subjects (non-users). For k-nearest neighbor searching the number of the closest cases (k-nearest neighbors where k was an odd integer) classified into the class of the same subject (legitimate user) was compared to a priori probability of the class of this legitimate user as to be shown later. Classification was determined on the basis of such comparison. Of course, if the nearest neighbors were classified to be from the same subject as a case to be tested, this was the right recognition, but otherwise wrong. Closeness was computed by using Euclidean distance between saccades on the basis of four features computed. For linear and quadratic discriminant analysis and naïve Bayesian classification we ran tests by using majority vote for a test subject's saccades between two classes: a legitimate user and non-users.

Since the number of the saccades of a subject was small, we applied the leave-one-out principle suitable for the classification of small test sets: one by one, each saccade was the only case in a test set and other cases forming the cases of a training set to which the test case was compared.

For each test of condition 1 there were 12 large amplitude saccades (one of them alternately as a test case) from a legitimate user (every subject in turn) and one saccade from each of 18 other subjects (non-users) in alternative A. If a decision had been a pure guess, it would have been wrong, since its probability was greater, 18/29, of two possibilities. The task of the 11 saccades was to represent the class of a legitimate user in the feature space formed by the four real-value features and that of other 18 saccades to represent the class of non-users. Since obviously the latter with more subjects than one should include a larger part in the feature space, it was reasonable to incorporate more saccades into the class of non-users. For alternative B the above probability was 38/49, higher than that of alternative A. These a priori probabilities show that for condition 1 alternative A was easier for verification than alternative B, but for condition 2 this was vice versa.

According to the leave-one-out procedure we tested all saccades of each subject, one saccade by one, by computing Euclidean distances between the one selected and all other saccades of the same subject and then between the one and those of the other subjects (non-users). Following nearest neighbor searching technique we computed k nearest neighbors (saccades), k in {1, 3, 5, 7, 9, 11}. The k nearest neighbors were sorted according to whether they represented a legitimate user or non-users. Since there were $k \geq 1$ nearest neighbors due to be searched for, for condition 1 we classified a test saccade to represent a legitimate user provided that

$$\frac{x}{k \times t} > \frac{a-1}{a-1+b \times c} \tag{3}$$

where a equals the number of the saccades of a legitimate user in a training set, b those for each of c non-users, x the number of candidate saccades (nearest neighbors) from the class of a legitimate user, k the number of nearest neighbors searched for and t the number of tested saccades per subject. For condition 2 this formula was similar, but its comparison was opposite ($\leq$).

For other methods than nearest neighbor searching the majority of the test saccades determined the class: a legitimate user or non-users. We then computed the same tests using linear and quadratic discriminant analysis and naïve Bayesian rule.

We used feature values as both unnormalized and normalized into scale [0, 1]. Nevertheless, both alternatives yielded rather similar verification outcomes. All the test programs described were programmed with Matlab R2010a™ (MathWorks Inc., USA).

Table 3
Correct verifications of two data sets (EOG and VOG) with data normalization and two test conditions of (1) a legitimate user vs. non-users and (2) impostors vs. a legitimate user in percentages for the different odd numbers of nearest neighbors (saccades) found. Alternative A included smaller training sets than alternative B, in which the part of non-users was increased.

Verification results [%]	Number of nearest neighbors k searched for					
Data set and test condition	$k=1$	$k=3$	$k=5$	$k=7$	$k=9$	$k=11$
	Alternative A					
EOG(1)	76 ± 6	78 ± 8	85 ± 8	81 ± 8	78 ± 10	72 ± 6
VOG(1)	68 ± 7	78 ± 5	81 ± 7	84 ± 5	82 ± 4	81 ± 5
EOG(2)	61 ± 9	70 ± 8	64 ± 10	64 ± 9	57 ± 9	57 ± 5
VOG(2)	66 ± 8	70 ± 6	66 ± 10	61 ± 9	53 ± 7	54 ± 7
	Alternative B					
EOG(1)	72 ± 9	74 ± 4	80 ± 6	86 ± 5	84 ± 10	84 ± 12
VOG(1)	69 ± 7	69 ± 5	76 ± 6	80 ± 6	79 ± 8	81 ± 7
EOG(2)	59 ± 7	65 ± 9	61 ± 8	64 ± 11	63 ± 12	61 ± 7
VOG(2)	65 ± 7	65 ± 9	67 ± 6	65 ± 6	64 ± 11	63 ± 10

4. Results

The two test conditions and alternatives were run according to the procedures of the preceding section for both eye movement data sets including measurements from 40 subjects (different people in the sets). Using each subject's saccades $p=10$ test runs were repeated by randomly varying saccades of non-users used in training sets. The outcomes of nearest neighbor searching in percentages are given in Table 3 with the data normalization

Table 4
Correct verifications of two data sets (EOG and VOG) without data normalization and two test conditions of (1) a legitimate user vs. non-users and (2) impostors vs. a legitimate user in percentages for the different odd numbers of nearest neighbors (saccades) found. Alternative A included smaller training sets than alternative B, in which the part of non-users was increased.

Verification results [%]	Number of nearest neighbors k searched for					
Data set and test condition	$k=1$	$k=3$	$k=5$	$k=7$	$k=9$	$k=11$
	Alternative A					
EOG(1)	76 ± 5	82 ± 6	82 ± 8	80 ± 10	78 ± 9	71 ± 4
VOG(1)	70 ± 5	74 ± 7	80 ± 5	84 ± 7	83 ± 8	83 ± 5
EOG(2)	64 ± 7	66 ± 11	65 ± 8	62 ± 5	61 ± 11	58 ± 9
VOG(2)	69 ± 6	71 ± 7	68 ± 6	61 ± 8	56 ± 8	56 ± 9
	Alternative B					
EOG(1)	72 ± 9	75 ± 3	79 ± 7	85 ± 6	83 ± 10	83 ± 12
VOG(1)	69 ± 6	69 ± 6	77 ± 4	80 ± 7	79 ± 7	81 ± 7
EOG(2)	57 ± 10	65 ± 9	62 ± 10	66 ± 9	65 ± 9	59 ± 11
VOG(2)	61 ± 9	66 ± 9	66 ± 7	67 ± 8	63 ± 14	62 ± 8

and in Table 4 without this. The greatest k was equal to 11 since we used 12 saccades to represent each subject.

For test alternative A the results of test condition 1 simulating the verification for a legitimate user vs. non-users yielded 85% of EOG test subjects and 81% of VOG test subjects correct on the average for the best nearest neighbor searching test set-ups. Then their false rejection rates (FAR) also called Type I error or false negative rate were 15% and 19%. The results of test condition 2 simulating a legitimate user vs. non-users tested with the impostors seemed to be a more difficult classification problem according to the results in Tables 3 and 4. On the basis of k equal to 3 the average results of 70% correct with nearest neighbor searching were obtained. Then false accept rate also called Type II error or false positive rate was around 30%.

Results of linear and quadratic disrciminant analysis and naïve Bayesian classification are shown in Tables 5 and 6. Mostly the EOG results were better than the VOG results. Throughout Tables 5 and 6 for condition 2 the average results were approximately 10% better than the corresponding best results of nearest neighbor searching in Tables 3 and 4. Naïve Bayesian rule gave poor results, less than 50%, for the VOG data of alternative B and condition 1, but resulting in high average accuracies for condition 2. The latter phenomenon is as expected for the binary classification, since the a priori probability in Alternative B favored the larger non-users' class. The use of data normalization did not affect the results crucially although the scales of the features varied greatly.

Let us remember that we have to look at the results of conditions 1 and 2 of the same test set-up at the same time, because these are the opposite sides of testing: either a legitimate user or a random impostor. There were only small differences (a few percent) between the results of the corresponding test set-ups in Tables 3 and 4. Between test results in Tables 5 and 6 differences were 0–3%. The results gained with or without data normalization did not virtually differ from each other.

Table 5
Correct verifications of two data sets (EOG and VOG) with data normalization and two test conditions of (1) a legitimate user vs. non-users and (2) impostors vs. a legitimate user in percentages by linear discriminant and quadratic analysis and naïve Bayesian rule classification.

Verification results [%] Data set and test condition	Classifier: Linear discriminant analysis	Quadratic discriminant analysis	Naïve Bayesian rule
	Alternative A		
EOG(1)	89 ± 5	86 ± 9	90 ± 3
VOG(1)	68 ± 5	74 ± 7	66 ± 7
EOG(2)	82 ± 9	93 ± 5	91 ± 5
VOG(2)	72 ± 6	82 ± 7	84 ± 5
	Alternative B		
EOG(1)	86 ± 7	87 ± 8	79 ± 3
VOG(1)	63 ± 5	72 ± 7	44 ± 3
EOG(2)	91 ± 4	96 ± 5	99 ± 1
VOG(2)	88 ± 4	93 ± 5	99 ± 1

Table 6
Correct verifications of two data sets (EOG and VOG) without data normalization and two test conditions of (1) a legitimate user vs. non-users and (2) impostors vs. a legitimate user in percentages by linear discriminant and quadratic analysis and naïve Bayesian rule classification.

Verification results [%] Data set and test condition	Classifier: Linear discriminant analysis	Quadratic discriminant analysis	Naïve Bayesian rule
	Alternative A		
EOG(1)	89 ± 4	86 ± 9	90 ± 2
VOG(1)	66 ± 6	72 ± 5	68 ± 9
EOG(2)	83 ± 7	92 ± 6	90 ± 6
VOG(2)	74 ± 9	82 ± 5	89 ± 5
	Alternative B		
EOG(1)	87 ± 7	86 ± 7	78 ± 3
VOG(1)	64 ± 4	72 ± 8	45 ± 3
EOG(2)	91 ± 6	97 ± 4	99 ± 2
VOG(2)	87 ± 7	90 ± 6	99 ± 1

5. Conclusions

Our measurements included two types of eye movement signals, those recorded electro-oculographically (EOG) and those from image recordings of the videocamera system or video-oculography (VOG). The advantage of the former was its high sampling frequency (400 Hz), while that of the latter was fairly noiseless eye movement signals. The disadvantages were noisiness often present in the former and the low sampling frequency interpolated in the latter. Not only concentrating on the latter, which is naturally the motivation for the biometric verification and recognition for practical reasons, we also used EOG signals to compare their results and to estimate what increasing sampling frequencies of forthcoming VOG systems could produce.

In most situations the EOG measurements achieved better results on the average than the VOG measurements. We suppose that the higher original sampling frequency of the EOG signals resulted in better verification results when this property produces more accurate feature values compared to far lower sampling frequency of the VOG signals. Data normalization did not seem to affect results. As expected, alternative B was an easier approach for condition 2 than 1.

For Alternative A the best EOG results of 90% (condition 1) and 91% (condition 2) were obtained with naïve Bayesian rule, but the EOG results of quadratic discriminant analysis were virtually similar. Correspondingly, the best VOG results of 74% and 82% were obtained with quadratic discriminant analysis. For Alternative B the best EOG results of 86% (condition 1) and 97% (condition 2) were also computed with quadratic discriminant analysis as. Similarly this yielded the best VOG results of 72% and 93%. The results obtained are competitive with the following results. Eye movement identification rates were given in [22] average false accept rates of 1.4–17.5% and average false rejection rates 12.6–35.6% depending on classifiers were gained. For the former, nearest neighbor searching of k equal 7 gave the best results. For the latter, naïve Bayesian classification gave them. Average results of 4.8% for false accept rate and 9.4% for false rejection rate were later presented [23]. In [24] they obtained false accept rates of 5.4% and false rejection rates of 56.6%. In [6], the feature of distance between eyes produced the best results, 90% accuracy for identification. However, test alternatives including velocity values of eye movement signals without the previous eye distance gave lower values. Since the use of eye movement features only were not tested [6], these values cannot actually be compared to our results. The numbers of the subjects were 12 [6], 9 [22], 47 [23] and 41 [24].

Intervals between the stimulation movements in our VOG measurements were fairly long, almost 3 s at their largest. However, in principle intervals could be well cut down to approximately 1 s used, e.g., in [28,31]. Because around 30 saccades, from which perhaps 2/3 are of large amplitude saccades as 40°–60°, are sufficient to be applied to a verification process, it is possible to decrease the test duration to 60 s, perhaps even so short as 45 s. The actual verification process can be computed virtually in real time because of the fast classification methods and a small number of signal data. We used these targets described especially, because we knew that this was a natural and easy test set-up for subjects on the basis of our long-term research in medical informatics. Some "effort" is needed for any biometric technique. For example, to take a face or iris image one has to look at a camera. If eye movement verification data could be analyzed faster than two-dimensional image data, an eye movement test can be comparable as to the duration of the processing time with other biometric alternatives.

To the best of our knowledge, no other attempts have been published, in addition to the referred articles, to apply eye movements for the verification or recognition of subjects. Of course, there are several chances to extend our study, e.g., by collecting more data, and to improve our techniques. We applied the nearest neighbor searching, linear and quadratic discriminant and naïve Bayesian rule in the classification. These are perhaps some of the simplest classification methods. Their important advantage is that they are very straightforward to program and, in particular, fast for small data sets, which is essential for the use of such almost real-time computational tasks as biometric verification and recognition.

In principle the eye movement verification could be used in a device including a computer and eye movement camera system. Since saccades of the same subject can vary in the course of time, at least long time, their feature values could be occasionally updated to the computer or even regularly after a successful verification. Machine learning methods applied to our classifications are effective to adapt along with slightly changing data. Thus such a method would be used to verify a legitimate user. For instance, the maximum velocities of 10° saccades recorded with the infrared reflection technique from 58 subjects varied $\sim$2% on average between two weeks [33]. However, the maximum velocity feature is, as to our experience, the most sensitive of those used. It is calculated along with the noise-sensitive first derivative of an eye movement signal. Instead, the other features are directly calculated from an actual signal.

Fatigue, alcohol and aging are known to affect eye movements [25–39]. In order to prevent effects of fatigue one minute recording time for verification is certainly short enough, because far longer, such as several minutes, are used in medicine without visible fatigue. The abrupt effect of alcohol dosage may affect saccades. However, really greatly affecting fatigue or alcohol dosage would cripple behavior so clearly that the use of any biometric verification means would be clumsy in general. Aging is a slow process. The adaptive system of machine learning methods could follow it easily if it changed saccades. There are diseases that affect saccades, but they are infrequent [36–38]. In fact, in the present study we also used the otoneurological patients' EOG saccades affected by diseases that altered saccades on the average [31]. Notwithstanding this, both patients and healthy subjects of the EOG data were verified fairly reliably at their best. To conclude these issues have merely marginal influence from the practical viewpoint.

In principle the technique presented might be used even with small devices as mobile phones including a small eye movement camera in the future. If the width of the screen were 13 cm and the distance from a user's eye 15 cm, the angle between the edges of the screen would be approximately 47°, which would be sufficient for the verification purpose according to our tests described.

In the future we will continue our study by extending our data and especially by confirming our present results with other classification methods and test set-ups simulating a legitimate user's as well impostor's use. For instance, support vector machines are efficient classification algorithms for complicated recognition problems. They are also computationally fast, a crucial property for the biometric verification. Moreover, eye movements are interesting for the biometric use since eye movements of someone else can hardly be imitated or faked.

Acknowledgments

The authors are thankful to prof. Ilmari Pyykkö from the Department of Otorhinolaryngology, Tampere University Hospital, and docent Timo Hirvonen, M.D., and Heikki Aalto, Ph.D., from the Department of Otorhinolaryngology, Helsinki University Central Hospital, Finland, for medical guidance in eye movements, physiology and otoneurology, and help in recording signals. The research of the second author was supported by the Center for International Mobility, Finland, and by Tampere Doctoral Program in Information Science and Engineering (TISE) and that of the third author was supported by the Academy of Finland (grant 115609).

Appendix 1. verification procedures

Condition 1:
For $i=1,\ldots,p$ **do**// p iterations are accomplished to calculate averages
  For $j=1,..,n$ **do**// n subjects are tested, each of them alternately as legitimate one
    For $k=1,\ldots,a$ **do**// leave-one-out procedure
      Create a training set by taking $a-1$ saccades of a subject j except the kth saccade and choose randomly b saccades from every of c random subjects l (non-user), $l=1,\ldots,c$, $c \leq n-1$ $(l \neq j)$.
      Create a test set by taking the kth saccade of subject j (excluded from the training set) as a single test case.
      Test by classifying with some method and **check** whether its verification outcome was right (legitimate user) or not.
    End
  End
  Calculate the numbers of right and wrong verifications.
End
Calculate the means and standard deviations of right and wrong verifications of p iterations.
Condition 2:
For $i=1,\ldots,p$ **do**// p iterations are accomplished to calculate averages
  For $j=1,\ldots,n$ **do**// leave-one-out procedure
    Create a training set by taking a saccades of subject j and select randomly $2b$ saccades from every of $d=c/2$ (c even) subjects k (non-user) randomly extracted, $k=1,\ldots,d$, $c \leq n-1$ $(k \neq j)$.
    For $l=d+1,\ldots,c$, $c \leq n-1$, $l \neq j$ **do**// testing impostors
      Create a test set by randomly extracting an impostor's (the lth subject) saccade as a single test case.
      Test by classifying with some method and **check** whether its outcome was right or wrong.
    End
  End
  Calculate the numbers of right and wrong verifications
End
Compute the means and standard deviations of right and wrong verifications for p iterations.

References

[1] D.C. Leonard, A.P. Pons, S.S. Asfour, Realization of a universal patient identifier for electronic medical records through biometric technology, IEEE Trans. Inf. Technol. Biomed. 13 (4) (2009) 494–500.
[2] A. Ross, A. Jain, Information fusion in biometrics, Pattern Recognit. Lett. 24 (13) (2003) 2115–2125.

[3] A. Jain, Biometric recognition: how do I know who you are? in: Proceedings of 14th Scandinavian Conference on Image Analysis, Joensuu, Finland, Springer-Verlag, Berlin, LNCS 3540, 2005, pp. 1–5.
[4] R. Chellappa, P. Sinha, P.J. Phillips, Face recognition by computers and humans, IEEE Comput. (2010) 46–55.
[5] M.K.H. Leung, A.C.M. Fong, S.C. Hui, Palmprint verification for controlling access to shared computing resources, IEEE Pervasive Comput. 6 (4) (2007) 40–45.
[6] R. Bednarik, T. Kinnunen, A. Mihaila, P. Fränti, Eye-movements as a biometric, in: Proceedings of 14th Scandinavian Conference on Image Analysis, Joensuu, Finland, Springer-Verlag, Berlin, LNCS 3540, 2005, pp. 780–789.
[7] S.-D. Bao, C.C.Y. Poon, Y.-T. Zhang, L.-F. Shen, Using the timing information of heartbeats as an entity identifier to secure body sensor network, IEEE Trans. Inf. Technol. Biomed. 12 (6) (2008) 772–779.
[8] S.A. Israel, J.M. Irvine, A. Cheng, M.D. Wiederhold, B.K. Wiederhold, ECG to identify individuals, Pattern Recognit. 38 (2005) 133–142.
[9] G. Wübbeler, M. Stavridis, D. Kreiseler, R.-D. Bousseljot, C. Elster, Verification of humans using the elctrocardiogram, Pattern Recognit. Lett. 28 (2007) 1172–1175.
[10] Y. Wang, F. Agrafioti, D. Hatzinakos, K.N. Plataniotis, Analysis of human electrocardiogram for biometric recognition, EURASIP J. Adv. Sig. Proc. (2008), http://dx.doi.org/10.1155/2008/148658.
[11] J.M. Irvine, S.A. Israel, A sequential procedure for individual identity verification using ECG, EURASIP J. Adv. Sig. Proc. (2009), http://dx.doi.org/10.1155/2009/243215.
[12] F. Sufi, I. Khalil, Faster person identification using compressed ECG in time critical wireless telecardiology applications, J. Netw. Comput. Appl. 34 (1) (2011) 282–293.
[13] R.V. Yampolskiy, V. Govindaraju, Behavioural biometrics: a survey and classification, Int. J. Biometrics 1 (1) (2008) 81–113.
[14] J. Liang, C. Chen, H. Zhao, H. Hu, J. Tian, Gait feature fusion using factorial HMM, in: Behavioral Biometrics for Human Identification Intelligent Applications IGI Global, 2010, pp. 189-206.
[15] A.A.E. Ahmed, I. Traore, Mouse dynamics biometric technology, in: Behavioral Biometrics for Human Identification Intelligent Applications, IGI Global, 2010, pp. 207-223.
[16] C. Fookes, A. Maeder, S. Sridharan, G. Mamic, Gaze based personal identification, in: Behavioral Biometrics for Human Identification Intelligent Applications, IGI Global, 2010, pp. 237-263.
[17] C. Spampinato, Visual attention for behavioral biometric systems, in: Behavioral Biometrics for Human Identification Intelligent Applications, IGI Global,2010 pp. 290-316.
[18] A.T. Bahill, F.K. Hsu, L. Stark, Glissadic overshoots are due to pulse width errors, Arch. Neurol. 35 (1978) 138–142.
[19] A.T. Bahill, B.T. Troost, Types of saccadic eye movements, Neurology 29 (1979) 1150–1152.
[20] R.W. Baloh, V. Honrubia, Neurophysiology of the Vestibular System, F.A. Davis Company, Philadelphia, 1979.
[21] M. Nishigaki, D. Arai, A user authentication based on human reflexes using blind spot and saccades response, Int. J. Biometrics 1 (2) (2008) 173–189.
[22] P. Kasprowski, J. Ober, Eye movements in biometrics, in: Proceedings of Biometric Authentication: International Workshop, Prague, Czech Republic, Springer-Verlag, Berlin, LCNS 3087, 2004, pp. 248–258.
[23] A. Kapczyński, P. Kasprowski, P. Kuźniacki, Modern access control based on eye movement analysis and keystroke dynamics, in: Proceedings of International Multiconference on Computer Science and Information Technology, Wisla, Poland, 2006, pp. 477–483.
[24] O.V. Komogortsev, S. Jayarathna, C.R. Aragon, M. Mahmoud, Biometric identification via an oculomotor plant mathematical model, Proceedings of 2010 Symposium on Eye Tracking Research & Applications, Austin, TX, 2010, pp. 57–60.
[25] M. Juhola, V. Jäntti, I. Pyykkö, Effect of sampling frequencies on the maximum velocity of saccadic eye movements, Biol. Cybern. 53 (1985) 67–72.
[26] M. Juhola, A syntactic method for analysis of saccadic eye movements, Pattern Recognit. 19 (1986) 353–359.
[27] M. Juhola, Detection of nystagmus eye movements using a recursive digital filter, IEEE Trans. Biomed. Eng. 35 (1988) 389–394.
[28] H. Aalto, M. Juhola, Pyykkö, Visual stimulator using laser beam for eye-movement studies, Med. Biol. Eng. Comput. 27 (1989) 332–335.
[29] M. Juhola, Median filtering is appropriate to signals of saccadic eye movements, Comput. Biol. Med. 21 (1991) 43–49.
[30] T. Hirvonen, H. Aalto, M. Juhola, I. Pyykkö, A comparison of static and dynamic calibration techniques for signal of vestibulo-ocular reflex, Int. J. Clin. Monit. Comput. 12 (1995) 97–102.
[31] M. Juhola, H. Aalto, T. Hirvonen, Using results of eye movement signal analysis in the neural network recognition of otoneurological patients, Comput. Methods Programs Biomed. 86 (2007) 216–226.
[32] M. Juhola, H. Aalto, T. Jutila, T. Hirvonen, Signal analysis of three-dimensional nystagmus for otoneurological investigations, Ann. Biomed. Eng. 39 (3) (2011) 973–982.
[33] E. Bollen, J. Bax, J.G. van Dijk, M. Koning, J.E. Bos, C.G.S. Kramer, E.A. van der Velde, Variability of the main sequence, Invest. Ophthalmol. Vis. Sci. 34 (13) (1993) 3700–3704.
[34] A.T. Bahill, A. Brockenbrough, B.T. Troost, Variability and development of a normative data base for saccadic eye movements, Invest. Ophthalmol. Vis. Sci. 21 (1) (1981) 116–125.
[35] A.T. Duchowski, A breadth-first survey of eye tracking applications, Behavior Research Methods, Instruments Comput. 34 (4) (2002) 455–470.
[36] R.W. Baloh, W.E. Kumley, V. Honrubia, Algorithm for analyses of saccadic eye movements using a digital computer, Aviat. Space Environ. Med. 47 (5) (1976) 523–527.
[37] R.W. Baloh, A.W. Sills, W.E. Kumley, V. Honrubia, Quantitative measurement of saccades amplitude, duration, and velocity, Neurology 25 (11) (1975) 1065–1070.
[38] R.W. Baloh, H.R. Kondrad, A.W. Sills, V. Honrubia, The saccade velocity test, Neurology 25 (11) (1975) 1071–1076.
[39] A.T. Bahill, J.S. Kallman, J.E. Lieberman, Frequency limitations of the two-point central difference differentiation algorithm, Biol. Cybern. 45 (1982) 1–4.
[40] Y.Y. Gu, Y. Zhang, and Y.T. Zhang, A novel biometric approach in human verification by photoplethysmographic signals, in: Proceedings of 4th Annual IEEE Conference on Information Technology Applications in Biomedicine, UK, 2003, pp. 13–14.

Publication III

Biometric verification of a user based on eye movements

Youming Zhang and Martti Juhola

Biometric verification of a user based on eye movements

Youming Zhang and Martti Juhola

Computer Science, School of Information Sciences, 33014 University of Tampere, Finland

E-mail: Youming.Zhang@uta.fi, Martti.Juhola@sis.uta.fi

Abstract: The biometric verification of users of computers or other machines is usually performed with fingerprints, face images or even iris or palm images. Eye movements have seldom been studied for biometric verification, although in the future their use will perhaps extend from laboratory applications to integrated parts of computer interfaces. Eye movements have long been studied in medical and psychological applications. We have noticed that there are differences between saccade eye movements of individuals, even in a group of young people approximately of the same age. We measured saccades from 68 voluntary subjects by performing the same stimulation for each to obtain comparable data. We tested two verification conditions: (1) an authenticated user vs. all other subjects and (2) an impostor vs. an authenticated user and others. Thorough randomized classifications with discriminant analysis, *k-d* tree and *k* nearest-neighbour searching, decision trees and the naïve Bayesian rule provided good verification results at best, but the best results of all were obtained with logistic discriminant analysis.

Keywords: Eye movements; saccade; user verification; signal analysis; classification

Biographical notes: Youming Zhang, M.Sc., is a Ph.D. student of computer science. His research areas are signal analysis, data mining and biometrics. Martti Juhola, Ph.D., is a professor at the University of Tampere. His research interests include biomedical signal analysis, biometrics, pattern recognition, machine learning and data mining.

1 Introduction

Several computerized biometric authentication methods have been researched and developed in the past 20 years. Perhaps best known are fingerprints (Conti et al., 2010) and face images (Pentland and Choudbury, 2000; Frischholz and Dieckmann, 2000). In addition, retina, iris (Negin et al., 2000; Sun et al., 2005) and palm print (Veeramachaneni, Osadciw, and Varshney, 2005) scanning; voice signals have been studied as well. Not only the reliable verification of a subject is essential, but spoofing attacks and counterfeits must be detected. Fingerprints and face images have been studied frequently and are seen as promising. Nevertheless, there are also problems in reliably recognizing faces, for example, because of changing factors such as illumination, glasses and hairstyle. Iris images obviously separate individuals well, but both iris and retinal images are somewhat complicated to measure.

Palm vein patterns seem to be effective (Boatwright and Luo, 2007), since they are ostensibly different from one individual to the next. Voice (Frischholz and Dieckmann, 2000) may be interesting, but is sensitive to background noise and other circumstancial factors (Boatwright and Luo, 2007). The voices of same-sex parent and child can sometimes resemble each other. Voice biometric data sets are often relatively large, as is also the case with images. Therefore, if it is possible to find novel physiological measurements that contain smaller data quantities, a verification procedure might become simpler and perhaps computationally faster. Information originated from images frequently comprises a great number of features: e.g. a face image could require as many as 100 different features. As a result, image data analysis is typically a complicated process in respect of the demand for a rapid verification service. Of one-dimensional signals, EEG (Nakayama and Abe, 2012) and ECG (Israel and Irvine, 2012) have been researched with a view to use in biometrics, but obviously less than image data.

In this research, we followed the definition that the verification of the identity of a user of a computer or other device is separating a certain subject from the rest of a group. On the other hand, another person, an impostor, who might attempt to use the computer of the authenticated person, should not be able to log into the computer. Identification is defined as a more extensive task in which each subject of a group can be separated from anyone else in the group of C individuals, where C is the number of individuals in the group. We characterize the former to be two-class or binary classification and the latter C-class or multiclass (C>2) classification task. For example, the latter is needed for the identification of criminals. For a computer user recognition or equivalent task, however, verification is sufficient. A novel biometric technique should also be impossible to forge or steal. Passwords are easy to acquire if an impostor finds the written password of a user, tricks it out of him or her, or sees at the writing moment, for instance, which keys an authenticated user presses. We emphasize that our research concerns verification, not identification.

Eye movements are an interesting and potential objective for biometric recognition purposes. They have been investigated for decades in various medical fields (Bahill, Brockenbrough and Troost, 1981; Bollen et al., 1993; Fricker and Sanders, 1975; Schmidt et al., 1979), in related areas (Joyce, 2002; Juhola, Jäntti and Pyykkö, 1985; Juhola, Aalto and Hirvonen, 2007; Salvucci and Goldberg, 2000; Tweed and Vilis, 1990), and in psychology (Allik, Rauk and Luuk, 1981; Underwood, Bloomfield and Clews, 1988). In the recent years, eye movements have also been studied with a view to developing human-computer interfaces (Hyrskykari, 2006; Majaranta, 2009). It is useful to attempt to utilize such cumulated knowledge. For instance, we can draw the conclusion that mostly time domain rather than frequency domain variables have been used to evaluate the influence of diseases and physiological disorders on saccadic eye movements or to detect these movements.

Our interest is directed towards saccades, since these eye movements are especially relevant for our current research. The brainstem is partly responsible for controlling the saccades of a human being. Some variables, for example latency, reaction time or angular velocity, are used to characterize the form and properties of saccades measured in the respect of time (Abrams, Meyer and Kornblum, 1989; Sparks, 2002). They have been extensively and increasingly studied in many areas such as physiology and psychology, but not in biometrics. Saccade latency has been explored in children and adults, and children's latencies have been found to be longer than those of adults (Yang, Bucci and Kapuola, 2002). Infants between four and eight months old were studied to determine the latency values of their saccades, which became shorter just during these four months (Gredebäck, Örnkloo and von Hofsten, 2006). Saccade trajectory deviations were studied for their association with visual attention, with the latency and endpoint of a saccade observed to determine its trajectory (Van der Stigchel, 2010). In this area, the remote distractor effect is frequently studied. Saccades may be delayed, with longer latencies than usual, by the appearance of distractor in a subject's visual field, even though the target location of the gaze is fully fixed (McIntosh and Buonocore, 2012). Effects of suddenly appearing visual stimulations or distractors on saccades were studied, and various effects were found, depending on when and where in the visual field the distractor appeared (Edelman and Xu, 2009). Saccades have also been used to investigate visual memory, with subjects whose eye movements were recorded to evaluate their observations after deletions and changes of objects in a scene (Henderson and Hollingworth, 2003). In addition, the processing of shape information of subjects in human peripheral visual fields has been studied with saccades: Nandy and Tjan (2012) looked at problems of visual crowding, i.e. the inability to identify shapes of targets in peripheral vision.

Eye movements can be measured sufficiently precisely with small eye-movement cameras so that saccades are possible to detect reliably. Saccades are the fastest eye movements, but probably also the simplest to detect and recognize. Of other types of eye movements, smooth pursuit could be considered, but they are longer and often include small-amplitude corrective saccades (the brain "automatically" correcting the direction of movement). Nystagmus is a repeated "sawtooth" reflexive movement that must first be stimulated in various ways: e.g. by seating the subject in a rotating chair that is then stopped abruptly. Its stimulation is complicated from a practical point of view, and nystagmus responses from the same individual may vary considerably. Especially since it is involuntary, it cannot be used here. Actually, saccades are obviously the most suitable type. They are very natural for us, since looking at almost anything in our surroundings involves mostly saccades. To read, we use a sequence of saccades, moving our gaze from one unit of a few letters to the next, sometimes returning to the previous unit or changing rows. Saccades are seen as paramount to observing the surroundings in several situations that depend on visual information. For example, sight is the most important sense in traffic.

Eye movements, particularly saccades (including fixations), are also being studied for human-computer interfaces, which means that eye movement camera technologies are advancing as well. It follows naturally, then, to look at using eye movements in the biometric verification of subjects, although very few attempts have as yet been made to use eye movements for biometric purposes. In other studies (Kapczyński, Kasprowski and Kuźniacki, 2006; Kasprowski and Ober, 2004), researchers recorded saccade signals and used them to compute cepstrum signals. They classified their signal analysis results using naïve Bayesian rules, nearest-neighbour searching, decision trees and support vector machines. Pupil size, gaze velocity and distance between subjects' eyes were also used (Bednarik et al., 2005), in which the fast Fourier transform and principal component analysis were run for the saccade signals measured;

nearest-neighbour searching was used for the classification of subjects. Altogether, the technique (Bednarik et al., 2005) was mainly based on distances between the eyes of subjects using image analysis, since the other aforementioned variables did not provide promising outcomes. Thus, this technique cannot really be applied with eye movements. In a recent article (Komogortsev et al., 2010), the authors introduced an approach on the basis of a computational model of the oculomotor plant derived from an earlier model (Bahill, 1980). These techniques model the function of the six muscles that rotate the eyeball during eye movements. The input of their recognition procedure was formed from parameter vectors of the model, and they used nearest-neighbour searching and decision trees provided by the C4.5 method for classification. They reported better results, i.e. fewer false acceptance rates (FAR) and false rejection rates (FRR) than earlier studies (Kasprowski and Ober, 2004) when nearest-neighbour searching was used for both. However, their test results indicated that the C4.5 decision tree method was not promising (Komogortsev et al., 2010). Recently, a new technique for biometric identification based on eye movements was published in which face images were shown to subjects whose eye movements were monitored at the same time (Rigas, Economou and Fotopoulos, 2012). They used minimum spanning trees derived from graphs of the fixation points on the plane and classified using nearest-neighbour searching and support vector machines.

In the present article, we introduce a method to employ saccades in verification of a user. The method is very different from those in previous studies, since our approach is based on physiological variables of saccade eye movements, such as those that have long been applied in medical investigations (Bahill, Brockenbrough and Troost, 1981; Bollen et al., 1993; Fricker and Sanders, 1975; Schmidt et al., 1979) and related areas (Gredebäck, Örnkloo and von Hofsten, 2006; Joyce, 2002; Juhola, Jäntti and Pyykkö, 1985; Juhola, Aalto and Hirvonen, 2007; Yang, Bucci and Kapuola, 2002).

The present research was started concurrently with other recently published articles in which we used smaller data sets measured with the same device as the present data set combined with data sets measured electro-oculographically with skin electrodes (Zhang, Rasku and Juhola, 2012; Juhola, Zhang and Rasku, 2013). The latter, old measuring method was interesting for comparison purposes because it enabled the use of a high sampling frequency. In addition, we made a third measurement series with the same device as now (Zhang and Juhola, 2012). However, measuring set-ups and the majority of classification methods and their training arrangements as used in those former studies were different from those applied in the present research. Some of them, such as linear and quadratic discriminant analysis, were also utilised this time for comparison. In the present research, we started our signal analysis phase by comparing three different interpolation techniques to “artificially” increase the originally low sampling frequency.

2 Saccade eye movement measurements

We measured saccade signals using a two-camera eye movement detection system (Visual Eyes, Micromedical Technologies, UK) with an image resolution of 320×240 and a sampling frequency (frames per second) of 30 Hz. As was the case before, the system detected positions of a pupil in images and used them to compute eye movements. Both horizontal and vertical movements could be registered, but to make the approach as straightforward as possible we used horizontal movements only. This is practical for a subject and makes a verification test faster. On the other hand, the larger the data set we have at our disposal, the more information we can get from a subject’s eye movements. The 30 Hz sampling frequency was low compared to other sampling frequencies used for eye movement cameras, i.e. 50 Hz, 60 Hz, 120 Hz or even 1000 Hz. One aim of our research was to simulate the use of web cameras, whose sampling frequencies are usually low. It would be useful if user verification could be employed in this context.

Figure 1. A subject following the stimulation movement (horizontally jumping light dot) with his gaze.

The same stimulation series was run for each subject to make tests comparable to each other. Subjects were asked to watch a jumping small red light (Figure 1). The light was an abruptly switched-on LED in the black bar in front of a seated subject. The same distance from the LED bar was kept for the chair and every subject, whose alertness was checked. Neither alcohol nor medication were permitted during a measurement day prior to testing. Only one LED of the row was switched on at a time. When it was switched off by the measurement computer, another was switched on so that the subject moved his or her gaze between them, making a saccade to the new position after each change. Stimulation amplitudes (angles) and intervals between stimulation movements were varied so that a subject could not anticipate a movement, which is important in obtaining valid responses to genuine stimulation movements rather than extraneous movement without any stimulation. Anticipation would occur if a saccade preceded the temporally closest stimulation or followed this earlier than 0.120 s, which is considered the minimum latency (Juhola, Aalto and Hirvonen, 2007). A subject could not learn the stimulation series even after several repetitions because of variations in stimulation directions, intervals and amplitudes and the number of saccades to be made. Such stimulation set-ups have been used in medical applications for decades (Henriksson et al., 1980; Juhola, Jäntti and Pyykkö, 1985; Juhola, Aalto and Hirvonen, 2007).

With the present approach, we simulated a situation in which a computer user would sit down to start the computer and to wait for it to boot up. This was assumed to end with the stimulation light jumping on the screen as described above roughly 20 times to measure the user’s eye movements and using neither user identifier nor password to verify him or her as either an authenticated user or not. A human being is curious. It is inherent for humans to follow the “events” on a screen, especially when nothing else catches their attention. Naturally, the whole process should be as quick as possible.

To enable good circumstances for prospective data analysis, it was necessary to repeat several large-amplitude saccades in the stimulation series. Relatively large saccade amplitudes, preferably at least 40°, were required because large stimulation amplitudes can produce sufficient differences between variable values of saccades in different subjects (Juhola, Aalto and Hirvonen, 2007). There may also be some intraindividual variation for some saccade variables, at least maximum velocity (Bollen et al., 1993). Notwithstanding this, interindividual variation seemed to be more prominent because we could detect individuals from among the whole set of subjects in our verification process. Results indicated that maximum velocities showed no statistically significant differences when 5 ° and 10 ° saccades of 58 healthy subjects were measured on two occasions with two weeks in between (Bollen et al., 1993). This supported the thought that saccades from the same individual do not vary greatly over the course of weeks. Thinking our verification simulation of a computer user, this is a long enough time, since variable values of saccades of an authenticated user could be collected to a buffer from every accepted login. A buffer including perhaps the last ten measurements would be updated along with each accepted login, which would adapt to slow average changes.

Since the sampling frequency was low, only 30 Hz, we approximated saccades using linear and other interpolation techniques in order to raise the number of samples per second as high as 33-fold to correspond to approximately 1000 Hz. In principle, 300-400 Hz would be enough to precisely express important physiological features of saccades such as maximum angular velocity (Juhola, Jäntti and Pyykkö, 1985). Nevertheless, to measure saccades very accurately, even 1000 Hz is also used (Bahill, Brockenbrough and Troost, 1981) in laboratory studies. The high precision was desirable in order to attempt to increase the success rate of verification. On the other hand, the 33-fold interpolation is heavy, which might even overstate some values. Recently (Wierts, Janssen and Kingma, 2008), it was shown that although a low sampling frequency of 50 Hz typical to eye movement cameras was applied, accurate maximum velocities of saccades were obtained for saccades with amplitudes 5 °-28 °. Because amplitudes of our saccades employed were around twice their maximum, after all this should mean a favourable situation. The result (Wierts, Janssen and Kingma, 2008) was obtained with a modern eye movement camera, but our old one (Juhola, Jäntti and Pyykkö, 1985) with noisy electro-oculographic signals compared to signals measured with videocameras.

Since our objective was to verify an authenticated user, it was necessary to separate him or her from the others, called non-users, using the values of their saccade variables. Our main aim was not to attempt to achieve as correct and precise variable values as possible (with a high sampling frequency of a special device) as such for them, but rather to calculate estimates sufficient to distinguish them from each other. Actually, it is no easy question to determine precisely what the correct values are, as they depend on many factors such as measuring method and device and the computational methods used.

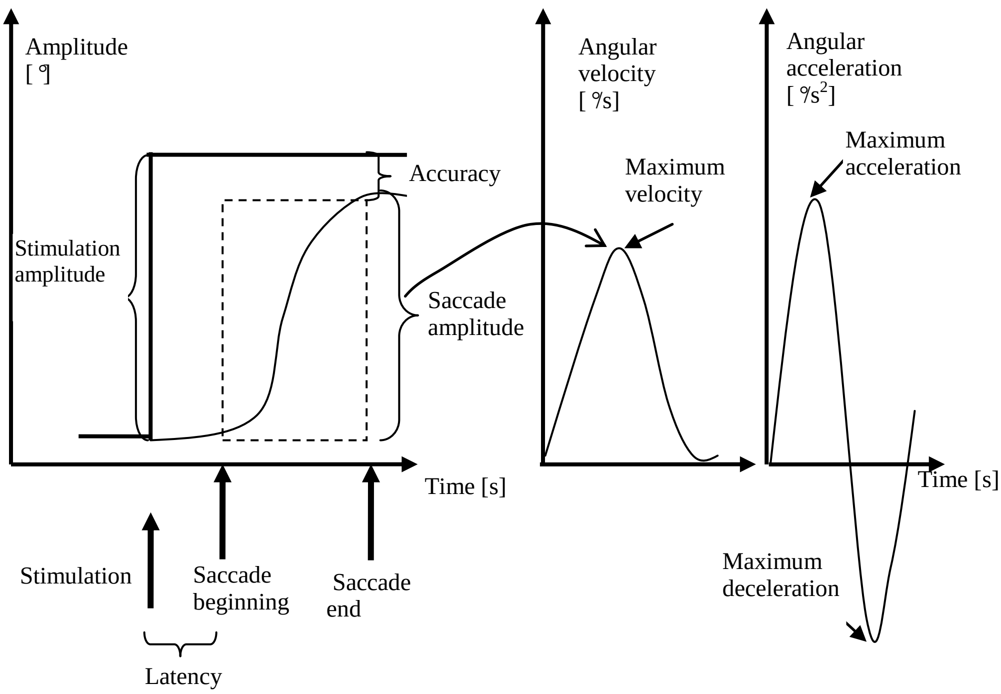
3 Signal detection and data preparation for verification

The beginning and end of each saccade were detected as precisely as possible. At first, the angular velocity of a saccade signal was approximated as the first derivative of a signal. As usual with eye movement signals, the beginning of a saccade was found when velocity magnitude began to increase and its end correspondingly when its velocity magnitude began to decrease close to a small value. The velocity threshold value of 50 °/s was employed; all saccade variables were computed from responses to fast step-like stimulation movements. It usually takes an average of about 0.2 s to generate a saccade (Juhola, Aalto and Hirvonen, 2007; Yang, Bucci and Kapuola, 2002). During this time, visual information, after observation of the stimulation, is transferred from the eyes to the brain and processed there, and control information is transferred to the two sets of six ocular muscles that move the eyes.

We took the largest 20 saccades from each subject by using the same stimulation series for each. Five measurements were repeated for each subject, and four saccades larger than 40 ° were taken from each of five measurements. Using several measurements from every subject, our motivation was also to test and ensure that this would not impair verification results. In principle, slight calibration variations between different measurements could affect values of saccade variables, which could result in the larger intraindividual variation of saccade values than within a single measurement of a subject. Altogether, 68 voluntary subjects (16 females and 52 males) with the mean age of 25±5 years were measured. The objective was to test quite young people close to the same age so that our data source would form a homogeneous basis so that we could test as difficult verification cases as possible. Namely, we assumed that age could affect saccades.

Saccade variables calculated on the basis of accepted saccades larger than about 40 ° are described in Figure 2. All these variables were computed in time domain, since we knew that most of them have commonly been used in various medical tests since the 1970s at least. These variables reflect the physiological condition of subjects with respect to their ability to perform saccades and follow their surroundings rapidly.

Figure 2. A hypothetical saccade curve and its stimulation. Variables computed for verification: saccade amplitude, accuracy, latency, maximum angular velocity (approximated first derivative of saccade curve), maximum angular acceleration and maximum angular deceleration (approximated second derivative).



Means and standard deviations in the accepted saccades of all 68 subjects were as follows: 0.265±0.056 s for latency, 1010±300 °/s for maximum velocity, 47±13 ° for amplitude, 5±9 ° for accuracy, 46700±22500 $°/s^2$ for maximum acceleration and 43800±24000 $°/s^2$ for maximum deceleration. Before computation of the results, cubic spline interpolation was applied to the saccades. When means and standard deviations varied among the individuals measured, this denoted an opportunity to separate between the subjects. All signal and verification computations were performed using Matlab R2010a™ (MathWorks Inc., USA).

4 Verification computation

We formed two verification conditions to experiment with the data collected. The purpose of Condition 1 (Figure 3) was to test whether an authenticated user could be detected as such. Condition 1 corresponded to a test of false rejection. The data set used for this machine learning task was created by taking all 20 saccades of an authenticated user and various samples from other subjects (called non-users) to test whether results might depend on the ratio of the sizes of these two subsets. One subset represented the saccades of an authenticated user and the other subset corresponded to those of the non-users. In the space of the six saccade variables, the non-users represented the area outside an authenticated user's saccades. For the purpose of testing, every subject was alternately in the role of an authenticated user and in the group of those other non-users represented.

Figure 3. Verification Condition 1 in a hypothetical two-dimensional variable space. An authenticated user's saccades should be more similar to each other than to those of randomly selected saccades of the other subjects (non-users), whose task is to represent the variable space outside the authenticated subject's saccades.

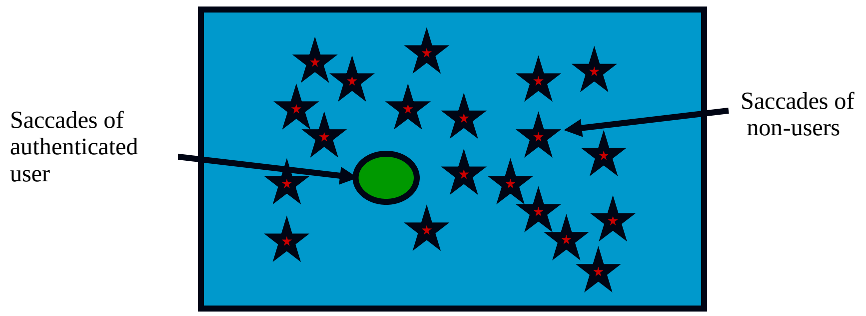
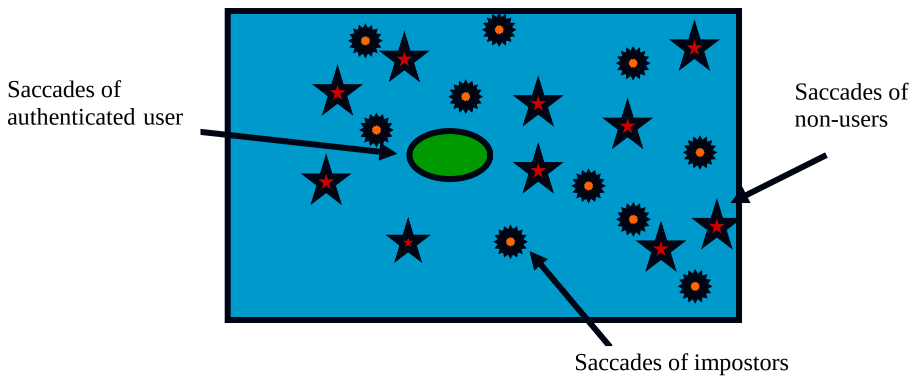


Figure 4. Verification Condition 2 in a hypothetical variable space. Impostors' saccades should be more similar, on the average, to those of non-users in the training set than the saccades of an authenticated user.



Verification Condition 2 (Figure 4) corresponded to the situation of an impostor attempting to log in: in other words, a test of false acceptance. We then divided the subjects into three subsets: an authenticated user, the subset of non-users to build the part outside him or her in the variable space of a training set, and a distinct subset representing impostors. Each of the last two subsets was roughly a half of the whole set, the former being 30 and the latter 37 subjects at a time. Impostors had to be outside each training set, because we did not "know" their saccades as we knew those of non-users used to form that training set together with the saccades of an authenticated user.

We ran all tests by applying the leave-one-out technique that is suitable in machine learning validations with relatively scarce data, since the size of each training set was then maximized to n-1 cases (saccades) out of the total of n. One by one, each saccade from the whole set selected was the only test case of the current test round, and all other selected saccades formed the training set. Each of 20 saccades of an authenticated user was tested versus those 19 of the same user and those of the others in the training set in Condition 1. The result of a test was correct if a test saccade of an authenticated user resembled more his or her own saccades than those of non-users: in other words, if the test saccade was classified as being in its own class.

In Condition 2, impostors' saccades were classified against all of the training set in the hope that they would resemble, on average, more the saccades of non-users than those of the authenticated user in the training set. Saccades taken from an impostor were tested and their values calculated, whether their verification results were correct or not.

Before actual verification tests, we were interested in experimenting with different techniques for interpolation of saccades. The simplest choice was to run with linear interpolation. However, we could expect it to be too straightforward because of the curvature of saccades and therefore it was best to also test other alternatives: a cubic spline curve and a piecewise cubic Hermite interpolating polynomial. For these we used linear discriminant analysis to find which of the interpolation techniques would result in the best verification results. For actual tests we used discriminant analysis, nearest-neighbour searching with exhaustive searching or *k-d* search trees, decision trees and the naïve Bayesian rule.

Every test set-up was repeated t=10 times by randomly picking up b saccades from c non-users, other than an authenticated user, whose all a=20 saccades were involved in the whole set. We ran 10 times n=68 (each was the authenticated user in turn) test series for both verification conditions. The next procedures describe our test implementations.

To make classification majority vote was computed: the majority of test saccades of a subject determined a decision, the class of an authenticated user or that of non-users. For nearest-neighbour searching with exhaustive searching or *k-d* trees the following formula was applied to Condition 1 since k nearest neighbours were searched for each test saccade:

$$x/k\,r > (a-1)/(a-1+b\,c)$$

Numbers a and b were mentioned above, x is the number of the saccades classified into an authenticated user's class, k the number of nearest neighbours searched for and r the number of saccades tested (=a for Condition 1 and =n-c-1 for Condition 2). For Condition 2 the inequality was opposite ($\leq$).

Verification Condition 1:
h=0

while $h< t$; $h=h+1$ {this is iterated to test several random selections}
$i=0$; **while** $i<n$; $i=i+1$ {n is the number of subjects in the entire data set}
$l=0$; **while** $l<a$; $l=l+1$ {a equals the number of saccades of subject i}
To construct a training set:
Choose a-1 saccades of a user, i, other than the lth saccade **and select randomly** b saccades from each of c subjects j (non-users), $j=1,\ldots,c$, $j\neq i$, $c\leq n-1$.
To construct a test set:
Run leave-one-out: **Choose** a user's lth saccade (excluded in training) to be the test saccade.
Classify using method M.
end
Check whether either correct or incorrect verification was obtained: either an authenticated user (correct) or some of non-users (incorrect) obtained more hits.
end
Compute the numbers of correct and incorrect verifications.
end
Compute the means of correct and incorrect verifications for t iterations.

Verification Condition 2:
$h=0$
while $h< t$; $h=h+1$ {iterated to test several random selections}
$i=0$; **while** $i<n$; $i=i+1$ {n is the number of subjects in the data set}
To construct a training set:
Choose a-1 saccades of user i **and select randomly** b saccades from each of c subjects j (non-users) **randomly selected**, $j=1,\ldots,c$, $j\neq i$, $c<n-1$.
$q=c+1$; **while** $q<n$, $q=q+1$; $q\neq i$, {$n-c-1$ impostors are tested}
To construct a test set:
Run leave-one-out: **Randomly pick up** an impostor's saccade (the qth subject) to be the test saccade.
Classify using method M.
End
Check whether either correct or incorrect verification was obtained: either an authenticated user (incorrect) or some of non-users (correct) received more hits.
end
Compute the numbers of correct and incorrect verifications.
end
Compute the means of correct and incorrect verifications for t iterations.

The two procedures could be united, but we ran separate tests due to the relative scarcity of subjects. Performing tests on the basis of the separate procedures, we were able to vary training sets more extensively than we could with one procedure including both test types. Namely, Condition 2 required both training saccades of non-users and test saccades of impostors: in other words, two separate sets, in addition to the saccades of an authenticated user. Condition 1 did not incorporate impostors, yielding more extensive opportunities for the selection of non-users.

5 Computational results

At first we computed the classification results from linear discriminant analysis after all the saccade signals were interpolated using the techniques mentioned above. The motive here was to find the best interpolation technique for large-amplitude saccades measured using the current device. Our effectiveness criterion was the accuracy of correct classifications for the two verification conditions described. Accuracy was computed on the basis of the outermost loops of the preceding verification procedures as follows. For n subjects and t iterations, the mean accuracy A is

$$A = \frac{\sum_{h=1}^{t} \frac{d_h}{n}}{t} \cdot 100\%$$

where d_h stands for the number of correct classifications of iteration h in each procedure.

The mean results of the correct verifications obtained are shown in Figure 5. False acceptance and false rejections rates are sometimes used, and can be calculated straightforwardly by subtracting accuracies from 100% in Condition 1 for false rejections and in Condition 2 for false acceptances.

The five sizes of training sets given in Figure 5 correspond to different divisions between the numbers of saccades of an authenticated user and those of non-users. Each subject had 20 saccades measured. In addition to an authenticated user, there were 67 other subjects. First, one saccade was taken randomly from each of a=20 non-users randomly selected (ratio 19:20) for a training set of 39. Second, b=1 saccade from each of c=40 non-users in Condition 1, or b=2 saccades from c=20 non-users in Condition 2 (ratio 19:40) for a training set of 59. Third, from 60 non-users, either one (19:60), two (19:120) or three (19:180) saccades were selected randomly for training sets of 79, 139 or 199 for Condition 1. In Verification Condition 2, there were either two (19:60), four (19:120) or six (19:180) randomly selected saccades from each of 30 non-users. Note that the saccades of 37 other subjects serving as impostors formed the data source for a test set. Still, for training sets of 39 or 59, the number of test subjects was 47, since we used each subject in two of the three roles (authenticated user, non-user or impostor) to utilize the data maximally for tests.

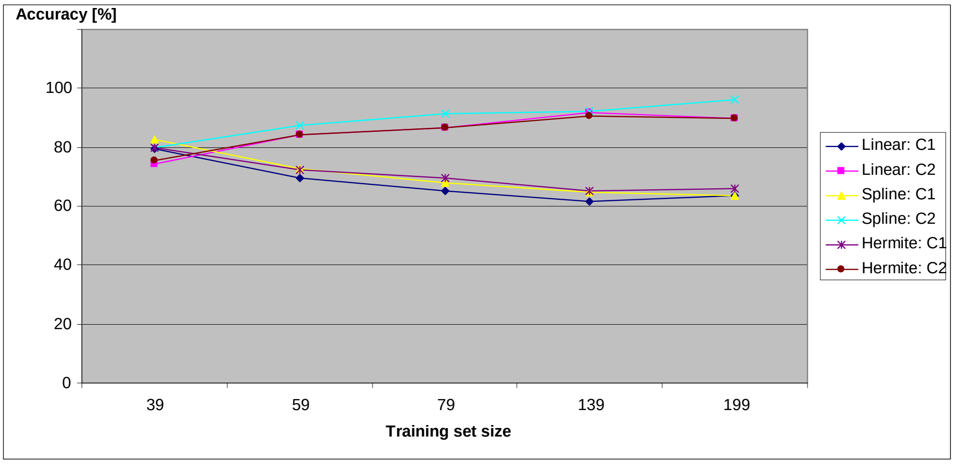
For Condition 1, all 20 saccades from each subject were tested as if they were from an authenticated user. For Condition 2, the saccades of each impostor were run. Since random selections were used for the subsets of non-users and impostors, all these tests were repeated ten times to vary the contents of the two subsets. For Condition 1, there were $68 \times 20 = 1360$ tests, which were repeated ten times to calculate the means for each size of training sets and method. Correspondingly, for Condition 2 there were $68 \times \{47|37\} = \{3196|2516\}$ tests repeated ten times.

Let us consider the results shown in Figure 5 and the way we evaluated their meaning in general. We have to assess verification results x_i and y_i by maximizing the results of both verification conditions at the same time:

$$\max_i \{x_i\} \wedge \max_i \{y_i\}, i \in \{\text{training set sizes}\}$$

It is important that an authenticated user is accepted with as great a degree of certainty as possible according to x_i of Condition 1, and that an impostor is rejected reliably according to y_i of Condition 2. We can handle this maximization by calculating, for example, the sum of the accuracies of both conditions:

Figure 5. Mean accuracies of correct classifications for Verification Conditions 1 (C1) and 2 (C2) after interpolating saccade signals using linear, cubic spline and cubic Hermite techniques before linear discriminant analysis classifications for five training sets between an authenticated user's saccades and those of non-users.



$$s = \max_i \{x_i + y_i\}, i \in \{\text{training set sizes}\}$$

Of course, we pursued a situation where it would be as difficult as possible for an impostor to succeed in logging in. For this reason, we increased the number of non-users' saccades when thinking that there should be more possible representatives of non-users than one authenticated user. Thus, non-users ought to cover a clearly larger area in the variable space formed by the six saccade variables selected than the part covered by an individual authenticated user.

According to the results in Figure 5, differences were small between the interpolation techniques. When *s* of the cubic spline interpolation was a few per cent better than the other on the average, we chose it for actual verification tests.

Our main results were obtained by computing tests for Verification Conditions 1 and 2 with different classification methods. First, we employed linear, quadratic and logistic discriminant analysis and nearest-neighbour searching with exhaustive searching or with *k-d* trees. Nearest-neighbour searches were run with *k* (number of nearest neighbours) equal to 1, 5, 9 and 13, for which a classification decision was computed according to the inequality given above to classify a saccade case in one of two classes: authenticated user or a subset of non-users. Figure 6 shows results of discriminant analysis. Figures 7 and 8 present those of nearest-neighbour searching. Figure 9 includes results of decision trees performed with pruning and the naïve Bayesian rule.

In Figure 6, the standard deviations of the means presented were 1.1%-4.9%. The linear and quadratic discriminant analysis methods were better than logistic discriminant analysis for training set sizes other than 39, for which the latter was better and the best of all methods throughout Figures 6-9. However, for other training set sizes, it greatly favoured Condition 2 and mostly failed in Condition 1.

Figure 6. Mean accuracies of correct classifications for subjects in Verification Conditions 1 (C1) and 2 (C2) with five sizes of training sets: linear (LinDA), quadratic (QDA) and logistic (LogDA) discriminant analysis.

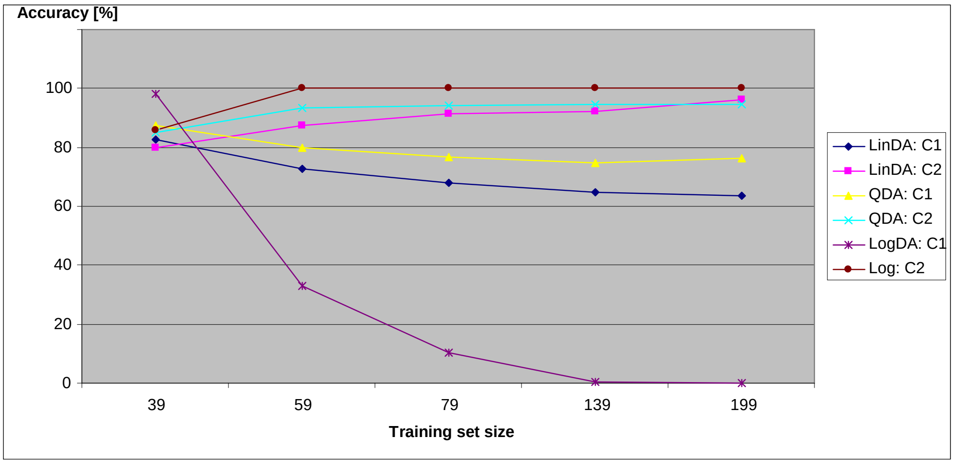
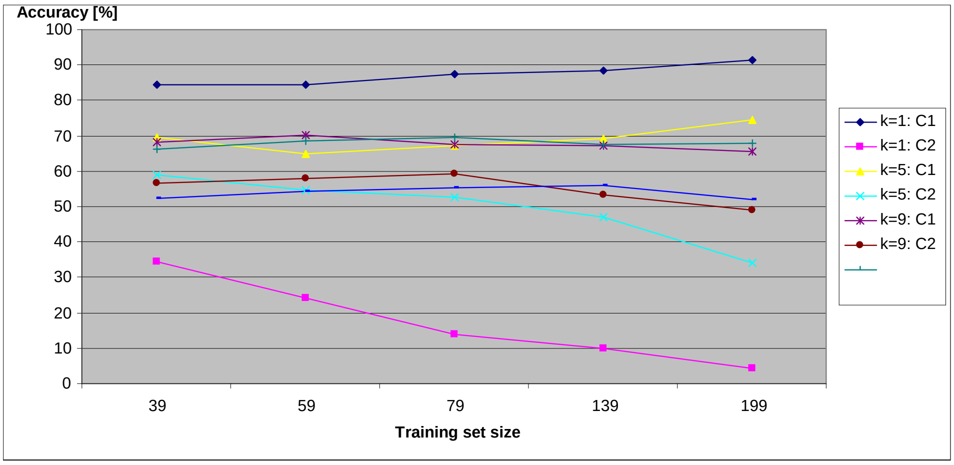


Figure 7. Mean accuracies of correct classifications for subjects in verification Conditions 1 (C1) and 2 (C2) with five sizes of training sets: exhaustive nearest-neighbour searching with *k* equal to 1, 5, 9 or 13.



Linear and quadratic discriminant analysis methods succeeded better in Condition 2. The discriminant analysis methods were clearly superior to exhaustive nearest-neighbour searching in Figure 7 and a few percent better than *k-d* tree searching in Figure 8. In Figures 7 and 8, standard deviations were 1.1%-6.1% and 1.4%-5.9%. In Figure 7, the best result was obtained with *k* equal to 5 and a training set size of 39. The results were fairly stable for other *k* values except 1 and for other sizes of training sets. In Figure 8, the best result was given by *k* equal to 1 and a training set size of 39. Otherwise, the results varied slightly. Results were impaired at *k* values greater than 1, whereas in Figure 7 this depended on verification conditions. Both searching methods were better for

Condition 1 than Condition 2, which was the opposite of the case with most pairs in Figure 6. In Figure 9, standard deviations were 1.1%-4.9% and decision trees were superior to the naïve Bayesian rule for a training set size of 39, but otherwise inferior. Both methods were considerably better in Condition 2 for the other training set sizes.

Figure 8. Mean accuracies of correct classifications for subjects in Verification Conditions 1 (C1) and 2 (C2) with five sizes of training sets: *k-d* tree searching with *k* equal to 1, 5, 9 or 13.

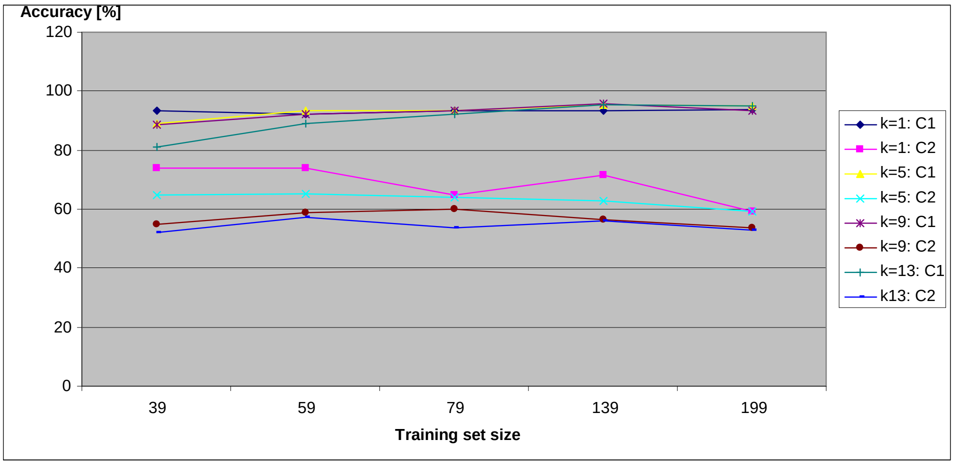
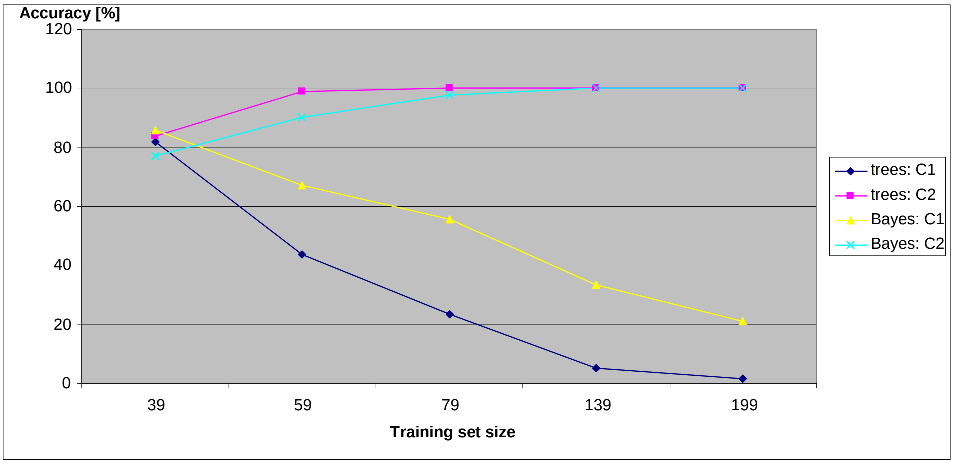


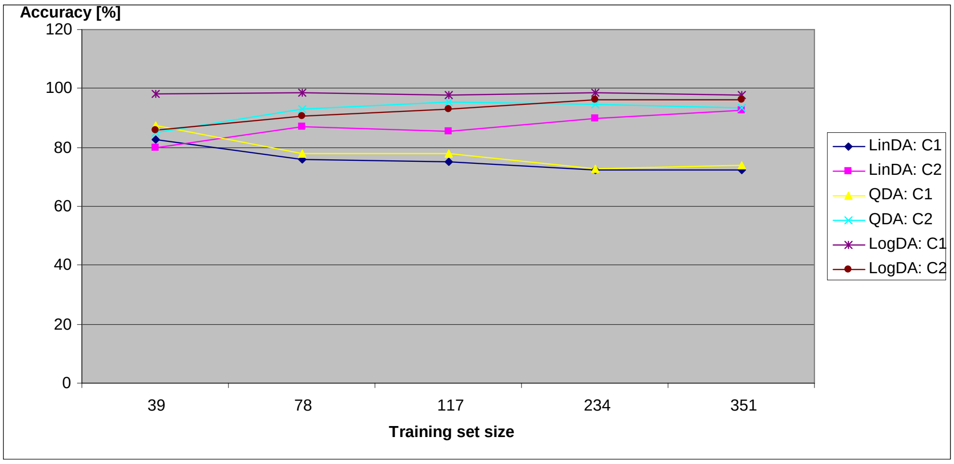
Figure 9. Mean accuracies of correct classifications for subjects in verification Conditions 1 (C1) and 2 (C2) with five sizes of training sets: the naïve Bayesian rule and decision trees.



As regards the result values in all figures, we emphasize that these values indicate how many of the test subjects were classified correctly as authenticated users for Verification Condition 1 or as impostors for Condition 2. Thus, the values do not denote the numbers of saccades correctly classified, although decisions are based on such information.

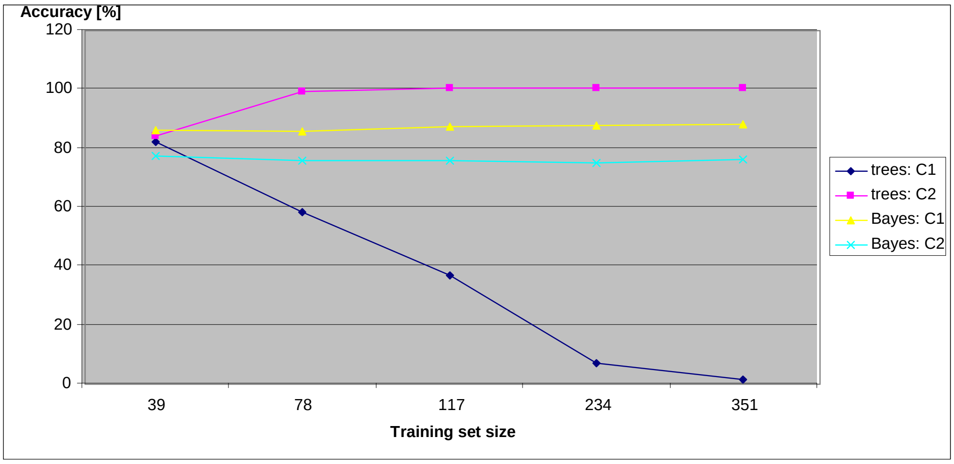
Up to now, sizes of training sets other than 39 and 59 did not seem to yield the highest results when both verification conditions were taken into account at the same time. The results of logistic discriminant analysis were the best, but they were highly prone to poor results in Condition 1 with other training set sizes than 39. For this reason, we continued our testing by balancing (Swingler, 1996) the sizes of the classes of an authenticated user and non-users in a training set. For a training set size other than 39 (ratio 19:20), we inserted 19 copies of the saccades of each authenticated user once, twice, five or eight times to include 38, 57, 114 or 171 cases in an authenticated user's class. These extensions balanced a priori probabilities between the two classes. Note that our original idea of using a skewed distribution for the two classes was adopted from the assumption that the class of non-users representing several non-users should be larger than that of an authenticated user in the variable space (Figures 3 and 4). The tests were repeated the same as previously, but nearest-neighbour searching tests were left out. Above, they gave worse outcomes than those of the other methods. In addition, with respect to an authenticated user's class, our technique of balancing by copying would have reduced most nearest-neighbour searching situations to a scenario similar to that of *k* equal to 1. Figures 10 and 11 below show the results obtained: their standard deviations were 1.2%-6.1% and 0-5.2% respectively.

Figure 10. Mean accuracies of correct classifications for subjects in Verification Conditions 1 (C1) and 2 (C2) with five sizes of training sets: linear (LinDA), quadratic (QDA) and logistic (LogDA) discriminant analysis.



As shown in Figure 10, the results of linear and quadratic discriminant analysis methods did not change considerably, mostly remaining within ±3% (a minimum of -5.9% and a maximum of 8.5%), compared with the results in Figure 6 before balancing. Instead, the results of logistic discriminant analysis for Verification Condition 1 and the training set sizes ranging from 78 to 351 in Figure 10 were essentially improved compared with the corresponding rows in Figure 6, but the results of Condition 2 did not decrease much: 3.7-9.6%. The results of logistic discriminant analysis were clearly better than the other method in Figure 10. The best in all figures (6-11) was logistic discriminant analysis for the training set size of 234. In Figure 11, the results of the naïve Bayesian rule for the training set sizes of 78 to 351 were greatly bettered for Condition 1 compared with Figure 9, but also 15-20% impaired for Condition 2. At the same time, the effects on the results of decision trees were minor; they could not be properly classified in Condition 1.

Figure 11. Mean accuracies of correct classifications for subjects in Verification Conditions 1 (C1) and 2 (C2) with five sizes of training sets: decision trees and the naïve Bayesian rule.



6 Discussion and conclusion

The results presented in Figures 5-9 revealed the phenomenon of Condition 1 becoming easier to verify as the class of non-users in a training set grew smaller. On the other hand, Condition 2 was easier to verify, the larger that class was. The reason was simple. Calculating a priori probabilities on the basis of these subsets of saccades, the a priori probability of obtaining a correct verification in Condition 1 is greatest (49%) with a training set size of 39 and smallest (10%) with a training set size of 199. For Condition 2 the situation is the opposite: 51% and 90%. In principle, one should look at the selection that satisfies both requirements as well as possible. Balancing the number of saccades of an authenticated user close to that of non-users in Figures 10 and 11 served this aim while also improving the results of logistic discriminant analysis and the naïve Bayesian rule.

Logistic discriminant analysis yielded the best results. Among other methods, linear and quadratic discriminant analysis methods were the second-most promising ones, but they were inclined to produce good results only for one of the two conditions. This is a typical problem for binary classification in general. Nearest-neighbour searching and naïve Bayesian rules produced poorer results and decision trees the poorest results of all.

It is, naturally, essential to eliminate the possibility of an impostor occasionally succeeding in logging into a computer, so it was best to choose the ratio for training sets that provided the highest possible expectation of preventing impostor success. At the same time, verification of an authenticated user has to be reliable. In this sense, the best results are in Figure 10. Since the training set sizes of 234 (ratio 114:120) and 351 (ratio 171:180) satisfied these two opposite requirements for both verification conditions while applying logistic discriminant analysis, they were the best choices.

The results of the few other publications on this research theme are not directly comparable because their methods, data and objectives were different from ours. Their average false acceptance rates of 1.4-17.5% and the average false rejection rates of 12.6-

35.6% depended on a different classification method for nine subjects (Kasprowski and Ober, 2004), average false acceptance rates of 4.8% and the average false rejection rates of 9.4% for 47 subjects (Kapczyński, Kasprowski and Kuźniacki, 2006), and 90% accuracy for 12 subjects (Bednarik et al., 2005). The last study's results were, however, obtained mainly on the basis of image analysis with distances between eyes, not with eye movements. Recently reported were false acceptance rates of 5.4% and false rejection rates of 56.6% with 41 subjects tested (Komogortsev et al., 2010). The results gained with minimum spanning trees were approximately 70% correct with nearest-neighbour searching and a support vector machine after applying the equal error rate technique (Rigas, Economou and Fotopoulos, 2012). However, please note that none of these results cannot be directly compared with ours: their approaches were very different; they used different eye movement recording systems and test set-ups; and they obviously performed identification in the sense defined in Section 1. In the results of our present research, false rejections were obtained from the first condition and false acceptances from the second condition by reducing their accuracy values from 100%.

To summarize, our results indicate that it is possible to distinguish an authenticated user from a set of other subjects, and that a training set collected from an authenticated user must be sufficiently large, preferably more than 100 saccades. Class sizes for an authenticated user and for non-users representing the area in the variable space outside an authenticated user should be approximately equal. In a case such as ours with the current data, copying the scarce saccade data of an authenticated user could be used to balance the two classes successfully. Furthermore, our results showed that the variables selected were good for the purpose of biometric verification and for the test set-ups used in our present research. In a signal analysis sense, all these variables were of the time domain type. We did not use variables based on frequency, for instance, given by power spectrum. A weak point of these variables might be the possible instability over the course of a longer period of time: say, days, weeks or more. Researchers have investigated their variability only infrequently (Bollen et al., 1993; Schmidt et al, 1979), so we intend to study this substantive issue in the future by performing pertinent repeated measurements for groups of the same subjects.

Since good accuracies were gained, we consider the user verification procedure introduced to be promising and worth investigating further. Naturally, we intend to collect data from more subjects, study other classification methods, and test alternatives. Obviously, it is possible to design other ways to use saccades for the verification of an authenticated user. A good feature of using eye movements is that they can hardly be stolen or emulated.

Acknowledgements

The first author received the support of the Tampere Doctoral Programme in Information Science and Engineering, Tampere, Finland. The authors are grateful to Professor Ilmari Pyykkö of the Department of Otorhinolaryngology at Tampere University Hospital for physiological advice on eye movements.

References

Abrams, R.A., Meyer, D.E. and Kornblum, S. (1989) 'Speed and accuracy of saccadic eye movements: characteristics of impulse variability in the oculomotor system', *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 15, No. 3, pp. 529-543.

Allik, J., Rauk, M. and Luuk, A. (1981) 'Control and sense of eye movement behind closed eyelids', *Perception*, Vol. 10, pp.39-51.

Bahill, A.T. (1980) 'Development, validation and sensitivity analyses of human eye movement models', *CRC Critical Reviews in Bioengineering*, Vol. 4, pp.311-355.

Bahill, A.T., Brockenbrough, A. and Troost, B.T. (1981) 'Variability and development of a normative data base for saccadic eye movements', *Investigative Ophthalmology & Visual Science*, Vol. 21, pp.116-125.

Bednarik, R., Kinnunen, T., Mihaila, A. and Fränti, P. (2005) 'Eye-movements as a biometric', in *14th Scandinavian Conference on Image Analysis*, Joensuu, Finland, Lecture Notes in Computer Science, Vol. 3540, Springer-Verlag, pp.780-789.

Boatwright, M. and Luo, X. (2007) 'What do we know about biometrics authentication?' in *Proceedings of Information Security Curriculum Development Conference '07*, ACM, 5 p.

Bollen, E., Bax, J., van Dijk, J.G., Koning, M., Bos, J.E., Kramer, C.G.S and van der Valde, E.A. (1993) 'Variability of the main sequence', *Investigative Ophthalmology & Visual Science*, Vol. 34, No. 13, pp.3700-3704.

Conti, V., Militello, C., Sorbello, F., and Vitabile, S. (2010) 'An adaptive multimodal biometric management algorithm', *IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews*, Vol. 40, No. 4, pp.384-395.

Edelman, J.A. and Xu, K.Z. (2008) 'Inhibition of voluntary saccadic eye movement commands by abrupt visual onsets', *Journal of Neurophysiology*, Vol. 101, No. 3, 1222-1234.

Fricker, S.J. and Sanders, J.J. (1975) 'Velocity and acceleration statistics of pseudorandomly timed saccades in humans', *Vision Research*, Vol. 15, pp.225-229.

Frischholz, R.W and Dieckmann, U. (2000) 'BioID: A multimodal biometric identification system', *IEEE Computer*, Vol. 33, No. 2, pp.64-69.

Gredebäck, G., Örnkloo, H. and von Hofsten, C. (2006) 'The development of reactive saccade latencies', *Experimental Brain Research*, Vol. 173, pp. 159-164.

Henderson, J.M. and Hollingworth, A. (2003) 'Eye movements and visual memory: detecting changes to saccade targets in scenes', *Perception & Psychophysics*, Vol. 65, No. 1, 58-71.

Hyrskykari, A. (2006) *Eyes in attentive interfaces: Experiences from creating iDict, a gaze-aware reading aid*. PhD thesis, Dissertations in Interactive Technology, No. 4, University of Tampere, Finland.

Israel, S.A. and Irvine, J.M. (2012) 'Heartbeat biometrics: a sensing system perspective', *International Journal of Cognitive Biometrics*, Vol. 1, No. 1, pp.39-65.

Joyce, C.A., Gorodnitsky, I.F., King, J.W. and Kutas, M. (2002) 'Tracking eye fixations with electroocular and electroencephalographic recordings', *Psychophysiology*, Vol. 39, pp.607-618.

Juhola, M., Jäntti, V. and Pyykkö, I. (1985) 'Effect of sampling frequencies of the maximum velocity of saccadic eye movements', *Biological Cybernetics*, Vol. 53, pp.67-72.

Juhola, M., Aalto, H. and Hirvonen, T. (2007) 'Using results of eye movement signal analysis in the neural network recognition of otoneurological patients', *Computer Methods and Programs in Biomedicine*, Vol. 86, pp.216-226.

Juhola, M., Zhang, Y. and Rasku, J. (2013) 'Biometric verification of a subject through eye movements', *Computers in Biology and Medicine*, Vol. 43, pp. 42-50.

Kapczyński, A., Kasprowski, P. and Kuźniacki, P. (2006) 'Modern access control based on eye movement analysis and keystroke dynamics', in *Proceedings of International Multiconference on Computer Science and Information Technology*, Wisla, Poland, pp.477-483.

Kasprowski, P. and Ober, J. (2004) 'Eye movements in biometrics', in *Biometric Authentication: International Workshop*, Prague, Czech Republic, Lecture Notes in Computer Science, Vol. 3087, Springer-Verlag, pp.248-258.

Komogortsev, O.V., Jayarathna, S., Aragon, C.R. and Mahmoud, M. (2010) 'Biometric identification via an oculomotor plant mathematical model', *Proceedings of the 2010 Symposium on Eye Tracking Research & Applications*, Austin, Texas, pp.57-60.

Majaranta, P. (2009) *Text entry by eye gaze*. PhD thesis, Dissertations in Interactive Technology, No. 11, University of Tampere, Finland.

McIntosh, R.D. and Buonocore, A. (2012) 'Dissociated effects of distractors on saccades and manual aiming', *Experimental Brain Research*, Vol. 220, pp. 201-211.

Nakayama, M. and Abe, H. (2012) 'Performance of single-trial classifications of viewed characters using EEG waveforms', *International Journal of Cognitive Biometrics*, Vol. 1, No. 1, pp.10-25.

Nandy, A.S. and Tjan, B.S. (2012) 'Saccade-confounded image statistics explain visual crowding', *Nature, Neuroscience*, Vol. 15, No. 3, pp. 463-469.

Negin, M., Chmielewski, Jr., T.A., Salganicoff, M., Camus, T.A., Cahn von Seelen, U.M., Venetianer, P.L., and Zhang, G.G. (2000) 'An iris biometric system for public and personal use', *IEEE Computer*, Vol. 33, No. 2, pp.70-75.

Pentland, A., and Choudbury, T. (2000) 'Face recognition for smart environments', *IEEE Computer*, Vol. 33, No. 2, pp.50-55.

Salvucci, D.D. and Goldberg, J.H. (2000) 'Identifying fixations and saccades in eye-tracking protocols', *Proceedings of the Eye Tracking Research and Applications Symposium*, New York, ACM Press, pp. 71-78.

Schmidt, D., Abel, L.A., Dell'Osso, L.F: and Daroff, R.B. (1979) 'Saccadic velocity characteristics: Intrisinc variability and fatigue', *Aviation, Space, and Environmental Medicine*, Vol. 50, pp.393–395.

Sparks, D.L. (2002) 'The brainstem control of saccadic eye movements', *Nature, Neuroscience*, Vol. 3, pp. 952-964.

Van Stigchel, S. (2010) 'Recent advances in the study of saccade trajectory deviations', *Vision Research*, Vol. 50, pp. 1619-1627.

Sun, Z., Wang, Y., Tan, T. and Cui, J. (2005) 'Improving iris recognition accuracy via cascaded classifiers', *IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews*, Vol. 35, No. 3, pp.435-441.

Swingler, K. (1996) *Applying Neural Networks, A Practical Guide*, Academic Press, London, UK.

Tweed, D. and Vilis, T. (1990) 'Geometric relations of eye position and velocity vectors during saccades', *Vision Research*, Vol. 30, No. 1, pp.111-127.

Underwood, G., Bloomfield, R. and Clews, S. (1988) 'Information influences the pattern of eye fixations during sentence comprehension', *Perception*, Vol. 17, pp.267-278.

Veeramachaneni, K., Osadciw, L.A. and Varshney, P.K. (2005) 'An adaptive multimodal biometric management algorithm', *IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews*, Vol. 35, No. 3, pp.344-356.

Wierts, R., Janssen, M.J.A. and Kingma, H. (2008), Measuring saccade peak velocity using a low-frequency sampling rate of 50 Hz', *IEEE Transactions on Biomedical Engineering*, Vol. 55, No. 12, pp. 2840-2842.

Yang, Q., Bucci, M.P. and Kapoula, Z. (2002) 'The latency of saccades, vergence, and combined eye movements in children and in adults', *Investigative Ophthalmology & Visual Science*, Vol. 43, No. 9, pp. 2939-2949.

Zhang, Y., Rasku, J. and Juhola, M. (2012) 'Biometric verification of subjects using saccade eye movements', *International Journal of Biometrics*, Vol. 4, No. 4, pp. 317-337.

Zhang, Y. and Juhola, M. (2013) 'On biometric verification of a user by means of eye movement data mining', *The Second International Conference on Advances in Information Mining and Management*, Venice, Italy, Xpert Publishing Services, pp. 85-90.

Publication IV

On Biometric Verification of a User by Means of Eye Movement Data Mining

Youming Zhang and Martti Juhola

 In *Proceedings of the Second International Conference on Advances in Information Mining and Management (IMMM2012)*, 85-90, Venice, Italy, October 2012.

On Biometric Verification of a User by Means of Eye Movement Data Mining

Youming Zhang and Martti Juhola
Computer Science, School of Information Sciences
33014 University of Tampere
Tampere, Finland
Youming.Zhang@uta.fi, Martti.Juhola@sis.uta.fi

***Abstract*—In biometric verification, a signal, image or other dataset is measured from a subject to detect him or her to be or not to be an authenticated subject such as the user of a computer. So far, biometric verification has mainly been on the basis of fingerprints or face images, infrequently other images, e.g., iris. We studied the idea to apply fast eye movements called saccades to verify an authenticated user from among other subjects. We recorded eye movement signals with eye movement cameras using a suitable visual stimulation for a subject. By means of machine learning methods, we classified a subject's eye movements to verify whether one was an authenticated user. We employed multilayer perceptron networks, radial basis function networks, support vector machines and logistic discriminant analysis for classification. The best accuracy results obtained were approximately 90% and showed that it is possible to verify a subject according to saccade eye movements.**

Keywords-biometric verification; eye movements; saccades; multilayer percetron neural networks; radial basis function networks; support vector machines; logistic discriminant analysis

I. INTRODUCTION

So far, various biometric data sources have been used to verify a subject. Mostly fingerprints [1, 2] and face images [3] are applied to this task. Other images measured from subjects such as iris images [2, 4] are also studied. In addition to these two-dimensional data sources, one-dimensional signals are also used, e.g., voice signals [5]. Usually, these datasets contain an abundance of data and several variables are computed from them to ground the verification procedure on variable values of different subjects. Data mining tasks needed here may be complicated because of complex data.

Eye movements are a new potential alternative for biometric verification. Eye movements have been researched for decades in medicine. During the past 15 years eye movements have become an important research objective for human-computer interfaces. Along with these applications efficient eye movement cameras have been developed. Since there is long-term experience in the signal analysis of eye movements, for example [6-8], for biomedical and physiological applications, it was a direct development to attempt to utilize them for biometric verification of a subject simulating a computer user. Note that verification corresponds to the binary classification between two classes: an authenticated user and other subjects.

There are a few different eye movement types such as saccade, nystagmus, smooth pursuit and vestibulo-ocular reflex eye movements [7]. Probably the most frequent of all are saccades which are made while looking at surroundings or reading a text. In addition, they are very fast, in fact the fastest movements of man. They are easy to visually stimulate and their recording does not require more time than a few minutes for our tests. Those other eye movement types would require longer recordings or more complicated stimulation arrangements [7]. For these reasons, we chose saccades to be our data sources here, particularly after observing differences between saccades of individuals [7].

Up to now, a couple of attempts only have been published about this idea to use eye movements for biometric verification. In one research [8] they recorded saccade eye movement signals to compute cepstrum from these and classified signal analysis outcomes by using naïve Bayesian method, nearest neighbour searching, decision trees as well as support vector machines. In another research [9] they used a computational oculomotor model on the parameters of which verification was based using nearest neighbour searching and decision trees. Our approach differs from those since we use physiological variables computed from eye movement signals. Most of these variables have been employed for long in biomedical investigations [6,7].

II. EYE MOVEMENT DATA

We recorded saccade eye movements with a two-camera system (Visual Eyes, Micromedical Technologies, UK). Its resolution is 320×240 and sampling frequency or frames per second 30 Hz. The camera system recognized positions of each pupil from successive images of a video stream to detect eye movements. The system records horizontal and vertical signals, but we used the horizontal direction only. We wanted to keep the arrangement as simple as possible for stimulation design so that this was simple for a subject in order to avoid complex stimulations. Furthermore, using simple stimulations means that long recordings are not necessary which is important to see this biometric verification idea as sensible. On the other hand, the more data from each individual, the easier it is perhaps to separate him or her from the group of other subjects. The sampling frequency of 30 Hz was low compared to other typical ones used in eye movement camera systems such as 50 or 60 Hz, occasionally even higher like 200 Hz. Nevertheless, it was

 ISBN: 978-1-61208-227-1

interesting to see whether this low sampling frequency allowed verification. Perhaps using a higher frequency in the future could only better results because of more accurate variable values to be computed. The system included one camera for each eye embedded in the mask attached tightly with a headband. The one of lower noise level of two eye movement signals was used for verification. Usually, both are almost identical.

We used the same stimulation series for every subject. This is, of course, the essential detail for biometric verification so that we can assume that every subject has followed the same stimulation by his or her gaze and we can classify them according to their eye movements. Each subject saw a horizontally jumping LED light dot in front of him or her. The stimulation component of the eye movement recording system included a horizontal LED bar in which one LED was switched on for a while, then switched off and another switched on immediately, and so on, by varying the LED to be next switched on. This way different gaze angles were formed. Intervals between light dot jumps were varying to make them random for a watching subject. Since intervals of 1-3 s were short and varying, the spectator could learn neither them nor varying stimulation angles. Varying, “random” intervals are important to minimize anticipations of a subject while waiting for the next stimulation movement. Anticipation would occur if latency or reaction time from the beginning of a stimulation movement to the beginning of its response, saccade, were shorter than 0.120 s seen as a minimum latency in the physiological sense [7]. It takes some time for the brain to observe a movement and control the response to move the eyes.

The present stimulation arrangement was used to simulate the beginning of a computer session where a user would first sit down to start the machine and to wait for its initialization. We can imagine that the eye movement stimulation would be run immediately after the initialization by stimulating a subject with a few dozen stimulation movements on the screen of a computer or mobile device. Thereafter, the verification procedure would be run.

We used saccades with the largest stimulation amplitudes of around 48° only since saccades of such large amplitudes contain greater differences between subjects than those with small amplitudes [7]. Great differences between subjects aid in verification. Nonetheless, there were smaller stimulation angles between large to give a random character between stimulations from a spectator’s viewpoint. Consequently, we obtained 20 large amplitude saccades from every subject. Values of saccade variables depend on saccade amplitudes. Thus, we used merely the saccades of the largest stimulation amplitude,

For the sake of the low sampling frequency of 30 Hz, we interpolated every signal with a cubic spline method up to 1000 Hz. The purpose here was to simulate a sampling frequency of the newest, expensive high resolution eye movement cameras and, most of all, to estimate values of eye movement variables more precisely than enabled by the original signals sampled at 30 Hz.

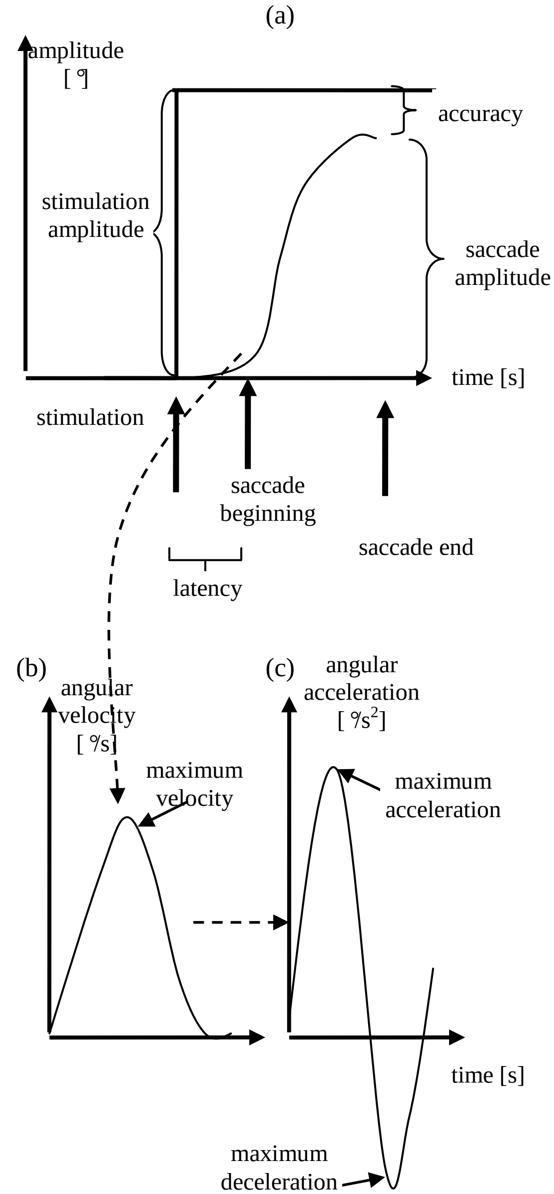


Figure 1. (a) The step (broken line) is a stimulation movement produced by a horizontally jumping light dot from the left (down in the figure) to the right (up). A saccade as a response follows it after a latency. The difference between amplitudes determines a negative accuracy, because the saccade amplitude is smaller here. A positive accuracy is also possible, but is more infrequent than negative. In our tests these values were used as absolute. Accuracy, amplitude and latency were three variables used. (b) From the saccade signal the first derivative approximation of the velocity curve is computed from which (c) the second derivation of the acceleration curve is approximated. The maximum velocity, maximum acceleration and maximum deceleration were other three useful variables to be computed.

III. Signal Analysis and Data Preprocessing

Fig. 1 depicts an ideal saccade and its stimulation as a schema. The first signal analysis task is to detect the exact beginning and end of every stimulation movement and those of the following response eye movement, saccade.

 ISBN: 978-1-61208-227-1

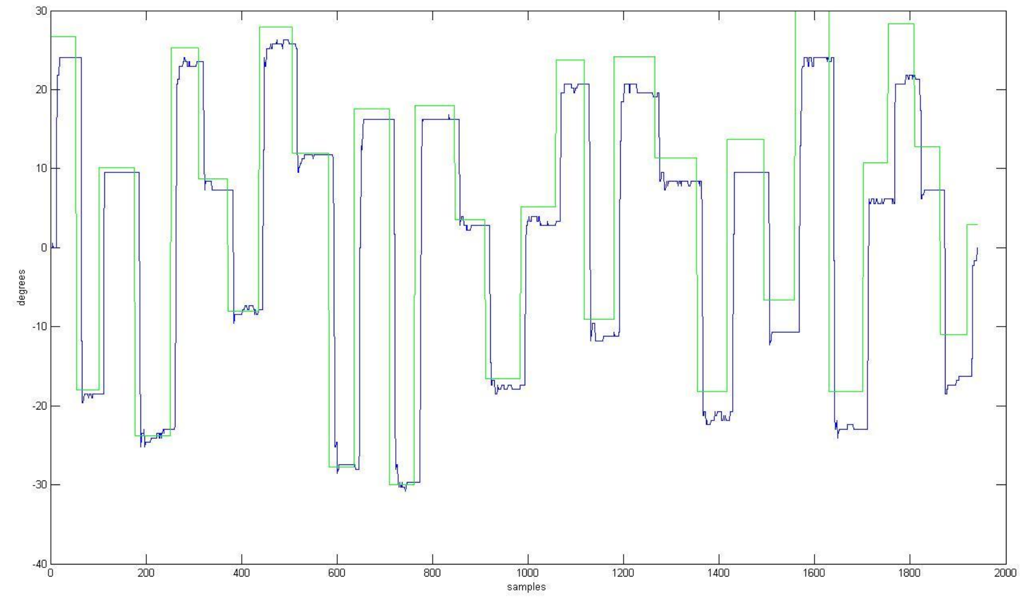


Figure 2. A smooth (green) stimulation signal of 64 s sampled at 30 Hz and its (blue) response with saccades.

The former is easy to detect since it is a clear step in a signal. The latter may rarely be somewhat corrupted by noise or artifacts such as blinks; See Fig. 2, including horizontal saccades.

If a saccade is inaccurate, its amplitude clearly differs from that of its stimulation. The brain can rapidly produce a corrective saccade with a small amplitude to correct the gaze closer to the objective. One cannot sense this correction movement, but it is "automatic". We did not include possible, quite infrequent corrective saccades, but determined the accuracy of a saccade along with the primary saccade as usual. A response to its stimulation movement had to resemble a real saccade sufficiently to be accepted for further use in signal analysis. In principle a subject might not occasionally follow the target with the gaze. This would yield no saccade at all. Anticipation as a too early eye movement including a latency value less than 0.120 s or even a saccade before a stimulation would be rejected as no actual responses to stimulations. The quality of signals given by the camera system was high with low noise. Thus, rejections of eye movements from signals were infrequent, no more than a few per cent of all saccades.

The same stimulation movements (Fig. 2) were run for every recording, so that eye movements of subjects were comparable with each other. A stimulation series included four stimulations with the largest amplitude of 48°. Five recordings were run successively from every subject giving 20 large saccades for a subject.

The five recordings of each subject formed our data for biometric verification tests. The stimulation series also contained saccades of smaller amplitudes between those four large to make the stimulation series more random-like for a subject not able to guess the direction or amplitude of a stimulation movement or an interval between two successive stimulations. Intervals were 1-3 s within a recording of 64 s.

After the interpolation of signals, the first derivative and second derivative were computed with approximation formulas such as two-point central difference differentiation [8] from each eye movement signal. A saccade beginning was found provided that absolute velocity values rapidly increased above a threshold of 50°/s and the corresponding saccade end was found when velocity decreased back below that threshold. After detecting a saccade and ensuring that it was valid according to latency criterion, etc., all its variable values were computed and stored: amplitude, accuracy, latency and maximum velocity, acceleration and deceleration.

During recordings, a sitting, alert, relaxed subject was asked to follow the stimulation light dot by the gaze. In all, we recorded five successive recordings from healthy 132 subjects from whom 33 were females and 99 males. Mean and standard deviation of their ages were 26.2±7.2 years. Neither alcohol nor medications were used during 24 h before a measurement. We wanted to test mainly young subjects in a pretty homogeneous dataset to create a strict testing basis. Age, alcohol or medications may have influence on values of saccade variables. Means and standard deviations were the following: amplitude 48.0±13.4°, accuracy 3.2±8.4°, latency 0.269±0.057 s, maximum velocity 1038±322 °/s, maximum acceleration 47591±23166 $°/s^2$ and maximum deceleration 44845±24745 $°/s^2$.

IV. VERIFICATION PROCEDURE

In biometric or whatever user verification, we have to prepare two opposite conditions: a subject attempting to log in is either authenticated user or impostor. Thus, we built our test procedure to take these two conditions into account.

When machine learning algorithms are used, we have to construct a training set and its corresponding test set. The content of these two sets are varied on the basis of available data. In the current case, our eye movement data were quite limited. Although there were several subjects, the bottle neck for tests was the small number 20 of saccades of the largest amplitude. Therefore, we implemented two experimental test settings called Alternatives 1 and 2. In the one of them for every subject there were either three recordings (12 saccades) in a training set and the rest of two recordings (q=8 saccades) in the corresponding test set. In the other there were four recordings (16 saccades) in a training set and one recording (q=4 saccades) in its test set. Since from every subject there were five recordings all in all, we obtained c=10 different combinations of a training set and a test set from five recordings for the former Alternative 1 and c=5 combinations for the latter Alternative 2. These were prepared for every of n=132 subjects. Our aim was to test our data as broadly as possible as conventional while applying data mining methods for classification.

Our verification task (Fig. 3) comprised two classes. Therefore, it was best that the number of saccades of an authenticated user and that of other subjects now called nonusers were not very imbalanced. We had either m=12 (Alternative 1) or m=16 (Alternative 2) saccades of an authenticated user in a training set. We then took one saccade randomly from either 2m=24 or 32 nonusers to test Condition 1 (an authenticated user) and, in addition, still one saccade to represent an impostor from q=8 or 4 other random subjects. Nonusers and impostors were naturally represented by different random subjects from among n-1=131 subjects (an authenticated user excluded). At first, we implemented tests with this approach since we may assume that randomly selected subjects represent a more extensive area in the variable space than one authenticated. Nonetheless, we noticed that better results could be obtained by once copying the saccades of an authenticated user to balance the class size of an authenticated user's class and that of nonusers to be equal 2m. Copying once m saccades of the former increased the density of these saccades in a dataset.

In the verification procedure, the following symbols are also employed. All tests were repeated r=10 times since there were random choices of saccades of nonusers and impostors and also random initializations, among others, in multilayer perceptron networks. To test the remaining q saccades were taken to a test set where q was equal to 8 (Alternative 1) or 4 (Alternative 2). Symbols TP and FN equal the numbers of true positive and false negative decisions in classifications and FP and TN those of false positive and true negative decisions. On the basis of the two former, a decision for a subject is made whether a test subject is an authenticated user (Condition 1). Correspondingly, the two latter are used for a decision whether a test subject is an impostor (Condition 2).

$C1_1=C2_1=C1_2=C2_2=0$; % counters for correct classifications of authenticated users and those of impostors

For h=1:r % iterations of the main loop
  For i=1:n % one by one as an authenticated user
    $TP_2=TN_2=FP_2=FN_2=0$ (Alternative 2);
    For j=1:c % c combinations of recordings
      Take m saccades from 3 (Alternative 1) or 4 (Alternative 2) recordings of an authenticated user to a training set;
      Copy these m saccades in the training set;
      Take randomly 2m nonusers and one saccade from each and add these saccades to a training set;
      Train a model with 4m saccades of two classes: an authenticated user and nonusers;
      $TP_1=TN_1=FP_1=FN_1=0$ (Alternative 1);
      For j=1:q % tests of Condition 1
        Classify a test saccade of an authenticated user into either correct class
        $TP=TP+1$
        or incorrect class
        $FN=FN+1$;
      End
      For k=1:q % tests of Condition 2
        Classify a test saccade of an impostor into either correct class
        $TN=TN+1$
        or incorrect class
        $FP=FP+1$;
      End
      % Follow majority vote for decision
      If $TP_1 \geq FN_1$ **then** $C1_1=C1_1+1$ (Alternative 1);
      If $TN_1>FP_1$ **then** $C2_1=C2_1+1$ (Alternative 1);
    End
    % Follow majority vote for decision
    If $TP_2 \geq FN_2$ **then** $C1_2=C1_2+1$ (Alternative 2);
    If $TN_2>FP_2$ **then** $C2_2=C2_2+1$ (Alternative 2);
  End
End

(Alternative 1)
Accuracy of authenticated users=100 % · $C1_1/(r \cdot n \cdot c)$
Accuracy of impostors=100 % · $C2_1/(r \cdot n \cdot c)$
(Alternative 2)
Accuracy of authenticated users=100 % · $C1_2/(r \cdot n)$
Accuracy of impostors=100 % · $C2_2/(r \cdot n)$

Figure 3. Verification procedure for authenticated users (Condition 1) and impostors (Condition 2). Two different test settings are called Alternatives 1 and 2.

V. CLASSIFICATION RESULTS AND DISCUSSION

The main data mining task was to classify test saccades into two classes: an authenticated user or nonusers. There were n=132 subjects and r=10 main iterations in the verification procedure yielding 13200 decisions in Alternative 1 and 1320 decisions in Alternative 2.

TABLE I. CLASSIFICATION ACCURACIES OF MLP NETWORKS WITHOUT NORMALIZATION: MEANS AND STANDARD DEVIATIONS IN PERCENTS (*ON* EQUALS THE NUMBER OF OUTPUT NODES AND *C* CONDITIONS 1 AND 2)

Accuracies for two test alternatives, output node numbers *ON* and conditions *C*						
Alter-native	***ON***	***C***	***Number of hidden nodes***			
			4	6	8	10
1	1	1	71.8±0.8	70.8±0.9	**71.0±1.6**	70.4±0.8
1	1	2	64.9±0.7	65.2±1.8	**66.5±1.4**	66.4±0.9
1	2	1	72.1±1.2	71.8±1.3	**72.2±0.8**	71.7±1.1
1	2	2	66.8±1.5	66.8±1.5	**66.8±1.7**	67.0±1.5
2	1	1	78.5±3.3	79.6±3.4	78.9±2.5	**78.5±2.5**
2	1	2	74.2±3.1	78.0±3.4	79.8±3.6	**80.8±2.6**
2	2	1	81.9±2.8	**82.6±3.2**	82.2±3.5	79.8±2.8
2	2	2	77.8±1.9	**78.6±3.0**	78.6±2.7	79.2±3.0

TABLE II. CLASSIFICATION ACCURACIES OF MLP NETWORKS WITH NORMALIZATION AND ALTERNATIVE 2: MEANS AND STANDARD DEVIATIONS IN PERCENTS

Accuracies for output nodes and conditions					
Output nodes	***Condition***	***Number of hidden nodes***			
		4	6	8	10
1	1	81.1±2.7	78.4±3.9	**78.9±2.6**	78.4±2.5
1	2	75.8±2.4	79.5±3.2	**81.1±3.6**	80.5±3.2
2	1	80.5±1.8	80.1±4.2	80.0±4.5	**79.6±2.6**
2	2	77.3±2.4	81.7±3.6	80.2±4.3	**82.7±2.5**

We applied multilayer perceptron (MLP) networks [9] with 6 input nodes (6 variables), 4, 6, 8 or 10 hidden nodes and 1 or 2 output nodes for two classes. A validation error was used for MLP networks. It automatically stopped training after 9 or 10 epochs to avoid overtraining. Since we used the backpropagation algorithm in Matlab (MathWorks Inc., USA) also used for all tests of our research, we experimented with its training procedure variations including the adaptive learning rate, Powell-Beale restarts, batch gradient descent with momentum and Levenberg-Marquardt algorithm [10]. For actual tests we used the last method that yielded slightly better results than those of the other.

At first, we investigated possible differences between test results of Alternatives 1 and 2. Since the number of 5 recordings (20 saccades) of each subject was small subject to build training and test sets in data mining, it was important to test more than one alternative. However, the scarcity of the data did not allow more alternatives than the aforementioned two. We also varied the number of output nodes from 1 to 2. On the basis of the best results written in Bold in Tables I and II 2 output nodes produced accuracies 1-4% superior to those of 1 node.

TABLE III. CLASSIFICATION ACCURACIES OF LOGISTIC DISCRIMINANT ANALYSIS AND SVM WITH NORMALIZATION: MEANS AND STANDARD DEVIATIONS IN PERCENTS.

Accuracies for two test alternatives *A* and conditions *C*						
A	***C***	LogDA	***SVM kernels***			
			Linear	2nd deg.	3rd deg.	Gaussian
1	1	78.5±105	80.0±0.5	75.6±1.0	69.6±1.2	**84.9±0.7**
1	2	65.7±1.7	62.2±1.3	63.6±1.7	61.8±1.7	**73.0±1.3**
2	1	86.6±1.7	88.0±2.9	82.7±2.1	74.1±2.5	**92.1±1.9**
2	2	77.4±3.4	73.9±4.8	77.1±2.9	73.3±4.5	**84.8±1.9**

The scales of the variables markedly differed from each other. We tested MLP networks without and with normalization into interval [0,100]. The accuracies obtained without or with normalization had virtually no differences on an average. The results of the former are showed in Table I. Those of the latter are in Table II with Alternative 2 only, since Alternative 2 with the larger training set than with Alternative 1 indicated to be 7-13% better in Table I. The similar observation was gained for all later results. Note that while evaluating results we always have to look at both conditions at the same time, because they both are equally critical objectives. Note also that 50% is seen as a baseline result for Conditions 1 and 2. Because there are two classes of equal size, a random guess between them would be correct with probability 0.5. The number of the hidden nodes from 6, 8 or 10 yielded the best results for the pairs of Conditions 1 and 2.

We ran support vector machines (SVM) with the linear, quadratic, third degree polynomial and radial basis function (Gaussian) kernels. Table III shows results for SVM kernels and logistic discriminant analysis (LogDA). We ran tests for all four SVM kernels and logistic discriminant analysis by using both Alternatives 1 and 2 with and without normalization. Alternative 2 again generated higher results than Alternative 1. The use of normalization according to Table III did not affect average results seemingly at all compared with those not presented without normalization, mostly less than ±1%. SVM with the radial basis function (Gaussian) kernel was the best choice here, but differences were small compared with a few other kernels.

TABLE IV. CLASSIFICATION ACCURACIES OF RBF NETWORKS WITH NORMALIZATION: MEANS AND STANDARD DEVIATIONS IN PERCENTS

Accuracies for two test conditions				
Condition	***Spread and goal***			
	15 0.05	15 0.08	20 0.08	20 0.1
1	75.4±4.1	77.8±0.1	83.4±2.6	**88.5±1.8**
2	92.6±1.6	94.7±1.6	88.9±3.9	**88.9±1.9**

Ultimately, we exploited RBF networks by running system parameters of spread 10, 15, 20, 25, 30, 35, 40, 45 and 50, and goal 0.005, 0.02, 0.03, 0.05, 0.08 and 0.1. The best combinations of these were spread equal to 15 or 20 and goal equal to 0.05, 0.08 or 1.0. Final results of RBF networks

 ISBN: 978-1-61208-227-1

are presented in Table IV. For the RBF networks, our data required normalization, because our tests (not presented here) without it favoured Condition 2 and almost entirely failed with Condition 1. Thus, the results in Table V were computed with normalization and using Alternative 2.

Since our final objective to develop a biometric verification procedure on the basis of eye movements included a criterion that computing time should be fast, it is important to look at running times of the preceding tests. There were $132 \times 10 \times 5$=6600 models trained for every test type or structure (cell) in the case of Alternative 2. For Alternative 1 there were $132 \times 10 \times 10$=13200 models trained, correspondingly. The training and test time of an MLP network was around 0.5 s on an average. For RBFs that time of one network was around 4 s and for SVMs and LogDA less than 0.05 s. Let us remember that these execution times also included training not always necessary to do while applying a data mining method in actual applications, except when the system is used for the first time and then adaptively, say, after a successful login. In any case, even the use of the slowest method here was fast enough. Of course, additional computation is needed before the data mining phase to perform signal analysis. Still, this is also very fast, because its time complexity is linear and the length of eye movement signals is short, no more than a few thousand samples, say 1-3 minutes. Consequently, the running time would be minimal compared to such a recording time. At the beginning, in the course of a recording the eye movement camera system also makes image processing, but this is also close to real time. The camera system used consisted of only an initial calibration when taken into use. Thus, calibration required no additional processing time here.

VI. Conclusion

The MLP networks produced their best results with Alternative 2, 2 output nodes, 6 hidden nodes in Table I and 10 hidden nodes in Table II. The use of normalization did not improve the results obtained which were around 8% poorer than the best of SVMs and RBFs in Tables III-V. The Gaussian kernel was the best choice with SVMs. RBFs were very sensitive to normalization needed apart from the other being very insensitive to normalization.

The best results obtained were fairly good as 89% of the best results in Tables IV and V. We may assess that the best realistic accuracies based on various biometric verification references are around 95%. Thus, the results of this quite novel way to perform a biometric verification task are promising although more research has to be made to improve verification accuracies. A clear chance here is to collect a larger set of recordings from each individual. There were only five recordings with four large saccades per a subject. Forming a larger training set from each subject than now it is quite probable that we are able to improve classification results based on data mining methods. To compare with other scarce results presented thus far, our results were equal or better than various values 50-90% given in [11, 12].

The eye movement camera system used included a low sampling frequency of 30 Hz (frames per second). Still, verification was fairly successive. The low sampling frequency was, however, interesting since it was similar to that often used in cheap web cameras. We may except that in the future eye movement cameras are installed in computers or mobile devices to follow a user's gaze for various human-computer interface tasks [13]. If their sampling frequencies will be higher, e.g., 200 Hz, biometric verification with eye movements may well be realistic.

Acknowledgement

The authors thank prof. Ilmari Pyykkö, M.D., from the Department of Otorhinolaryngology, Tampere University Hospital, Finland, for advice subject to eye movements. The first author acknowledges the support given by Tampere Doctorial Program in Information Science and Engineering.

References

[1] X. Tan, B. Bhanu, and Y. Lin, "Fingerprint classification based on learned features," IEEE Trans. Syst. Man Cybern, Part C: Appl. Rev., vol. 35, no 3, pp. 287–299, August 2005.

[2] V. Conti, C. Militello, F. Sorbello, and S. Vitabile, "A frequency-based approach for features fusion in fingerprint and iris multimodal biometric identification systems," IEEE Trans. Syst. Man Cybern, Part C: Appl. Rev., vol. 40, no 4, pp. 384–395, July 2010.

[3] K. Venkataramani, S. Qidwai, and B. V. K. Vijayakumar, "Face authentication from cell phone camera images with illumination and temporal variations," IEEE Trans. Syst. Man Cybern, Part C: Appl. Rev., vol. 35, no 3, pp. 411–418, August 2005.

[4] Z. Sun, Y. Wang, T. Tan, and J. Cui, "Improving iris recognition accuracy via cascaded filters," IEEE Trans. Syst. Man Cybern, Part C: Appl. Rev., vol. 35, no 3, pp. 435–441, August 2005.

[5] R.W. Frischholz and U. Dieckmann, "BioID: A multimodal biometric identification system," IEEE Computer, vol. 33, no 2, pp. 64–69, February 2000.

[6] M. Juhola, "A syntactic method for analysis of sacadic eye movements," Pattern Recogn., vol 19, pp. 353-359, 1986.

[7] M. Juhola, H. Aalto, and T. Hirvonen, "Using results of eye movement signals in the neural network recognition of otoneurological patients," Comp. Meth. Progr. Biomed.. vol 86, pp. 216-226, 2007.

[8] S. Usui and I. Amidror, "Digital low-pass differentation for biological signal processing," IEEE Trans. Biomed. Eng., vol. 29, pp. 686-693, 1982.

[9] S. Haykin, Neural Networks, A Comprehensive Foundation, Second Edition, Prentice Hall, 1999.

[10] H. Demuth, M. Beale and M. Hagan, Neural Network Toolbox™ 6 User's Guide, The MathWorks, 2009.

[11] P. Kasprowski and J. Ober, "Eye movements in biometrics," Biometric Authentication: Int. Workshop, Springer-Verlag, LNCS, vol. 3087, pp. 248-258, 2004.

[12] O. V. Komogortsev, S. Jayarathna, C. R. Aragon, and M. Mahmoud, "Biometric identification via an oculomotor plant mathematical model," Proc. 2010 Symp. Eye Tracking Research & Applications, pp. 57-60, 2010.

[13] A. Holzinger, R. Geierhofer, and G. Searle, "Biometric signatures in practice: A challenge for improving human-computer interaction in clinical workflows," A. M. Heinecke, H. Paul (Eds.), Mensch & Computer, Oldenbourg Verlag, München, Germany, pp. 339-347, 2006.

Publication V

Biometric verification of a subject with eye movements with special reference to temporal variability in saccades between a subject's measurements

Youming Zhang, Jorma Laurikkala and Martti Juhola

Biometric verification of a subject with eye movements, with special reference to temporal variability in saccades between a subject's measurements

Youming Zhang, Jorma Laurikkala and Martti Juhola*

School of Information Sciences, 33014 University of Tampere, Finland

E-mail: Youming.Zhang@uta.fi, Jorma.Laurikkala@uta.fi, Martti.Juhola@cs.uta.fi

*Corresponding author

Abstract: We recently studied the application of saccadic eye movements, measured with video cameras, to biometric verification using subjects who receive identical stimulation. The properties of a subject's saccades may vary between measurements over the course of time, so to be useful as a means of biometric verification, the temporal variability of saccades should not distort verification results significantly. We investigated the effects of such variability by repeating the same test several times with the same groups of subjects. We found that temporal variability had only a minor effect on verification results when intervals were from a few hours to two months. Compared with the classification accuracies of approximately 90% of our earlier studies when measurements were run immediately one after another, our present verification accuracies were a few percent lower. In contrast, a long interval of approximately 16 months reduced the accuracies considerably. Our results indicate that reasonably short intervals between a subject's saccade measurements do not hinder verification based on them, while very long intervals between logins can pose a problem. Since most common electronic devices, such as computers and mobile phones, are used at frequent intervals, the analysis of saccadic eye movements seems to be viable technique for enabling biometric verification.

Keywords: Biometric verification, eye movements, saccades, classification, data analysis, temporal variability of saccades

Biographical notes: Youming Zhang, M.Sc., is a postgraduate in computer science. He took his B.Sc. degree at Xiamen University, China, in 2004, and M.Sc. degree at the University of Science and Technology Beijing, China, in 2007. He studies biometric verification and classification.

Jorma Laurikkala, Ph.D., is a lecturer in computer science. He took his M.Sc. degree at the University of Kuopio, in 1995, and his Ph.D. degree at the University of Tampere, in 2001. His research interests cover data analysis.

Martti Juhola, Ph.D., is a professor at the University of Tampere. His research interests include biomedical signal analysis, biometrics, pattern recognition and data analysis. He was an assistant, researcher and lecturer at the University of Turku, 1980-1992 and professor at the University of Kuopio, Finland, 1992-97.

1 Introduction

Fingerprints are probably the most extensively used biometric method to verify or identify a subject (Chang et al., 2011; Chen et al., 2006; Jain et al., 2000). Face images have frequently been studied for the biometric purposes (Kukharev et al., 2011; Torres, 2004; Wright et al., 2009). Other biometric data sources, such as iris images (Abdullah et al., 2011) and palm images (Prasad, 2010), have also been employed. Furthermore, these can be combined to enable multimodal means of improving verification (Mane and Jadhav, 2009).

We follow the definition of verification where a subject, for example the user of a computer or other device, has to be recognized to be the authenticated subject of that device. Anyone else should be recognized as an impostor. Identification is understood to mean a more complex task where any subject has to be separated from all others in a group. Thus, the former is a two-class classification task and the latter a multiclass task with far more than two classes.

We are interested in verification enabled by saccadic eye movements, which has been studied very infrequently thus far. The origin of our research was our extended eye movement studies connected to medical informatics, signal, and data analysis in the area of otoneurological research on human balance problems (for example, see Juhola, 1988; Juhola et al., 2011). An imperative factor has been the technical development of eye movement cameras during the last 10-20 years (Morimoto and Mimica, 2005; Duchowski, 2002). It is quite possible that in the future eye movement cameras will also be used for human-computer interaction for practical applications in addition to a keyboard, mouse, touchpad and other means. During recent years, the research of eye movement cameras has rapidly grown for this purpose (Duchowski, 2002). Consequently, employing eye movements for subject verification seems to be topical.

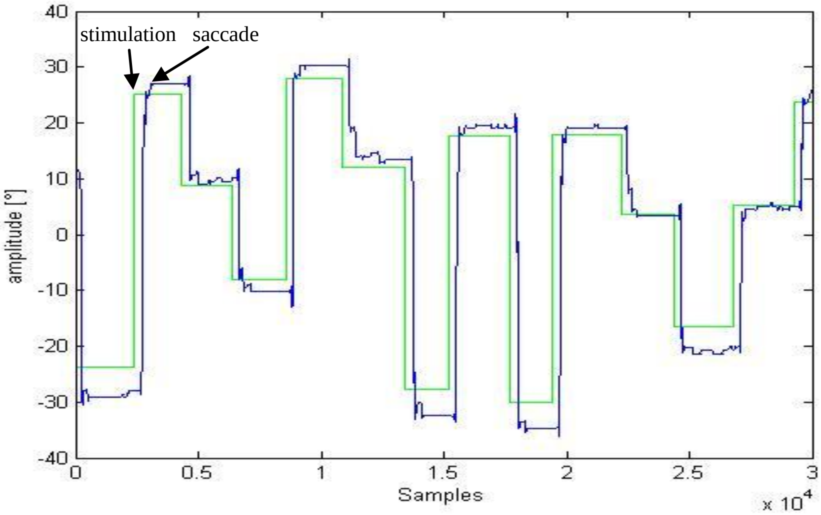
For biometric verification purposes, a voluntary response is good for a simple, standard-type stimulation presented similarly to anyone attempting to log in. It must be similar for every attempt so that an authenticated user can be detected and distinguished from possible impostors; this is a classification task.

One-dimensional physiological signals have rarely been studied for biometric verification or identification. Still, electrocardiogram (ECG) signals have recently been increasingly studied in this context (Shen et al., 2011; Wang et al., 2008). One could assume that its problem is the variation in heartbeat, but this can be alleviated with a suitable normalization of heartbeat waveforms (Lourenco et al., 2011).

Research on eye movement signals for biometric purposes began a few years ago. Kapczyński et al. (2006) and Kasprowski and Ober (2004) computed the cepstrum of a signal and classified its results with the naïve Bayes technique, nearest neighbour searching, decision trees and support vector machines. A mathematical model of the oculomotor system was used (Komogortsev et al., 2012) in which the parameters of the model were classified. Scanpaths in reading were recently used (Holland and Komogortsev, 2011), and different stimulation types were also considered (Holland and Komogortsev, 2012). Face images were used for stimulation (Rigas, 2012) where fixation points of a subject's gaze on a face image stimulation were detected in the plane of the horizontal and vertical axes. Minimum spanning trees were then computed between fixation points and these trees were applied to biometric identification by classifying subjects with nearest neighbour searching and support vector machines.

We began our research on the biometric use of saccadic eye movements by investigating the idea of using saccades for user verification of a computer or other device including a system to record a spectator's eye movements. In addition to signals acquired with a video camera system, we also employed electro-oculographic signals measured earlier at a higher sampling frequency, but otherwise in the corresponding way so that we were able to compare the results (Zhang et al., 2012; Juhola et al., 2013). We continued the research by using various classification methods and various ways to form training and test sets for classification (Zhang and Juhola, 2012). All the time we have collected more data from study (Juhola et al., 2013) to study (Zhang and Juhola, 2012) in order to increase the numbers of subjects and measurements.

Figure 1 A (blue) saccadic eye movement signal of 30 s as a response to its smooth (green) stimulation signal after interpolation from 30 Hz to 1000 Hz. The stimulation movements (angles) preceded the eye movements when the subject followed the stimulation light dot by gaze.



In the present research, we continued to develop our methodological work by testing classification methods not yet applied to this data. Most of all, we studied the effect of various time spans between the measurements of individual subjects. If such an interval is hours, days or more, the features of saccades could change. Our fundamental task was to clarify how much such possible variations might distort the saccadic biometric verification. Medical and physiological literature indicates that eye movements of a subject, as responses for the same stimulation type and angle, vary between repetitions (Smeets and Hooge, 2003) and, consequently, over the course of time. Nonetheless, the majority of such investigations seem to consider variability within a single measurement or a few instantly successive measurements. Opinions of how, and by how much, they change in the case of saccades seem to be different. According to Smeets and Hooge (2003), the main source of variability in saccades depended on the measurement devices employed. It was observed that maximum velocities of saccades depended on measurement systems when saccades were recorded with two different devices at the same time (Juhola et al., 1985).

Medical or physiological studies on the variability of saccades between days are rare. Intra-individual variability of maximum velocities of 20°, 40°, and 60° saccades recorded electro-oculographically were found to be statistically different at 9 a.m., 12 a.m., and 4 p.m. and on three separate early autumn days, the intervals of which were not mentioned (Schalén et al., 1984). However, all calculations of maximum velocities were made manually with a pencil and ruler, because recordings were received in the form of rolls of paper from the ink plotter of the recording device. Calculations based on manual approximation are not as equally precise as modern computer-based work. In particular, there were only six subjects, which is a very small sample from a statistical perspective. On the other hand, no statistically significant differences were reached when the average maximum saccadic velocities of 58 healthy subjects were measured repeatedly over a span of two weeks (Bollen et al., 1993). There is intra-individual and inter-individual variability in saccades with the same stimulation. While the former is considerable (Bollen et al., 1993; Zhang et al., 2012), the latter seems to be more influential for the separation of subjects, because we have succeeded in running such tests that yielded good verification accuracies (Zhang et al., 2012; Zhang and Juhola, 2012; Juhola et al., 2013). Other eye movement verification or identification experiments mentioned above support this result. Notwithstanding these, it was a rather obvious problem if the intra-individual variability of saccades decreased the accuracy of biometric verification when the time span between the measurements is longer than a few minutes or less.

2 Acquisition of eye movement data

2.1 Eye movement camera system used

We conducted our eye movement recordings with a head-mounted video camera system (Visual Eyes, Micromedical Technologies, UK). Its resolution was 320×240. The system did not require calibration for each measurement, but was calibrated only once, after its installation. Since its sampling frequency (frame rate per second) was only 30 Hz, we interpolated every signal up to 1000 Hz. Interpolation (estimation) from 30 Hz to an upper frequency was necessary, since information content of saccades can reach 50-100 Hz (Bahill et al., 1981; Juhola et al., 1985) suggesting sampling frequencies above 200 Hz. Sampling frequencies up to 1000 Hz have been applied (Bahill et al., 1981) to measurements of saccadic eye movements. Our motivation to use the low sampling frequency of 30 Hz was to simulate situations where a possible eye movement camera close to properties of simple web cameras (or front cameras of cellular phones subject to sampling frequencies) could detect eye movements in future technologies. If they included low sampling frequencies, it would be interesting to study here whether it is sensible to apply a low sampling frequency such as 30 Hz and interpolate these signals to have a higher, “artificial” sampling frequency producing estimates for saccades. Of course, the higher original sampling frequency would be better, but interpolation was, after all, very useful in our previous works with the same video camera system (Juhola et al. 2013; Zhang et al., 2012; Zhang and Juhola, 2012). Recently, results were presented that the sampling frequency of 50 Hz enabled fairly accurate computation for maximum velocity values of 5° only saccades (Weirts et al., 2008). Although the bandwidth of saccades exceeds its Nyquist frequency (25 Hz), obviously their main influence is below it and, thus, also below 70 Hz (Bahill et al., 1981).

Even if an interpolated signal was not exactly the same as the actual higher sampling frequencies, it was sufficient for our verification purposes. As such, a frequency of 30 Hz would obviously have been too low to attempt successful verification. Our previous investigations showed that the estimation of saccades by means of interpolation helped to separate an authenticated user from among other subjects. We also wanted to use the same eye movement camera system as before because our intention was to compare our new results with earlier results. There was a new essential trait in our present research: the possible influence of temporal variability of saccades in biometric verification.

The video camera system included a built-in image processing system to search for the pupil of a subject's eye to compute saccades according to the positions of the pupil. The system required no separate calibration except when the system was installed. Since there were two video cameras, one for each eye, two horizontal signals were used at every measurement. The better of the two, i.e., the one with less possible noise or fewer artefacts, was selected. The amplitude resolution was better than 1°.

2.2 Stimulation

In the system there are two cameras, one for each eye. To make the stimulation as simple and easy for a subject as possible, only horizontal eye movements were utilised. Figure 1 contains part of a saccade signal and the corresponding stimulation signal. The moving gaze of a subject followed a light dot (LED) along a black bar. The distance between the target on the LED bar and the subject was 0.74 m for all measurements. The computer-controlled light dot jumped to another place on the bar, in other words, one LED was switched off and another switched on. The angle formed by them in the visual field was random to the subject. Such angles were constant when the distance of the eyes to the bar was constant. Variation of time intervals between the jumps of the light dot made the stimulation movements seem random for a subject. Still, this was a fixed series of stimulation movements because it is essential that the same stimulation sequence is presented for every subject. This way, responses of a different subject and those of the same subject can be compared with each other. The sequence was so complicated that a subject could not learn it even if repeated.

A simulated log on was as follows. The stimulation was used jointly with an occurrence of a subject starting a computer session by logging on to a computer. The log on required no written password, but instead the computer recognized its authenticated user by recording the user's saccades at the beginning. The computer showed the same stimulation sequence of light dot jumps on its screen, and the user's task was to look at the stimulation dot which was moving approximately once every two seconds for one minute. Both stimulation amplitude (lengths of jumps of the light dot) and time intervals between jumps were varied, and most amplitudes were large. In this test, they were 48° at their widest. The large amplitudes helped obtain inter-individual variability from the subjects (Henriksson et al., 1980). Only the largest saccade amplitudes were used for verification. Saccades of smaller amplitudes were only used to make the stimulation sequence seem random for the subject.

2.3 Eye movement measurements

We recorded saccades from two sets of subjects, one set consisting of 22 and the other 12 subjects. Every subject was asked not to move his or her head during measurements to keep the angles of the saccades stable. Because the duration of measurements (signals) was 64 s, each of which contained only four large saccades of 48°, five such signals were recorded from every subject per session. Altogether, we took 20 saccades of 48° from every subject. Signals were interpolated to raise its (artificial) frequency up close to 1000 Hz. Naturally, an interpolated signal is not precisely the same as an actual measurement with a 1000 Hz sampling frequency, but it can be used as an estimate. We applied cubic spline interpolation since after having compared this to linear and cubic Hermite interpolation, we obtained accuracies a few per cent higher (not presented for brevity) in verifications between an authenticated user (subject) and other subjects.

The set of 22 subjects consisted of 14 males and 8 female. Their mean age was 29 ± 8 years. The other set included only 12 subjects. Originally 19 subjects took part in the first measuring session, but 7 of them, having moved away, did not attend the latter session. Consequently, those 7 absent were omitted from our research. The mean age of the 12 was 36±12 years. The group included 8 males and 4 females. There were 7 subjects who were involved in both subject sets. The eyesight of all subjects who did not wear spectacles was checked before recordings. Two subjects had age-associated presbyopia, and they were involved in both data sets. In addition, one subject's eyes were rather astigmatic and the subject was also myopic. Furthermore, this subject used allergy medication (antihistamine and cortisone) throughout the involvement in the measurement sessions of the 22-subject data set, but nobody else had used medication or alcohol the day before each measurement. This subject was also a member in the set of 12 subjects. Some medications (van der Meyden et al., 1989), alcohol (Jäntti et al., 1983; Lehtinen et al., 1979) and age (Munoz et al., 1998) are known to affect values of saccade variables. Nonetheless, we did not discard the measurements of this subject as an outlier because we did not think it sensible to regard any subject as an outlier in a biometric verification task, as would the subject might have been, for instance, in medical research. In a way, this made the research even more challenging.

For the set of 22 subjects, measurements were performed on five separate days. Every measuring day consisted of two measurement sessions, the one in the morning and the other in the afternoon with a span of approximately five hours. This was arranged because circadian rhythm is sometimes seen as a plausible influence (Schalén et al., 1984; Fransson et al., 2008), given the assumption that eye movements could be affected by time of day, e.g., between morning and afternoon. To study different spans between measurement sessions, an interval of 7 days was used between the first and second measurements for all subjects. For the second and third measurement sessions, a longer time was chosen, varying from 15 to 35 days with 26±8 days on average between subjects. For the third and fourth measurement sessions, there were also varying intervals from 18 to 59 days, with 31±11 days on average between subjects. Ultimately, there was one day between the fourth and fifth measurement sessions. All in all, these covered an average of almost 10 weeks. For the set of 12 subjects, the span was much longer, approximately 16 months. These consisted of one daily measurement session only. The numbers of the subjects associated with our longitudinal research were limited for the sake of several repeated tests and practical reasons connected to attendance of the subjects. These types of longitudinal saccade test series have not perhaps ever been made for the same subjects with the same stimulations and measuring devices; this is surely the first time for biometric verification.

3 Signal data preprocessing and analysis

3.1 Recognition of saccades and stimulation movements from signal data

After interpolation, the first task in signal analysis was to recognize the beginning and end of every saccade so that variables were possible to compute accurately from the saccades. After interpolation, the first derivative, angular velocity of eye movements, was approximated on the basis of a simple two-point differentiator (Juhola et al., 2013). A threshold of 50°/s was used for velocity in order to search for the beginning and, thereafter, 10°/s for the end of every saccade (Figure 1). The greater threshold was used for beginnings than ends of saccades because we had to accurately detect beginnings for the computation of latency values. The greater threshold value helped to pass slightly noisy samples and, thus, avoid resulting in incorrectly or inaccurately detected beginnings of saccades. On the other hand, ends of saccades were detected with the lower threshold because normally the latter half of a saccade curve is less steep than the former half. This is seen from Figure 1 and later observed from our results that magnitudes of maximum decelerations are smaller than those of maximum accelerations.

3.2 Analysis of saccades

Stimulation signals were analysed to find at which time each stimulation movement, i.e., a jump of the light dot, started. An actual response to a stimulation movement had to be later than 0.120 s after the stimulation. This lower bound was used since the brain of a human being requires a short time to recognize the target and to control a movement of gaze. If there were any earlier response, this would be rejected as a probably anticipated eye movement, not a response to the stimulation movement.

The data of a subject's measurement session included 20 large amplitude saccades (stimulation 48°): 4 saccades in 5 consecutive measurements made one after another with approximately one minute intervals. After the detection of saccades, the variables of latency, amplitude, accuracy, maximum velocity, maximum acceleration and maximum deceleration (Figure 2) were calculated from every valid large saccade of the subject. Latency or reaction time is the time between the beginnings of a saccade and the corresponding stimulation movement. Amplitude accuracy is the difference of the amplitudes (angles) of a stimulation movement and saccade. The maximum of velocity values was computed (Figure 2). Angular acceleration and deceleration were computed as the approximated second derivative of a saccade. We did not use the seventh possible variable, duration, which is the time difference between the beginning and end of a saccade, because we found it slightly less useful than the others, based on its smaller ratio of inter-individual and intra-individual variability (Zhang et al., 2012).

To explore the separation ability of the features, we calculated ratios of inter-individual and intra-individual variability by using standard deviations as follows: subscript *j* corresponds to the *j*th variable, *n* to the number of subjects, $\bar{u}_{ij}$ to the mean of variable *j* of subject *i*, $\bar{a}_j$ the mean of variable *j* for all subjects' saccades, u_{kj} the value of variable *j* of saccade *k* for subject *i*, and n_i the number of the saccades for subject *i*. We may expect that the higher the ratio, the better the separating variable. Results are shown in Table 1 promising reasonable chances for separation between subjects.

$$R_j = \frac{\sqrt{\dfrac{1}{\sum_{i=1}^{n} n_i} \sum_{i=1}^{n} \sum_{k=1}^{n_i} (u_{kj} - \bar{a}_j)^2}}{\dfrac{1}{n} \sum_{i=1}^{n} \sqrt{\dfrac{1}{n_i} \sum_{k=1}^{n_i} (u_{kj} - \bar{u}_{ij})^2}}$$

The results R_j in Table 1, computed for the larger set of subjects, indicated that the variables would be useful for the separation of subjects. Naturally, using simply standard deviations in this way only produces a crude estimate.

4 Verification procedure and tests

Two test conditions were performed to study the verification of an authenticated user on the basis of saccadic eye movements. We applied two classes to the first test condition:

saccades of an authenticated user and those of other named non-users. In addition, we employed a third group of subjects called impostors. Impostors were different from an authenticated user and non-users. The third group was then a test set for the second test condition.

Figure 2 A hypothetical saccade on the left producing six variables computed: saccade amplitude, accuracy, latency, maximum angular velocity, maximum angular acceleration and deceleration.

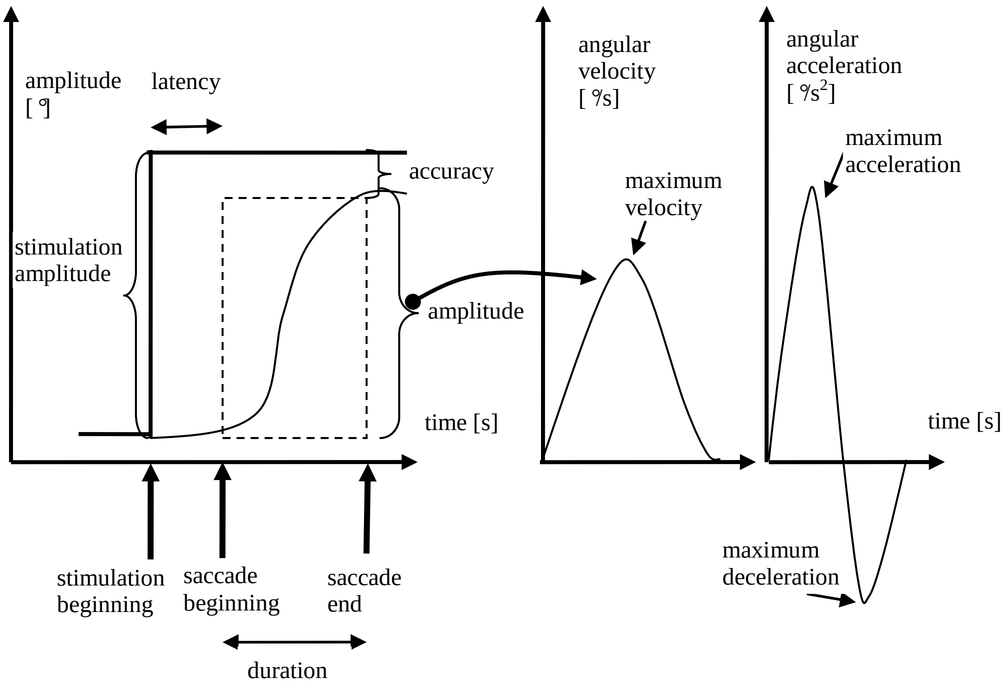


Table 1 Means and standard deviations of variables for the data set of 22 subjects and their ratios between inter-individual variability and intra-individual variability.

Amplitude °	*Accuracy °*	*Latency s*	*Maximum velocity °/s*	*Maximum acceleration °/s²*	*Maximum deceleration °/s²*
Mean and standard deviation					
54±15	6±11	0.27±0.06	1106±336	50343±23501	47331±25298
Ratio R_j of inter-individual and intra-individual variability					
1.13	1.13	1.09	1.13	1.06	1.05

The following procedure represents the pseudocode of the verification test for the larger set of n=22 subjects. For the smaller set of 12 subjects, the principle was similar, but there were 2 measurement sessions instead of 10 of the former. When classification accuracy values of Conditions 1 and 2 are subtracted from 100%, we obtain a false rejection rate (FRR) and a false acceptance rate (FAR) (in this order) frequently used for biometric identification and verification. However, we did not limit our tests with an equal error rate (EER) to have an opportunity to see results of various situations.

```
% counters for successful classifications of authenticated users and those of impostors
% (C1 for Condition 1 and C2 for Condition 2)
C1=C2=0;
For h=1:r  % iterations of the main loop
        For i=1:n % one by one as an authenticated user
            For j=1:c % c measurement sessions
                Take m saccades from s=c-1 measurement sessions (excluding the jth)
                of an authenticated user to a training set;
                Copy these m saccades in the training set;
                Take randomly 2m saccades from p non-users and add these saccades
                to a training set;
                Train a model with 4m saccades of two classes: an authenticated user
                and non-users;
                TP=TN=FP=FN=0;
                For k=1:q1 % tests of Condition 1
                        Classify a test saccade of an authenticated user into either
                        correct class
                        TP=TP+1
                         or incorrect class
                        FN=FN+1;
                End
                For l=1:q2 % tests of Condition 2
                        Classify a test saccade of an impostor into either correct class
                        TN=TN+1
                        or incorrect class
                        FP=FP+1;
                End
                % Follow majority vote for decisions
              If TP≥FN then C1=C1+1;
              If TN>FP then C2=C2+1;
            End
        End
End
Accuracy of authenticated users=100 % · C1/(r·n·c)
Accuracy of impostors=100 % · C2/(r·n·c)
```

An authenticated user was verified in Condition 1. The pseudocode showed the building of a training set by using saccades from 9 measurement sessions of an authenticated user and its test set of 20 saccades of one measurement session for classification. This was iterated for 22 subjects and repeated several times by varying randomly which saccades were selected from s=9 measurement sessions to the training set. For Condition 2, subjects excluding an authenticated user were set into either non-users or impostors. Because the other set of 12 subjects was so small, we took an authenticated user and non-users from it, but randomly selected impostors from an entirely separate set of 132 subjects whose data was measured earlier, but precisely in the same way (Zhang and Juhola, 2012).

In our earlier research (Zhang and Juhola, 2012), we found out that it is reasonable to apply a training set containing approximately as many training saccades of both an authenticated user and non-users. Since there are far more saccades available from the latter than from a single authenticated user, we balanced the number of the former by doubling its training saccades and took the same number from the set of non-users.

Here, we first explain the tests for the set of n=22 subjects. There were five times two measurement sessions made for every subject. Thus, we simulated verification so that one of c=10 measurement sessions formed a test set of $q1$=20 saccades, and its training set was randomly picked from the other s=9 measurement sessions. For Condition 1, there were m=9·20=180 saccades doubled from an authenticated user and 2m=10·36=360 from p=10 randomly taken non-users in the training set. Thus, $q1$=20 saccades of every subject as an authenticated user were tested against b=36 saccades from other p=10 subjects (non-users) randomly chosen from n-1=21 subjects. This was run one by one for n=22 subjects. This way there were 22·10·20=4400 saccades tested for 22·10=220 models trained. For Condition 2 one saccade was taken randomly from $q2$=n-p-1=11 impostors after selecting an authenticated user and p=10 non-users. There were now 22·10·11=2420 saccades tested for 22·10=220 models trained. All these tests were repeated r=10 times for other classification methods except learning vector quantization, for which the previous models were run once only because of their relatively long execution times (a few hours).

Tests for the longer period set of 12 subjects were performed almost in the same manner. However, there were only two measurement sessions instead of 10 as in the preceding set-up. As one alternative, the data of the former session was used as the source of training sets and that of the latter session as test sets. For the other alternative, these roles were changed. The former measurement session contained three measurements, giving 12 large amplitude saccades from every subject. The latter consisted of five measurements from a subject, i.e., 20 large saccades. Accordingly, we used n=12 subjects, c=2 measurement sessions and $q1$=$q2$={12 | 20} depending on two session days in this order. There were p=n-1=11 non-users and one saccade from $q2$ impostors from an outside set of subjects. For the training sets we took 4 saccades randomly from each of p non-users, altogether 44 for the first alternative, but only 2 saccades from p non-users, total 22, for the second alternative. In this way, we kept a ratio of 1:2 between the saccades of an authenticated user and non-users in any training set, but balanced this by doubling the saccades of an authenticated user. The doubling was performed because of scarcity of authenticated user saccades. All in all, there were 12 subjects and 10 iterations, producing 120 models trained differently for both opposite alternatives. Thus, there were 120·12=1440 and then 120·20=2400 saccades tested for Conditions 1 and 2.

We performed classifications by applying linear, quadratic and logistic discriminant analysis, naïve Bayes rule, multilayer perceptron networks (MLP), radial basis function networks (RBF), learning vector quantization (LVQ) and support vector machines (SVM). Variable values were used as either normalized into interval [0,1] or without normalization. According to the previous means of the variables in Section 3, the scales of the variables varied considerably. Therefore, normalization could be influential on the results for some classification methods.

5 Results

Computation was implemented and executed with Matlab R2010a™ (MathWorks Inc., USA). Results are given as classification accuracies (true positive rates) as a percentage, i.e. how many classifications were correct compared to all cases tested. If false rejection rates are desired (false negative rate FNR), these are formed by decreasing an accuracy value from 100% for Condition 1. False acceptance rates (false positive rate FPR) can be calculated similarly from Condition 2.

5.1 The results of the larger data set from 10 measurement sessions.

We ran multilayer perceptron networks (MLP) with a Levenberg-Marquardt learning algorithm and used 4, 6, 8 or 10 hidden nodes and 1 or 2 output nodes. Before the tests, variable values were normalized into [0,1]. The results in Table 2 were better than those (not presented) without using normalization. Note that results of both test conditions have to be considered at the same time; we pursue as high accuracies as possible for both. The combination of 10 hidden nodes and 2 output nodes gave the best results.

Next, we ran radial basis function networks (RBF). This time normalization was necessary, since results without it would have been poor. The results are shown in Table 3. The best are only shown from various pairs of parameter combinations.

Table 2 Means and standard deviations of accuracies as a percent for multilayer perceptron networks with data normalization. The best value pair is given in bold face.

	1 output node		*2 output nodes*	
Hidden nodes	*Condition 1*	*Condition 2*	*Condition 1*	*Condition 2*
4	85±2	77±3	86±1	75±3
6	86±1	77±4	85±2	81±3
8	84±2	81±3	85±1	81±4
10	84±3	80±5	**87±1**	**81±3**

Table 3 Means and standard deviations of accuracies as a percentage for radial basis function networks with its various parameter values of goal and spread and with data normalization. The best value pair is given in bold face.

	Spread 15		*Spread 20*	
Goal	*Condition 1*	*Condition 2*	*Condition 1*	*Condition 2*
0.11	83±2	89±3	84±2	87±2
0.12	**84±1**	**90±3**	85±2	86±3
0.13	84±2	89±2	86±2	85±2

Then we classified with learning vector quantization algorithm, linear, quadratic and logistic discriminant analysis, naïve Bayes rule and support vector machines with four kernels. Since learning vector quantization was very slow (hours) compared with the other classification methods, it was calculated only once (in the main loop of the verification procedure, *r* was equal to 1) and no standard deviations were computed. For support vector machines, several different parameter combinations were computed and the best given in Table 4. Normalization was used only for learning vector quantization tests. For the other methods it did not bring improvement. As for the tests in Tables 2-4, the support vector machines with the RBF kernel yielded the best results of all, but the best alternative among the radial basis function networks was very close to it.

Table 4 Means and standard deviations of accuracies as a percentage for learning vector quantization (LVQ), discriminant analysis, naïve Bayes rule and support vector machines with data normalization. The best value pair is given in bold face.

Condition 1	*Condition 2*	*Condition 1*	*Condition 2*
LVQ: step 30, rate 0.005		LVQ: step 50, rate 0.001	
84	57	80	55
Linear discriminant analysis		Quadratic discriminant analysis	
82±2	74±4	87±1	65±4
Logistic discriminant analysis		Naïve Bayes	
81±3	74±4	84±2	58±5
Support vector machines			
linear		Quadratic	
82±1	74±3	87±1	78±3
polynomial 3rd degree		RBF, σ=2.5	
88±2	81±3	**88±2**	**86±3**

The results in Tables 2-4 corresponded to a situation in which classification was based on majority votes, in other words, a threshold of 0.5. To portray results more extensively, we computed ROC curves. In Figure 3 there are curves for discriminant analysis and naïve Bayes rule and in Figure 4 those of support vector machines. These indicate that support vector machines with the RBF kernel was the best.

5.2 The results for the set of 12 subjects where the span between the measurement sessions was 16 months

We ran fewer classification methods than above by employing only linear and quadratic discriminant analysis and support vector machines (Table 5). The best result was gained with quadratic discriminant analysis. We did not only maximize the average of the two conditions, but also chose such pair of values that were as close to each other as possible. For verification, both conditions should be maximized at the same time. We see that all results favoured strongly Condition 2, but the results of Condition 1 were moderate. We may assume that this was caused by gradual changes of subjects' saccade variable values over the long period of 16 months. This is, however, merely an assumption, because the number of subjects is too small to produce statistically definite conclusions.

Ultimately, we dealt with our two data sets statistically in order to see whether there were any statistically significant differences for the values of the saccade variables between morning and afternoon (day of time), and between the five measurement days. First, we describe results for the set of 22 subjects. Figure 5 describes the 95% confidence intervals of the subjects' latency values. The sixth subject was the myopic allergy medication-user. There is no doubt that the measurements of this subject differ from the others.

Figure 3 ROC curves for linear, quadratic and logistic discriminant analysis and for naïve Bayes rule. TPR refers to true positive rate and FPR to false positive rate.

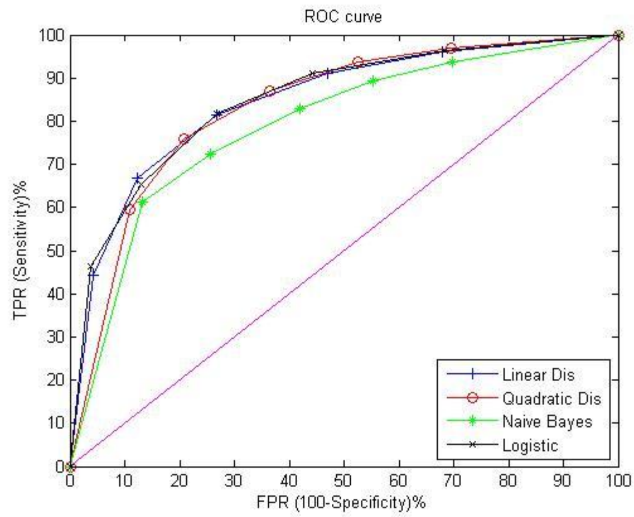


Figure 4 ROC curves for support vector machines with linear, quadratic, polynomial and RBF kernels. TPR refers to true positive rate and FPR to false positive rate.

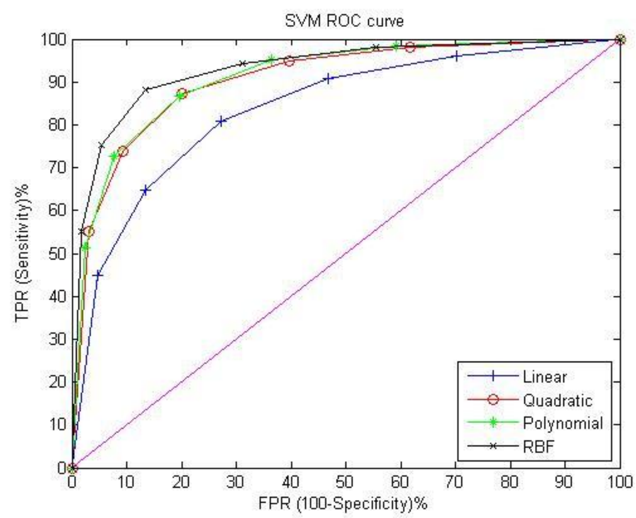


Table 5 Means and standard deviations of accuracies for the set of 12 subjects as a percentage for discriminant analysis and support vector machines with data normalization. The best value pair is given in bold face.

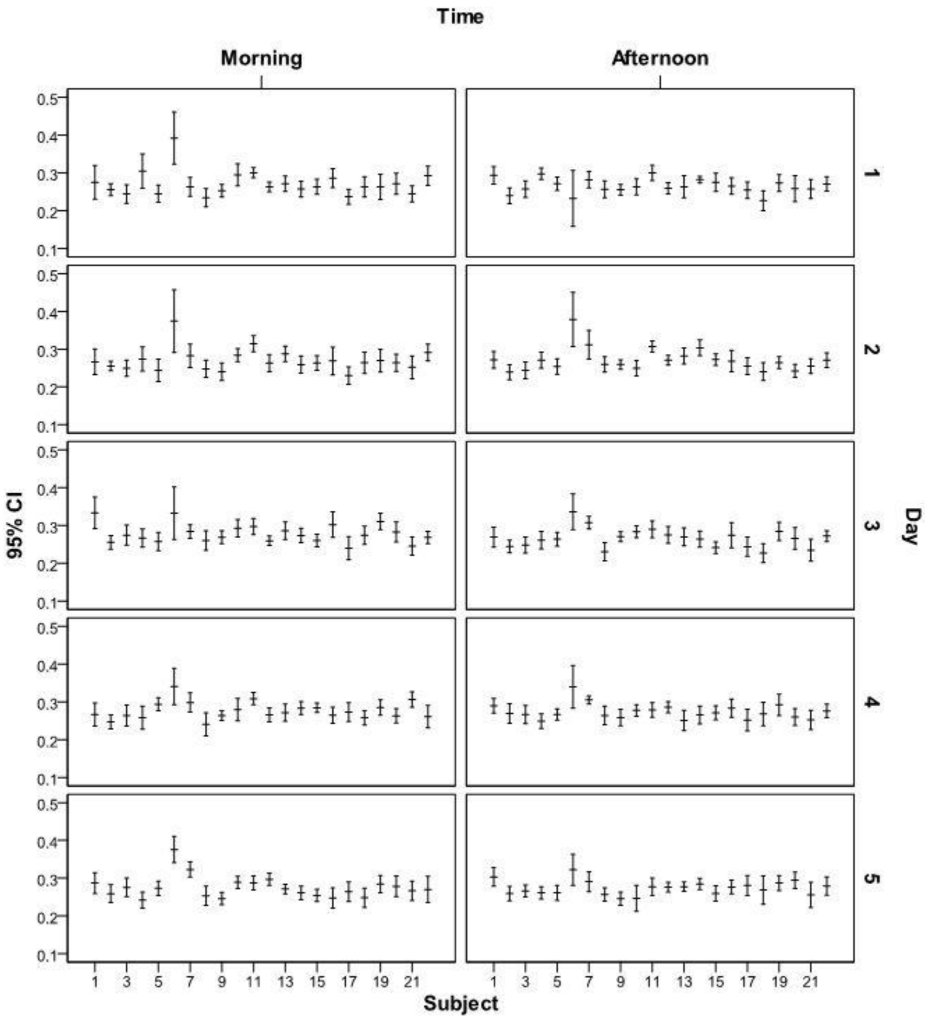
Iteration j	*Condition 1*	*Condition 2*	*Condition 1*	*Condition 2*
	Linear discriminant analysis		Quadratic discriminant analysis	
1	34±6	83±12	32±4	95±6
2	51±6	65±12	16±5	94±6
	Support vector machines			
	Linear		Quadratic	
1	45±4	72±11	**43±10**	**76±11**
2	52±12	65±9	52±11	58±14
	polynomial 3rd degree		RBF, σ=2.5	
1	34±14	77±9	43±8	73±10
2	42±16	67±14	43±10	68±9

Although there is naturally some intra-individual and inter-individual variability between days, and morning and afternoon, quite similar intervals of latency values were obtained for the same subject throughout all measurement sessions. This can be seen for the 22 subjects in Figure 5. The sixth subject had exceptionally large confidence limits, but this was caused by poor sight and medication. As for the other five variables, their confidence limits followed the corresponding trends as in Figure 5. All these support our assumption about the current data that intra-individual variability does not harm essentially biometric verification with saccades and inter-individual variability is large enough to enable verification.

Repeated measures ANOVA was performed on all of the 22 subjects and all but one subject (the subject with exceptionally poor eyesight). The statistical significance (α = 0.05) of the main effects of day and time of day, and the interaction between day and time of day, were assessed both with multivariate and univariate testing. Wilks' Lambda was applied in the multivariate testing. Since Maulcy's test of sphericity indicated that the assumption of sphericity was not met, the degrees of freedom were adjusted using Greenhouse-Geisser correction in the univariate testing. Multivariate tests showed that the main effect of day was significant on maximum acceleration ($p = 0.02$) in the data set of the 22 subjects and on maximum acceleration ($p = 0.03$) and maximum deceleration ($p = 0.04$) in the data set of the 21 subjects. The other main effects, as well as all of the interactions, were insignificant. There were no significant results in the univariate testing.

Next, we consider statistical results of the set of 12 subjects with the span of approximately 16 months between two measurement sessions. The sixth subject with poor sight and allergy medication was the same as in the data set of 22 subjects. The Mann-Whitney U test was used to assess the differences between the earlier and later measurements, both in the whole data and in individual subjects. A non-parametric method was used because the intra-individual groups were quite small and their sizes unequal (12 and 20 saccade measurements). The significance level $\alpha = 0.05$ was corrected with the Bonferroni method to $0.05 / 6 \approx 0.008$ to counter the effects of multiple comparisons.

Figure 5 Confidence intervals (CI) of the latency values [s] of the 22 subjects for five days and two measurement sessions (morning and afternoon) of each day. The locations in order from the left to the right represent the same subjects for all sessions. Note the sixth subject's bar of every measurement session. It represents the medication-using subject with poor sight.



The analysis of the whole data showed that there were statistically significant ($p < 0.008$) differences between the groups in all of the variables, except in latency, both with and without Bonferroni correction.

Next, we explored the data of individual subjects (Table 6). A total of six subjects had no significant differences between the earlier and the later measurements. One subject had four significant differences in four variables, while three subjects had significant differences in three variables. The other two subjects had one and two significant differences. The earlier maximum velocity values of five subjects differed significantly from those of the later measurements. Accuracies and maximum decelerations of three subjects were significantly different. These results are in accord with the analysis of the whole data, where the most significant differences were found in the maximum velocity ($p < 10^{-7}$), accuracy ($p < 10^{-6}$) and the maximum deceleration (p = 0.0003) variables. Note that the total of pairs (cells) of a subject and variable is 72 in Table 6, but 24 of these obtained significant differences without Bonferroni correction and 16 with this. To summarise, one third or less than one fourth (with Bonferroni) of the pairs of a subject and variable encountered significant changes after a period of approximately 16 months.

Table 6 Mann-Whitney U test results (p values) for the set of 12 subjects. Symbols #* mean the number of the variables with significant differences at α = 0.05 without Bonferroni correction and #+ those with (or without) correction for both rows and columns.

Subject	*Latency*	*Maximum velocity*	*Amplitude*	*Accuracy*	*Maximum acceleration*	*Maximum deceleration*	*#*/#+*
1	0.22	0.0482*	0.77	1.00	0.98	0.076	1/0
2	0.89	0.070	0.26	0.24	0.86	0.80	0/0
3	0.17	0.27	0.076	0.0118*	0.33	0.13	1/0
4	0.0053+	0.63	0.95	0.09	0.86	0.48	1/1
5	0.72	0.0005+	$<10^{-6}$+	$<10^{-8}$+	0.95	0.26	3/3
6	0.92	0.14	0.058	0.0263*	0.21	0.41	1/0
7	0.0019+	0.0048+	0.83	0.0063+	0.099	0.0005+	4/4
8	0.98	0.0023+	0.26	$<10^{-7}$+	0.0236*	0.076	3/2
9	0.65	0.0072+	0.35	0.064	0.0020+	0.0072+	3/3
10	0.29	0.0092*	0.099	0.60	0.16	0.69	1/0
11	0.25	0.0482*	0.29	0.69	0.099	0.0482*	2/0
12	0.51	0.0001+	0.24	0.0359*	0.0011+	0.0027+	4/3
#*/#+	2/2	8/5	1/1	6/3	3/2	4/3	∑=24/16

6 Conclusion and discussion

According to Tables 2-4, the support vector machines with the RBF kernel produced the best results and radial basis functions with suitable parameter values were almost equally high for the set of 22 subjects. Regarding the set of 22 subjects, it was somewhat surprising that these results were so good compared with our previous results with the same eye movement camera system and stimulation, for which we had not yet experimented with measurement sessions from different times of day and different days (Juhola et al., 2013; Zhang et al., 2012; Zhang and Juhola, 2012). Our previous best results were only 1-4% higher than those presented, although they were measurement sessions where either three or five measurements were run immediately one after the other. These results clearly support the hypothesis that values of saccade variables do not change intra-individually highly on average between morning and afternoon and, moreover, between days or even a few weeks. Although there was some intra-individual variability, it was not statistically significant except, at $\alpha = 0.05$, for maximum acceleration and deceleration of multivariate tests between days only, and not for times of days or both, and not at all for univariate tests. It is of consequence why specifically acceleration and deceleration were found to differ between days. These variables are the most sensitive to changes in saccade signals since they approximate the second derivatives of saccades.

The original low sampling frequency of 30 Hz might have been the cause why very few statistically significant differences were obtained. Perhaps interpolation did not help to reveal sufficiently the presumably small variability of saccade variable values. To study this detail more precisely in the future, we are going to use other eye movement camera systems with higher sampling frequencies. It is known that frequency content of saccades can reach 70 Hz, perhaps even 100 Hz (Bahill et al., 1981; Juhola et al., 1985). After all, in the present study we were especially interested in investigating whether and how it is possible to verify an authenticated user with saccades (after interpolating their signals) although the original sampling frequency was low, 30 Hz.

It must be remembered that the total span for the set of 22 subjects was almost 10 weeks. Therefore, the span was fairly long, but did not deteriorate verification accuracies. The number of subjects was not large, but it was statistically tolerable to suppress random effects. The statistical results were, after all, a minor part of our research. The major and by far most important result was the relatively good biometric verification accuracies; the best obtained were 86% for an authenticated user (Condition 1) and 88% for impostors (Condition 2) with support vector machines. If we aim at 95 %, which can obviously be reached with several biometric means such as fingerprints, and remember that we used long intervals up to weeks in length between measurement sessions, our results are promising.

The smaller set of subjects included only 12 people. The span between its measurement sessions was very long at 16 months. This made it difficult to verify an authenticated user correctly, but greatly aided verification of impostors. This and the Mann-Whitney U results in Table 6 denote the probable cause that saccades had been changed intra-individually so much that an authenticated user was difficult to classify, which, on the other hand, alleviated the opposite task, classification of impostors.

The results gained cannot be compared to results of other research, since no prior comparative, extensive tests for biometric verification have been done to the best of our knowledge. Recently, two measurements with an interval of 20 minutes were repeated after a week for each of the 32 subjects (Holland and Komogortsev, 2011). However, this research was very different from ours and the number of measuring days was only two. In other areas, such as medicine or psychology, some corresponding tests have been made with saccades, but, naturally, they considered intra-individual and inter-individual variability, not as classification between subjects.

In the future, we are going to study more subjects with various spans. We shall also use eye movement video camera systems with sampling frequency and other properties higher than the one now used.

To summarize, we see that our present and previous results show good opportunities to apply saccade measurements to user verification. Although the data sets of 22 and 12 subjects were small, we implemented versatile repeated test sequences with far longer spans than minutes or even hours. The long span of 16 months may show that such a period is too long for the use of saccades for user verification. Nevertheless, it is marginal in the sense that computers are typically used daily. After very long a time since the preceding authentication, an authenticated user should probably “train” the computer anew by inputting a new, current eye movement data set of one’s own to maintain the verification capacity of his or her device.

Acknowledgements

We are grateful to Prof. Ilmari Pyykkö of the Department of Otorhinolaryngology, Tampere University Hospital, Finland, for physiological advice on saccades and assistance in recording signals. The work of the first author was supported by Tampere Doctoral programme in Information Science and Engineering (TISE) which is greatly appreciated.

References

Abdullah, M. A. M., Alu-Dulaimi, F. H. A., Al-Nuaimy, W. and Al-Ataby, A. (2011) 'Efficient small template iris recognition system using wavelet transform', *International Journal of Biometrics and Bioinformatics*, Vol. 5, No. 1, pp. 16-27.

Bahill, A. T., Brockenbrough, A. and Troost, B. (1981) 'Variability and development of a normative data base for saccadic eye movements', *Investigative Ophthalmology & Visual Science*, Vol. 21, No. 1, pp. 116-125.

Bollen, E., Bax, J., van Dijk, J. G., Koning, M., Bos J. E., Kramer, C. G. S. and van der Velde, E. A. (1993) 'Variability of the main sequence', *Investigative Ophthalmology & Visual Science*, Vol. 34, No. 13, pp. 3700-3704.

Chang, S., Secker, J. Xiao, Q., Reid, B., Bergeron, A. and Almuhtadi, W. (2011) 'Artifical finger detection by spectrum analysis', *International Journal of Biometrics*, Vol. 3, No. 4, pp. 376-389.

Chen, X., Tian, J. and Yang, X. (2006) 'A new algorithm for distorted fingerprints matching based on normalized fuzzy similarity measure', *IEEE Transactions on Image Processing*, Vol. 15, No. 3, pp. 767-776.

Duchowski, A. T. (2002) 'A breadth-first survey of eye tracking applications', *Behavior Research Methods, Instruments & Computers*, Vol. 34, No. 4, pp. 455-470.

Fransson, P. A., Patel, M., Magnusson, M., Berg, S., Almbladh, P. and Gomez, S. (2008) 'Effects of 24-hour and 36-hour sleep deprivation on smooth pursuit and saccadic eye movements', *Journal of Vestibular Research*, Vol. 18, No. 4, pp. 209-222.

Henriksson, N. G., Pyykkö, I., Schalén, L. and Wennmo, C. (1980) 'Velocity patterns of rapid eye movements', *Acta Otolaryngologica*, Vol. 89, pp. 504-512.

Holland C. and Komogortsev, O. V. (2011) 'Biometric identification via eye movement scanpaths in reading', *International Joint Conference on Biometrics*, IEEE, USA, 8 p.

Holland, C. and Komogortsev, O.V. (2012) 'Biometric verification via complex eye movements: The effects of environment and stimulus', *IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems, pp. 39-46.*

Jain, A. K., Prabhakar, S., Hong, L. and Pankanti, S. (2000) 'Filterbank-based fingerprint matching', *IEEE Transactions on Image Processing*, Vol. 9, No. 5, pp. 846-859.

Juhola, M., Jäntti, V. and Pyykkö, I. (1985) 'Effect of sampling frequencies on computation of the maximum velocity of saccadic eye movements', *Biological Cybernetics*, Vol. 53, pp. 67-72.

Juhola, M. (1988) 'Detection of nystagmus eye movements using a recursive digital filter', *IEEE Transactions on Biomedical Engineering*, Vol. 35, 389-395.

Juhola, M., Aalto, H., Jutila, T. and Hirvonen, T. (2011) 'Signal analysis of three-dimensional nystagmus for otoneurological investigations', *Annals of Biomedical Engineering*, Vol. 39, No. 3, 973-982.

Juhola, M., Zhang, Y. and Rasku J. (2013) 'Biometric verification of a subject through eye movements', *Computers in Biology and Medicine*, Vol. 43, No. 1, pp. 42-50.

Jäntti, V., Lang, A. H., Keskinen, E., Lehtinen, I. and Pakkanen, A. (1983) 'Acute effects of intravenously given alcohol on saccadic eye movements and subjective evaluations of intoxication', *Psychopharmacology*, Vol. 79, No. 2-3, pp. 251-255.

Kapczyński, A., Kasprowski, P. and Kuźniacki, P. (2006) 'Modern access control based on eye movement analysis and keystroke dynamics', *International Multiconference on Computer Science and Information Technology*, Wisla, Poland, pp. 477-483.

Kasprowski, P. and Ober, J. (2004) 'Eye movements in biometrics', *Biometric Authentication: International Workshop*, *Lecture Notes in Computer Science*, Prague, Czech Republic, Springer-Verlag, Vol. 3087, pp. 248-258.

Komogortsev, O. V., Karpov, A., Price, L. R. and Aragon, C. (2012) 'Biometric authentication via oculomotor plant characteristics', *Proceedings of 5th IAPR International Conference on Biometrics*, IEEE, pp. 413-420.

Kukharev, G., Tujaka, A. and Forczmanski, P. (2011) 'Face recognition using two-dimensional CCA and PLS', *International Journal of Biometrics*, Vol. 3, No. 4, pp. 300-321.

Lehtinen, I., Lang, A. H., Jäntti, V. and Keskinen, E. (1979) 'Acute effects of alcohol on saccadic eye movements', *Psychopharmacology*, Vol. 63, No. 1, pp. 17-23.

Lourenco, A., Silva, H. and Fred, A. (2011) 'Unveiling the biometric potential of finger-based ECG signals', *Computational Intelligence and Neurosciences*, 8 p., doi:10.1155/2011/720971.

Mane, V. M. and Jadhav, D. V. (2009) 'Review of multimodal biometrics: applications, challenges and research areas', *International Journal of Biometrics and Bioinformatics*, Vol. 3, No. 5, pp. 90-95.

van der Meyden, C. H., Bartel, P. R., De Sommers, K., Blom. M. and Pretorius, L. C. (1989) 'Effects of clobazam and clonazepam on saccadic eye movements and other parameters of psychomotor performance', *European Journal of Clinical Pharmacology*, Vol. 37, No. 4, pp. 365-369.

Morimoto, C. H. and Mimica, M. M. R. (2005) 'Eye gaze tracking techniques for interactive applications', *Computer Vision and Image Understanding*, Vol. 98, pp. 4-24.

Munoz, D. P., Broughton, J. R., Goldring, J. E. and Armstrong, I. T. (1998) 'Age-related performance of human subjects on saccadic eye movement tasks', *Experimental Brain Research*, Vol. 121, No. 4, pp. 391-400.

Prasad, S. M., Govindan, V. K. And Sathidevi, P. S. (2010) 'An efficient wavelet-based palmprint verification approach', *International Journal of Wavelets, Multiresolution and Information Processing*, Vol. 8, No. 6, 961-985.

Rigas, I., Economou, G. and Fotopoulos, S. (2012) 'Biometric identification based on the eye movements and graph matching techniques', *Pattern Recognition Letters*, Vol. 33, pp. 786-792.

Schalén, L., Pyykkö, I., Juhola, M., Magnusson, M., Jäntti, V. and Henriksson, N. (1984) 'Intra-individual variation in oculomotor performance in man', *Acta Otolaryngology*, Suppl. 446, pp. 212-217.

Shen, J., Bao, S.-D., Yang, L.-C. and Li, Y. (2011) 'The PLR-DTW method for ECG based biometric identification', *33rd Annual International Conference of the IEEE EMBS*, Boston, Massachusetts, USA, pp. 5248-5251.

Smeets, J. B. J. and Hooge, I. T. C. (2003) 'Nature of variability in saccades', *Journal of Neurophysiology*, Vol. 90, pp. 12-20.

Torres, L. (2004) 'Is there any hope for face recognition?', *Proceedings of the 5th International Workshop on Image Analysis for Multimedia Interactive Services*, Lisbon, Portugal.

Wang, Y., Agrafioti, F., Hatzinakos, D. and Plataniotis, K. N. (2008) 'Analysis of human electrocardiogram for biometric recognition', *EURASIP Journal on Advances in Signal Processing*, 11 p., doi:10.1155/2008/148658.

Weirts, R., Janssen, J.A. and Kingma, H. (2008) 'Measuring saccade peak velocity using a low-frequency sampling rate of 50 Hz', *IEEE Transactions on Biomedical Engineering*, Vol. 55, No. 12, pp. 2840-2842.

Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S. and Ma, Y. (2009) 'Robust face recognition via sparse representation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 2, pp. 210-227.

Zhang, Y., Rasku, J. and Juhola, M. (2012) 'Biometric verification of subjects using saccade eye movements', *International Journal of Biometrics*, Vol. 4, No. 4, pp. 317-337.

Zhang, Y. and Juhola M. (2012) 'On biometric verification of a user by means of eye movement data mining', *The Second International Conference on Advances in Information Mining and Management*, Venice, Italy, Xpert Publishing Services, pp. 85-90.

Publication VI

On applying signals of saccade eye movements for biometric verification of a subject

Youming Zhang and Martti Juhola

On applying signals of saccade eye movements for biometric verification of a subject

Youming Zhang and Martti Juhola

{Youming.Zhang, Martti.Juhola}@uta.fi

Abstract. We began to develop signal analysis and classification method to distinguish or verify an authenticated user of a computer from possible other subjects called impostors. So far, we have shown that this is possible even with very low a sampling frequency of 30 Hz. In the present study, we continued our research after applying an eye movement camera system with a higher sampling frequency, 250 Hz. Our results obtained this time were better than those earlier with average classification accuracy of approximately 90%. Our current best average classification accuracies reached 95-96% with multilayer perceptron networks, radial basis function networks and support vector machines.

Keywords: Biometric verification, eye movements, saccade, signal analysis, classification, machine learning, data mining

1 Introduction

Various images from subjects have been typical data sources in biometric verification and identification. Maybe the fingerprints [18,22] and face images [4,14] are researched most frequently and these have also been implemented in some commercial applications. Moreover, iris [12], palm [11] and even ear [19] images are studied for biometric purposes. Finally, these alternatives are combined in order to use multimodal identification and to improve identification results [4]. One-dimensional signals have rather infrequently been studied for biometric purpose. However, at least electrocardiogram (ECG) signals have been researched for identification [16], although applying them is not so easy, because recordings require skin electrodes attached to suitable positions on the body to collect ECG signal of good quality. To distinguish between subjects, high quality of data is always crucial. Otherwise, subtle differences between them may be very difficult to achieve.

In the following we adhere to the definition of biometric verification that we attempt to recognize an authenticated subject from among a set of subjects and, on the other hand, to recognize any one else not to be an authenticated subject, but an outsider called an impostor. Thus, this classification task involves two classes. Biometric identification not researched in the current study consists of probably a more complex classification task of many classes where any subject has to be discerned from a group of *n* subjects resulting in *n*-class classification and this group size *n* may be very large. To be an authenticated user of a computer, a cell phone or some device containing a computer requires verification, which, as binary classification, is simpler than multiclass classification of *n* subjects for identification.

The beginning of the present research was originated from our long-term work on signal analysis and pattern recognition of eye movements for medical purposes, especially balance studies in otoneurology [5,6]. It was salient that recording devices for eye movement signals have been developed remarkably in the past decades [3]. Introduction of eye movement videocamera systems in the 1990s also brought about research ideas to apply these in human-computer interfaces of computers. Even simple cell phones nowadays include a front camera system directed towards a user. In the future such camera systems will obviously be employed to follow a user's gaze, i.e., what he or she looks at the screen of a computer, cell phone or possible other device and this gaze information is utilized by and with computers. Our idea for these camera systems is to develop such a method with which a computer is able to verify its authenticated user and reject logins of impostors when starting it.

Saccades are rapid eye movements that are performed while moving one's gaze from a target to another. Perhaps, most of our eye movements are saccades. Other types of eye movements are of more complex forms, for instance, smooth pursuit tracking movements that are made when a moving target is followed by the gaze and the velocity of this target is not very high, say, less than 50º/s, because a clearly faster object cannot be tracked with smooth pursuit, but with much faster saccades. Nystagmus is a reflexive or involuntary movement which is repeatedly made and needs a suitably moving stimulation arrangement. Sitting in a train and following rapidly changing and sufficiently near views through a train window causes a spectator to make optokinetic nystagmus. The curve profile of a saccade in Fig. 1 is simpler than those of smooth pursuit or nystagmus eye movement. Saccades are easier to stimulate than other eye movements.

For the biometric purpose studies on eye movement signals were begun a few years ago. The cepstrum of an eye movement signal as calculated and its results were classified with naïve Bayes technique, nearest neighbour searching, decision trees and support vector machines [8,9]. A computational model for the oculomotor system was implemented, where parameter values of the model were classified [10]. Face images were used for stimulation, where fixation points of a subject's gaze on an image stimulation were detected in the plane of horizontal and vertical axes [15]. Minimum spanning trees were formed between fixation points and were applied to biometric identification by classifying subjects on the basis of nearest neighbour searching and support vector machines.

Originally, we began our biometric verification investigations on the basis of saccades by taking advantage of an eye movement camera system with a low sampling frequency (30 Hz) and, secondly, by utilizing our earlier eye movement signals measured with skin electrodes of an electro-oculographic (EOG) system [7]. The advantage of the latter was its high sampling frequency (400 Hz), but the character of EOG signals was typically very noisy compared with signals of modern camera systems.

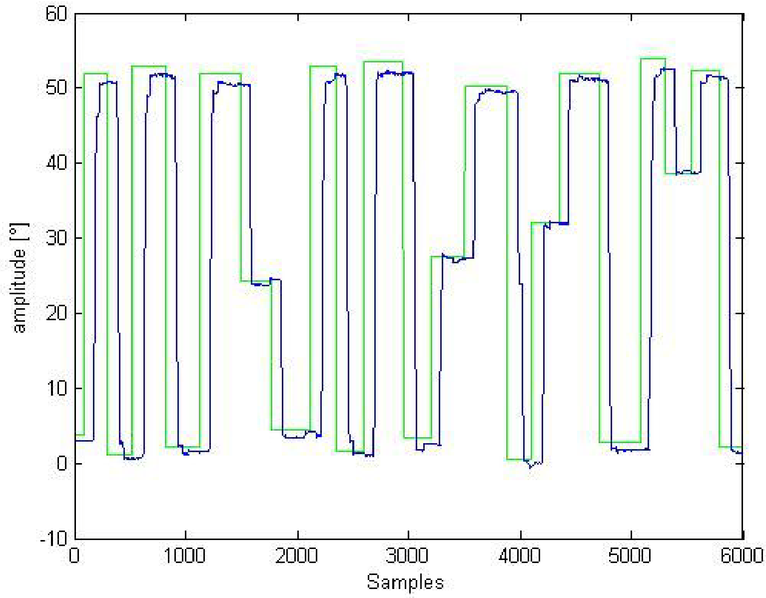


Fig. 1. An example of a (blue) saccade signal and with its (green) smooth stimulation signal in which stimulation angles preceded their responses, saccades. The signal segment was 24 s long.

Recently we tested various classification methods and ways to build training and test data sets collecting signals from more frequent numbers of subjects [20,21] than in [7]. It was naturally important to increase the total of subjects measured with the same way all the time in order to make classification tests more versatile and more convincing by including more people with their more or less different saccades. These preliminary investigations showed that saccades are sufficiently dissimilar between subjects, but also sufficiently similar within individuals to make a verification task possible with them. These are called interindividual and intraindividual variabilities.

Research results and viewpoints about similarity or variability of saccades within and between individuals seem to vary between researches accomplished infrequently in medical sciences, physiology and psychology. Although natural intraindividual variability is typically found [2,17], when saccades with the same stimulation angles were repeatedly measured from subjects, this variability was not, after all, statistically significantly different between a group of 58 subjects even after two weeks time. Further, it was found that the main source of variability of saccades was different recording devices used [17]. Earlier, we noticed the same for maximum velocities of saccades, an important variable, when measured eye movements at the same time with two different devices [5].

In the present research, we continued our biometric verification research for theme how an improvement of a sampling frequency for eye movement signals may affect results. Previously, we applied the eye movement camera system of 30 Hz [7,20,21]. This was interesting, because using so low a sampling frequency we, in fact, showed that in principle our verification procedure was possible to implement even at almost the lowest sampling frequency applied that is frequently utilized in cheap and simple web cameras. Naturally, higher sampling frequencies are also interesting, because we then have always better possibilities to collect more accurate signal information from eye movements. Better results could be expected because it is self-evident that better properties of camera system can reveal features of saccades more accurately than those more moderate devices. Instead, a question aroused whether any subtle differences between saccades of subjects possibly revealed by better recording devices do aid to improve verification results that are based on classifications made with variable values computed from saccades. By catering for our prior results obtained from EOG recordings of 400 Hz [7], we might expect improvements since the results given by EOG signals, although noisy, were somewhat better than those of 30 Hz with the videocamera system [7]. For the latter, suitable interpolation methods were deployed to raise the (artificial) sampling frequency up to 400 Hz or 1000 Hz [7,20,21].

2 Eye movement signals recorded and variables computed

Eye movement recordings were conducted with a videocamera system named EyeLink (SMI, Berlin, Germany). Its sampling frequency is 250 Hz and enables the maximum angle of ±30º in the horizontal direction. All eye movement cameras are based on pupilometry, measuring, as accurately as possible, pupil locations in successive images of a videostream and computing their differences. Accuracy of this device is 0.1º for pupil locations. A subject wore the system attached to the headband so that cameras, one for each eye, were in front of him or her and slanting downwards in the visual field not covering view on the screen of computer (Fig. 2). The distance of a subject's eyes to the screen of computer was constant 45 cm, which was kept for all subjects seated in a chair set in a fixed location. Before every recording the device required calibration as most eye movement recording systems. All recordings were performed painstakingly and precisely in the same way.

For actual recordings, a computer-controlled stimulation sequence of small jumping light dot was abruptly moved from a location to another in the horizontal direction in the middle of the screen. At the same time the computer recorded an eye movement signal of a subject looking at the jumping dot. This is a typical way for medical eye movement tests. Every recording took 60 s and included stimulation movement of intervals from 0.8 s to 1.5 s with an average of approximately 1 s. A half of stimulation angles were less than the maximum angle of 51º used for the verification task. Those smaller were used nothing but to make stimulation sequence more random-like for a subject. Since both stimulation angles and intervals between these varied in a way seemingly random, the sequence of around 60 stimulation movements could be learnt by no one even if it had been repeated several times. Thirty saccades as responses to the largest stimulation angle of 51º were used for the verification task. When values of saccade variables may alter slightly between measurements for identical stimulation angles, four successive recordings were executed for every one.

Thus, 120 large amplitude saccades were recorded from each subject. The number of subjects was 55. There were 41 males and 14 females with the average age of 26±6 years. The recordings were accomplished in a laboratory room of the cellar level equipped with shields of high quality against electric distortion, noise and tremble.

Fig. 2. A subject performing calibration in which eye movement cameras are located so that each of them precisely catches the pupil in an image seen on the screen.

The stimulation sequence was constant for every recording and every subject, because the main idea was to investigate verification tests, i.e., we had to have the exactly same stimulation for every one to enable both intraindividual and interindividual computation and waiting for such situation that a subject's saccades were more similar with each other than with those of other subjects on the average. It was also useful to collect fairly many saccades of the same large stimulation angle so that any subject was well represented in the data set in statistical sense. In our earlier studies [7,20,21] we had smaller numbers of saccades per subject because of different recording devices and arrangements. Nevertheless, the larger training subset from every subject we govern, the better the starting point for classification.

The type of stimulation sequence was as simple as possible to guarantee that it was easy to follow and did not cause fatigue though repeated consecutively four times for a subject. Further, as simple stimulation as possible is reasonable approach since if it were more complicated, we might need more repetitions to collect enough data to distinguish between subjects. A complicated stimulation set-up might produce failed tests, but our simple stimulation yielded none.

3 Data handling

At first, eye movement signals were filtered with a median filter of length 3. Although the camera system contained two cameras, one for each eye, only one of them was utilized, signal of which contained less noise. Normally, quality of these signals was very high. It is essential to detect beginnings and ends of saccades as exactly as possible. These underlie values of saccade variables to be calculated for our verification task. See their breakdown delineated in Fig. 3.

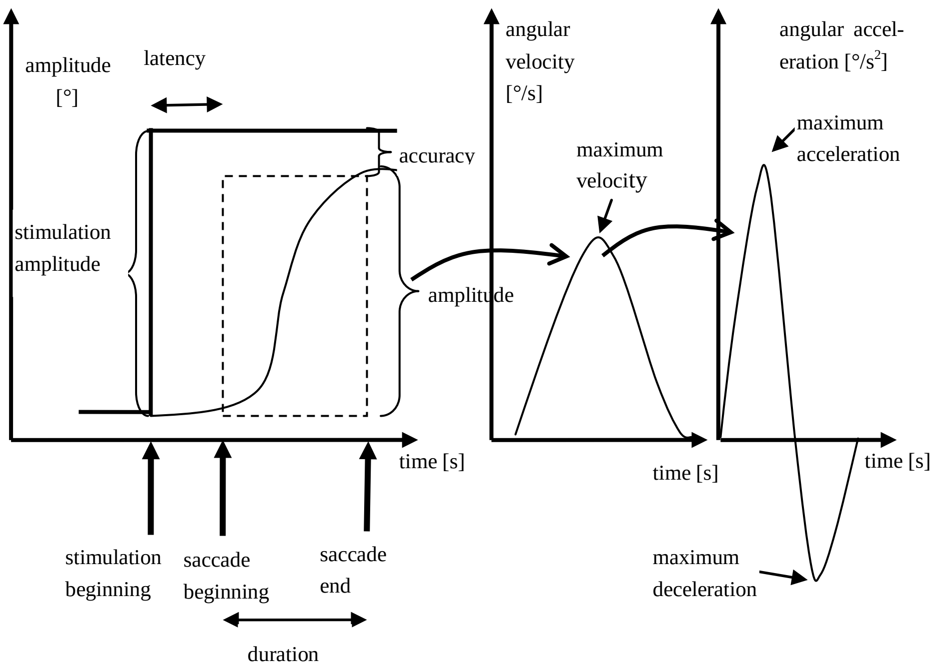


Fig. 3. A schema for a stimulation angle and its response, saccade, on the left achieving six variables: amplitude, accuracy, latency, maximum angular velocity, maximum angular acceleration and maximum angular deceleration of the saccade. The saccade curve upwards in the figure corresponds to a horizontal eye movement from the left to right. Other possible variables such as duration here were not used for computation.

At first, the first derivative is approximated by computing angular velocity values of an eye movement signal. Beginning and ends of all saccades are detected by means of velocity values along with the presentation in Fig. 3. When magnitudes of velocity values increase above the threshold of 50º/s, its beginning was registered. Correspondingly, when magnitudes of velocity values decreased below 30º/s, its end was found. The latter was smaller than the former since saccades are typically steeper at the beginning and somewhat more gently sloping at the end which can be observed from Fig. 1.

The stimulation signals were also analysed, and beginnings and ends of stimulation angles were easy to detect being as steep steps according to Figs. 1 and 3. Such saccades would be valid for the further computation being later than approximately 0.1 s after their stimulations. Any saccade prior to that lower bound was seen as anticipation, because one's brain requires some time to observe a movement in the visual field and to control to move the gaze near a new location of stimulation light dot.

The 4 recordings of every subject contained 4 times 30 large amplitude saccades. After detecting their beginnings and ends precisely, their variable values were computed as outlined in Fig. 3. Latency is the difference in time between the beginning of a saccade and its stimulation beginning. Using the amplitudes of a stimulation angle and its saccade, accuracy is computed. The maximum velocity of a saccade is given by its velocity curve. Ultimately, the maximum acceleration and deceleration of a saccade are searched for from the approximated second derivative.

To predict separation capability of six variables described, we computed ratios of interindividual and intraindividual variabilities by applying the formula given below.

$$R_j = \frac{\sqrt{\dfrac{1}{\sum_{i=1}^{n} n_i} \sum_{i=1}^{n} \sum_{k=1}^{n_i} (u_{kj} - \bar{a}_j)^2}}{\dfrac{1}{n} \sum_{i=1}^{n} \sqrt{\dfrac{1}{n_i} \sum_{k=1}^{n_i} (u_{kj} - \bar{u}_{ij})^2}}$$

Subscript j equals the jth variable, j=1,…,6, n the number of subjects, $\bar{u}_{ij}$ the mean of variable j of subject i, $\bar{a}_j$ the average of variable j for the saccades of all subjects, u_{kj} the value of variable j of saccade k for subject i and n_i the number of the saccades of subject i. We expected that the higher the ratio, the better separating variable. Results are presented in Table 1 indicating reasonable opportunity for separation of subjects.

Table 1. Means and standard deviations of variables for the data of 55 subjects and ratios between interindividual and intraindividual variabilities.

Amplitude [°]	Accuracy [°]	Latency [s]	Maximum velocity [°/s]	Maximum acceleration [°/s^2]	Maximum deceleration [°/s^2]
Mean and standard deviation					
47±11	1±9	0.25±0.05	631±184	33991±22830	17587±19350
Ratio R_j of interindividual and intraindividual variabilities					
1.68	1.43	1.11	2.09	1.91	1.96

4 Verification procedure applied

There are two opposite objectives in verification of subjects: to detect an authenticated subject as accurately as possible and to observe possible impostors as effectively as possible. We call these as Conditions 1 and 2, respectively, that are tested with the procedure by employing the eye movement recordings described above.

```
% Verification: counters for correct classifications of
% authenticated users and those of impostors
% (C1 for Condition 1 and C2 for Condition 2)
C1=C2=0;
For h=1:r  % iterations of the main loop
  For i=1:n % one by one as an authenticated user
    For j=1:a % a measurements
      Take m saccades, apart from the jth, from b=a-1
      measurements of an authenticated user to a training
      set;
      Copy m saccades to have 2m in the training set;
      Take randomly 2m saccades from p non-users and add
      these saccades to a training set;
      Train a model with 4m saccades of two classes: an
      authenticated user and non-users;
      TP=TN=FP=FN=0;
      For k=1:c % tests for Condition 1
        Classify a test saccade of an authenticated user
        into either correct class
        TP=TP+1
        or incorrect class
        FN=FN+1;
      End
      For l=1:d % tests for Condition 2
        Classify a test saccade of an impostor into
        either correct class
        TN=TN+1
        or incorrect class
        FP=FP+1;
      End
    % Compute majority vote for decisions
    If TP≥FN then C1=C1+1;
    If TN>FP then C2=C2+1;
    End
  End
End
Mean accuracy of authenticated users=100 %   C1/(r·n·a)
Mean accuracy of impostors=100 %·  C2/(r·n·a)
```

The procedure is on the basis of typical machine learning tests how classification problems are considered by dividing a data set into training and test sets to be varied largely in order to achieve statistically relevant results via numerous training and test set pairs sampled from the available data. Arrangements for classification are described exactly in the following section alongside with the verification procedure to be given as follows.

The verification procedure given in pseudocode above was executed with the data set of 55 subjects. It deals with both Conditions 1 and 2. Its main loop executes r iterations for the whole process. Because of random selections within the verification procedure, several repetitions are required from which means and standard deviations are computed as final results. Three disjoint subsets of subjects are needed. For Condition 1 there is a single authenticated subject at a time and some other subjects taken randomly from the whole set. The latter are called non-users, who are thought not to be users of a computer or other device and whose saccades are used to represent the part in the variable space outside an authenticated subject. For Condition 2, a third subset is still required, impostors apart from non-users, in order to test verification. Both conditions are needed because we may assume that an occasional subject attempting to log in a computer can be either an authenticated subject or an impostor as well.

5 Test arrangements

In Condition 1 of the verification procedure, one by one each subject was the authenticated subject. From other n-1=54 subjects p=30 random subjects formed the subset of non-users. The rest of 24 subjects were the subset of impostors. Every subject possessed 30 large amplitude saccades from 4 successive measurements, 120 in total. Thus, m=90 saccades of 3 measurements of an authenticated subject (doubled to be 2m=180 saccades) were set to a training set, but c=20 randomly chosen saccades from those 30 of one measurement were set to the corresponding test set. Further, 2m=180 saccades were randomly chosen from p=30 non-users to the training set. For Condition 2, we always employed the same construction, but test cases were from the subset of 24 impostors. Here, there were d=20 saccades taken from the subset of impostors. According to the verification procedure, $r \cdot n \cdot a$=10·55·4=2200 different models were built. With these 20 saccades were tested for both Condition 1 and Condition 2 entailing 44000 tests for each.

We ran tests by means of multilayer perceptron networks with several training algorithms, radial basis function networks, logistic discriminant analysis and support vector machines. We chose these methods on the basis of our previous [7,20,21], where these typically gave the best results for our earlier data sets measured with two different devices, not the same as now. Subject to multilayer perceptron networks, we then used Levenberg-Marquardt training algorithm only. This time, we wanted to liken their several, various training algorithms as well. For these networks, 6 input nodes as saccade variables, 1 output node and 6, 8, 10 or 12 hidden nodes of 1 layer were exploited.

According to Table 1, the scales of the different saccade variables vary considerably. Therefore, normalization into [0,1] of their values could be needed before inputting data values to classification programs. Nonetheless, on the basis of our tests, we observed that radial basis function networks were the only that required them in order to produce reasonable results. Meanwhile, for results of the other methods its use was meaningless and was not used as to results to be presented in the following section.

6 Test results

The section describes test results gained. All computations were implemented in Matlab R2010a (MathWorks Inc, USA). Results are presented as classification accuracies in percent, i.e., how many tests gave a correct decision between an authenticated subject and non-users. For Condition 1, the former alternative was right and for impostors of Condition 2 the latter. False rejection and false acceptance rates are often exerted. They can be directly derived from accuracy values: False rejection or false negative rate is attained by subtracting accuracy of Condition 1 from 100%. Correspondingly, false acceptance or false positive rate can be computed by subtracting accuracy of Condition 2 from 100%.

The results in Table 1 should not be compared directly with the corresponding results of other researches, e.g., those of our earlier [20], since the recording devices were different. Different recording devices and types of recording devices seem to affect somewhat [1,5,17]. It has to be understood that virtually all variable values of saccades depend on their amplitudes, for example, the greater amplitude, the greater the maximum velocity. Amplitudes of saccades, of course, depend on stimulation angles. We used large stimulation angles because it is possible to obtain more variability between subjects between large, say, over 30º than small such as 5º-30º. Nevertheless, the crucial result here was that there was clear interindividual variability when the same stimulation angle, test set-up and recording device were used for all the subjects.

First, we computed all results by using the original data sampled at the frequency of 250 Hz. Second, we computed the same tests with the same data that was interpolated up to 1000 Hz before any signal analysis or other computation at the beginning. We included the latter, interpolation alternative in our tests, because we found it necessary in our previous studies on biometric verification with saccades [7,20,21] in which we applied an eye movement camera system with the low sampling frequency of 30 Hz whereas now it was much higher, 250 Hz.

Results of eight training algorithms of multilayer perceptron networks are given in Table 2. To reduce excessive numbers of result tables, we present merely the best choice for each network in regard with the number of hidden nodes. In addition, two other training algorithms were also run, Batch Gradient Descent (gd) and Variable Learning Rate Backpropagation (gdx), but since their results were around 10% inferior to those of the other, they were left out. Note that to search for the best result in the following tables, we have to look at pairs of Conditions 1 and 2, because both of them are due to be as high as possible. Levenberg-Marquardt training algorithm gave decidedly the best outcomes, particularly for Condition 2.

Table 2. Means and standard deviations of classification accuracies for multilayer percetron networks in percent: Resilient Backpropagation (rp), Fletcher-Powell Conjugate Gradient (cgf), Polak-Ribiére Conjugate Gradient (cgp), Conjugate Gradient with Powell/Beale restarts (cgb), Scale Conjugate Gradient (scg), BFGS Quasi-Newton (bfg), One Step Secant (oss) and Levenberg-Marquardt (lv). The best pair is marked in bold face.

Training algorithm and hidden nodes		Condition 1 and 250 Hz	Condition 2 and 250 Hz	Condition 1 and 1000 Hz	Condition 2 and 1000 Hz
rp	8	91±2	87±2	86±2	81±2
cgf	12	91±1	86±3	87±2	80±2
cgp	10	91±1	85±2	87±2	78±4
cgb	6	89±2	87±2	86±2	79±2
scg	10	89±2	85±3	85±1	79±2
bfg	12	91±2	87±3	87±2	80±3
oss	8	89±2	84±2	83±3	79±3
lm	6	**95±1**	**95±2**	91±1	92±2

Table 3. Means and standard deviations of classification accuracies for radial basis function networks in percent. The best pair is marked in bold face.

Parameters spread and goal		Condition 1 and 250 Hz	Condition 2 and 250 Hz	Condition 1 and 1000 Hz	Condition 2 and 1000 Hz
20	0.08	97±1	93±1	95±1	87±2
20	0.1	**96±1**	**95±2**	96±1	91±2
20	0.12	96±1	94±2	96±1	93±2
15	0.08	96±1	94±2	95±1	90±2
15	0.1	**96±1**	**95±2**	95±1	91±1
15	0.12	**96±1**	**95±2**	95±1	92±2

Next, we conducted tests of radial basis function networks with different values of parameters goal and spread. These results are presented in Table 3. The best results were obtained with three pairs of the parameter values, but differences between others were small, a few percent.

Next, we ran tests with logistic discriminant analysis that succeeded well, usually better than linear or quadratic discriminant analysis in our previous studies [7,20,21]. Finally, we experimented with support vector machines using various kernels. Results are shown in Table 4.

Table 4. Means and standard deviations of classification accuracies for logistic discriminant analysis and support vector machines (SVM) in percent. Kernel functions of SVM are linear, quadratic, polynomial of degree 3 and radial basis function (RBF). The best pairs are marked in bold face.

	Condition 1 and 250 Hz	Condition 2 and 250 Hz	Condition 1 and 1000 Hz	Condition 2 and 1000 Hz
logistic discriminant analysis	91±1	89±2	91±1	88±2
kernel				
linear	90±1	86±2	90±1	85±3
quadratic	97±1	91±3	96±1	90±2
polynomial, degree 3	**98±1**	**93±1**	97±1	91±1
RBF, σ=2.5	**96±1**	**95±1**	96±1	93±1

All the preceding results in Tables 2-4 represented a situation where classification was based on majority votes, i.e., with a threshold values of 0.5 (equal error ratio). In order to administer results still more extensively, ROC curves are portrayed in Fig. 4 for results of support vector machines that were among the best results in Tables 2-4. It indicates how the kernels of radial basis function (with parameter value σ=2.5) and polynomial of degree 3 manifested the best performance.

7 Conclusion and discussion

The results of the preceding section indicated multilayer perceptron networks implemented with Levenberg-Marquardt training algorithm, radial basis function networks and support vector machines with the polynomial kernel of degree 3 or RBF kernel to be the most efficient methods to solve the main task of our verification procedure, classification between saccades of an authenticated subject and those of others.

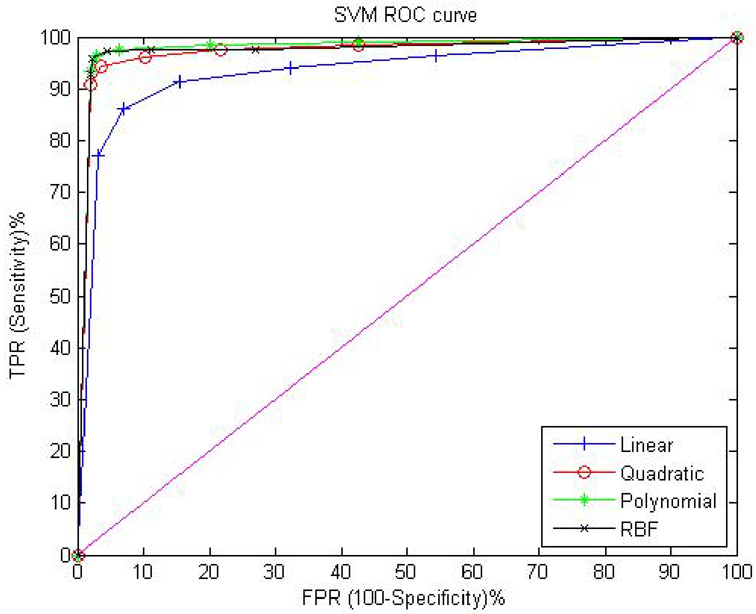


Fig. 4. ROC curves presented with true positive rates (TPR) and false positive rates (FPR) in percent for support vector machines with respect to linear, quadratic, polynomial of degree 3 and radial basis function (RBF) kernel with σ=2.5.

High accuracies of 93-98% were obtained in the present tests, in other words, 2-7% error rates. When accuracies were 90% at their best in our previous studies [7,20,21] and the original sampling frequency was only 30 Hz even if interpolated to be much higher, there is a clear betterment at hand. The obvious reason is the higher sampling frequency. This was now 250 Hz instead of the earlier 30 Hz that was interpolated up to approximately 400 Hz or 1000 Hz. Nevertheless, interpolation of the present data from 250 Hz to 1000 Hz did even impair results somewhat, 3-8% for multilayer perceptron networks, 0-6% radial basis function networks, 0-1% for logistic discriminant analysis and 0-2% for support vector machines. This empowers us to conclude that there is no reason to interpolate to acquire a higher, artificial sampling frequency for the data of the present recording system since the original sampling frequency suffices. Apparently, 250 Hz is high enough. The conclusion is natural while recalling that information content of large angle saccades is known to reach 50-100 Hz [1,5]. Nyquist frequency, a half of sampling frequency, ought to be enough here, in theory. In practice, however, it has to be taken into account a higher bound and, thus, 250 Hz is enough for verification task.

Another reason for the present results may be that, this time, our recording system required calibration for every recording session and subject, while using the other eye movement videocamera system earlier this applied precalibration made only once.

Nonetheless, according to our experience with several eye movement recording devices of different types, we deem this less important issue than the great difference of the sampling frequencies of 30 Hz and 250 Hz.

Inasmuch as the results obtained by other researchers who have studied biometric verification or identification are originated from very different tests and methods [8-10,15], we are not able to directly compare to their results. However, looking at sheer classification accuracies or error rates, ours are at least as good as theirs and partly better.

In the future we are going to explore the use of saccade eye movement signals due to be measured in different times of days, days and perhaps even weeks. It is arguable how much and how in general variable values of saccades vary in the course of time. On the one hand, their values are seen to have intraindividual variability during days or weeks, but on the other hand it was seen [2] that, e.g., maximum velocities did not differ statistically significantly after two weeks.

To draw the main conclusion, it is a sensible and good idea to verify a subject on the basis of saccade eye movements provided that an eye movement camera system is included in the device to be used. It does not require two cameras as were used here; only one eye can be measured.

Acknowledgements. The authors thank prof. Kari-Jouko Räihä with his research group, especially researchers Oleg Spakov and Päivi Majaranta, for their aid in acquiring eye movement signals. The first author is grateful to Tampere Doctoral Programme in Information Science and Engineering (TISE) for support.

References

1. Bahill, A.T., Brockenbrough, A., Troost, B.: Variability and development of a normative data base for saccadic eye movements. Investigative Ophthalmology & Visual Science, 21(1), 116-125 (1981).
2. Bollen, E., Bax, J., van Dijk, J.G., Koning, M., Bos J.E., Kramer, C.G.S., van der Velde, E.A.: Variability of the main sequence. Investigative Ophthalmology & Visual Science, 34(13), 3700-3704 (1993).
3. Duchowski, A.T.: A breadth-first survey of eye tracking applications. Behavior Research Methods, Instruments & Computers, 34(4), 455-470 (2002).
4. Ghobakhlou, A., Zhang, D., Kasabov, N. An evolving neural network model for person verification combining speech and image. In: N.R. Pal et al. (eds.) ICONIP 2004, LNCS 3316, pp. 381-386. Springer, Heidelberg, (2004).
5. Juhola, M., Jäntti, V., Pyykkö, I.: Effect of sampling frequencies on computation of the maximum velocity of saccadic eye movements. Biological Cybernetics, 53, 67-72 (1985).
6. Juhola, M., Aalto, H., Hirvonen, T.: Using results of eye movement signal analysis in the neural network recognition of otoneurological patients. Computer Methods and Programs in Biomedicine, 86, 216-226 (2007).
7. Juhola, M., Zhang, Y., Rasku J.: Biometric verification of a subject through eye movements. Computers in Biology and Medicine, 43(1), 42-50 (2013).

8. Kapczyński, A., Kasprowski, P., Kuźniacki, P.: Modern access control based on eye movement analysis and keystroke dynamics. In: International Multiconference on Computer Science and Information Technology, Wisla, Poland, 477-483 (2006).
9. Kasprowski, P., Ober, J.: Eye movements in biometrics. In: Biometric Authentication, International Workshop, Lecture Notes in Computer Science, Prague, Czech Republic. LNCS vol. 3087, pp. 248-258. Springer, Heidelberg, (2004).
10. Komogortsev, O. V., Karpov, A., Price, L. R., Aragon, C.: Biometric authentication via oculomotor plant characteristics. In: Proceedings of 5th IAPR International Conference on Biometrics, IEEE, pp. 413-420 (2012).
11. Lin, C.-L., Fan, K.-C. Biometric verification using thermal images of palm-dorsa vein patterns. IEEE Transactions on Circuits and Systems for Video Technology, 14(2), 199-213 (2004).
12. Ling, L.L., de Britto, D.F. Fast and efficient iris image segmentation. Journal of Medical and Biological Engineering, 30(6), 381-391 (2010).
13. Liu, L., Dai, T.-S.: A reliable fingerprint orientation estimation algorithm. Journal of Information Science and Engineering, 27, 353-368 (2011).
14. Oh, B.-S., Toh, K.-A., Teoh, A.B.J., Kim J. Combining local face image features for identity verification. Neurocomputing, 74, 2452-2463 (2011).
15. Rigas, I., Economou, G., Fotopoulos, S.: Biometric identification based on the eye movements and graph matching techniques. Pattern Recognition Letters, 33, 786-792 (2012).
16. Shen, J., Bao, S.-D., Yang, L.-C., Li, Y.: The PLR-DTW method for ECG based biometric identification. In: 33rd Annual International Conference of the IEEE EMBS, Boston, Massachusetts, USA, pp. 5248-5251 (2011).
17. Smeets, J.B.J., Hooge, I.T.C.: Nature of variability in saccades. Journal of Neurophysiology, 90, 12-20 (2003).
18. Tan, X., Bhanu, B. Fingerprint matching by genetic algorithms. Pattern Recognition, 39, 465-477 (2006).
19. Tavassoli, S., Yaqubi, M. An efficient preprocessing technique for noise reduction in ear verification system, International Journal of Computer Applications, 28(1), 34-40 (2011).
20. Zhang, Y., Rasku, J., Juhola, M.: Biometric verification of subjects using saccade eye movements. International Journal of Biometrics, 4(4), 317-337 (2012).
21. Zhang, Y., Juhola M.: On biometric verification of a user by means of eye movement data mining. In: The Second International Conference on Advances in Information Mining and Management, Venice, Italy, Xpert Publishing Services, 85-90 (2012).
22. Zhou, J., Chen, F., Gu, J. A novel algorithm for detecting singular points from fingerprint images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(7), 1239-1250 (2009).