

Remote Biometric Verification for eLearning Applications: Where We Are

Pietro S. Sanna^(✉) and Gian Luca Marcialis

Department of Electrical and Electronic Engineering,
University of Cagliari, Cagliari, Italy
{pietro.sanna,marcialis}@diee.unica.it
<http://pralab.diee.unica.it/en/Biometrics>

Abstract. In recent years there has been a huge spread of online learning methods - eLearning - that have allowed learners of any type to enjoy well-structured lessons and articulated training course, with solid teaching methodologies. Many solutions and platforms have extremely useful and powerful features for the student; the innovative methodologies have brought unexpected results to all levels, multiple contexts, and areas of interest. In the e-learning there is an ongoing revolution that determines a strong impact on the training system on different fields: methodological, educational, social and technological. However, the identity of an individual - in eLearning - has always been the basis of some controversies. In this regard, a critical aspect concerns the User Identity verification and recognition. Currently there is an element of attention, the problem of examination is still to be done in the presence of the teacher in order to be permanently accredited. Presently are already used various user authentication and identification techniques, such as conventional User ID and password, challenging questions, and some biometric techniques. From this point of view, biometric technologies that have the feature of verifying their personal identity without being “stolen” or “lend” can provide a possible solution. Despite this fact, a few approaches have been proposed for eLearning applications. The purpose of this paper is to provide a summary of the state of the art of biometric methods applied to current eLearning solutions.

Keywords: eLearning · Biometrics · Pattern classification · Authentication · User identification · Learning Management System · MOOC · Gamification · Keystroke dynamics · Stylometry

1 Introduction

The development and dissemination of instruments and didactic methodologies in eLearning area has had a strong acceleration in recent years. Not only with regard to the technological aspects in the development of the systems and also in content management, but even more in the optimization of communication and online teaching models, in deepening cognitive engineering, in delivery typologies and in the defined standards, etc.

eLearning is in fact a process that combines content, technology and cognitive aspects, closely related to the teaching methodology.

The global industry of training and development is complex and changing, and the eLearning approach currently involves all educational scopes, from primary and secondary school to academic education - in turmoil on all eLearning issues - to long life learning aspects up to the professional and business upgrade.

Learners include broader age ranges, cultural and social. They are involved in the use of systems and eLearning contents increasingly rapid and far removed from the new emerging trend. Currently the interaction with the learner relies on the paradigm of speed and mobility, consequently, on the scarcity of the available time. In addition, digital technologies for teaching are undergoing profound changes, including social training, mobile learning micro-learning, the recent development of *MOOC* in academia and business, and more. Consider the very recent introduction of *gamification* techniques now spread, to new methods of Flipped Classroom or Digitized Classroom [1].

The lack of organization strategy in the definition of an online learning program, of expert human resources in monitoring the users participation and satisfaction, as well as the problems to deal with in the development and implementation of eLearning projects, among others, are still elements of discussion and resistance to change management.

In particular, the identity of an individual - in eLearning courses - has always been the basis of some controversies by those who are critical for distance education. A critical aspect on which we carried out an initial analysis concerns the User Identity Verification in eLearning, with particular reference to the arrangements for evaluating the learner on an online training course: using biometrics for identification and recognition.

The ability to authenticate and identify the learner throughout the whole phase of the examinations and forms of assessment in general is still matter of discussions. Moreover, all modes leading to invalidate the examination test, such as the use of unforeseen media (books, tablet, mobile phone, other web browser, etc.) or the interaction of people unrelated to the test, in the same environment or remotely, are among the basic issues when allowing a remote examination to the students.

Adopting biometric techniques to this aim, could be a possible solution beside conventional techniques such as *User ID* and *Password*. Some biometric traits have been proposed, such as voice recognition, video and synchronous facial monitoring, keystroke dynamics, and others.

In the following Sections, we synthesize the salient characteristics of eLearning environments, the related state of the art, and the proposed biometric traits for the user identification for the distance evaluation (test and examination) of students. We also take into account the limitations of a single trait, which leads to the use of multiple biometrics. The paper ends with some discussions about pros and cons of biometric authentication in eLearning applications, even by considering the current technological point of view.

2 eLearning Technologies

Observing the landscape of online education, in the broader sense (institutions, organizations, public and private market) it is possible to notice that there is large number of organizations that operate in this segment: large corporations offer online training services (e.g. *Coursera*, *Lynda*, *EDX*, *Codecademy*, *Litmos*, etc.) and institutions such as universities and research centers.

Technologies which allow training have gradually evolved, becoming increasingly complex. Mobile technologies, such as cloud and social tools have made possible new approaches to formative experiences.

Current technologies and instruments have to be “tuned” with learners’ habits, minimizing the friction of the training process. Users know quite well popular software on the market, for which, for example, the use of an eLearning software should require any training. Moreover, users learn online on a daily basis (e.g., how to prepare recipes, assembling furniture or learn languages), using standard and widely disseminated educational conventions. The role of gamification, augmented and virtual reality - that increases engagement - and eLearning motivation is unquestionably essential in this process [2].

An overly complex technology represents a serious obstacle for training and education; technology should be really exciting for users to learn better and faster. The history of distance learning starts in the early 1900s. One of the first forms of distance learning took place through correspondence courses with teaching materials, exercises, papers sent to students by postal services. Then, learners provided their feedback in the form of filled questionnaires and documents to the teachers for the examination.

With the era of technology and internet models, paradigms and contents, the distance learning approach changed significantly [3] *Computer-Based Training (CBT)* led to the evolution of learning management with audio/video modalities and then with CD-ROM technology. Afterwards, in 1990s, the *Web-Based Training (WBT)* had a wide diffusion with the Internet advent until years 2000, in which the organizations began to combine Instructor Lead Training (ILT) with WBT by defining the *Blended and Information Training*. Today, discussions about *Collaborative, Social and Talent-driven Learning* occur, as Innovative and new learning design to address knowledge gaps.

Clearly, Internet is a key technology element for optimal results, as well as the offline access. As already said, mobility allows a training process without solution of continuity, always and everywhere and in different conditions and with any device. Today’s users live in symbiosis with their devices.

To date another key element is the micro-learning, with short and targeted contents, highly effective in blocking the learner’s attention. This is a support and reinforcement of the traditional training.

In recent years, the number of eLearning systems strongly increased, with a number of new features and a variety of options to choose from. In particular, Learning Management System - LMS and Content management are evolving, by exploiting the standardization of contents (eg. SCORM standard and its

evolutions), the support of multiple sources (eg. YouTube, TED, Slideshare, MOOCs, etc.) and the interaction capabilities of students.

On overall, the eLearning process combines content, technology and cognitive aspects, the latter closely related to teaching methodology.

As mentioned in the report “eLearning: La rivoluzione in corso e l’impatto sul sistema della formazione in Italia”, Aspen, 2014, about technology, we can consider three main areas: (A) Learning Management System (LMS); (B) proprietary software; (C) applications for mobile systems (“apps”) [1]. The area (B) relates to some great universities and research centres, which realize its own platform for teaching.

Generally they do not use web applications, but specific software shared among the different actors involved. The area (C) is the most recent and promising. The application market for mobile systems (tablets and smart phones) is expanding significantly, because apps are flexible tools, cheap, and do not require significant network performance. IT companies are pushing much the realization of mobile applications for education, such as educational games [1].

Currently, the area (A) is the most outstanding. Numerous commercial and open source solutions are available, each one having its strengths and weaknesses, methodologically different and technologically good. They exhibit very interesting features in content management, in specific functionalities, in teaching material management, exams, profiling of different user categories, and more.

All these functions are made possible through LMS that is mainly aimed at the management of learners and learning activities. A standard LMS is characterised by a minimal set of feature such as learner tools (communication, productivity, student involvement), support tools (administration, course delivery, curriculum design), and technical tools (hardware/software, licensing and pricing) [4].

Many evaluation tools have been developed to carry out the comparison of eLearning solutions and LMS, often dividing LMS in the groups of commercial and open Source, considering features as technical and support tools, adaptability, usability and international widespread, innovative functions and paradigms.

Furthermore, in the case of an LMS equipped with tools for content creation, as in the majority of cases, it is called *Learning Content Management System (LCMS)*, LMS belong to the broader category of Content Management System (CMS). According to the philosophy of education and learning Instructional Designers define, the product will assume the prerogative that will deviate slightly from the classic definition of LMS [7–11].

Some of the most popular LMS/LCMS platforms or with characteristics it is important to mention: *Moodle* considered being the best overall even if with some limitations. It is preferred in academia and PMI. *Claroline* available in over 30 languages worldwide, *BlackBoard*, which is a commercial VLE. And again, *Ilias*, *ATutor*, *Sakai*, with different adaptability features [4, 5].

It may be noted that in 2012 there were more than 500 producers of eLearning Platforms (Bersin & Associates). The market is fragmented, in fact only five producers hold a market share of over 4% and *Moodle* hold a market share of 30% in the Public sector and Education.

Recently, the so-called *MOOCs solutions* (*Massive Open Online Courses*) increasingly affirmed worldwide as a major on-line training system [1,6]. In particular, MOOCs characterize a different paradigm:

- Short duration and pace of videos (multimedia teaching contents)
- Educational material organized in a flexible and dynamic by the teacher
- Delivery with intervention of experts in Virtual Classroom
- Frequent evaluation and self evaluation of learning
- Individual works (deliveries) evaluated by the teachers or discussed and evaluated between peers.

eLearning contents are usually designed and realized in order to be independent from the system used in delivery phase, and then interoperable, accessible and reusable. The content module is made up of Learning Objects - LO, short-term, rapid and reproducible at any time and with a definite teaching goal. To ensure such interoperability and traceability between platforms, they have been defined and perfected specific standards.

The most known and consolidate is the *Shareable Courseware Object Reference Model - SCORM* that is a set of technical standards developed for eLearning software products, it enables interoperability: the model determines how online Learning Content and LMS communicate each other.

A trend that may become a huge part of eLearning is known as *Tin Can* that was the Project Name; *Experience API* is the Name of the Standard. Tin Can was born with the aim of collecting information on learning of learners and it uses specific data structure called *LRS - Learning Record Store* - unique for each user. LRS allows to collect and transmit the set of events of every learner, and makes it possible to collect lot of data about the experiences of a user, online and offline in the background.

3 The Personal Verification Issue: The Biometric Perspective

As observed some years ago, eLearning courses represent a very challenging field for educators: students can operate in a highly independent manner that removes or significantly reduces the instructor control, by providing the opportunity to cheat [12,13]. It is also believed that in most online courses, students engage in a greater number of episodes of cheating than students who attend traditional courses [12,14,15].

We may note that from 2014, 5.5 million students took at least one online course. To ensure integrity in online education, the distance learning industry is working on strategies to verify the identity of distance learning students, with the aim to improve the certainty that students really did the assigned work to get his credits [16,18,19]. The rising status of online education requires new user authentication and identification parameters to preserve the credentials of institutional integrity and students code of conduct.

Determining the identity of students is a crucial issue which educators have to deal with on any eLearning module. This moves to the “front and center” of the classroom experience during testing and determining the originator of written exams. Moreover, determining user identity in the virtual classroom is also linked to user progress and student aid eligibility [12].

So, in recent years, various approaches and methodologies to User Authentication for online learning environment have been tested and some simple solutions are now integrated into a few online learning platforms. There is a strong interest in finding a reliable and cost-effective protocol to safely assess students engaging in learning [18, 19].

Nowadays there is great variety of eLearning paradigms, and different assessment methodologies available, but one strategy will not fit all situations. Verifying user identity in every situation is not realistic, practical and cost effective, and again, the students satisfaction is a strong requirement which shouldn't interfere with learning and a student's privacy [16].

With regard to this problem, biometric identification techniques are inherently designed to recognize particular characteristics of our bodies and our behavior, considering that body and behavioral dynamic changes although slight are inevitable. Consequently, biometric technologies require a reliable and robust operating environment [12].

Biometric traits fall into the *something you are* paradigm, in contrast to the widely used *something you know* (passwords and PINs) and *something you have* (smart cards, tokens) paradigms.

It is possible to consider two categories of biometric traits:

- Physiological characteristics, such as retina, iris or facial characteristics, fingerprints, hand or palm geometry.
- Behavioral characteristics such as signature, keystroke dynamics and gait, voice [22].

By considering the cost of online identification and verification, sophisticated systems and solutions can be integrated into the process [12, 13]. A single biometric technology cannot meet all requirements; different biometrics have pros and cons, and each one has specific features and properties. Moreover, they exhibit a different degree of individuality [17].

In order to integrate a biometric trait for supporting the identity verification process of an online course: (i) the user verification identity must be performed continuously, during time (for example, during the examination); (ii) we can consider a closed environment, with a few of identities to be verified; (iii) users must be appropriately trained.

A first preliminary investigation was done in [20]. Facial recognition was proposed, because it allows a real-time verification of actual presence, low manufacturing costs and fair degree of reliability. A modular system implemented *detection and recognition operations*, able to verify the presence of learners beyond the screen and enable the authentication process.

At the same time, facial detection allowed to verify the contemporary presence of more individuals and unregistered people in the environment. Beside face, voice and keystroke dynamics were proposed. Voice Recognition was useful for detecting the presence of others who interact with the candidate, whilst the keystroke dynamics identified the candidate by typing frequency (continuous identity verification).

The combination of these three biometric traits significantly increases the percentage of recognition of candidates. Implementation costs are low and hardware equipment and sensors are usually already assembled on ICT devices [20].

Among the three cited traits, keystroke dynamics, that is, the way of typing, was particularly the focus of another work [13]. This investigation was carried out with the aim of developing a robust system to authenticate students taking online examinations. Keystroke dynamics focuses on the user's typing style by monitoring "the keyboard inputs thousands of times per second in an attempt to identify users based on habitual typing rhythm patterns" [13].

There is extensive evidence regarding the reliability of keystroke dynamics to accurately determine the user's identity. Keystroke dynamics is inexpensive compared to other biometric technologies: the capture device is the keyboard [13]. Two keystroke verification techniques were investigated:

- Static technique: it only works at specific time intervals.
- Continuous technique: it monitors the typing behavior throughout the interaction process; it is ideal to monitor environments where the fatigue or attention recognition, or continuous proof of the user must be performed.

Beside keystroke dynamics, the stylometry is a behavioural biometric trait which determines the authorship of manuscripts from the authors' linguistic styles. Typically statistical linguistic features are used at the word and syntax level. Keystroke and stylometry are appealing: not intrusive, inexpensive, continuing verification - dynamic verification.

Keystroke and stylometry were proposed in [21] for developing a robust system to authenticate students taking online examinations. This is also one of the few works where experiments are carried out in order to quantify the effectiveness of the proposed approach.

Stylometry seems to be a useful addition to the process because the student could type the answers to the test while someone provides suggestions to the student that simply types the words of the coach without worrying about converting the linguistic style into his own [21].

The keystroke and the stylometry systems consisted of a data collector, a feature extractor, and a pattern classifier. Data were collected from 30 students of a spreadsheet modeling course (17 males and 13 females). A feature vector was extracted from keystroke and stylometry traits. The feature extractor parses each file creating both keystroke and stylometry feature vectors for later processing. Experiments were carried out by separating the two traits.

In the keystroke dynamics experiment, 239 features have been employed, with means and standard deviations of the timings of key press durations and transitions, and the percentage utilization of specific keys, as follows [21]:

- 78 duration features - 39 means and 39 standard deviations - individual letter and non-letter keys, and groups of letter and non-letter keys (see Fig. 1)
- 70 type - 1 transition features - 35 means and 35 standard deviations - transitions between any combination letters/non letters or groups of them (see Fig. 2)
- 70 type - 2 transition features - 35 means and 35 standard deviations - as type-1 transition features with different method of measurement (see Fig. 2)
- 19% features - non-letter keys and mouse clicks
- 2 keystroke input rates: total time to enter the text/total number of keystrokes and mouse events, total time to enter the text minus pauses greater than half second/total number of keystrokes and mouse events.

The following figures present the hierarchy trees for duration categories and transition categories [21]:

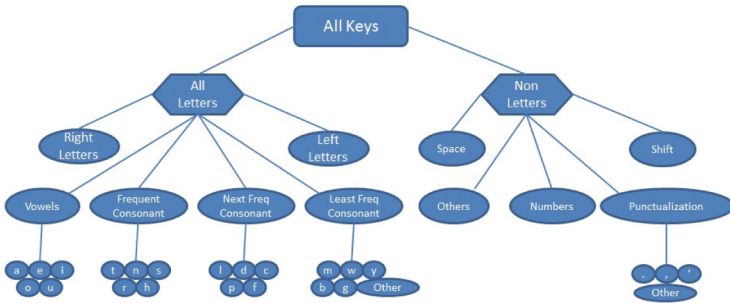


Fig. 1. Hierarchy tree - 39 duration categories - this figure quotes from [21]

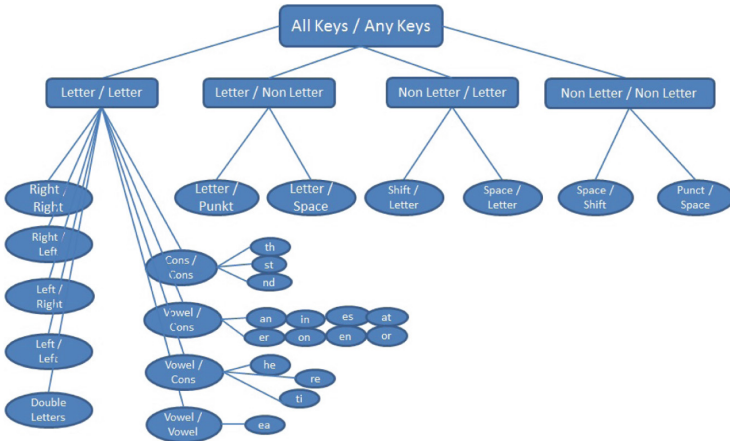


Fig. 2. Hierarchy tree - 35 transition categories (type 1 and type 2) - this figure quotes from [21]

In the stylometry experiment, a set of 228 linguistic features was used: 49 character-based, 13 word-based and 166 syntax-based features. “The features were normalized to be relatively independent of the text length – e.g. the number of different words (vocabulary)/total number of words was used rather than simply the number of different words” [21]. The choice of features shows typical differences in the student population, in fact it can be seen that some students have knowledge and use of extensive vocabulary and terminology, for others this is somewhat reduced.

For both keystroke dynamics and stylometry, the classification procedure is based on the *k-nearest-neighbor* algorithm with the Euclidean distance; it classified the unknown difference vectors, with a reference set composed of the differences between all combinations of the claimed user’s enrolled vectors (within-person) and the differences between the claimed user and every other user (between-person) [21].

Two closed-system experiments were carried out on each of the keystroke and stylometry systems. It has operated on a relatively small database, data was collected by 30 students of the course. But with these requirements of the application that dimension may be acceptable: a small sets of subjects is the universe of interest, the remainder being outside of this universe. The following performance was achieved, given in terms of percentage of correctly verified users, on keystroke dynamics and stylometry (and combined) systems:

– **Keystroke System:**

99.96% on the 3000 keystroke half test sample

100% on the 6000 keystroke full test sample.

– **Stylometry System:**

74% on test input of 500 words

78% on test input of 1000 words.

In the two experiments the answers have been combined to obtain reasonably sized biometric samples; and each experiment has been based on four online short-answer tests of 10 questions each.

First experiment: each sample consisted of five test answers (half answers of each test), obtaining 8 overall samples per student, since each of the four tests contained ten questions for a total of 40 questions.

Second experiment: each sample consisted of ten answers (all the answers of a test) defining 4 samples per student.

Table 1 synthesizes design and results of both experiments. Observe that keystroke dynamics based system has achieved significantly superior performance over the stylometry based system. Keystroke dynamics and stylometry are both behavioral biometrics, but they work at different cognitive levels. The stylometry mainly involves aspects of syntax and semantics, and their is much more complex than that required by the keystroke dynamics [21].

Table 1. Design and results of experiments reported in [21]

Experiment	Data Samples	Keystr EER	Stylo EER
1	– 8 Samples/Student – 5 Answers Combined	0.04 %	~ 26 %
2	– 4 Samples/Student – 10 Answers Combined	0.00 %	~ 22 %

4 Discussions

Although biometrics have become very popular in the last years, some studies have proposed biometrics for eLearning applications, to achieve positive results in the identification of the user, in the continuous especially.

It is important to weigh up costs and ease of implementation, taking into account that now all the devices are equipped with a camera, microphone and keyboard, and then it seemed natural to choose biometric recognition techniques such as face, voice and keystroke dynamics.

In addition, biometrics can be used in eLearning to evaluate the degree of the candidate’s attention. The different proposed combinations can provide excellent results in terms of efficiency and effectiveness.

Certainly eLearning technology and systems have achieved a high architectural and educational level. Current paradigms and technologies introduced new methodologies in collaborative learning, rapid and blended learning but still there is a gap in the management of continuous user identification process during the examination and assessment of learning.

This last aspect seems to be still a problem of not simple solution. Biometric techniques are sometimes invasive or costly or comfortably articulated during the implementation phase during an examination. There is also a psychological factor and the unevenness of the environments seems to make the process of continuous identification quite complex.

We believe that there isn’t currently a criterion for completeness of biometric technologies, since we do not yet have truly specific mature technologies that are applicable to the eLearning world. It can be observed that discussed experiments, where done, were performed in a favorable, very specific context with verification in the continuous and during the examination phase by fully collaborative students. The results were certainly encouraging but significant efforts must still be done. The, positive, side effect was that the search for specific biometric verification methods to eLearning applications increased the interest on behavioral biometrics such as the keystroke dynamic and the stylometry. These traits, coupled with others, as the fingerprint and the face, are worthy to be considered as a good basis for developing specific approaches for personal verification to be performed continuously over time, in order to detect the presence and the identity of the student during the exam remotely.

References

1. Beltrametti, M., Lattanzi, R., Coppi, M., Genovese, P.V., Gilbert, P.: eLearning - la rivoluzione in corso e l'impatto sul sistema della formazione in Italia, Aspen (2014)
2. Leavoy, P., Biraghi, A.: The Importance of the Learner Experience in eLearning, Docebo (2017)
3. Campanella, S., Dimauro, G., Ferrante, A., Impedovo, D., Impedovo, S., Lucchese, M.G., Modugno, R., Pirlo, G., Sarcinella, L., Stasolla, E., Trullo, C.A.: eLearning platforms in the Italian Universities: the technological solutions at the University of Bari. WSEAS Trans. Adv. Eng. Educ. **5**(1) (2008)
4. Cavus, N., Zabadi, T.: A comparison of open source learning management systems. Soc. Behav. Sci. **143** (2014)
5. Ajlan, S.: A comparative study between eLearning features (2012)
6. CRUI: MOOCs MASSIVE OPEN ON-LINE COURSES - Prospettive e Opportunità per l'Università italiana (2015)
7. Botturi, L., Canottoni, L., Lepori, B., Tardini, S.: Fast prototyping as communication catalyst for eLearning design (2008)
8. Dick, W., Carey, L., Carey, J.O.: The Systematic Design of Instruction, 8th edn. Pearson (2015)
9. Molenda, M.: In search of the elusive ADDIE model. Perform. Improv. **42**(5) (2003)
10. Morrison, G.R., Ross, S.M., Kemp, J.E.: Designing Effective Instruction, 4th edn. Wiley, New York (2004)
11. Abkulut, Y.: Implications of two well-known models for instructional designers in distance education: Dick-Carey versus Morrison-Ross-Kemp. Turk. Online J. Distance Educ.-TOJDE **8**(2) (2007)
12. Poullet, K., Douglas, D.M., Chawdhry, A.: Verifying user identities in distance learning courses: do we know who is sitting and submitting behind the screen? Issues Inform. Syst. **15**(1) (2014)
13. Monrose, F., Rubin, A.D.: Keystroke dynamics as a biometric for authentication. Future Gener. Comput. Syst. **16** (2000)
14. Frost, J., Hamlin, A., Barczyk, C.: A survey of professor reactions to scenarios of academic dishonesty in American universities. J. Bus. Inq.: Res. Educ. Appl. **6**(1), 11–18 (2007)
15. Grijalva, T.C., Kerkvliet, J., Nowell, C.: Academic honesty and online courses. Coll. Stud. J. **40**(1) (2006)
16. Jortberg, M.: Methods to verify the identity of Distant Learning Students. Axiom White Paper (2009)
17. Jain, A., Bolle, R., Pankanti, S.: Biometrics – Personal Identification in Network Society. Kluwer Academic Publishers, Dordrecht (1999)
18. Bailie, J.L., Jortberg, M.A.: Distance learning student authentication: verifying the identity of online students. ACXIOM - Third Annual Report on Identity in Distance Learning, White Paper (2008)
19. Bailie, J.L., Jortberg, M.A.: Online learner authentication: verifying the identity of online users. MERLOT J. Online Learn. Teach. **5**(2) (2009)
20. Mastronardi, G., Buonamassa, G., Patruno, T., Pierri, C.: Tecniche Biometriche Combinate nei Processi di eLearning e eVoting (2014)
21. Monaco, J.V., Stewart, J.C., Cha, S.-H., Tappert, C.C.: Behavioral biometric verification of student identity in online course assessment and authentication of authors in literary works (2013)
22. Gagbla, G.K.: Applying Keystroke Dynamics for Personal Authentication (2005). <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.101.9517&rep=rep1&type=pdf>