Identity Verification Using a Kinematic Memory Detection Technique

Merylin Monaro, Luciano Gamberini and Giuseppe Sartori

Abstract We present a new method that allows the identification of false self-declared identity, based on indirect measures of the memories relating the affirmed personal details. This method exploits kinematic analysis of mouse as implicit measure of deception, while the user is answering to personal information. Results show that using mouse movement analysis, it is possible to reach a high rate of accuracy in detecting the veracity of self-declared identities. In fact, we obtained an average accuracy of 88 % in the classification of single answers as truthful or untruthful, that corresponds overall to 9.7/10 participants correctly classified as true tellers or liars. The advantage of this method is that it does not requires any knowledge about the real identity of the declarant.

Keywords Identity verification · Lie detection · Memory detection

1 Introduction

Nowadays the security concerning the identity has become a very sensitive issue. In particular, the increase of terrorist attacks in the last decades imposes the need to recognize declarants of false identity. Usually migrants from Middle East entering Europe or USA do not have any documents and personal details are frequently self-declared. Among them, a high number of terrorists giving false identities are believed to be hidden. Because terrorists move across countries using fake identities, the identity detection is now a major target in anti-terrorism [1].

M. Monaro (✉) · L. Gamberini · G. Sartori
University of Padova, Human Inspired Technolgy Research Centre, via Luzzati 4, 35122 Padua, Italy
e-mail: merylin.monaro@studenti.unipd.it

L. Gamberini
e-mail: luciano.gamberini@unipd.it

G. Sartori
e-mail: giuseppe.sartori@unipd.it

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_11

Deception is cognitively more complex than truth telling and this higher complexity reflects itself in a lengthening of the reaction times (RT) during a response [2]. According to literature, two memory detection techniques based on RT have been proposed to identify liars. These are the autobiographical Implicit Association Test (aIAT) [3] and the RT-based Concealed Information Test (RT-CIT) [4]. These techniques may be used also as tools for identity verification [5].

RT based techniques have a number of advantages compared to the traditional psychophysiological techniques to detect deception, as the polygraph [6]. First, RT are not subjected to strong individual and environmental changes, such as in the case of physiological parameters. Secondly, these techniques are inexpensive and suitable to be used on large scale. However, these techniques are not without limitations. Even though RTs are implicit measures, during the aIAT or CIT examination the lie detection purpose is explicit (overt detection of deception). Furthermore, RT based techniques only studied the latency in the response, so the liar has to check only this unique parameter to falsify the evidence. Finally, the use of these methods requires a prior knowledge about the information that has to be checked as true or false. In fact, both aIAT and CIT require that the true identity (or the true memory) is available, while in most real applications the true identity, as the migrant's case, is unknown to the examiner. This feature limits the practical application of RT based verifications, even if their efficiency is proved.

The analysis of movements during the response has already been shown to present a series of advantages, since it allows to capture the cognitive complexity in stimulus processing by the registration of a variety of indicators including not only the reaction time. Recently, researchers have shown that kinematic analysis can be used as implicit measure of the cognitive processes underlying a task [7]. Several authors, as [8–10], measured hand movements during choice tasks on a screen to understand the dynamics of a wide range of psychological processes. They described as a simple hand motion can reflects in real-time the progress of the underlying cognitive processing. Therefore, hand-motor tracking can provide a good trace of mind processes.

Because cognition is largely involved in the process of lie [11], it is reasonable to think that the analysis of hand movements can be a good implicit measure to study the cognitive mechanisms involved in lying. A first and precursive study about the kinematic as signatures of deception was presented in [12]. The authors compared motor trajectories while subjects were engaged in an instructed lie task. Participants were required to respond truthfully or lying to the presented sentences by a visual cue. Authors used the Nintendo Wii Remote to record subjects' responses. Results reported that deceptive responses could be distinguished from truthful ones on the basis of several parameters, including the motor onset time, the overall time required for responding, the trajectory of the movement and kinematic parameters such as velocity and acceleration. In Ref. [13], the authors studied mouse movements in an insurance fraud online context. Their results suggest that liars had an increasing in the distance of movements, a decreasing in the speed of movements, an increasing in the response time, and a more number of left clicks. In [14] authors proposed a pilot study to identify guilty individuals involved in specific insider threat activities. They

analysed mouse movements while participants compiled an online survey similar to the Concealed Information Test (CIT). Their preliminary observations showed that guilty insiders had a different motion pattern when answering the key-item as compared to the answering of non-key-items, which was indicative of an increased cognitive activity while deceiving.

Concerning the identity verification, there are also several studies in literature that applied mouse movements analysis to biometric user authentication or identification in informatics fields [15]. However, these methods require necessarily a certain level of knowledge about the alleged user and a user-specific training, in order to be able to recognize him/her or the liar.

The goal of this work is to present a new identity check technique based on mouse movements recording, to identify false self-declared identities without knowing anything about the real identity of the declarant. This method consists in a memory detection technique, which investigates the truthful or untruthful nature of the memory for the personal information declared, using implicit measures from mouse movements. In other words, we employed kinematic analysis of the mouse movements to identify implicit signatures of deception.

2 Method

2.1 Participants

40 participants were recruited at the Department of General Psychology in Padova University. The sample consisted of 17 males and 23 females. Their average age was M = 25 (SD = 4.6), and their average education level was M = 17 (SD = 1.8). Because they use the mouse differently, left-handed subjects were excluded. All subjects agreed on the informed consent before the experiment.

2.2 Experimental Procedure

The experiment was implemented using *MouseTracker* software [16].

During the experimental procedure, participants were asked to answer 3 *yes* or *no* questions about their personal information, clicking with the mouse on the correct alternative response on the computer screen (Fig. 1 shows an example). 20 participants answered truthfully, while the others were instructed to lie about their identity according to a false autobiographical profile.

The 20 liars were instructed to learn a false identity from an Italian standard Identity Card, where a photo of the subject were attached, and which contained false personal data (an example of ID Card is reported in Appendix). After the

Fig. 1 Example of the task presented to the subjects

learning phase, participants recalled the information in the ID card for two times. Between the two recalls, they held a mathematical distracting task. On the other hand, the truth tellers performed a mathematical task and revised their real auto-biographical data only once before starting the experiment.

During the experimental task, three different kinds of questions were presented to participants, in random order. *Expected questions*: 6 questions about information explicitly trained from liars during the learning and recall phases (e.g. date of birth). *Unexpected questions*: 6 questions related to the identity but not explicitly rehearsed before the experiment (e.g. age). Liars can get this information by applying a reasoning to the learned data. For example, if I know that I was born in April 1989, I can conclude that I am 26 years old. *Control questions*: 4 questions about personal characteristics that could not be denied. These are information regarding evident physical traits that cannot be hidden to the examiner, as the gender.

Each of these 16 questions was presented two times, one time the subject had to answer *yes* and in the other one the participant had to give a *no* response, for a total of 32 questions. In this way, truth tellers answered sincerely at all questions, whereas liars answered lying on *expected* and *unexpected* questions that required a *yes* response. Liar's answers to *control* questions and to *expected* and *unexpected* questions, which required a *no* response, were truthful. An example of questions is reported in Appendix.

To view each question, participants had to click on the Start button in the lower part of the screen. Then they chose the answer clicking on the response boxes positioned in the two top corners of the screen.

2.3 Data Analysis

For each answer, the motor response was recorded using *MouseTracker* software. Because each recorded trajectory have a different length, in order to permit averaging and comparison across multiple trials, each motor response was time-normalized. By default, *MouseTracker* performs a time normalization in 101

time steps using linear interpolation. Thus, each trajectory had 101 time-steps and each time-step had a corresponding *x* and *y* coordinate.

We analysed signatures of deception in terms of the shape of each movement trajectory and the location of the trajectory over time. We also quantified the trajectory properties on dimensions of velocity, stability, and direction. In particular, we collected the following features:

- *Number of errors*: number of incorrect answers.
- *Initiation time*: time between the appearance of the question and the beginning of the mouse movement.
- *Reaction time*: time from the appearance of the question to the click on the answer box.
- *Maximum deviation*: the largest perpendicular deviation between the actual trajectory and its idealized trajectory.
- *Area under the curve*: the geometric area between the actual trajectory and the idealized trajectory.
- *Maximum deviation time*: time to reach the point of maximum deviation.
- *x-flip*: number reversals of direction along the *x*-axis.
- *y-flip*: number reversals of direction along the *y*-axis.
- *X*, *Y coordinates over the time*: position of the mouse along the axis over the time. Specifically we choose to use for the analysis only Y coordinate data, for time-steps 18, 29, 30. This is because already from a preliminary visual analysis the two experimental groups clearly differed only in position of the mouse along the *y*-axis over the time.
- *Acceleration over the time*: acceleration of the mouse along the axis over the time. We calculated acceleration along *y*-axis for time intervals 18–29 and 29–30.

These features were used to train different machine learning classifiers on all subject responses.

3 Results

A preliminary visual analysis showed a significant difference in kinematic responses between liars and truth tellers. The average maximum deviation (MD) for liars is 0.33 (SD = 0.42), while for truth tellers is 0.15 (SD = 0.28). The area under the curve is wider in liars (AUC = 0.6, SD = 1.1), than truth tellers (AUC = 0.22, SD = 0.5). Figure 2 shows the average trajectory for liars and truth tellers. Furthermore, it represents the position of the mouse along *x* and *y*-axis during the response time for liars and truth tellers. In addition, liars made a greater number of errors than truth tellers (error frequency for liars = 84, error frequency for truth tellers = 7).

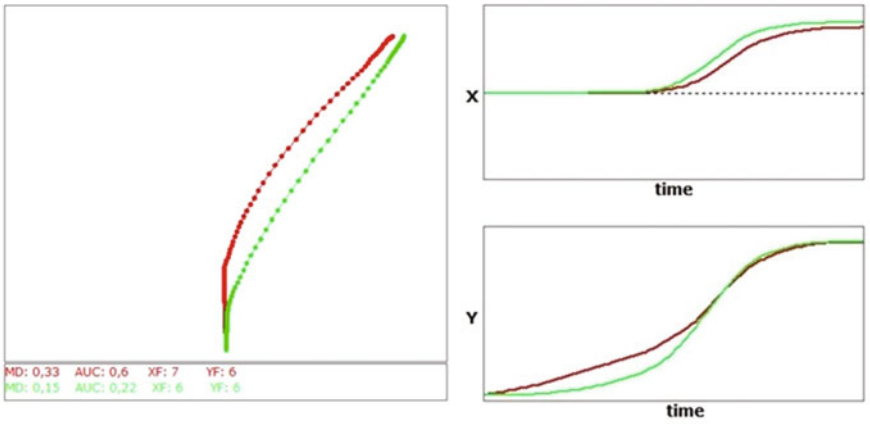

Fig. 2 *Left* average trajectory for liars (in *red*) and truth tellers (in *green*). *Right* Position of the mouse along *x* and *y* axis during the response time for liars (in *red*) and truth tellers (in *green*)

In a first step, a 10-fold cross-validation Random Forest classifier was run on all dataset (1280 stimuli). We obtained an accuracy around 90 % in the classification of the single answer as truth or lie.

Secondly, the efficiency of the classification was evaluated on 10 test sets of 10 subjects each one. The 10 test sets were extracted from the original dataset of 40 subjects using the following rules: each test set contained 5 liars and 5 truth tellers; each subject appeared in the 10 test sets a minimum of 2 and a maximum of 3 times. In this way, each test set included 320 stimuli gathered from 10 subjects. Data from the remaining 30 subjects (960 stimuli) were employed to build the model. Results for the classification of each training-test couple is reported in Table 1. Using a

Table 1 Simple logistic classifier accuracy for 10 training set and 10 test set including all stimuli

Training-test set couples	Accuracy for cross-validation on training set (%)	Accuracy in test set (%)	Number of subjects correctly classified
1	84.37	65.62	7/10
2	77.81	75.31	9/10
3	82.08	72.5	7/10
4	77.7	77.81	9/10
5	81.35	74.68	8/10
6	77.81	75.31	8/10
7	77.06	76.56	7/10
8	80	80.31	8/10
9	8.16	63.75	7/10
10	76.85	73.75	8/10
Mean	79.91	73.56	7.8/10

Table 2 Simple logistic classifier accuracy for 10 training set and 10 test set including as stimuli only *expected* and *unexpected* questions that required a *yes* response

Training-test set couples	Accuracy for cross-validation on training set (%)	Accuracy in test set	Number of subjects correctly classified
1	83.88	74.16	8/10
2	80.55	72.5	9/10
3	79.44	82.5	9/10
4	82.5	80.83	9/10
5	85.55	70.83	8/10
6	80.27	84.16	10/10
7	80	83.33	9/10
8	80.27	80	9/10
9	83.05	73.33	8/10
10	79.44	78.33	9/10
Mean	81.49	77.99	8.8/10

Simple Logistic classifier, we obtained an overall accuracy of 73.56 % in classifying a single stimulus as truthful or untruthful. From the classification of single answer as true or false, according to a majority vote system, we traced the classification of the single subject as liar or truth teller. On the single participant, we reached an average accuracy of 7.8/10 participants correctly classified as true tellers or liars, with a minimum accuracy of 7/10 and a maximum of 9/10.

We repeated this procedure for training and testing the classifier on the answers in which only truth tellers responded sincerely and liars cheating (*expected* and *unexpected* questions that required a *yes* response). 10 training sets and 10 test sets were created as above. This time, each test set included 120 stimuli gathered from 10 participants and each training set included 360 stimuli obtained from 30 participants. Classification results are shown in Table 2. Training a Simple Logistic classifier, we obtained an accuracy around 78 % in the classification of the stimuli as sincere or deceitful, which means that 8.8/10 participants were correctly classified as true tellers or liars, with an accuracy ranging from 8/10 to 10/10.

Finally, we built a model including in the training set also the answers of all 40 participants, in which both liars and truth tellers answered truthfully (*control* questions and *expected* and *unexpected* questions that required a *no* response). Each test set included the answers of 10 participants in *expected* and *unexpected* questions that required a *yes* response. Therefore, each one of the 10 training sets included 1160 stimuli, and each test set included 120 stimuli. Using a Random Forest classifier, we reached an average accuracy of 88.08 % in the classification of single answers as truthful or untruthful, that corresponds overall to 9.7/10 participants correctly classified as true tellers or liars, with a minimum accuracy of 8/10 and a maximum of 10/10. These data are reported in Table 3.

Table 3 Random forest classifier accuracy for 10 training set and 10 test set including *control* questions, *expected* and *unexpected* questions of all 40 participants in the training set

Training-test set couples	Accuracy for cross-validation on training set (%)	Accuracy in test set (%)	Number of subjects correctly classified
1	91.37	87.5	10/10
2	92.58	76.66	8/10
3	90.86	90	10/10
4	91.63	86.66	10/10
5	90.77	90.83	10/10
6	91.63	89.16	10/10
7	91.12	87.5	10/10
8	90.77	95.83	10/10
9	92.15	90.83	10/10
10	92.4	85.83	9/10
Mean	91.52	88.08	9.7

4 Conclusions

This work shows that using mouse movement analysis, it is possible to reach a high rate of accuracy in detecting the veracity of self-declared identities. The accuracy of the classification is very high not only for the single subject, but also for the single answer.

As already shown in literature [17], the presence of *unexpected questions* induce in liars an increase in cognitive load. This increase reflected itself in a different pattern of the kinematic response that became distinguishable from the truth teller pattern.

We believe that this approach can have several advantages compared to the RT based techniques mentioned above. First, kinematic indices can be recorded in a hidden way while the user interacts with the device and not being aware of what we are observing. Secondly, the detection of these indices is inexpensive, easily obtainable and does not require any equipment in addition to what the subject is already using during the interaction with the computer. This method is potentially very well adapted to the detection of deception also in the context of web, because it do not require the presence of an examiner and can be run automatically, quickly and anywhere. Furthermore, the use of mouse kinematic instead of the simple RT pushing a key on keyboard in order to record responses has a number of advantages. While button press may only permit to record RT, to use a mouse allows to capture the cognitive processes and their complexity by the registration of a large set of indicators, which include not only the reaction time. For this reason, the technique is promising also concerning resistance to countermeasures. The large number of characteristics of movement seem, in principle, difficult to control entirely via efficient countermeasures to lie detection.

Appendix

Fake ID Document for Facilitating Rehearsal of Faked Identity by Liars. The document reproduced an Italian standard Identity Card. It contains the following information: last name, first name, date of birth, city of birth, citizenship, city of residence, residence address, marital status, occupation, height, hair color, eye color.

List of Questions Presented to Subjects

Topic	Example for *yes* answer	Example for *no* answer
Control questions		
Gender	Are you female?	Are you male?
Skin color	Is your skin white?	Is your skin brown?
Hair color	Do you have blond hair?	Do you have black hair?
Citizenship	Are you an Italian citizen?	Are you a French citizen?
Expected questions		
First name	Is Alice your name?	Is Maria your name?
Last name	Is Rossi your last name?	Is Bianchi your last name?
Year of birth	Were you born in 1989?	Were you born in 1986?
Month of birth	Were you born in April?	Were you born in August?
City of residence	Do you live in Limena?	Do you live in Caserta?
Residence address	Do you live at Vespucci street?	Do you live at Marconi street?
Unexpected questions		
Age	Are you 26 years old?	Are you 23 years old?
Zodiac sign	Is Aries your zodiac sign?	Is Leo your zodiac sign?
Region of birth	Were you born in Veneto?	Were you born in Campania?
Province of birth	Were you born in Padova province?	Were you born in Caserta province?
Region of residence	Do you live in Veneto?	Do you live in Campania?
Chief town of residence region	Is Venezia the chief town of your residence region?	Is Napoli the chief town of your residence region?

References

1. The University of Texas at Austin. http://news.utexas.edu/2015/12/07/the-direct-link-between-identity-theft-and-terrorism
2. Walczyk, J.J., Igou, F.P., Dixon, A.P., Tcholakian, T.: Advancing lie detection by inducing cognitive load on liars: a review of relevant theories and techniques guided by lessons from polygraph-based approaches. Front. Psychol. **4**, 14 (2013)
3. Sartori, G., Agosta, S., Zogmaister, C., Ferrara, S.D., Castiello, U.: How to accuratly detect autobiographical events. Psychol. Sci. **19**(8), 772–780 (2008)
4. Verschuere, B., Ben-Shakhar, G., Meijer, E.: Memory Detection: Theory and Application of the Concealed Information Test. Cambridge University Press, Cambridge (2011)
5. Verschuere, B., Kleinberg, B.: ID-check: online concealed information test reveals true identity. J. Forensic Sci. **61** (2015)
6. Ganis, G., Keenan, J.P.: The cognitive neuroscience of deception. Soc. Neurosci. **4**(6), 465–472 (2009)
7. Freeman, J.B., Dale, R., Farmer, T.A.: Hand in motion reveals mind in motion. Front. Psychol. **2**, 59 (2011)
8. Calcagni, A., Lombardi, L.: Dynamic fuzzy rating tracker (DYFRAT): a novel methodology formodeling real-time dynamic cognitive processes in rating scales. Appl. Soft Comput. **24**, 948–961 (2014)
9. Song, J.-H., Nakayama, K.: Hidden cognitive states revealed in choice reaching tasks. Trends Cogn. Sci. **13**(8), 360–366 (2009)
10. McKinstry, C., Dale, R., Spivey, M.J.: Action dynamics reveal parallel competition in decision making. Psychol. Sci. **19**(1), 22–24 (2008)
11. Vrij, A.: Deception Detection: Current Challenges and New Approaches. Wiley, Oxford (2014)
12. Duran, N.D., Dale, R., McNamara, D.S.: The action dynamics of overcoming the truth. Psychon. Bull. Rev. **17**(4), 486–491 (2010)
13. Hibbeln, M., Jenkins, J., Schneider, C., Valacich, J., Weinmann, M.: Investigating the effect of fraud on mouse usage in human-computer interactions. In: International Conference on Information Systems (2014)
14. Valacich, J.S., Jenkins, J.L., Nunamaker, Jr., J.F., Hariri, S., Howie, J.: Identifying insider threats through monitoring mouse movements in concealed information tests. In: Hawaii International Conference on System Sciences. Deception Detection Symposium (2013)
15. Jorgensen, Z., Yu, T.: On mouse dynamics as a behavioral biometric for authentication. In: ASIACCS '11 Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security, pp. 476–482 (2001)
16. Freeman, J.B., Ambady, N.: MouseTracker: software for studying real-time mental processing using a computer mouse-tracking method. Behav. Res. Methods **42**, 226–241 (2010)
17. Vrij, A., Leal, S., Granhag, P.A., Mann, S., Fisher, R.P., Hillman, J., Sperry, K.: Outsmarting the liars: the benefit of asking unanticipated questions. Law Hum. Behav. **33**(2), 159–166 (2009)