Gender Identification through Handwriting: an Online Approach

Gennaro Cordasco*, Michele Buonanno*, Marcos Faundez-Zanuy‡,
Maria Teresa Riviello*, Laurence Likforman-Sulem§, and Anna Esposito*

*Department of Psychology, Universit`a degli Studi della Campania "L. Vanvitelli", and
International Institute for Advanced Scientific Studies (IIASS), Italy,
Email: *gennaro.cordasco@unicampania.it.*

‡Escola Superior Politecnica, TecnoCampus Mataro-Maresme, 08302 Mataro, Spain.

§Télécom ParisTech, Université Paris-Saclay, 75013 Paris, France.

Abstract— The present study was designed to identify writer's gender trough online handwriting and drawing analysis. Two groups – one of 126 males (mean age 24.65, SD=2.45) and the other of 114 females (mean age =24.51, SD=2.50) participants were involved in the experiment. They were asked to perform seven writing and drawing tasks utilizing a digitizing tablet and a special writing device. Seventeen writing features grouped into five categories (Pressure, Ductus, Time Space and Inclination) were considered. The experiment's results show that it is possible to distinguish between male and female writers investigating their performance while copying a house drawing (task 2), writing words in capital letters (task 3) and writing a complete sentence in cursive letters (task 7), in particular focusing on Ductus (number of strokes) and Time categories of writing features.

Keywords—Handwriting, Drawing, Online analysis, Gender recognition.

I. INTRODUCTION

Handwriting is a human common activity that can be considered distinctive of a person, i.e. everyone has her/his own unique style of handwriting which turns out differently from others. This is the reason why handwriting is often used as the basis to identify people's different characteristics such as personality [1] [7], neurodegenerative diseases [5], emotional states [3] [8] and also age, nationality and gender [9].

The prediction of this last kind of writer's soft biometrics information (gender, age, ethnicity, among others) [10] is useful in demographic investigations and forensic applications. As an example, in the business marketing area, the use of demographics helps to identify the number of people to which companies could potentially target their products or services [11]. In the forensic area, for anonymous handwritten text, if the writer's gender and age are known, one can restrict the investigation to a small population [12].
In terms of demographic properties, differences regarding gender have become the subject of numerous researches [13] [14].

Therefore, detecting gender through handwriting can be useful for research in any field where gender detection is needed.

In [15], an attempt was made to examine and analyze the various minute features possessed in handwriting in order to discriminate between male and female writers. Over 130 samples of handwriting, 65 from males and 65 from female individuals, aged from 18 to 30 years, were examined. They were all given a pangram to write on a white sheet of paper with a classical Reynolds jetter dot pen. The handwriting samples of male and female volunteers were observed for the presence of total 27 features of handwriting, which were divided into two groups: macro features (such as slant, word spacing, dispersive writing and cursive writing) and micro features (such as dot over "i", hook at the start of "c", hook at the end of "c", hook at end of "d", hook at end of "e", hook at end of "h"). The statistical z-test formula was used to signify the results of the study showing that there are significant differences between the handwriting of a male and a female and hence a handwriting sample can be examined for gender identification purposes.

A similar method of handwriting analysis to determine writers' gender was exploited in [6]. This study considered a total of 130 handwriting samples (by 65 males and 65 females aged from 18 to 30 years). Handwriting samples were taken on A4 paper of London Letter. Macro features and micro features were examined in all the handwriting by using a feature extraction method and z-tests. The results confirmed the effectiveness of the two methods in identifying gender from handwriting.

In [4] a study to predict the gender of individuals from scanned images of their handwritings was reported. Among the different discriminating features of male and female writings, they focused on the slant/orientation, roundedness/curvature, neatness/legibility and writing texture in our study. These features were used to train two classifiers: ANN and SVM. The effectiveness of these features in predicting the gender of the writer of a given sample was evaluated on two databases: the QUWI and the MSHD with text samples in different languages. The proposed technique realized interesting results in predicting gender from handwriting.

In all the above researches, the effort to detect writer's gender is based on a "static" analysis of handwriting with an inevitable loss of information on "dynamic" writing features.

The present paper reports on a dynamic evaluation of male and female "online" handwriting and drawing, with the aim of investigating which quantitative and easy computable features may enable identifying writer's gender.

II. HANDWRITING ANALYSIS FOR GENDER IDENTIFICATION

A. *Handwriting Analysis*

Progress in technological development, especially the introduction of graphic and digital tables and pencil, makes it possible to analyze *online* handwriting and consider dynamic information such as in-air motions, pressure, pen inclination, etc.

In this study, an INTUOS WACOM series 4 digitizing tablet was used for data collection; it should be noted that the tablet in question originally uses a digital pen that does not provide any feedback. Therefore, a special writing device named Intuos Inkpen and a white sheet of paper, (DIN A4 normal paper) laid on the tablet, were used to enable the subject to visualize the stroke making the tasks as natural as possible. Apart from the information about the stroke on the paper, the system is able to capture in-air movements (that is when the ink pen is very close to the paper).

During each experimental task, the following information is continuously captured with a frequency of 125 Hz:

1. position in x-axis;
2. position in y-axis;
3. time stamp;
4. pen status (in-air = 0 or down = 1);
5. azimuth angle of the pen with respect to the tablet;
6. altitude angle of the pen with respect to the tablet;
7. pressure applied by the pen on the paper.

Digital strokes are stored in a comma separated values (CSV) file.

III. PARTICIPANTS AND METHODOLOGY

In this section the involved participants, the experimental setting and procedure are described.

A. *Participants*

The experiment involved a total of 240 subjects split into two groups – one of 126 males (mean age 24.65, SD=2.45) and the other of 114 females (mean age =24.51, SD=2.50). The participants were volunteers recruited at Università degli Studi della Campania "Luigi Vanvitelli' in Caserta (south Italy). All subjects are right-handed and were asked to perform the 7 tasks described in the following.

B. *Tasks and Features*

Figure 1 depicts an A4 paper with all the 7 tasks filled by one participant:

1) copy of a two-pentagon drawing;

2) copy of a house drawing;

3) writing of four Italian words in capital letters (BIODEGRADABILE (biodegradable), FLIPSTRIM (flipstrim), SMINUZZAVANO (to crumble), CHIUNQUE (anyone));

4) loops with left hand;

5) loops with right hand;

6) clock drawing;

7) writing of the following phonetically complete Italian sentence in cursive letters (I pazzi chiedono fiori viola, acqua da bere, tempo per sognare: Crazy people are seeking for purple flowers, drinking water and dreaming time).

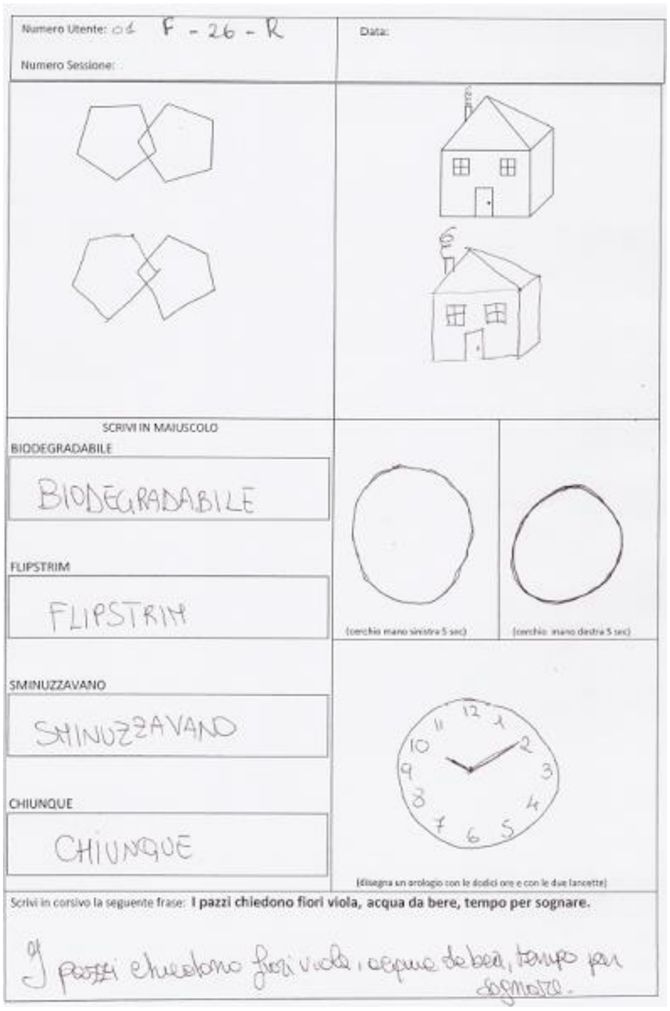

Figure 1: A sample of tasks filled by one participants.

It is worth remembering that the *online* acquisition of data during the experiment enables us to collect much more information than can be deduced from the traits that appear for instance in Figure 1. Indeed, additional information such as acceleration, velocity, instantaneous trajectory angle, instantaneous displacement, time features, and ductus-based features can be inferred. In addition, the system is able to capture in-air movements (when the pen is close to the paper), which do not appear on the paper but have been proved to be useful [16] [17]. Figure 2 depicts on-paper (pen-down) and in-air (pen-up) traits acquired during task 1 (copy of a two-pentagon drawing).

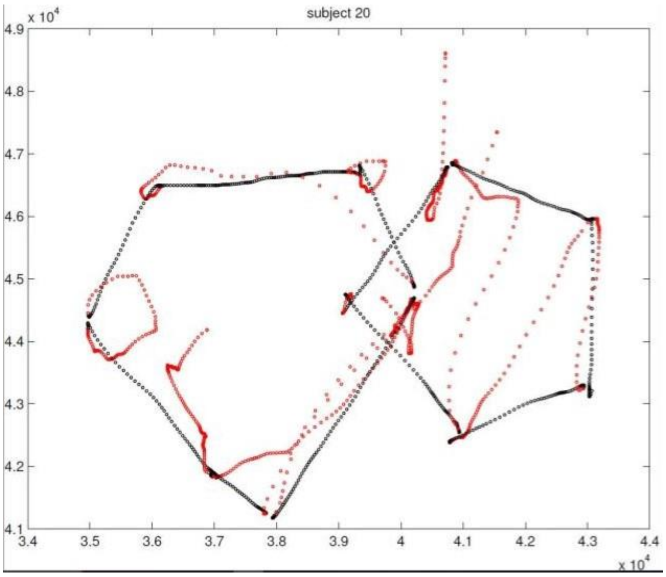

Figure 2 Sample with on-paper strokes (black) and in-air strokes (red).

In this paper we adopt the approach defined in [8] partitioning the pen states into three categories:

- **up**, recorded with pen status 0;
- **down**, recorded with pen status 1;
- **idle**, not recorded but recognizable using timestamps.

Specifically, we considered 17 features grouped into 5 categories:

- **Pressure**, based on the pressure applied by the pen on the paper during a specific task:
  1) P_{min}; minimum pressure;
  2) P_{max}; maximum pressure;
  3) P_{avg}; average pressure;
  4) P_{sd}; standard deviation of pressures;
  5) P_{10}; lower 10_{th} percentile of pressures;
  6) P_{90}; lower 90_{th} percentile of pressures.
- **Ductus,** based on the number of strokes, in each pen state, performed during the task:
  7) N_{up}; number of in-air strokes;
  8) N_{down}; number of on-paper strokes;
  9) N_{idle}; number of idle strokes (strokes away from the tablet).
- **Time**, based on the time spent, in each pen state, to complete the task:
  10) T_{up}; time spent on in-air strokes;
  11) T_{down}; time spend on on-paper strokes;
  12) T_{idle}; time spent in the idle state;
  13) T_{total}; total time elapsed for the task.
- **Space**, based on the space used by the strokes:
  14) S_{bb}; sum of spaces used by on-paper strokes. For each stroke we compute the smallest axis aligned box containing the stroke and sum its area;
  15) S_{avg}; average lengths of empty spaces between consecutive on-paper strokes;
  16) S_{total}; sum of the lengths of empty spaces between consecutive on-paper strokes.
- Inclination, based on the inclination of strokes:
  17) I_{avg}; average inclination of on-paper strokes.

IV. RESULTS

Results are computed through different repeated measures ANOVA analyses. In particular, an ANOVA was performed for each group of handwriting features (category) and for each handwriting task. Each ANOVA considers participants' gender as between subjects factors, whereas participant's scores on each handwriting feature, belonging to the considered category, as within subjects factors. For instance, the ANOVA for the category Ductus and the Task 2 (e.g., Copy of a house drawing) uses the three ductus features scores (e.g., N_{up}, N_{down} and N_{idle}) computed on the data acquired during the task 2 as within subjects factors. The significance level was set at $\alpha<.05$ and differences among means were assessed through Bonferroni's post hoc tests.

In the following, for each category, the significant results are presented:

A. Ductus

a) Task 2: Copy of a house drawing: It exists a significant difference between the two groups [F (1;238) <8.03; p=0.005]. The male (M=26.558, SD=0.96) significantly differs from the female group (M=22.623, SD=1.006) (p= 0.005)

b) Task 3: Writing in capital letters: It exists a significant difference between the two groups [F (1;238) <7.54; p=0.006]. The male (M=45.548, SD=0.518) significantly differs from the female group (M=43.482, SD=0.545) (p= 0.006)

c) Task 7: Writing sentence in cursive letters: It exists a significant difference between the two groups [F (1;238) <25.072; p=0.000]. The male (M=35.024, SD=0.684) significantly differs from the female group (M=30.053, SD=0.719) (p=0.000).
In particular, this difference is due to the scores in the following features:

- N_{up}: males (M=46.32, SD=12.11) significantly differ (p=0.000) from females (M=40.30, SD=9.54),
- N_{down}: males (M=46.56, SD=12.04) significantly differ (p=0.000) from females (M=38.84, SD=9.61).

For the female group, this difference is due to the scores in the following features:

- N_{up} (M=40.30, SD=9.54) significantly differ (p=0.000) from N_{down} (M=38.84, SD=9.61) scores,
- N_{up} (M=40.30, SD=9.54) significantly differ (p=0.000) from N_{idle} (M=11.02, SD=6.68) scores,
- N_{down} (M=38.84, SD=9.61) significantly differ (p=0.000) from N_{idle} (M=11.02, SD=6.68) scores.

For the male group, this difference is due to the scores in the following features:

- N_{idle} (M=12.19, SD=6.12) significantly differ (p=0.000) from N_{up} (M=46.32, SD=12.11) scores,
- N_{idle} (M=12.19, SD=6.12) significantly differ (p=0.000) from N_{down} (M=46.56, SD=12.04) scores.

The results show significant effects of gender for the Ductus category relating to the Copy of a house drawing (2), Writing in capital letters (3) and Writing a sentence in cursive letters (7) tasks. To complete task 7, females use fewer strokes than males, both in the air (N_{up}) and on the sheet (N_{down}), in accordance with [18].

B. Time

a) Task 2: Copy of a house drawing: It exists a significant difference between the two groups [F (1;238) <6.225; p=0.013]. The male (M=19576.050, SD=696.372) significantly differs from the female group (M=17055.162, SD=732.107) (p= 0.013).

In particular, this difference is due to the scores in the following features:

- T_{up}: males (M=15.411, SD=7390.047) significantly differ (p= 0.016) from females (M=13267.31, SD=6243.363),
- T_{down}: males (M=18916.52, SD=8273.096) significantly differ (p= 0.039) from females (M=16904.68, DS=6532.322),
- T_{total}: males (M=39155.45, SD=17157.582) significantly differ (p= 0.013) from females (M=34112.83, SD=13756.778),

For the female group, this difference is due to the scores in the following features:

- T_{up} (M=13267.31, SD=6243.363) significantly differ (p =0.000) from T_{down} (M=16904.68, DS=6532.322) scores,
- T_{up} (M=13267.31, SD=6243.363) significantly differ (p =0.000) from T_{idle} (M=3935.83, SD=3563.602) scores,
- T_{down} (M=16904.68, SD=6532.322) significantly differ (p =0.000) from T_{idle} (M=3935.83, SD=3563.602) scores.

For the male group, this difference is due to the scores in the following features:

- T_{up} (M=15411.63, SD=7390.047) significantly differ (p =0.000) from T_{down} (M=18916.52, SD=8273.096) scores,
- T_{up} (M=15411.63, DS=7390.047) significantly differ (p =0.000) from T_{idle} (M=4820.60, SD=4248.236) scores,
- T_{down} (M=18916.52, SD=8273.096) significantly differ (p =0.000) from T_{idle} (M=4820.60, SD=4248.236) scores.

b) Task 3: Writing in capital letters: It exists a significant difference between the two groups [F (1;238) <14.510; p=0.000]. The male (M=17594.173, SD=365.921) significantly differs from the female group (M=15571.752, SD=384.698) (p= 0.000).

In particular, this difference is due to the scores in the following features:

- T_{up}: males (M=15066.42, SD=50800.972) significantly differ (p= 0.000) from females (M=12563.32, SD=3583.539),
- T_{total}: males (M=35189.12, SD=9460) significantly differ (p= 0.000) from females (M=31145.04, SD=6568.947).

For the female group, this difference is due to the scores in the following features:

- T_{up} (M=12563.32, SD=3583.539) significantly differ (p =0.000) from T_{down} (M=15030.64, SD=2625.135) scores,
- T_{up} (M=12563.32, DS=3583.539) significantly differ (p =0.000) from T_{idle} (M=3548.01, SD=3398.160) scores,
- T_{down} M=15030.64, SD=2625.135) significantly differ (p =0.000) from T_{idle} (M=3548.01, SD=3398.160) scores.

For the male group, this difference is due to the scores in the following features:

- T_{up} (M=15066.42, SD=50800.972) significantly differ (p =0.000) from T_{idle} (M=4226.46, SD=3267.214) scores,
- T_{down} (M=15894.69, SD=4029.221) significantly differ (p =0.000) from T_{idle} (M=4226.46, SD=3267.214) scores,

c) Task 7: Writing sentence in cursive letters: It exists a significant difference between the two groups [F (1;238) <5.062; p=0.025]. The male (M=15367.496, SD=441.371) significantly differs from the female group (M=13926.660, SD=464.020) (p= 0.025).

In particular, this difference is due to the scores in the following features:

- T_{up}: males (M=11079.85, SD=5615.823) significantly differ (p= 0.000) from females (M=8482.26, SD=3476.438),
- T_{total}: males (M=30736.23, SD=10186.987) significantly differs (p= 0.025) females (M=27854.86, SD=9593.057).

For the female group, this difference is due to the scores in the following features:

- T_{up} (M=8482.26, DS=3476.438) significantly differ (p =0.000) from T_{down} (M=15793.83, SD=2853.772) scores,
- T_{up} (M=8482.26, SD=3476.438) significantly differ (p =0.000) from T_{idle} (M=3575.68, SD=7278.985) scores,
- T_{down} (M=15793.83, SD=2853.772) significantly differ (p =0.000) from T_{idle} (M=3575.68, SD=7278.985) scores.

For the male group, this difference is due to the scores in the following features:

- T_{up} (M=11079.85, SD=5615.823) significantly differ (p =0.000) from T_{idle} (M=3859.21, SD=3470.047) scores,
- T_{up} (M=11079.85, SD=5615.823) significantly differ (p =0.000) from T_{down} (M=15794.69, SD=3699.818) scores,
- T_{down} (M=15794.69, SD=3699.818) significantly differ (p =0.000) from T_{idle} (M=3859.21, SD=3470.047) scores.

In summary, the results show significant effects of gender for the Time category, concerning, even in this case, the Copy of a house drawing (2), Writing in capital letters (3) and Writing a sentence in cursive letters (7) tasks: females perform fewer strokes in the air (T_{up}) than males in all the considered tasks.

V. Conclusion

The intent of the present study was to identify gender through online analysis of individuals' handwriting.

The experiment's results show that it is possible to distinguish between male and female writers investigating their performance while copying a house drawing (task 2), writing Italian words in capital letters (task 3) and writing of a phonetically complete Italian sentence in cursive letters (task 7), in particular focusing on Ductus and Time categories.

Considering the positive results achieved, it would be interesting to expand this analysis by enrolling subject of different ages and different levels of education.

Acknowledgment

The research leading to these results has received funding from the European Union Horizon 2020 research and innovation programme under grant agreement N. 769872 (EMPATHIC) and N. 823907 (MENHIR). Partial supports were received from the Italian project SIROBOTICS, MIUR, PNR 2015-2020, DD 1735 July 13 2017 and the Italian project ANDROIDS, Università della Campania "Luigi Vanvitelli" programma V:ALERE 2019, funded with D.R. 906 del 4/10/2019.

References

[1] Gavrilescu, M., & Vizireanu, N. (2018). Predicting the Big Five personality traits from handwriting. EURASIP Journal on Image and Video Processing, 2018(1), 57.

[2] Huang, B. Q., Zhang, Y. B., & Kechadi, M. T. (2007, October). Preprocessing techniques for online handwriting recognition. In Seventh International Conference on Intelligent Systems Design and Applications (ISDA 2007), pp. 793-800.

[3] Likforman-Sulem, L., Esposito, A., Faundez-Zanuy, M., Clémençon, S., and Cordasco, G. (2017). EMOTHAW: A novel database for emotional state recognition from handwriting and drawing. IEEE Transactions on Human-Machine Systems, 47(2), pp. 273-284.

[4] Siddiqi, I., Djeddi, C., Raza, A., & Souici-Meslati, L. (2015). Automatic analysis of handwriting for gender classification. Pattern Analysis and Applications, 18(4), pp. 887-899.

[5] Taleb, C., Khachab, M., Mokbel, C., & Likforman-Sulem, L. (2017, April). Feature selection for an improved Parkinson's disease identification based on handwriting. In 2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR), pp. 52-56.

[6] Upadhyay, S., Singh, J., & Shukla, S. K. (2017). Determination of Sex Through Handwriting Characteristics. Int. J. Cur. Res. Rev.| Vol, 9(13), 11.

[7] Esposito A., Amorese T., Buonanno M., Cuciniello M., Esposito A. M., Faundez-Zanuy M., Likforman-Sulem L., Riviello M. T., Troncone A., and Cordasco G. (2019). Handwriting and Drawing Features for Detecting Personality Traits. In 10th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Naples, Italy, pp. 79-84.

[8] Cordasco G., Scibelli F., Faundez-Zanuy M., Likforman-Sulem L., and Esposito A. (2017). Handwriting and Drawing Features for Detecting Negative Moods, Quantifying and Processing Biomedical and Behavioral Signals. WIRN'17. Smart Innovation, Systems and Technologies, Vol 103.

[9] Al Maadeed S. and Hassaine A. (2014). Automatic prediction of age, gender, and nationality in offline handwriting. EURASIP Journal on Image and Video Processing, 2014(1), 10.

[10] Jain A.K., Dass S.C., Nandakumar K. (2004). Soft Biometric Traits for Personal Recognition Systems. In: Zhang D., Jain A.K. (eds) Biometric Authentication. ICBA 2004. Lecture Notes in Computer Science, vol 3072. Springer, Berlin, Heidelberg.

[11] Patel J. and Bansal D. (2018). Effect of demographic variables on e-marketing strategies: A review. In International Journal of Academic Research and Development, Vol. 3(1), pp. 311-321.

[12] Bouadjenek N., Nemmour H. and Chibani Y. (2015). Histogram of Oriented Gradients for writer's gender, handedness and age prediction. International Symposium on Innovations in Intelligent SysTems and Applications (INISTA), Madrid, 2015, pp. 1-5.

[13] Li G. (2018). Gender-Related Differences in Credit Use and Credit Scores. FEDS Notes 2018-06-22, Board of Governors of the Federal Reserve System (U.S.).

[14] Huynh C., Brunelle E., Halámková L., Agudelo J., Halámek J. (2015). Forensic Identification of Gender from Fingerprints. Analytical chemistry. 87. 10.1021/acs.analchem.5b03323.

[15] Sahu M., Yadav A. and Isukapatla A. (2017). Examination of Handwriting for Gender Identifying Features. International Journal of Current Research. 9, pp.45923-45928.

[16] Sesa-Nogueras E., Faundez-Zanuy M. and Mekyska J. (2012). An information analysis of in-air and on-surface trajectories in online handwriting. Cognitive Comput., vol. 4, no. 2, pp. 195-205.

[17] Rosenblum S., Parush S. and Weiss P. (2003). The in-air phenomenon: temporal and spatial correlates of the handwriting process. Perceptual Motor Skills, vol. 96, pp. 933-954.

[18] Sokic E., Salihbegovic A. and Ahic-Djokic M. (2012). Analysis of off-line handwritten text samples of different gender using shape descriptors. In 2012 IX International Symposium on Telecommunications (BIHTEL), pp. 1-6).