How you type is what you type: Keystroke dynamics correlate with affective content

Héctor López-Carral*, Diogo Santos-Pata*, Riccardo Zucca* and Paul F.M.J. Verschure*†
*Synthetic, Perceptive, Emotive and Cognitive Systems (SPECS) Lab,
Institute for Bioengineering of Catalonia (IBEC),
Barcelona Institute of Science and Technology (BIST), Barcelona, Spain
†Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain
Email: pverschure@ibecbarcelona.eu

***Abstract*—Estimating the affective state of a user during a computer task traditionally relies on either subjective reports or analysis of physiological signals, facial expressions, and other measures. These methods have known limitations, can be intrusive and may require specialized equipment. An alternative would be employing a ubiquitous device of everyday use such as a standard keyboard. Here we investigate if we can infer the emotional state of a user by analyzing their typing patterns. To test this hypothesis, we asked 400 participants to caption a set of emotionally charged images taken from a standard database with known ratings of arousal and valence. We computed different keystroke pattern dynamics, including keystroke duration (dwell time) and latency (flight time). By computing the mean value of all of these features for each image, we found a statistically significant negative correlation between dwell times and valence, and between flight times and arousal. These results highlight the potential of using keystroke dynamics to estimate the affective state of a user in a non-obtrusive way and without the need for specialized devices.**

***Index Terms*—keystroke, keyboard, typing, arousal, valence, affect**

I. Introduction

Estimating the affective state of users while interacting with computers attracted much interest in recent years due to its potential for enhancing Human-Computer Interaction (HCI). This has been the focus of the affective computing field for the last few decades, with the expectation of having an impact in many fields that depend —increasingly so— in HCI, such as education, robotics, or human health, among others. Furthermore, progress in affective computing could also help to advance our knowledge of emotions and human cognition [1].

Typical experimental approaches for inferring affective states include subjective reports, in which users are repeatedly asked to describe or rate how they are feeling, or estimates from measures such as physiological responses, facial expressions or body gestures. Subjective reports are a traditional tool in a large portion of affective sciences. Multi-item scales [2] or pictorial tools [3] can be easily administered both in paper or digital formats. Meanwhile, estimations from bodily responses or expressions require devices that have to be either worn by the users or placed close to them. Examples of this include inferring emotional states by detecting changes in electrodermal activity (EDA) [4], [5] and heart rate variability (HRV) [6], [7]. Facial expression analysis typically requires a classification between a set of discrete basic emotions [8].

These methods, although valid in assessing affective states, present several issues. Subjective reports interrupt the regular user's workflow and present known limitations in terms of validity and reliability [9], [10]. In the case of bodily responses, different devices are required to be working alongside the computer, in many cases in direct physical contact with the user (e.g., electrodes placed on the skin), which can be intrusive to the user and expensive due to the economic cost of these devices.

A way to overcome these problems would be using components that are already available when interacting with a computer. Devices that do not require any unusual or particular action from the user, while still able to provide relevant correlates of internal states. Several studies have previously attempted to infer specific user's traits by analyzing the way a person types on a keyboard. The initial observation of unique typing rhythms across individuals [11] promoted in the last three decades great interest in the study of keystroke dynamics, particularly in the field of user authentication [12]. Several studies have succeeded in authenticating users with high accuracy based on a variety of classification algorithms [13], [14], suggesting that the relevant component in typing is not only the content typed but also how it is typed [15].

More recently, some studies have also provided evidence of the possibility of using keystroke dynamics to estimate emotional states [16]. In a recent field study, typing rhythms were coupled with periodic self-reports to classify between a series of discrete emotional states [17] with high accuracy. In a different study, keystroke patterns were successfully used to estimate the level of individual stress (induced by a mental arithmetic test), as self-reported in a pre- and post-questionnaire, as well as heart rate variability [18].

To the best of the authors' knowledge, no previous study has investigated the possibility of adopting keystroke dynamics to discriminate the affective features of presented stimuli that participants are describing. This would imply that the emotional content of these stimuli is actively affecting the participant's way of typing in subtle ways that could be detected through the analysis of their typing patterns. In this study, we show that this is indeed a possibility. To do this,

978-1-7281-3888-6/19/$31.00 ©2019 IEEE

we asked 400 participants to caption a set of images rated in terms of continuous values of arousal and valence, instead of discrete emotions. The arousal and valence dimensions of affect follow the circumplex model of affect [19], which is generally used in experiments related to affect induction and detection [20].

II. Methods

We performed an online study where participants ($N = 400$) were asked to observe a series of 46 images and to type a description for each one of them.

A. Participants

Participants were recruited using the Amazon Mechanical Turk (MTurk) service. To guarantee high-quality responses, we restricted participation to volunteers with at least 50 tasks previously completed and an approval rate of over 90 % and with a proficient level of English. No personal information from the participants (such as names or IP addresses) was recorded. As expected from this pool of participants, there was considerable variability in demographic backgrounds, with a mean age of 37.51 ($SD = 12.17$). 60.4 % of participants were male. 47.11 % reported spending over 5 hours per day typing on a keyboard, 38.35 % between 3 and 5 hours, and 14.54 % fewer hours.

B. Experimental Protocol

Before starting the task, a page informed the participants about the experimental protocol and their right to recess at any moment. Subsequently, they were required to fill a set of questions including demographics (age, gender, education level, and primary language), keyboard experience (number of hours per day spent typing) and keyboard layout used. After this, the main task started.

The main task consisted of visualizing a sequence of 46 images and providing a description of each one. In each trial, an image was presented for 2 seconds, and it was followed by a text field in which the participants were instructed to type a free description of the image seen using a minimum of 4 words.

The images were selected from the Open Affective Standardized Image Set (OASIS) [21], a set of 900 open-access images available for online use with normative ratings of arousal and valence. From the whole set, 46 images were extracted to cover the entire range of arousal and valence ratings. To select the images, we first binned the dataset into a 7 by 7 matrix along the two dimensions of valence and arousal (OASIS ratings are expressed using a 7-point Likert scale), and we selected two images at random from the 23 bins which contained at least 2 images. Thus, 46 images were finally selected (see Fig. 1 which were randomly presented to every participant.

C. Data Collection and Processing

During the task session, all keyboard events performed on the provided text field have been logged. Specifically, this

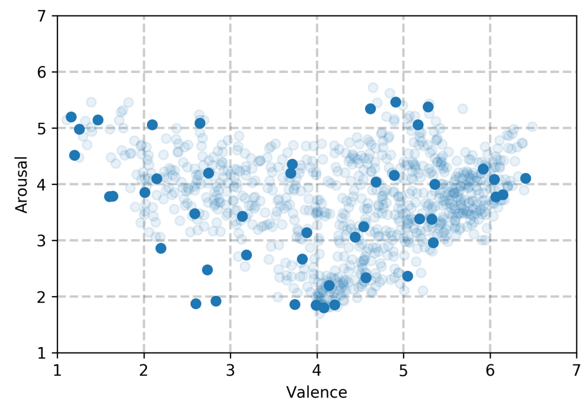

Fig. 1. The selection process of the 46 images included in the experimental dataset. Two images were extracted at random from each of the 23 non-empty bins in which the 900 OASIS image set was previously divided [21]. This selection covers the entire range of arousal and valence. Dark circles represent the selected images, while semitransparent ones represent the rest of the images in the original set.

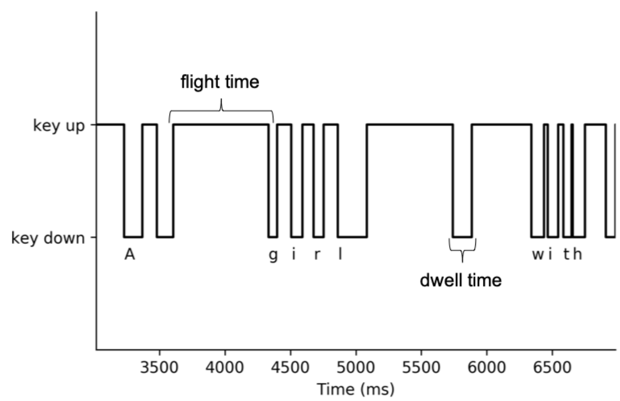

Fig. 2. Visualization of keystrokes during an example image description. An example of the graphical representation of one instance of flight time and dwell time is provided in the annotations.

includes both key-down (or key-press) and key-up (or key-release) events (see Fig. 2). For each event, three pieces of information were stored: the type of event (key up or down), the key that was pressed, and the timestamp (in milliseconds). We also recorded the time in which each image and the text field were presented to the participants, as well as when they clicked to proceed to the next image. All these data were collected in a file for each participant in JSON format. The mean duration of the experiment was 14.52 minutes, and the mean length of the image descriptions was 5.93 words.

After collection, data were inspected to ensure that there were no instances of image descriptions that were copied and pasted, missing data due to connectivity issues or other problems that might affect the integrity of the acquired data. To do this, we reconstructed each final image descriptions provided by the participants from the individual keystroke events.

For each image description provided by the participants, we derived a series of features. Two standard features in keystroke dynamics are the duration and the latency [14], [15], [17]. Keystroke duration (also known as *dwell time*) represents the time that a single key was pressed in an instance (time since key-down until key-up). Keystroke latency (or *flight time*) represents the elapsed time between two sequential key presses (time since key-up until next key-down). Additionally, we computed the number of error corrections (presses of the backspace key), the total time to write each description, the time since the presentation of the text field until the participants started typing, and the time since the participants finished typing until they pressed the button to continue. See Table I for a summary.

TABLE I
FEATURES EXTRACTED FROM PARTICIPANT'S BEHAVIOR WHILE TYPING

Feature	**Description**
Dwell time	Keystroke duration, time a key was pressed
Flight time	Keystroke latency, time between two key presses
Backspace count	Number of error corrections
Time total	Time since text field was available until submission
Time start	Time to start typing since text field appears
Time end	Time since finishing typing until submission

For each of the derived features, we computed the Spearman's rank correlation coefficient between each feature and the reported arousal and valence ratings.

III. Results

In order to assess how affective properties of perceived content modulate typing behavior, we tested the interplay between the previously described features (see Table I) and the affective ratings of the stimuli as provided in the image set.

We first asked how valence affects the typical typing behavior of individuals. To do so, we extracted each participant average flight times and grouped participant's scores per valence binned category. Statistical testing revealed a significant negative correlation between valence and typing flight times, with shorter flight times consistently associated to high valence ratings (Fig. 3, top-left, Spearman-test $r = -1.0$, $p = 0.0$). A similar analysis was then performed for each participant average of dwell times and again grouped the individuals' scores per valence binned category. Statistical testing revealed a significant interplay between valence and typing dwell times (Fig. 3, top-middle, Spearman-test $r = -1.0$, $p = 0.0$).

Next, we examined whether the content arousal score would reveal similar effects in the typing behavior of individuals.

As for valence analysis, we extracted each participant average flight times and grouped the individuals' scores per arousal binned category. Statistical testing revealed a significant interplay between arousal and typing flight times (Fig. 3, bottom-left, Spearman-test $r = -1.0$, $p = 0.0$). Similarly, we analyzed each participant average of dwell times and the individuals' scores per arousal binned category. Statistical testing revealed

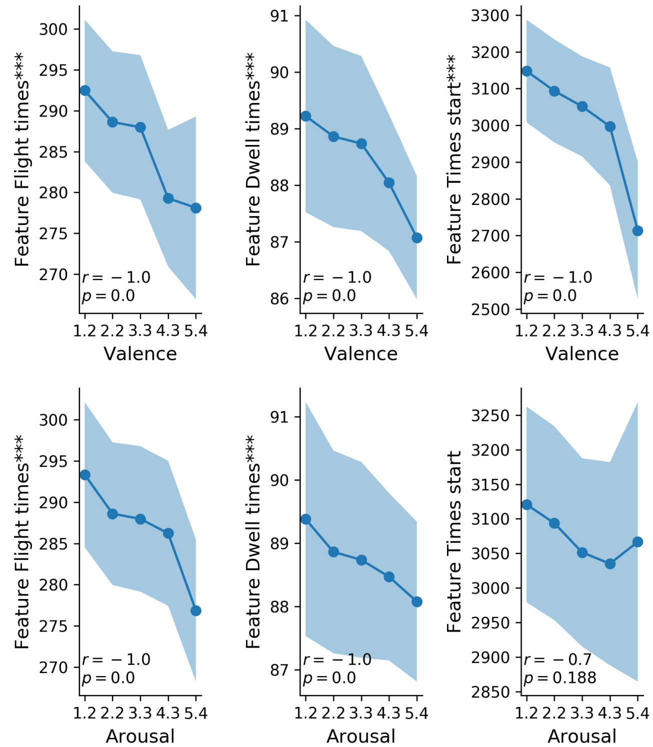

Fig. 3. Correlation analysis of each of the extracted features with valence and arousal considering bins of these affective ratings. The three features of interest are shown. We can observe highly significant correlations between both flight times and dwell times with both arousal and valence. Additionally, there is an highly significant correlation between the time to start typing since the text field was presented (Times start) and valence. There is no significant correlation between the start time and arousal ($r = -0.7$, $p = 0.188$). Times are expressed in milliseconds.

a significant interplay between arousal and typing dwell times (Fig. 3, bottom-middle, Spearman-test $r = -1.0$, $p = 0.0$).

So far we have reported how affective features modulate the typing behavior of individuals. However, in a natural environment, the valence rate of a stimulus does affect how quickly humans, and other animals, react to that same stimulus. To test whether, in a goal-oriented typing task, the onset of typing could be predicted by the affective content upon which participants were reporting, we grouped the individuals time to start typing scores accordingly to the stimuli affective rate. We observed a significant negative correlation between the onset of typing (initiation of behavior) and valence rate (Fig. 3, top-right, Spearman-test $r = -1.0$, $p = 0.0$).

These metrics capture how much individuals modulate their typing profile based on the perceived and reported content. Next, we asked whether the content alone could generalize the participants typing behavior. To do so, we extracted the values of each feature individually for each image calculating the participant's population mean of the different features per image. We found a significant correlation between dwell times and valence ($r = -0.293$, $p = 0.0479$), as well as a highly significant correlation between flight times and arousal ($r =$

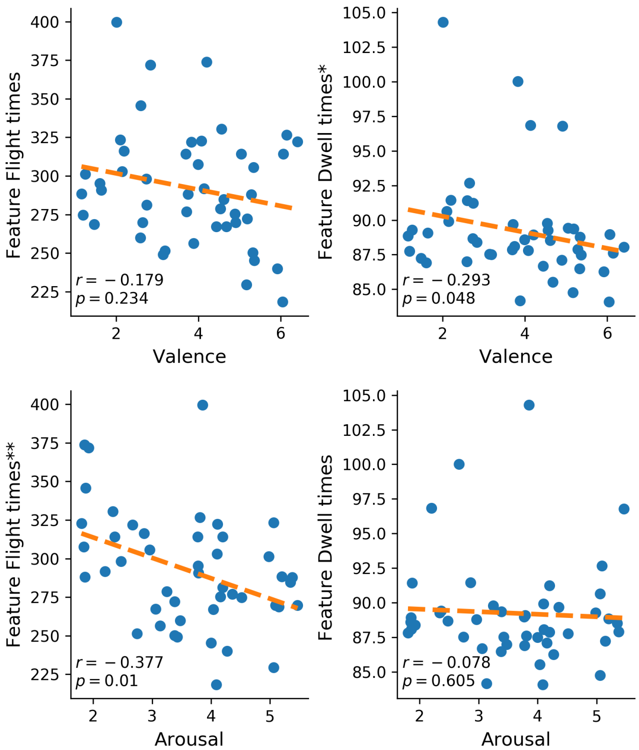

Fig. 4. Correlation analysis of the mean value of the extracted features for each image with valence and arousal. Only the two features of interest are shown. We can observe a significant correlation between dwell times and valence, and between flight times and arousal. Times are expressed in milliseconds.

-0.377, $p = 0.009$) (see Fig. 4). Therefore, suggesting an overall modulating effect between content and typing behavior.

IV. Discussion and Conclusion

Being able to reliably estimate the affective state of a user while interacting with a computing device would significantly improve the interaction process, with machines that can be more reactive and adaptive. Such capability could be beneficial for diverse fields such as education, human-robot interaction, digital health, and others [1].

However, typical approaches for inferring these emotional traits rely on either subjective reports (e.g., [3]) or on the usage of specialized equipment (e.g., [4]), which can be intrusive and expensive. Therefore, a way to overcome these issues would be to use an automatic approach that would take advantage of a device that users would typically use, without interfering with their behavior.

A possibility for this is to use keystroke dynamics computed from the typing patterns of users on regular keyboards. Such an approach has been explored mainly in the field of digital authentication [15], and only recently it has been extended to the field of affective computing with promising results. However, current research within this field has relied still on subjective reports to validate the estimations or classifies emotional states within discrete states [17].

In this study, we analyzed the typing patterns of a large sample of participants that were asked to describe a set of images selected from the OASIS normative database for affective research [21].

We processed the recorded data in order to extract a series of keystroke features, including keystroke latency and duration, and timings, for each participant and each image. Analyzing the keystroke dynamics of the participants, we found highly significant negative correlations of both flight times and dwell times with both arousal and valence, as well as between time to start and valence. We then checked for generalization on the content itself, finding significant negative correlations between dwell times and valence, and between flight times and arousal.

These results show that keystroke dynamics do indeed correlate with both arousal and valence. Therefore, it could be possible to infer affective states from keyboard activity.

Furthermore, we achieved these results by merely exposing participants to emotionally charged images (for 2 seconds each) and asking them to describe them, without using any subjective report, thanks to the fact that the images were already rated. Each participant rated only 46 images, which shows that not a lot of typing information is required from an individual user.

Although we have found significant results using a limited amount of keystroke features, it is possible that we could have obtained relevant results by using more sophisticated features such as digraphs (combinations of two letters) or trigraphs (combinations of three letters) [17], [22], or by using a simultaneous combination of multiple features.

Our results highlight the potential for a more in-depth analysis. This could include sentiment analysis on the descriptions written by the participants, in order to test the correlation between those results and both the affective ratings provided in the image set and the keystroke features we computed. Furthermore, machine learning techniques could be employed to train a model capable of determining the affective state during the typing of a sentence by using the described keystroke features.

In conclusion, this study reveals the correlation between keystroke dynamics and affective content by using descriptions of images from a rated set. This showcases the potential of using keyboard activity in order to infer affective states, either in addition to other techniques (such as physiological signals) or as a replacement when they are not possible, with the benefit of being an unobtrusive and inexpensive method.

Acknowledgment

This research has been supported by the European Commission under contract H2020-787061 ANITA and by EIT Health under grant 19277 RGS@home.

References

[1] R. W. Picard, “Affective Computing,” MIT Media Laboratory, Tech. Rep., 1995.

[2] D. Watson, L. A. Clark, and A. Tellegen, “Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales,” *Journal of Personality and Social Psychology*, vol. 54, no. 6, pp. 1063–1070, 1988.

[3] M. M. Bradley and P. J. Lang, “Measuring emotion: The self-assessment manikin and the semantic differential,” *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.
[4] H. D. Critchley, “Review: Electrodermal Responses: What Happens in the Brain,” *The Neuroscientist*, vol. 8, no. 2, pp. 132–142, 2002. [Online]. Available: http://journals.sagepub.com/doi/10.1177/107385840200800209
[5] P. J. Lang, M. K. Greenwald, M. M. Bradley, and A. O. Hamm, “Looking at pictures: Affective, facial, visceral, and behavioral reactions,” *Psychophysiology*, vol. 30, no. 3, pp. 261–273, 1993.
[6] F. Agrafioti, D. Hatzinakos, and A. K. Anderson, “ECG pattern analysis for emotion detection,” *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 102–115, 2012.
[7] J. Selvaraj, M. Murugappan, K. Wan, and S. Yaacob, “Classification of emotional states from electrocardiogram signals: A non-linear approach based on hurst,” *BioMedical Engineering Online*, vol. 12, no. 1, 2013.
[8] B. Fasel and J. Luettin, “Automatic facial expression analysis: A survey,” *Pattern Recognition*, vol. 36, no. 1, pp. 259–275, 2003.
[9] A. A. Stone, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, and V. S. Cain, *The science of self-report: Implications for research and practice.* Psychology Press, 1999.
[10] K. V. Keefer, “Self-Report Assessments of Emotional Competencies: A Critical Look at Methods and Meanings,” *Journal of Psychoeducational Assessment*, 2015.
[11] R. Joyce and G. Gupta, “Identity authentication based on keystroke latencies,” *Communications of the ACM*, vol. 33, no. 2, pp. 168–176, 1990.
[12] P. S. Teh, A. B. J. Teoh, and S. Yue, “A Survey of Keystroke Dynamics Biometrics,” *The Scientific World Journal*, vol. 2013, pp. 1–24, nov 2013.
[13] F. Monrose and A. Rubin, “Authentication via keystroke dynamics,” in *Proceedings of the 4th ACM conference on Computer and communications security*. Association for Computing Machinery (ACM), feb 1997, pp. 48–56.
[14] F. Bergadano, D. Gunetti, and C. Picardi, “User authentication through keystroke dynamics,” *ACM Transactions on Information and System Security*, vol. 5, no. 4, pp. 367–397, nov 2003.
[15] F. Monrose and A. D. Rubin, “Keystroke dynamics as a biometric for authentication,” *Future Generation Computer Systems*, vol. 16, no. 4, pp. 351–359, 2000.
[16] A. Kolakowska, “A review of emotion recognition methods based on keystroke dynamics and mouse movements,” in *2013 6th International Conference on Human System Interactions (HSI).*, Sopot, Poland, 2013, pp. 548–555.
[17] C. Epp, M. Lippold, and R. L. Mandryk, “Identifying emotional states using keystroke dynamics,” in *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI ’11*, 2011, p. 715. [Online]. Available: http://dl.acm.org/citation.cfm?doid=1978942.1979046
[18] S. D. Gunawardhane, P. M. De Silva, D. S. Kulathunga, and S. M. Arunatileka, “Non invasive human stress detection using key stroke dynamics and pattern variations,” in *International Conference on Advances in ICT for Emerging Regions, ICTer 2013 - Conference Proceedings*. IEEE, 2013, pp. 240–247.
[19] J. Russell, “Affective space is bipolar,” *Journal of Personality and Social Psychology*, vol. 37, no. 3, pp. 345–356, 1979.
[20] H. Gunes, B. Schuller, M. Pantic, and R. Cowie, “Emotion representation, analysis and synthesis in continuous space: A survey,” *2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, FG 2011*, pp. 827–834, 2011.
[21] B. Kurdi, S. Lozano, and M. R. Banaji, “Introducing the Open Affective Standardized Image Set (OASIS),” *Behavior Research Methods*, vol. 49, no. 2, pp. 457–470, 2017.
[22] P. S. Dowland and S. M. Furnell, “A Long-Term Trial of Keystroke Profiling Using Digraph, Trigraph and Keyword Latencies,” in *Security and Protection in Information Processing Systems*, 2005, pp. 275–289.