Reaction Times and Performance Variability in Touch and Desktop Users During a Stroop Task

Gianluca Merlo(✉), Luna Jaforte, and Davide Taibi

Istituto per le Tecnologie Didattiche, Consiglio Nazionale delle Ricerche, 90146 Palermo, PA, Italy
gianluca.merlo@itd.cnr.it

Abstract. Reaction times (RTs) are a measure of the time elapsed between the sensory stimulation and the occurrence of a response. It depends upon many factors. This paper focuses on one of the possible sources of variability introduced by the device used to gather RTs data and compares RTs performance between subjects who completed a Stroop task through desktop or mobile devices. The research hypothesizes that mobile devices users' (a) have faster RTs in incongruent trials than the desktop ones, (b) have a lower error rate, and (c) perceive the task easier. The results showed significant differences in RTs, error rate, and perception of the difficulty of the task between devices but only the last two hypotheses are supported by data. Subjects under 40 years achieved a lower error rate than older people in the incongruent and neutral trials. Moreover, participants always perceived the task easier when they complete it through mobile devices. Some implications of the results are discussed, and a second ongoing within-subjects study is described.

Keywords: Reaction time · Mobile device · Stroop task

1 Introduction

Reaction times (RTs) have been used as a psychological task since the mid-19th century [1] and it is also called response time or response latency. RTs are a simple, and presumably the most widely used, measure of behavioral response in time units (usually in milliseconds), from a presentation of a given task to its completion. There are two typical and functional procedures to estimate response latency: choice reaction time and simple reaction time. The first one is more complicated because it requires the subject to make the appropriate response to one of some stimuli. The second one concerns making a response as fast as possible in reaction to a single stimulus [1].

According to Kosinski [2], many factors impact RTs in addition to the type of stimulus, the type of RT experiment, and stimulus intensity. For example, age, gender, arousal, left/right hand, direct/peripheral vision [2, 3] have a direct influence on RTs.

Previous studies showed that RT latencies and variability increase and become more variable with age [4–9]. RTs tend to slow down with age starting from the 20s for choice

C. Stephanidis et al. (Eds.): HCII 2022, CCIS 1581, pp. 186–193, 2022.
https://doi.org/10.1007/978-3-031-06388-6_25

tasks and 30s for simple [10] but, in general, the decline accelerates after the age of about 70 [11, 12]. Rabbitt, Osman, Moore, & Stollery [13] proposed that the increase in mean RTs with age is a result of increasing variability, and variability is, therefore, a more important component of cognitive aging.

The present research hypothesizes that another possible source of variability is introduced by the device used to gather RTs data. Many studies revealed that touch devices are faster [14] and more natural and convenient [15] than desktop computers. Moreover, users usually prefer touch-input devices because of a coordination factor. Touch input implicates eyes-hand direct coordination, unlike mouse which requires a greater effort of coordination: "the input is also the output device" [16]. The ordinary use of touchscreen is also associated with the cortical reorganization of the fingertips in the somatosensory cortex [17].

According to the above-mentioned theoretical background, this paper describes a research aimed to compare RTs performance between subjects who completed a Stroop task [18], one of the most widely used measures of cognitive functioning [19], through desktop or mobile devices. The Stroop task requires participants to name the color of ink in which a color name is printed while ignoring the written word. Indicating the color with which the congruent color-word stimuli were written takes less time than indicating the color of the incongruent color-word stimuli. Increased RTs during incongruent trials have been attributed to the interfering effects of the printed color name on the different colored ink.

The study hypothesizes that mobile devices users' (a) have faster RTs in incongruent trials than the desktop ones, (b) have a lower error rate, and, (c) perceive the task easier.

2 Methods

Data were collected from June 2020 to June 2021. Participants (general population aged 18 years or older) were contacted through the authors' network and posts on the most popular social networks. The compilation instructions informed the participants that the answers were anonymous and that there were no right or wrong answers. Informed consent was obtained from all subjects involved in the study. A privacy policy at the beginning of the questionnaire described to the participants the purposes and methods of treatment operated by the data controller on the personal data collected. The research was submitted to the ethics committee of the University Polyclinic "Paolo Giaccone" of Palermo and obtained a favorable opinion for its development.

2.1 Participants

The full sample comprised 397 subjects, with an overall mean of 41.24 years (SD = 11.85). Their age ranged from 18 to 74 years. Among all participants, 304 (76.57%) were female, 92 (23.17%) were male, and 1 (0.25%) other.

68.26% of the subject participated in the research through a mobile phone, while the rest (31.74%) did it through a desktop computer. A Wilcoxon signed-rank test showed that these groups did not differ in age ($Z = 19{,}062.50$, $p = 0.06$) or the proportion of male and female X^2 $(2, N = 397) = 4.37$, $p = 0.10$.

2.2 Stroop Task

A web-based computerized version of the Stroop color-naming task has been developed for the study. The Stroop task was coded using HTML5 and JavaScript programming. In particular, the web-based Stroop Task implemented the High-Resolution Time API that provides the time in sub-millisecond resolution, such that it is not subject to system clock skew or adjustments [20]. In order to make the Stroop Task usable from both desktop computers and mobile devices, the Bootstrap responsive frontend framework [21] has been used to design the task templates.

The Stroop Task included 21 training trials and 28 experimental trials for each type: congruent, incongruent, and neutral (displayed as a colored horizontal rectangle). Participants were instructed to focus their attention on a fixation cross at the center of the screen. Then, after 1000 ms, stimuli were presented against a white background. For all trial types, participants responded by choosing the color (red, green, blue, or yellow) of the stimulus by typing on the keyboard the letter corresponding to the color or touching a box matching the color displayed on the screen. Letters (only for desktop computers) and colors were displayed in a bar at the bottom of the screen (Fig. 1).

Fig. 1. Stroop Task interfaces for desktop and mobile devices

At the end of the Stroop task, a question about the difficulty was asked to participants with a response mode from 0 to 100.

2.3 Data Analysis

Starting from the results obtained to the Stroop task, mean reaction time and error rates for congruent, incongruent, and neutral trial types were calculated. The training trials are not included in the analysis. For statistical analyses, a logarithm transformation was applied to data to better approximate a normal distribution. The dependent variables of the study were the reaction time latency, the error rate, and, according to the work of Cain and colleagues [22], indexes of interference (mean RTs of incongruent trials - mean RTs of neutral trials) and facilitation (mean RTs neutral trials - mean RTs congruent trials). While the interference index refers to the additional time needed to respond when reading the word interferes with responding to the color, the facilitation index is a measure of the increased speed of reaction when color and word matched. The independent variables used in the statistical models were: age group (18–40, >40), device (desktop, mobile), and trial typology (congruent, incongruent, neutral). The age groups were chosen applying a median split transformation and this is consistent with the research of Deary and Der [6] who demonstrated that RTs tend to slow down with age 30s for simple tasks and that intraindividual variability in choice reaction time increases steadily from the mid-30s to the mid-60s. The post hoc comparisons showing significant effects related to the variability of RTs related to age and task are not reported because they are well documented in the literature. The interquartile range method was applied to detect and remove the outliers. All statistical analyses were performed with R software.

3 Results

3.1 Latency and Task Per Age Group

The distribution of response time violated both the assumption of normality and homogeneity of variances. A Shapiro-Wilk test was performed and showed that the distribution of response times departed significantly from normality ($W = 1.00$, p-value < 0.01). To analyze data according to a Task(3) × Age group (2) × Device (2) design, first, the Aligned Rank Transform (ART) procedure [23–25] was applied and then a factorial ANOVA based on the aligned-and-ranked responses was calculated. The ART techniques for analysis of variance are considered a powerful and robust nonparametric alternative when the assumption of a normal error distribution is violated [26]. The ART procedure and the post hoc pairwise comparisons were calculated with the support of the ARTool R package [27, 28]. The three-way interaction between age, task, and device was not significant, ($F_2 = 0.06$, $p > 0.05$). A significant interaction between the age group and the device used was found ($F_1 = 37.31$, $p < 0.001$). Significant main effect on reaction times for age group ($F_1 = 144.81$, $p < 0.001$), task ($F_2 = 39.60$, $p < 0.001$), and the device used to complete the Stroop task ($F_1 = 1.94$, $p < 0.05$) were found. The ART-C post hoc pairwise comparisons, corrected with Holm's sequential Bonferroni procedure, indicates that users under 40 years were faster in incongruent tasks when they completed the Stroop task via desktop devices ($p < 0.001$).

3.2 Error Rate and Task Per Age Group

The error rate distribution significantly deviates from a normal distribution ($W = 0.76$, p-value < 0.01). Applying the already mentioned ART procedure, a Task (3) × Age group (2) × Device (2) design was analyzed through a factorial ANOVA. A three-way interaction was identified between task, age group, and device ($F_2 = 3.30$, $p < 0.05$).

Statistically significant interactions between task and device used ($F_2 = 11.30$, $p < 0.001$), and age group and task ($F_2 = 11.86$, $p < 0.001$) were found. Simple main effects analysis showed that age group ($F_1 = 24.75$, $p < 0.001$), task ($F_2 = 43.44$, $p < 0.001$), and the device used have a statistically significant effect on error rate ($F_1 = 30.27$, $p < 0.001$). Post hoc pairwise comparisons, applied with the support of the ART-C align-and-rank procedure and corrected with the Holm method, revealed that subjects under 40 years have a lower error rate in the incongruent ($p < 0.001$) and neutral ($p < 0.01$) trials when they did the Stroop task from mobile.

3.3 Interference and Facilitation Indexes

Facilitation and interference indexes violated the assumption of normality ($W = 0.96$, p-value < 0.01; $W = 0.98$, p-value < 0.01). As a consequence, the ART procedure was applied to the analyzed variables. A two-way ANOVA revealed that there was not a statistically significant interaction between age group and device both for facilitation ($F_1 = 0.96$, $p > 0.05$) and interference ($F_1 = 0.09$, $p > 0.05$) indexes. Simple main effects analysis showed that the device used did have a statistically significant effect on both facilitation ($F_1 = 5.83$, $p < 0.05$) and interference indexes ($F_1 = 16.41$, $p < 0.001$). The age group had a significant main effect only on the interference index ($F_1 = 5.91$, $p < 0.05$). Post hoc pairwise comparisons, applied with the support of the ART-C align-and-rank procedure and corrected with the Holm method, revealed that the index of interference is greater in mobile device users in both age groups ($p < 0.05$).

3.4 Difficulty

The perceived difficulty distribution is positively skewed ($W = 0.94$, p-value < 0.01). After having applied the ART procedure on the analyzed variables, the interaction between the device used and age group was not significant ($F_1 = 2.21$, $p > 0.05$) but a simple main effect of the device used was found ($F_1 = 25.36$, $p < 0.001$). Post hoc pairwise comparisons revealed that mobile users perceived the task easier in both the analyzed age groups.

4 Conclusion

The study explored the variability of reaction times, error rates, and perception of difficulty comparing users who performed a Stroop task via desktop or mobile devices. The authors' hypotheses were only partially supported by the data. Although not always statistically significant, young users (<40 years) tend to be faster from desktop computers than from mobile devices, especially in incongruent tasks. This tendency is identifiable, also, in the interference index, higher in the just mentioned group of users. This result needs further investigation, and, for this reason, a second within-subject study is in progress. The within-subjects design has been chosen to expose every subject to both the mobile and desktop conditions and to control the possible bias introduced by the JavaScript library developed for the data gathering. Secondly, the second study is collecting data about the daily use and experience that subjects have with the devices.

Familiarity and ability to use the device probably influence the speed and accuracy of the task. The finding that young people are slower during incongruent tasks with mobile devices could also be due to diminished attentional capacity among heavy users of mobile devices [29]. For example, Leiva and colleagues reported that when the user's thoughts drift toward a smartphone-related activity, task completion in one app can be delayed by up to 400% when these endogenous interruptions appear [30].

In the study presented in this paper, as predicted, users made fewer errors when they completed the task via mobile devices. Age, task typology, and the device used affect the number of errors made during the color naming tasks. In particular, subjects under 40 years showed a significantly lower error rate in the incongruent and neutral trials when they did the Stroop task from mobile. A clear difference is also found in the perception of the task difficulty: users perceived the task easier when completed through mobile devices. The perception that tasks based on reaction times are easier when performed from mobile devices may increase their uptake as a cognitive screening tool for older people. RTs are widely used to evaluate cognitive performance in general [31] and even life expectancy [10], and intraindividual variability of reaction times in cognitive aging is a predictor of dementia [32] and chronic fatigue syndrome [33].

Finally, the second study will have to try to balance the proportion of males and females which is unbalanced in the present study due to convenience sampling.

References

1. Deary, I.J., Liewald, D., Nissan, J.: A free, easy-to-use, computer-based simple and four-choice reaction time programme: the deary-liewald reaction time task. Behav. Res. Methods **43**, 258–268 (2011)
2. Kosinski, R.J.: A literature review on reaction time. Clemson Univ. **10**, 337–344 (2008)
3. Chundakal, D., Kulkarni, P., Chavan, S., Prabhakar, R.: Effect of music therapy and frenkel exercise on reaction time in geriatric population-a comparative study. Ed. Adv Board. **15**, 17 (2021)
4. Lindenberger, U., Mayr, U., Kliegl, R.: Speed and intelligence in old age. Psychol. Aging **8**, 207 (1993)
5. Verhaeghen, P., Salthouse, T.A.: Meta-analyses of age–cognition relations in adulthood: estimates of linear and nonlinear age effects and structural models. Psychol. Bull. **122**, 231 (1997)
6. Ratcliff, R., Thapar, A., McKoon, G.: The effects of aging on reaction time in a signal detection task. Psychol. Aging **16**, 323 (2001)
7. Jevas, S., Yan, J.H.: The effect of aging on cognitive function: a preliminary quantitative review. Res. Q. Exerc. Sport **72**, 38–40 (2001)
8. Hultsch, D.F., MacDonald, S.W., Dixon, R.A.: Variability in reaction time performance of younger and older adults. J. Gerontol. B Psychol. Sci. Soc. Sci. **57**, P101–P115 (2002)
9. Der, G., Deary, I.J.: Age and sex differences in reaction time in adulthood: results from the United Kingdom health and lifestyle survey. Psychol. Aging **21**, 62 (2006)
10. Deary, I.J., Der, G.: Reaction time, age, and cognitive ability: longitudinal findings from age 16 to 63 years in representative population samples. Aging Neuropsychol. Cogn. **12**, 187–215 (2005)
11. Era, P., Sainio, P., Koskinen, S., Ohlgren, J., Härkänen, T., Aromaa, A.: Psychomotor speed in a random sample of 7979 subjects aged 30 years and over. Aging Clin. Exp. Res. **23**(2), 135–144 (2011). https://doi.org/10.1007/BF03351077

12. Pierson, W.R., Montoye, H.J.: Movement time, reaction time and age. J. Gerontol. **13**, 418–421 (1958)
13. Rabbitt, P., Osman, P., Moore, B., Stollery, B.: There are stable individual differences in performance variability, both from moment to moment and from day to day. Q. J. Exp. Psychol. Sect. A. **54**, 981–1003 (2001)
14. Zabramski, S., Shrestha, S., Stuerzlinger, W.: Easy vs. tricky: the shape effect in tracing, selecting, and steering with mouse, stylus, and touch. In: Proceedings of International Conference on Making Sense of Converging Media, pp. 99–103 (2013)
15. Forlines, C., Wigdor, D., Shen, C., Balakrishnan, R.: Direct-touch vs. mouse input for tabletop displays. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 647–656 (2007)
16. Stark, J., Burwitz, M., Braun, R., Esswein, W.: Cognitive efficient modelling using tablets. In: Enterprise Modelling and Information Systems Architectures (EMISA 2013) (2013)
17. Gindrat, A.-D., Chytiris, M., Balerna, M., Rouiller, E.M., Ghosh, A.: Use-dependent cortical processing from fingertips in touchscreen phone users. Curr. Biol. **25**, 109–116 (2015)
18. Stroop, J.R.: Studies of interference in serial verbal reactions. J. Exp. Psychol. **18**, 643 (1935)
19. MacLeod, C.M.: Half a century of research on the stroop effect: an integrative review. Psychol. Bull. **109**, 163 (1991)
20. Grigorik, I., Simonsen, J., Mann, J.: High resolution time level 2. https://www.w3.org/TR/hr-time-2/. Accessed 07 Aug 2021
21. Otto, J.M., Thornton, J., Rebert, C., Xhmikos, R., Lauke, P.H., Mazovetskiy, G.S.J.: Build fast, responsive sites with bootstrap. https://getbootstrap.com/. Accessed 07 Aug 2021
22. Cain, S.W., Silva, E.J., Chang, A.-M., Ronda, J.M., Duffy, J.F.: One night of sleep deprivation affects reaction time, but not interference or facilitation in a stroop task. Brain Cogn. **76**, 37–42 (2011)
23. Fawcett, R., Salter, K.: A monte carlo study of the f test and three tests based on ranks of treatment effects in randomized block designs. Commun. Statist. Simul. Comput. **13**, 213–225 (1984)
24. Higgins, J.J., Tashtoush, S.: An aligned rank transform test for interaction. Nonlinear World **1**, 201–211 (1994)
25. Salter, K., Fawcett, R.: The ART test of interaction: a robust and powerful rank test of interaction in factorial models. Commun. Statist. Simul. Comput. **22**, 137–153 (1993)
26. Mansouri, H., Paige, R.L., Surles, J.G.: Aligned rank transform techniques for analysis of variance and multiple comparisons. Commun. Statist. Theory Methods **33**, 2217–2232 (2004)
27. Wobbrock, J.O., Findlater, L., Gergle, D., Higgins, J.J.: The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 143–146 (2011)
28. Elkin, L.A., Kay, M., Higgins, J.J., Wobbrock, J.O.: An aligned rank transform procedure for multifactor contrast tests. In: The 34th Annual ACM Symposium on User Interface Software and Technology, pp. 754–768 (2021)
29. Egan, T.: The eight-second attention span. The New York Times, 22 (2016)
30. Leiva, L., Böhmer, M., Gehring, S., Krüger, A.: Back to the app: the costs of mobile application interruptions. In: Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 291–294 (2012)
31. Collardeau, M., Brisswalter, J., Audiffren, M.: Effects of a prolonged run on simple reaction time of well trained runners. Percept. Mot. Skills **93**, 679–689 (2001)

32. Hultsch, D.F., MacDonald, S.W., Hunter, M.A., Levy-Bencheton, J., Strauss, E.: Intraindividual variability in cognitive performance in older adults: comparison of adults with mild dementia, adults with arthritis, and healthy adults. Neuropsychology **14**, 588 (2000)
33. Fuentes, K., Hunter, M.A., Strauss, E., Hultsch, D.F.: Intraindividual variability in cognitive performance in persons with chronic fatigue syndrome. Clin. Neuropsychol. **15**, 210–227 (2001)