The 6th IEEE International Conference on E-Health and Bioengineering - EHB 2017
Grigore T. Popa University of Medicine and Pharmacy, Sinaia, Romania, June 22-24, 2017

Neurobiological and Neurocognitive Models of Vision for Touch Input on Mobile Devices

Maria Doina Schipor, Radu-Daniel Vatavu
MintViz Lab | MANSiD Research Center, University Ștefan cel Mare of Suceava
Suceava 720229, Romania, vmdoina@yahoo.com, vatavu@eed.usv.ro

Abstract **Touch input requires precise eye-hand coordination skills to be effective and efficient. In this work, we address touch input from the *new perspective of understanding human vision and eye-hand coordination mechanisms in the fields of neurobiology and psychology*, and we highlight implications of neurobiological and neurocognitive approaches of human vision for touch target acquisition on mobile devices. We draw relationships between several models of human vision and practical aspects of designing the visual appearance of touch targets, such as target size, shape, location, padding, and animation speed. Our work presents the interested community with relevant knowledge about the eye-hand coordination mechanisms that are responsible for effective and efficient touch input. Practitioners can employ our results to inform touch target design, and researchers can use our investigation as a starting point for new empirical explorations toward a better understanding of touch input on mobile devices.**

Keywords: touch input; visual processing; neurocognitive models; neurobiological models; mobile devices.

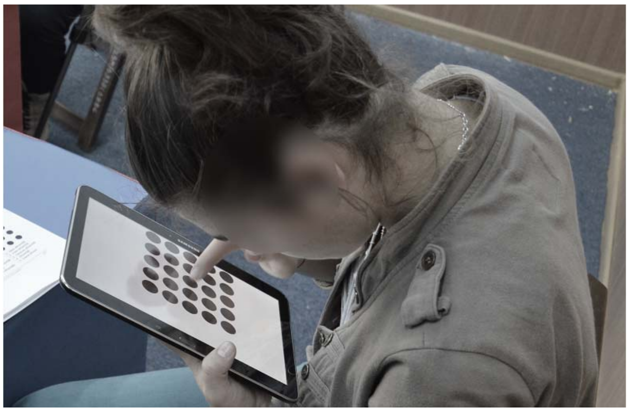

Fig 1. A person with low vision performing a complex touch input task on a mobile device: selecting a target in a dense context of many surrounding targets. As touch input requires precise eye-hand coordination continuously guided by visual stimuli, ineffective design (*e.g.*, small or dense targets) or low visual or motor abilities may cause less effective touch input.

I. INTRODUCTION

With the increasing importance of mobile *e*-Health apps running on smart devices, effective user interface design is key to accurate input and efficient visualization of *e*-Health data. While touchscreens represent the *de facto* interfaces for operating tablets, smartphones, and smart watches, the continuous strive for technological miniaturization to its physical limits in the semiconductor industry has been pushing users' visual and motor abilities to their limits as well.

Effective touch input is the direct result of delicate, fine-tuned processes of eye-hand coordination that involve complex networks of motor and sensory neurons, which continuously adjust the hand and fingers' motor action plan until the target is reached [1-3]. In the absence of a clear perception of visual stimuli, as in the case of people with low vision (see Figure 1), coping strategies are usually instantiated by other networks of neurons in the brain that find new ways to deal with visually-challenging tasks, such as acquiring small, dense, or hard-to-reach targets. The perceived quality of the interactive experience is inseparable of the extent to which the interface complies with each user's visual, motor, and cognitive abilities.

Despite considerable research on understanding touch for a variety of devices, user categories, and contexts of use [4,5], little attention has been given to the *neurobiological* and *neurocognitive* aspects involved during touch input. However, such aspects, reflected by models derived in fields such as experimental psychology, neurobiology, and kinesiology can provide new insights about how to design effective touch input on smart mobile devices. For instance, according to the "eye-hand coordination approach," any decrease in the quality of the perceived visual stimuli causes gaps in the visuomotor plan, with the direct result of less accurate hand and finger movements toward the target [1-3].

We address in this work the *relationship between models of vision developed in neurobiology and psychophysics and touch input on smart mobile devices.* We overview neurobiological and neurocognitive approaches for modeling human vision, and we point to several relevant aspects for touch input. We select three foremost models of visual processing [6-8] that we connect to key aspects of touch target design, such as target size, shape, location, padding, and animation speed.

II. RELATED WORK

There has been a considerable interest in the Human-Computer Interaction community to examine the factors that determine effective touch input on smart mobile devices [4,5]. The visual appearance of touch targets has been often manipulated experimentally to understand its impact on the effectiveness and efficiency of touch input, evaluated with objective measures of performance, such as response times and target acquisition success rates, or with subjective ratings of the perceived experience. Various user categories have been addressed in the literature [9,10] by adapting standard touch input techniques to their visual, motor, and cognitive abilities.

Target size represents one of the most examined factors for empirical evaluation of touch input. Extensive studies have emphasized the strong relationship between target size and users' accuracy with target acquisition. For example, Park *et al.* [11] showed that target sizes of 4 mm led to success rates of only 54%; Parhi *et al.* [12] measured success rates of 70% for target sizes between 3.8 mm and 9 mm, and of 97% for targets of 11.5 mm wide. Using a mobile game developed for Android, Henze *et al.* [13] collected more than 120 million touches for targets of various sizes. Their results showed 60% success rate for targets between 5 mm and 10 mm wide, and over 90% for targets that were larger than 15 mm.

The location of targets on the screen is also known to influence users' accuracy of acquiring them. Perry and Hourcade [14] evaluated this effect in an experiment in which the display was divided into twenty-five regions. They reported significant differences between these regions in terms of success rate and users' perceived pressing convenience, which were explained by the interference between the target's image and the hand and finger's image. Avrahami [15] showed that targets located near the edges of the screen led to increased reaction times. Henze *et al.*'s [13] observed higher success rates for acquiring targets located in the center of the screen.

Target padding (*i.e.*, the size of the stimuli-free area around a target) can also affect the accuracy of touch input, especially on the small screens of wearable devices. Overall, visual clutter is known to cause a decrease of user performance. Prior work showed that dense or far targets are uncomfortable to reach [16-18]. For instance, Colle and Hiszem [17] evaluated response times and error rates for edge-to-edge key spacing of 1 mm and 3 mm. Quantitative guidelines provided by Hoober [18] indicated 8 mm as the minimum required distance between a target's center and proximate visual stimuli.

III. Models of Human Vision

In this section, we examine models of human vision developed in the neurobiology and psychology literature that are relevant to understand touch input. We systematize our discussion according to the *neurobiological*, *cognitive*, and *neurocognitive* approaches of modeling human vision.

A. The neurobiological approach to understand vision

Neurobiological models of human vision focus on the roles of the anatomic structures responsible for converting the visible spectrum of electromagnetic waves into representations suited for internal processing in the brain. After passing through several transparent membranes of the eye, light is converted into electrical signals by the retina, which are transmitted along the optic nerve by means of electrochemical neurotransmission to the lateral geniculate nucleus and the primary and secondary visual cortex of the brain. One fundamental finding stemming from the neurobiological view of the human visual system originates from Hubel and Wiesel [6], who showed that the primary visual cortex is retinotopically mapped, *i.e.*, the spatial structure of the cortical area is in a direct topological correspondence with the surface of the retina. Subsequent examinations revealed that this mapping does not preserve the uniformity of the original image, *i.e.*, the parts of the image that are associated with the fovea and its surrounding regions have a larger representation in the visual map. Starting from this area, visual information is distributed to other parts of the occipital lobe that share a similar spatial structure [19].

Hubel and Wiesel [20] showed that neurons with the same visual function are grouped into a hierarchical architecture. Each layer of this hierarchy serves the visual cortex with specific features from the flow of incoming data from the external world: "simple" cells are involved in the detection of edges, lines, and spots; "complex" cells identify location, direction, and speed of the stimuli; and "hypercomplex" cells compute lengths and angles. Visual perception is the result of the interactions between more than ten distinct cortical areas [21] connected in a branching hierarchy [22]. Prior work has tried to explain the process of combining information from disparate, sketched maps to obtain a unified, synthetic representation of our visual reality. While Crick [22] invoked the contribution of a search reticular mechanism situated at the thalamic level, Kanwisher *et al.* [23] emphasized the role of the extrastriate visual associative cortex. The role of the thalamus for object recognition is reinforced by the way it mediates the bottom-up visual system and the top-down pathways in the form of an information filter, making sure that the visual cortex receives only relevant neuronal signals [24].

B. Cognitive approaches of human vision

Cognitive models organize visual perception into a set of processing modules [7] that transform an *input* into an *output*. The visual system takes as input two 2-D matrices (with values corresponding to the light rays that reach the retinas in both eyes), applies logical functions, and generates an output consisting of a 3-D internal model of the visual reality. One of the most widely-used cognitive models for vision was proposed by Marr [7]. Marr's model distinguishes between *primary processing* during the pre-attentional phase of vision (*i.e.*, the first 200 ms), which reports edge, texture, color, location, and speed information, and *secondary processing* that is involved in more complex tasks, such as object identification, for which it relates to information from the long-term memory. During the processing chain, there are three output types [7]:

- A primal, 2-D sketch that contains blobs, terminations, discontinuities, edges, groups, and curvilinear organization.
- A 2.5-D sketch revealing the orientation and rough depth of visible surfaces in a viewer-centered frame of reference.
- A 3-D model that describes shapes and their spatial organization in an object-centered frame of reference.

Marr's model [7] was upgraded by subsequent studies. For instance, Treisman and Gelade's "feature integration theory of attention" [25] worked with the assumption that object features, such as shape, color, lines, and curves, are processed in parallel, while experiments revealed the key role of attention in the generation of a synthetic image of the visual reality.

Touch Target Characteristic	Models of Human Vision	Relevant Connections
① Size	Marr, 1982	Small targets lead to a poor quality of the visual information collected during the pre-attentional phase.
② Shape	Treisman & Gelade, 1980	Object features, *e.g.*, lines and curves, are processed in parallel and grouped by attentional processes.
③ Color	Livingstone & Hubel, 1988	Colors make a visual stimulus more salient. Two adjacent colors produce simultaneous contrast.
④ Contrast	Marr, 1982 Chang *et al.*, 2015	Contrast improves visual sensitivity and object identification.
⑤ Location	Lambrecht, 2001; Kosslyn, 1999; Owsley, 2013 Yang & Song, 2015	Reaction time is increased by target "fovealisation" actions. Targets at unexpected locations activate the look-up structure and the top-down search processes. The aim-reaching movement for out-of-center targets requires extra feedback and recalibration.
⑥ Padding	Owsley, 2013 Card, 1999	Dense targets are perceived as distractors, for which the cognitive system must activate additional coping strategies, such as filtering or zooming.
⑦ Animation speed	Marr, 1982; Owsley, 2013 Helsen *et al.*, 2000 Archambault *et al.*, 2015	Moving targets require more sequential processing and increase the load of the visual system. Moving targets are perceived as distractors. Moving targets demand continuous visual and motor corrections for effective acquisition.

Table 1. Connections between several models of human vision and the visual appearance of touch targets.

C. The neurocognitive approach to understand vision

The goal of the neurocognitive approach is to mediate the previous two perspectives by connecting the input-output formalization of the cognitive models with the neurocognitive structures of the visual system [8]. One functional structure groups the retinotopically-mapped areas in a "visual buffer." A second structure handles the processing of relevant information from the visual field, *i.e.*, the attention window. Retained information is then forwarded to two parallel cortical pathways: the ventral and dorsal systems [26], whose outputs are logically connected with information from the long-term memory to match object properties with stored representations. When the information about an object is incomplete or unclear and, thus, unlikely to form a good match with data from the associative memory, the look-up structure is activated to formulate a hypothesis about that object's identity. The look-up structure searches for back-up information in the long-term memory to support the hypothesis, after which top-down search processes are activated to identify the object [8].

Acquiring a target involves both visual perception and hand and finger motor action, and several models have been proposed to relate to these aspects jointly. For instance, according to the "cybernetic model of eye-hand coordination" of Helsen *et al.* [2] and Grible *et al.* [3], the visual information of a target is used by the brain to develop an efficient motor plan to acquire it effectively. The key role of planning has also been examined in kinesiology [27]. In a first step, visual information is used by the brain to produce an initial movement of the hand. Then, eye saccades focus on the target to maintain the target in the foveal or near-foveal zone. Afterwards, information delivered by the visual system is used to fine-tune the hand's trajectory towards the target [28]. For difficult tasks, a switch from feedforward to feedback control is activated [1] to consciously adjust the motor plan [28].

IV. Implications for Touch Input

In this section, we discuss the implications of vision modeling for touch input by addressing several appearance aspects of a visual target on a touchscreen, *i.e.*, target size, location, padding to other targets, etc.; see Table 1.

The cognitive models of the visual system imply that the low success rates for acquiring small targets are a direct consequence of the poor quality of the visual information collected during the pre-attentional phase [7]. The lower the quality of the output regarding edges, texture, movement, color, and spatial information, the more affected the following processing steps will be. Moreover, visual impairments add to this effect, causing even more inaccurate output. In the context of the eye-hand coordination approach [2-3], any decrease in the quality of the perceived image determines gaps in the visuomotor plan with less accurate hand and finger movement.

According to the biological paradigm, a target image is best formed when the stimulus falls in or near the foveal zone. Otherwise, a series of reflex eye movements are activated to bring and keep the target inside the foveal area. Any extra actions in this process of target "fovealisation" increase the magnitude of users' reaction times [29]. Empirical findings on touch target acquisition from the literature are in accordance with predictions of the neurobiological models, since an object at an unusual location does not match well with existing data from the associative memory. In this case, the look-up structure and top-down search processes become active [8], causing a slower speed for processing visual information [30]. Other studies [31] showed that the activation of searching tasks based on higher cognitive skills, such as filtering and zooming, determine longer reaction times as well. Lower success rates for targets next to edges or corners of the screen [15] are explained by the eye-hand coordination model: to attain fast speeds, movement begins before all the visual information has been obtained. Thus, the aim-reaching movement for out-of-center targets demands more feedback and recalibration [32].

Cognitive models suggest that the performance of visual processing is influenced by the density of cues, while moving or dense targets are perceived as distractors in the framework of attentional models [30]. In this case, the cognitive system must activate coping strategies, such as filtering or zooming [33], which increase cognitive load and decrease performance. In dynamic environments, such as mobile video games, targets often move at various speeds and in various directions on the screen. These factors add an extra level of difficulty, because users need to rely on their visuomotor predictive control [34]. According to Marr's model [7], moving targets require more sequentially image processing, which increases the load of the visual system. In terms of the attentional models [30], moving targets are distractors that perturb the visual focus. From the perspective of the eye-hand coordination model [2,3], moving targets demand continuous visual and motor corrections until targets are reached, causing longer response times.

V. Conclusion

We proposed in this work a new perspective to examine touch input by relating to neurobiological and neurocognitive models of human vision. Our goal was to present the technical community with relevant knowledge from neurobiology and psychology about how visual stimuli are processed in the brain for effective and efficient eye-hand coordination. We hope that our investigation will foster new explorations in this direction: practitioners can employ our results to inform touch target design for various user categories [9,10,36,37], and researchers can use our investigation as a starting point for new empirical explorations towards a better understanding of touch input on smart mobile devices.

Acknowledgment

This work was supported from the project PN-II-RU-TE-2014-4-1187 financed by UEFISCDI, Romania.

References

[1] P.S. Archambault, S. Ferrari-Toniolo, R. Caminiti, and A. Battaglia-Mayer, "Visually-guided correction of hand reaching movements," *Vision Research*, vol. 110, pp. 244–256, 2015.

[2] W.F. Helsen, D. Elliott, J.L. Starkes, and K.L. Ricker, "Coupling of eye, finger, elbow, and shoulder movements during manual aiming," *Journal of Motor Behavior*, vol. 32, no. 3, pp. 241–248, 2000.

[3] P.L. Gribble, S. Everling, K. Ford, and A. Mattar, "Hand-eye coordination for rapid pointing movements," *Experimental brain research*, vol. 145, no. 3, pp. 372–382, 2002.

[4] A.K. Orphanides and C.S. Nam, "Touchscreen interfaces in context," *Applied Ergonomics*, vol. 61, pp. 116–143, 2017.

[5] S.A. Brasel and J. Gips, "Tablets, touchscreens, and touchpads: How varying touch interfaces trigger psychological," *Journal of Consumer Psychology*, vol. 24, no. 2, pp. 226–233, 2014.

[6] D.H. Hubel and T.N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *The Journal of Physiology*, vol. 195, no. 1, pp. 215–243, 1968.

[7] D. Marr, "Vision: A computational investigation into the human representation and processing of visual information," MIT Press, 1982.

[8] S.M. Kosslyn, "If neuroimaging is the answer, what is the question?," *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 354, no. 1387, pp. 1283–1294, 1999.

[9] L.G. Motti, N. Vigouroux, and P. Gorce, "Interaction techniques for older adults using touchscreen devices: a literature review," in *Proc. of the 25th Conference on l'Interaction Homme-Machine*, 2013, p. 125.

[10] J. Stephenson and L. Limbrick, "A review of the use of touch-screen mobile devices by people with developmental disabilities," *Journal of Autism and Developmental Disorders*, vol. 45, pp. 3777–3791, 2015.

[11] Y.S. Park, S.H. Han, J. Park, and Y. Cho, "Touch key design for target selection on a mobile phone," *Proc. of MobileHCI '08*, 2008, 423–426.

[12] P. Parhi, A.K. Karlson, and B.B. Bederson, "Target size study for one-handed thumb use on small touchscreen devices," in *Proc. of Mobile HCI '06*, 2006, pp. 203–210.

[13] N. Henze, E. Rukzio, and S. Boll, "100,000,000 taps: analysis and improvement of touch performance in the large," in *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, 2011, pp. 133–142.

[14] K.B. Perry and J.P. Hourcade, "Evaluating one handed thumb tapping on mobile touchscreen devices," *Proc. of Graphics Interface*, 2008, 57–64.

[15] D. Avrahami, "The effect of edge targets on touch performance," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015, pp. 1837–1846.

[16] Y. Chang, S. L'Yi, K. Koh, and J. Seo, "Understanding users' touch behavior on large mobile touch-screens and assisted targeting by tilting gesture," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015, pp. 1499-1508.

[17] H.A. Colle and K.J. Hiszem, "Standing at a kiosk: Effects of key size and spacing on touch screen numeric keypad performance and user preference," *Ergonomics*, vol. 47, no. 13, pp. 1406–1423, 2004.

[18] S. Hoober, "Common Misconceptions About Touch," *UXmatters, March*, vol. 18, 2013.

[19] D.J. Felleman and D.C. Essen, "Distributed hierarchical processing in the primate cerebral cortex," *Cerebral Cortex*, vol. 1, pp. 1–47, 1991.

[20] D.H. Hubel and T.N. Wiesel, "Brain and visual perception: the story of a 25-year collaboration," Oxford University Press, 2004.

[21] M. Bar, R.B. Tootell, D.L. Schacter, D.N. Greve, B. Fischl, J.D. Mendola, B.R. Rosen, A.M. Dale, "Cortical mechanisms specific to explicit visual object recognition," *Neuron*, 29(2), pp. 529–535, 2001.

[22] F. Crick, "Function of the thalamic reticular complex: the searchlight hypothesis," *Proceedings of the NAS*, vol. 81, no. 14, 1984.

[23] N. Kanwisher, J. McDermott, and M.M. Chun, "The fusiform face area: a module in human extrastriate cortex specialized for face perception," *Journal of Neuroscience*, vol. 17, no. 11, pp. 4302–4311, 1997.

[24] R.D. Wimmer, L.I. Schmitt, T.J. Davidson, M. Nakajima, K. Deisseroth, and M.M. Halassa, "Thalamic control of sensory selection in divided attention," *Nature*, vol. 526, no. 7575, pp. 705–709, 2015.

[25] A.M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol. 12, no. 1, pp. 97–136, 1980.

[26] M. Mishkin and L.G. Ungerleider, "Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys," *Behavioural Brain Research*, vol. 6, no. 1, pp. 57–77, 1982.

[27] J. Messier and J.F. Kalaska, "Comparison of variability of initial kinematics and endpoints of reaching movements," *Experimental Brain Research*, vol. 125, no. 2, pp. 139–152, 1999.

[28] J. Pelz, M. Hayhoe, and R. Loeber, "The coordination of eye, head, and hand movements in a natural task," *Experimental Brain Research*, vol. 139, no. 3, pp. 266–277, 2001.

[29] C.J. Van den Branden Lambrecht, "Vision models and applications to image and video processing," Springer Science Business Media, 2013.

[30] C. Owsley, "Visual processing speed," *Vision research*, vol. 90, 2013.

[31] S. Chokron and G.N. Dutton, "Impact of Cerebral Visual Impairments on Motor Skills," *Frontiers in Psychology*, vol. 7, 2016.

[32] Y. Yang and R. Song, "Difficulty-dependent trajectory planning during target-reaching movements," in *Proc. of the 37th Annual Int. Conf. of the IEEE Eng. in Medicine and Biology Society*, 2015, pp. 6675–6678.

[33] S.K. Card, J.D. Mackinlay, and B. Shneiderman, "Readings in information visualization: using vision to think," M. Kaufmann, 1999.

[34] K. Caeyenberghs, D. van Roon, K. van Aken, P. De Cock, C.V. Linden, S.P. Swinnen, B.C. Smits-Engelsman, "Static and dynamic visuomotor task performance in children with acquired brain injury," *The Journal of Head Trauma Rehabilitation*, vol. 24, no. 5, pp. 363–373, 2009.

[35] M. Livingstone and D. Hubel, "Segregation of form, color, movement, and depth," *Science*, vol. 240, no. 4853, pp. 740–749, 1988.

[36] R.-D. Vatavu, "Visual Impairments and Mobile Touchscreen Interaction: State-of-the-Art, Causes of Visual Impairment, and Design Guidelines," *Int. J. of Human-Computer Interaction*, 2017.

[37] R.-D. Vatavu, G. Cramariuc, D.M. Schipor, "Touch Interaction for Children Aged 3 to 6 Years: Experimental Findings and Relationship to Motor Skills," *Int. J. of Human-Computer Studies* 74, pp. 54-76, 2015.