

# Evaluating visual attention for multi-screen television: measures, toolkit, and experimental findings

Radu-Daniel Vatavu<sup>1</sup> · Matei Mancias<sup>2</sup>

Received: 29 November 2014 / Accepted: 6 May 2015  
© Springer-Verlag London 2015

**Abstract** New and emerging multi-screen television scenarios and applications need new evaluation procedures, methodologies, and tools to support multi-screen data analysis. In this work, we introduce a set of 12 measures to characterize viewers' visual attention patterns for multi-screen TV. Out of these, eight measures are computed directly from eye-tracking data, while the other four are evaluated using questionnaires. We applied our measures during a controlled experiment involving nine distinct screen layouts with two, three, and four TV screens, for which we report new findings about viewers' distributions of visual attention. For example, we found that viewers need an average discovery time up to 4.5 s to visually fixate four screens, and their subjective perceptions of what they watched and for how long they watched each screen are substantially accurate, i.e., we report Pearson's correlation coefficients up to  $r = .892$  with ground truth measured with eye-tracking equipment. We also analyze and discuss the evolution of our participants' distributions of visual attention over time from the perspective of our new set of measures. For example, we found that people perform significantly more transitions between screens during the first seconds of watching television, after which their

level of visual attention converges to a stable value. We complement the findings revealed by our objective eye gaze measurements with subjective data about participants' perceived cognitive work load and comfortability while watching more than one TV screen, and we measure viewers' capacities to understand and recall content delivered simultaneously on multiple screens. To foster new studies and explorations of viewers' visual attention patterns during multi-screen television watching, we release in the community an update for an existing software toolkit (i.e., VATic-TV, the Visual Attention Toolkit for TV, now at version v2) that automatically computes our measures from data delivered by standard eye-tracking equipment. We hope that our new set of measures and the companion software will benefit the community as a first step toward understanding visual attention for emerging multi-screen TV applications and, consequently, will help researchers and practitioners to design new TV applications that will better exploit viewers' visual attention patterns toward new, richer television experiences.

**Keywords** Visual attention · Interactive TV · Multi-screen TV · Multi-display · Eye gaze · Eye tracking · Measures · Experiment · Evaluation · TLX

---

✉ Radu-Daniel Vatavu  
vatavu@eed.usv.ro

Matei Mancias  
matei.mancias@umons.ac.be

<sup>1</sup> Department of Computer Science, Electronics, and Automation and Integrated Center for Research, Development, and Innovation in Advanced Materials, Nanotechnologies, and Distributed Systems for Fabrication and Control (MANSiD), University Stefan cel Mare of Suceava, 720229 Suceava, Romania

<sup>2</sup> University of Mons, 20, Place du Parc, 7000 Mons, Belgium

## 1 Introduction

Multi-screen systems are able to deliver their users more content and more opportunities for control as well as to provide new ways for users to enrich, share, and transfer multimedia content originating from various sources [6]. Due to such particular attractiveness, the multi-screen scenario has recently received thoughtful attention in the academic community for technical design and

implementation [34, 41, 45] as well as for evaluating user performance [14, 30, 31, 36]. In the context of the interactive TV, today's common implementation of the multi-screen concept is represented by the secondary screen scenario, with smartphones and tablets being frequently employed by their users during television watching [6, 8]. In a larger context, Vatavu and Mancas referred to multi-screen TV systems as “television potpourris” as they are hybrid systems of individual screens of different form factors arranged in space according to various screen layouts and that broadcast transmissions and display content of various genres from various sources [45].

However, besides obvious advantages delivered by multi-screen television, more screens demand higher cognitive load for viewers to understand what they watch and to correlate content from different video streams, as well as increased visual attention that needs to distribute across multiple displays situated at various locations in a three-dimensional space. Therefore, it is likely for multi-screen TV systems to increase viewers' visual and cognitive attention load up to the point when the television watching experience becomes no longer pleasant. Unfortunately, such important aspects regarding the distribution of viewers' visual attention in multi-screen TV systems have not been thoroughly addressed by the interactive TV community up to now. In fact, in a recent work investigating the factors affecting users' visual attention in multi-display interfaces, Rashid et al. [31] noted that “further research is needed to investigate the influence of different categories of content coordination on attention switching and task performance.” We believe that rigorous methodologies and techniques to conscientiously evaluate viewers' performance in such scenarios are mandatory, given the wide adoption of the second-screen scenario—for example, “eighty-four percent of smartphone and tablet owners say they use their devices as second-screens while watching TV at the same time” [26, p. 14].

While today's prevalent implementation of a multi-screen system for iTV takes the form of the tablet used in conjunction with the TV set [6, 8], for which previous work has examined problems of attention and distraction [2, 3, 7, 17], investigation of visual attention distributed across more than two screens is a more challenging task. This work represents an extension of previous methodology, toolkit, and results that we obtained in this direction [46]. In this work, we make one step further toward understanding viewers' visual attention patterns for multi-screen television and, in doing so, we provide the community with a set of general and reusable measures to characterize viewers' visual attention patterns in various experimental scenarios addressing various scientific research goals about visual attention and multi-screen television.

By following the taxonomy of Rashid et al. [31], we address in this work screen layouts with depth contiguity (i.e., all screens are at equal distance from the observer) and visual field discontinuity (i.e., all screens are in the same vertical plane, but spatially separated, which makes them appear as distinct displays instead of a visually contiguous screen). Such installations are now possible and were demonstrated in the community. For instance, recent developments in multimedia systems have led to practical technical designs for such scenarios involving more than two screens [41, 44] up to augmented reality systems that turn the entire room into a multimedia multimodal interface [22, 23]. The methodology that we introduce in this work, i.e., measures and software toolkit to evaluate viewers' visual attention patterns, is applicable for any number of screens and any screen layouts, and we demonstrate how to apply it for scenarios consisting of up to four TV screens.

Our contributions are as follows: (1) We introduce a set of eight quantitative measures that we compute from eye-tracking data and four qualitative, self-reported measures to characterize viewers' visual attention patterns for multi-screen television watching; (2) we employ our new set of measures to report new findings about viewers' distributions of visual attention for multi-screen television; for example, we found that viewers' subjectively perceived watching time per screen is substantially accurate, i.e., Pearson's  $r = .892$  between measured and self-reported screen watching time, and that viewers discover content displayed by all screens during a minimum time interval that depends quadratically on the number of screens; and (3) we provide a software update to an existing toolkit (VATic-TV [46], now version v2) to enable other researchers and practitioners to automatically compute our measures and analyze data collected during their specific experimental setups. We hope that this work will benefit the members of the interactive TV community who will employ our measures to further investigate viewers' visual attention patterns for emerging multi-screen TV applications and, consequently, will inform new designs delivering enriched multi-screen television experiences.

## 2 Related work

In this section, we review existing work addressing multi-screen environments, and we pay particular attention to applications of multi-screen systems to television, modeling of visual attention, and understanding users' task performance in multi-screen environments.

The most straightforward way to create a multi-screen environment is to define individual screens inside a large

video projection, with all screens controlled by the same computer [41, 45]. This approach originates from the way augmented reality systems have been traditionally implemented [38], with video projections displayed on top of the physical environment in order to digitally enrich users' interactive experiences in that environment. So far, video games have been the most common application of augmented reality for home entertainment [22, 23]. However, augmented reality research has also targeted television, with AROUND-TV [44] being one such example. Alternatively, multiple physical screens can be put together to create multi-screen environments by using software platforms that control the distribution of content across the screens. For example, Phone as a Pixel is one such platform that displays synchronized content on hundreds of displays, such as laptops, smartphones, and tablets [34]. Although the technical design and implementation of such systems are fascinating and challenging, we are not interested in this work in the technology used to build such systems, but rather in the impact that multi-screen systems have on viewers' performance. Our concern about performance is justified by the fact that more screens are able to deliver more content to viewers, but, at the same time, more screens may also have side effects on visual attention and cognitive load. Therefore, we also review in this section previous work that evaluated and modeled visual attention in general, but also in conjunction with the TV set.

## 2.1 Multi-screen environments and user performance

Multi-screen environments have been investigated before in terms of the attention demands imposed on their users [30, 31] and the effects of the extra attention load on users' task performance [14, 36]. For example, Rashid et al. [30] explored the cost of switching attention between the small display of a mobile device and a large screen, and reported decreased user performance because of the adaptation mechanisms that naturally occur when users need to continuously shift eye gaze between screens. By evaluating multiple user interfaces for content-seeking tasks on two screens, i.e., mobile device and large display, the authors found that participants performed best when they controlled the content on the large display using the mobile device, while distributing visual output on both displays made users exhibit the poorest performance. The evaluation results were used by the authors to propose guidelines for practitioners designing such hybrid two-screen systems, such as avoid showing the same visual elements of the interface across displays [30, p. 106].

Tan and Czerwinski [36] addressed the effect of visual separation between displays and their physical discontinuities, such as monitor bezels on users' performance.

They found that discontinuities do not affect task performance, but displaying content on screens positioned at different depths has small, yet detrimental effects on performance, i.e., a 10 % detriment of performance was reported by the authors for a task consisting in proofreading a text with notifications popping up outside the focal region of the task. These results were later confirmed by Bi et al. [4], who conducted controlled experiments to better understand the effect of tiled-monitor interior bezels on users' performance during various tasks, such as visual search, straight-tunnel steering, and target selection. The authors found that task time and error rate were not affected by the presence of bezels during visual search; however, participants were less accurate when the content they were asked to locate had been split across the interior bezels. These experiments also revealed that interior bezels hinder steering performance, but they do not affect users' performance during target selection. Wallace et al. [51] confirmed that users' performance during visual search tasks is not affected by the presence of interior bezels (despite a very small effect of bezel width) and used their findings to suggest design implications for user interfaces for tiled displays, such as use bezels as visual anchors for displaying interface elements. Other design implications were provided by Wallace et al. [52], who studied the impact of bezel presence, bezel width, and user-to-display distance on users' perceptions of the magnitude of displayed content, e.g., the authors observed that bezels wider than 0.5 cm introduce errors up to 7 % in the perception of magnitude of content.

Forlines et al. [14] observed that their participants performed worse during a visual search task when content was displayed at different rotation angles on four vertical screens than when the same content was presented on a single screen. The authors of that study concluded that participants' scanning of multiple views added to the time duration of the task, but not to its accuracy. Hutchings [21] conducted a Fitts's law experiment [13] in a multi-display environment and found that increasing gap sizes between displays made participants achieve lower throughput values. Findings from that study suggest that Fitts's law underestimates the evaluation of task difficulty for multi-display systems. Finally, Grüninger and Krüger [18] investigated interior bezels for stereoscopic displays and reported that users' perceptions of stereo content improves, while bezels get smaller and displays get larger. They also found that users' adaptation times to the stereoscopic effect can be negatively affected by the presence of interior bezels for tiled displays.

## 2.2 Visual attention

Attention is the cognitive process to selectively interpret information subsets while ignoring others, i.e., to

selectively focus on solely one aspect of the environment [1, p. 519]. Because of its many implications for the human life, attention has been studied in many different domains, such as psychology [40], neuroscience [28], and engineering [5, 39]. By definition, attention allows people to focus on a single task at one given time. Sohlberg and Mateer [35] identified five levels of attention, which are focused, sustained, selective, alternating, and divided attention. Focused attention represents a person's ability to respond discretely to some stimulus. Sustained attention occurs during a continuous, repetitive activity, and it represents a person's ability to maintain response in a consistent manner across the entire duration of the stimulus. Upper in the hierarchy of attention levels, selective attention describes response behavior that is consistent in the presence of distracting stimuli. Alternating attention represents a person's cognitive flexibility to sequentially shift the focus of attention between different tasks. The highest level of attention describes the ability of a person to respond simultaneously to different tasks of different cognitive requirements, i.e., divided attention [35].

In this work, we are only interested in visual attention, which is attention-triggered or demanded by the presence of visual stimuli. Visual attention has been modeled by cognitive psychologists with the spotlight and zoom-lens models [11, 12]. The spotlight model [11] describes visual attention in terms of focus (i.e., the region from which information is extracted and processed at high resolution), fringe (i.e., the low-resolution extraction of information at the boundaries of the focus region), and margin (i.e., the cutoff of the visual attention area). The zoom-lens model was introduced by Eriksen [12] to update the spotlight model by making it adaptable in size and thus to explain the trade-off in the efficiency of processing visual information, e.g., larger the focus, slower the processing will be.

Researchers have also modeled the ways in which the human brain attends to stimuli and processes information in what is known as bottom-up and top-down processing [37]. For example, some stimuli attract attention simply because of their stringent nature (e.g., a quick motion or a telephone ring captures a person's attention instantly), which makes our brain process information at a preconscious level. On the other hand, the top-down processing model describes the act of individuals controlling their attention toward the achievement of a specific goal. Finally, attention can be overt, corresponding to the situation in which eye gaze attends to some fixed region in space, but also covert, i.e., describing mental focus shifting to other tasks without the eyes necessarily moving [27]. Overt attention is sequential by using eye saccades (e.g., ballistic movements) and fixations (e.g., the eye gaze stops at some spatiotemporal stable area). People use overt attention to explore complex visual scenes and to direct eye

gaze toward interesting spatial locations. During the process of overt visual attention, eyes produce fixations (which are longer periods of time during which the eyes focus on the same area), smooth pursuit movements (i.e., the eyes follow smoothly moving targets), and saccades (i.e., ballistic eye movements that occur between two consecutive fixations). In contrast, covert attention can process several stimuli in parallel. Eye movements are also known to be influenced by endogenous factors, such as the task at hand, behavioral goals, and the motivational state, as well as by exogenous factors, which are represented by the local and global spatial properties of the visual scene. In general, humans are known to be able to simultaneously attend to  $7 \pm 2$  stimuli at once [25].

### 2.3 Visual attention and television

Researchers have found that individual looks at the TV set vary in duration and that people develop different watching strategies to follow content displayed on TV. For example, people may look at the TV set only at the right times, just enough to be aware of what is happening, while being engaged in some other activity, e.g., talking or working. When investigating such phenomena, Geerts et al. [16] found that the genre of television content correlates with how much people talk during watching TV and that the plot structure influences talking during social television watching. For example, the authors reported that out of all the eighteen genres they had investigated, people talked the most and shared content with others during twelve genres (67 % of all genre types): news, sport, soap opera, docu-soap, reality show, talk show, comedy series, quizzes, film, animation film, stand-up comedy, and music programs [16, p. 77]. Such studies about people's behaviors of attention in the context of television watching reveal the importance of top-down attention during the everyday TV watching experience.

Researchers also found that most looks at television are very short, e.g., 2 s, and can be described as mere glances [19]. Although surprising, these findings were explained and characterized with the "hazard look" function that gives the probability that looks persisting a given duration will terminate in the next half second. Once a look begins, that look is likely to terminate in the first second, with a hazard peak at 1–1.5 s. Hawkins et al. [19] investigated this phenomenon and identified several types of looks, i.e., monitoring looks, which take less than 1.5 s, orienting looks up to 5 s, engaged looks lasting between 6 and 15 s, and staring looks, which occur after 15 s of continuous watching. The average duration of a look reported in the study of Hawkins et al. [19] was 7 s, but the median was less than 2 s and only 15 % of all looks lasted longer than 15 s (p. 163). The hazard function models the

way looks gain more inertia as they last beyond the first second and characterizes the probability of the viewer to continue looking at television. Longer looks were explained as greater cognitive engagement of viewers that dedicated their continuous attention to the visual and audio stimuli provided by television.

To characterize the visual attention behavior of viewers during television watching, researchers have employed eye-tracking devices that accurately follow and report viewers' eye gaze. For example, Kallenbach et al. [24] found that text displayed on TV affects the patterns of visual attention, memory, and cognitive workload more than simple pictorial information does. Holmes et al. [20] examined the visual attention of people watching TV in a secondary screen scenario and reported that 30 % of the attention was allocated to the tablet. In that study, the secondary screen grabbed participants' attention even without expressly providing any update for its content. Also, looks at the TV set were found to be shorter than those reported in other research [19], i.e., looks lasted 2 s in average, while most looks took between 1 and 5 s. These results were explained by the authors by the fact that participants were anticipating interactive content on the tablet and, therefore, were checking its screen more frequently with short eye gazes [20, p. 399]. In a multi-screen sports study, Cummins et al. [9] found viewers' visual attention to vary function of screen size, game play (i.e., the action displayed on screen), and repeated exposure. The authors also reported that viewers had to adopt screen watching strategies to cope with the amount of different pictures displayed simultaneously by different screens.

Prior work has considered designs of interaction techniques with television that would not distract viewers' visual attention toward the controlling device itself. For instance, Vatavu [42, 43] considered freehand and whole-body gestures to control the functions of the TV set, which were informed by the agreement analysis results of a gesture elicitation study [48, 53]; Vatavu and Pentiu [47] proposed the interactive coffee table as a control device for TV; Vatavu and Zaiti [49] examined low-effort finger and hand gestures for controlling TV in the context of the lean-back media consumption paradigm; Wagner et al. [50] introduced the BodyScape design space for multi-surface interaction, and Dezfali et al. [10] proposed the PalmRC prototype for eyes-free interaction with the TV set.

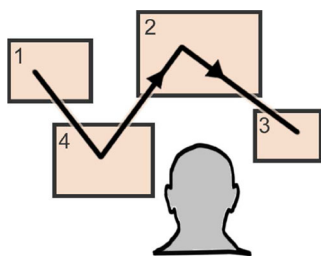
Rashid et al. [31] identified five factors that affect users' visual attention patterns for multi-display user interfaces, which are display contiguity, angular coverage, content coordination, input directness, and input-display correspondence. Display contiguity describes the spatial arrangement of displays in terms of the visual field being covered (i.e., displays appear contiguous, although they may be separated by bezels or they may be positioned at

various distances from the viewer) and depth (i.e., displays are located at the same distance from the observer, but not necessarily adjacent). The angular coverage of a multi-display system describes the size of the field of view from the location of the observer; in this regard, Rashid et al. [31] identified and discussed the panorama, field-wide, and fovea-wide configurations. Content coordination describes the relationship between content delivered on different displays, such as content may be cloned, extended, or generally coordinated. Input directness and input-display correspondence characterize the spatial separation between the viewer and the display in terms of interacting with content provided by the display. For example, input can be direct (e.g., touch), indirect (remote display), or hybrid. The taxonomy of multi-display environments enabled by these factors allows practitioners to design multi-display prototypes considering the visual attention demands imposed on users [31].

### 3 Visual attention measures for multi-screen television

By following closely the results of previous work in terms of experimental findings and modeling of visual attention [9, 20, 24, 30, 31, 46], we defined a set of eight *objective* measures to characterize viewers' visual attention behavior in the context of multi-screen television watching. All our measures can be computed from a sequence of point-time coordinates of the viewer's eye gaze, as delivered by any standard eye-gaze-tracking equipment:  $\{(x_i, y_i, t_i) \mid i = 1..n\}$ , where  $n$  represents the total number of points reported by the eye tracker during the monitored time interval. For example, the FaceLab eye gaze tracker that we employed for the experiments reported in this work (see next section) records eye gaze data at a rate of 60 Hz, resulting in approximately  $n = 3600$  data points for 1 min of continuous recording. Our objective measures for evaluating viewers' visual attention in multi-screen TV environments are:

1. *Discovery time (DT)* is defined as the amount of time required for the viewer to make a pass over all TV screens so that each individual screen is fixated visually at least once (see Fig. 1 for a visual illustration of this measure for a scenario involving four TV screens). When computing the values of DT, we consider that screen  $S$  has been fixated visually by the viewer if at least one eye gaze point  $(x_i, y_i)$  falls within the physical boundaries of screen  $S$ , i.e., within the rectangular shape that defines the screen. We then record the time stamp  $t_i$  of the first point, in chronological order, that satisfies this constraint. DT is



**Fig. 1** Graphical illustration of the discovery time (DT) and discovery sequence (DS) measures. Say the viewer looks at the four screens in the following order: 1–4–2–3. This sequence represents the value reported by the DS measure, while the time at which the viewer looks at the last screen in this sequence represents the value of DT

computed as the maximum of all the minimum time stamps  $t_i$  for each screen  $S$ :

$$DT = \max_S \left\{ \min_{i=1..n} \{t_i \mid (x_i, y_i) \in S\} \right\} \quad (1)$$

DT represents the minimum time that is imperative for the viewer to understand what is running on all the TV screens in order to inform what program to watch. We hypothesize that the discovery time is affected by various factors, such as the number of screens (i.e., more screens, larger the amount of time needed to visually fixate all), screens’ layout and their form factors, and the type of content that is displayed on those screens (e.g., more dynamic content on some screens may attract viewers’ eye gaze more in the detriment of other screens, resulting in larger discovery times).

2. *Discovery sequence (DS)* is defined as the sequence of screens that was traversed by the viewer’s eye gaze during the discovery time (see Fig. 1). We represent the values of this measure as the set:

$$DS = \bigcup_{i=1..n} \{S_j \mid t_i \leq DT \wedge (x_i, y_i) \in S_j\} \quad (2)$$

with the constraint that consecutive screens from this sequence must be different. The discovery sequence can be a permutation of all screens, e.g., 1–4–2–3 for configurations of four screens (see Fig. 1), or it may include the same screen more than once, e.g., 1–4–1–2–3 (which is the case of a permutation with repetitions) if that screen attracts the viewer’s visual attention again during the discovery time, before all the other screens have been fixated visually. The discovery sequence informs about viewers’ watching patterns during the discovery interval, but it can also be extended for other time intervals during television watching. For example, different sequences may be compared for different time intervals, such as the discovery period versus the period following 1 min later

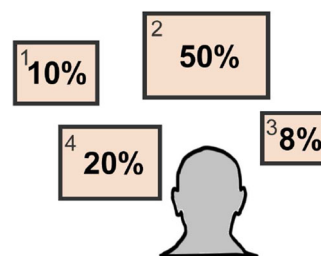
versus the sequence of screens from 10 min later. The factors that may affect the order of screens in this sequence are the layout of the screens and the type of content displayed on those screens.

3. *Screen watching time (SWT)* represents the percentage of visual attention allocated to each screen  $S$  during the monitored time interval:

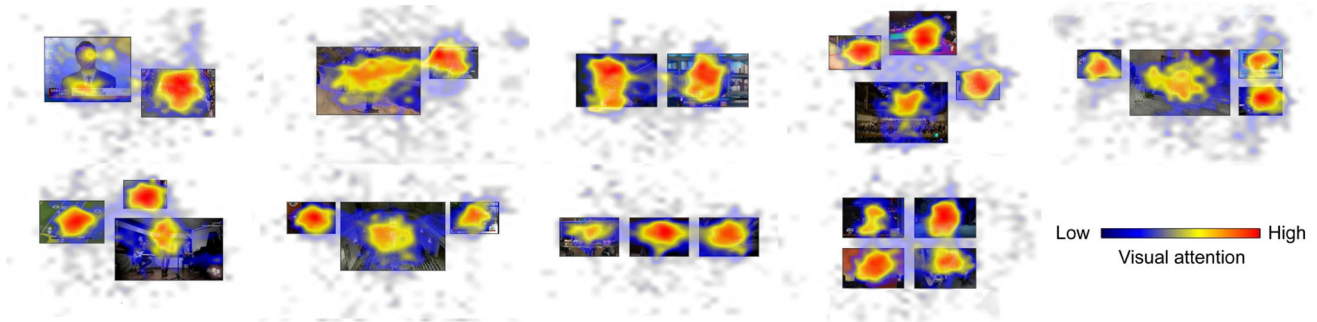
$$SWT(S) = \frac{|\{(x_i, y_i) \in S, i = 1..n\}|}{n} [\times 100\%] \quad (3)$$

where  $(x_i, y_i)$  represent eye gaze points, and  $||$  denotes the cardinal of the set of points. For example, the SWT distribution for a 4-screen layout may be uniform, with approximately 25 % of visual attention devoted to each screen, but it may also be nonuniform if, for example, some screens capture the viewer’s visual attention more, e.g., 50, 20, 10, and 8 % (see Fig. 2). SWT values may depend on the form factors, screen layout, and the content displayed by the screens, e.g., larger the screen or more attractive its content, more time will likely be devoted by viewers to consume that content. Screen watching times can also be visualized as heatmaps (see Fig. 3) that show localized concentrations of eye gaze for each screen and use colors to encode fixation magnitudes. As opposed to heatmaps, screen watching times deliver quick summary reports of the elaborated eye gaze distribution data provided by eye gaze heatmaps.

4. *Transition count (TC)* is defined as the number of eye gaze transitions between consecutively fixated TV screens that occurred during the monitored time interval (Fig. 4). The transition count measure may be affected by the number of screens, their arrangement in space, as well as by the content displayed on those screens. Larger TC values reflect more distributed attention that may be due to viewers’ following more programs at the same time or to the fact that visual attention is drawn by more screens simultaneously, which may impact negatively on viewers’

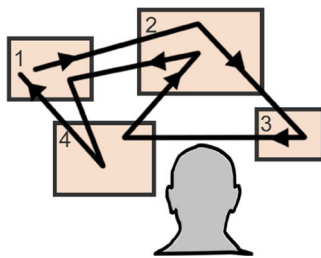


**Fig. 2** Graphical illustration of the screen watching time (SWT) measure. In this example, the second screen has received 50 % of the viewer’s visual attention during the monitored time interval, while the third screen only 8 %. Screen watching times can be further detailed using eye gaze heatmaps (see Fig. 3)



**Fig. 3** Screen watching times can be visualized in detail using eye gaze heatmaps. In this figure, visual attention heatmaps are shown for all the nine multi-screen TV layouts evaluated in the experiment

reported in this paper (see next section). Warmer colors (e.g., orange and red) denote more eye gaze looks at those regions



**Fig. 4** Graphical illustration for the transition count (TC) and transition speed (TS) definitions. In this example, the viewer’s eye gaze moves from screen 1 to screens 2, 3, and so on, until it reaches back to screen 1. The total number of transitions is  $TC = 7$ . Suppose that the monitoring interval is 30 s, then the transition speed is  $TS = 0.23 \text{ sec}^{-1}$

screens over time, and they may also correlate with cognitive load and watching fatigue. Note that real-world distance units, such as centimeters should be used when reporting EGTD values for screens with different pixel pitch densities. In this work, we compute EGTD values from the  $(x_i, y_i)$  eye gaze point coordinates returned by the eye tracker, as follows:

$$EGTD = \sum_{i=2}^n \left( (x_i - x_{i-1})^2 + (y_i - y_{i-1})^2 \right)^{1/2} \quad (5)$$

where  $(x_{i-1}, y_{i-1})$  and  $(x_i, y_i)$  represent consecutive eye gaze points in time.

7. *Eye gaze travel speed (EGTS)* represents the average speed of the viewer’s eye gaze, while it traveled between screens during the monitored time interval. We compute EGTS as the ratio between the distance traveled by the eye gaze (EGTD) and the elapsed time:

$$EGTS = \frac{EGTD}{\text{Time}} \quad (6)$$

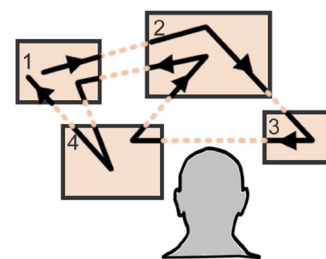
8. *Switch time (ST)* is defined as the percentage of time during which the viewer’s eye gaze travels between screens (see Fig. 5):

cognitive load and, consequently, on their television watching experience. The value of the transition count measure may also point to design flaws in the screen layout. For example, if viewers transition frequently between two screens located at the opposite sides of the layout, then their eye gazes pass through other screens that they do not necessarily intend to watch, increasing therefore the value reported by the TC measure.

5. *Transition speed (TS)* represents the average eye gaze speed at which viewers perform transitions between screens, defined as the ratio between TC and the duration of the time interval for which this measure is reported (Fig. 4):

$$TS = \frac{TC}{\text{Time}} \quad (4)$$

6. *Eye gaze travel distance (EGTD)* represents the total distance traveled by the viewer’s eye gaze during the monitored time interval expressed in the units reported by the eye-tracking equipment, such as pixels. Different EGTD values may reflect different visual attention patterns for viewers, different preferences for some



**Fig. 5** Graphical illustration for the switch time (ST) measure. In this example, the viewer’s eye gaze moves from screen 1 to screens 2, 3, and so on, until it reaches back to screen 1. While the eye gaze moves between screens, it also falls outside the screens for short periods of time (see the dotted lines in the figure). In this example,  $ST = 12 \%$

$$ST = \frac{|\{(x_i, y_i) \mid \exists S \text{ so that } (x_i, y_i) \in S\}|}{n} [\times 100 \%] \quad (7)$$

Switch time can also be defined in conjunction with the screen watching times measure by subtracting from 100 % the percentage of time spent by viewers looking at each screen:

$$ST = \left(1 - \sum_s SWT(S)\right) [\times 100 \%] \quad (8)$$

Large switch time values may point out flaws in the screen layout design, e.g., larger the ST value, further apart the screens are located one from the other in that particular layout.

We introduce the above measures to characterize viewers' television watching behavior in multi-screen TV environments in more nuanced ways, not easily accessible with generic eye gaze heatmaps and scan paths, and not available until now in the literature of the domain. For example, heatmaps (such as those shown in Fig. 3) are generally used to describe the spatial distribution of eye gaze, which is useful for investigating specific elements that attract viewers' attention *within* the same screen. However, our SWT measure reflects viewers' allocated watching time for the entire screen with one single value integrating the spatially distributed data of the heatmap, while TC and ST characterize attention switch *between* screens. Also, DT and DS are computed on top of the scan path to reflect viewers' specific behaviors occurring at specific moments, e.g., during the discovery of TV content to inform what to watch. Our measures are also flexible in terms of the units of measurement, a choice that we ultimately leave to the practitioner to make. For example, EGTD may be expressed in screen coordinates, such as pixels, or using real-world distance units, such as centimeters or inches. SWT and ST (and also PSWT, see next) are expressed in this work using percentages that normalize these measures with respect to the entire duration of the monitored time interval of the experiment. However, they could also be expressed using actual time units, e.g., seconds or minutes, should the practitioners employing them would actually need precise timing values of their viewers' television watching behaviors.

In the following, we define four *subjective* measures to characterize the perceived experience of viewers in terms of their distribution of visual attention in multi-screen TV environments:

9. *Perceived screen watching time (PSWT)* is defined as the percentage of the visual attention devoted to each screen during the monitored interval, as it was

perceived by viewers themselves. We show later in the paper how subjective PSWT correlates with objective SWT computed from actual eye-gaze-tracking data.

10. *Perceived comfort (PC)* is a subjective assessment of how comfortable the TV layout was for the viewer to watch. PC is measured on a 5-point Likert scale with values from 1 to 5 denoting very uncomfortable, uncomfortable, neutral, comfortable, and very comfortable perceptions.
11. *Maximum number of TV screens (MAX-TV)* is a subjective assessment denoting how many TV screens could be followed comfortably by the viewer at the same time.
12. *Content understandability (CU)* represents the capacity of the viewer to understand content delivered by multiple screens. CU is assessed by asking viewers questions about the content they watched and is measured as the percentage of correct answers. For example, in our experiment described in the next section, we asked one question of moderate difficulty for each TV screen in each layout.

## 4 Experiment

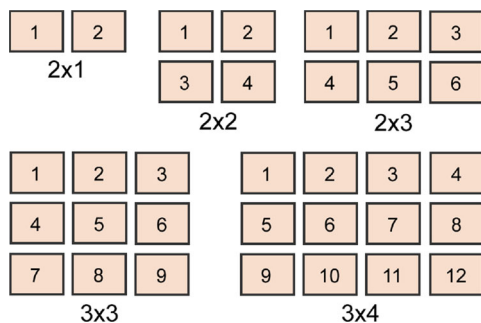
We conducted an experiment to exemplify how our set of measures can be applied in practice to study the effects that content displayed on multiple TV screens under various layouts have on viewers' distributions of visual attention. It is the main goal of this work to introduce new measures for characterizing viewers' visual attention patterns in the context of the interactive TV, which we want to demonstrate with a practical study. First, we inform the design of our experiment by running a preliminary study to understand a rough upper limit for the number of simultaneous TV screens (more than two) that could be followed at the same time. Next, we present an elaborate experiment design with multiple TV screens arranged in various layouts.

### 4.1 Preliminary study

Previous work showed that the number of screens, their form factors, and displayed content affect viewers' visual attention patterns in multi-display environments [9, 20, 24, 30, 31]. Therefore, we ran a preliminary experiment to inform on the upper limit of the number of TV screens that can be followed comfortably at the same time by viewers. Four participants were presented with five TV screen layouts composed of 2, 4, 6, 9, and 12 screens of

equal size arranged in tiled configurations of  $1 \times 2$ ,  $2 \times 2$ ,  $2 \times 3$ ,  $3 \times 3$ , and  $3 \times 4$  screens (see Fig. 6). Our choice for these layouts was inspired by previous work, such as Bi et al. [4] that used tiled screens ( $1 \times 1$ ,  $2 \times 2$ , and  $3 \times 3$ ) to evaluate users' performance during various interaction tasks. To prevent participants from visually privileging some screens over the others, all the screens displayed nonoverlapping sequences extracted from the same movie scene with sound turned off. All video sequences had 1 min in length. Each participant watched the movies individually (there was no social TV watching).

We found participants generally looking in the center of the tiled layouts as they tried to cover most of the information within their visual field (see Fig. 7 for visual attention heatmaps computed for the  $3 \times 3$  tiled layout). While the eye repartition remained well distributed in the case of two and four screens, we observed that the central screens took more importance and peripheral screens tended to be ignored for layouts with more than four screens, which confirms the spotlight model of Eriksen and Hoffman [11] for our specific screen layouts and television watching scenario. Also, our participants witnessed that more than 2–3 screens was too much to follow, because they were trying to make sense of the various sequences of the same movie, i.e., trying to put the pieces together. However, participants showed interest in watching multiple screens that would convey complementary information to a single, main screen. Therefore, findings at this stage revealed that the concept of a primary screen with an easily identifiable form factor (i.e., larger than all the other screens, for example) is important for viewers to easily understand the screen layout. These preliminary findings informed the design of our full experiment, for which we investigated in detail screen layouts composed of two, three, and four TV screens (see next).



**Fig. 6** Five tiled configurations of increasing complexity that we employed during a preliminary experiment to inform on the upper limit for the number of TV screens that can be comfortably followed by viewers at the same time

## 4.2 Participants

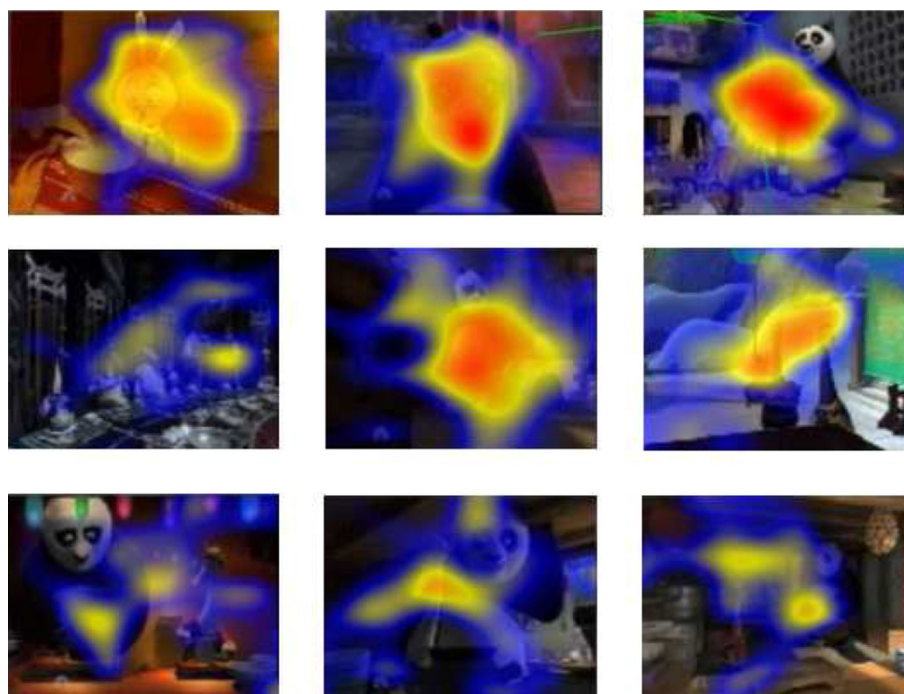
Ten volunteers (one female) participated in the experiment. Mean age was 27.9 years ( $SD = 3.7$  years). Although a small sample, the number of participants is more than enough to collect useful data and demonstrate how to apply our measures to reveal findings about visual attention; also, our sample size is consistent with other experiments conducted for visual attention and television [2, 24]. Our participants' self-reported daily average time for watching television was 1.5 h ( $SD = 2.1$  h). All participants had normal or corrected to normal vision.

## 4.3 Apparatus

TV screens were part of a large image ( $1.30 \times 0.87$  m) that was projected on a wall with a standard video projector (24.5 dpi). We adopted such a technical solution to implement our multi-screen setup due to its feasibility and practicality to easily configure arbitrary layouts of TV screens of various sizes and at various locations on a wall. The practical aspect of such a solution for rapid prototyping of multi-screen interactive TV systems and for running iTV experiments has been documented by Vatavu and Mancas [45] and implemented for multi-screen TV systems [41, 44]. Participants sat in a comfortable chair at a distance of 2.30 m from the projection. Given the fact that the projected screens did not fill the entire projected image, the maximum visual angle was  $25^\circ$  on the horizontal axis. The background of the projection was black, which gave participants the impression of multiple TV screens at the same depth on the same wall. Following the taxonomy of Rashid et al. [31], our screen layout possesses depth contiguity (i.e., all screens are at equal distance from the observer) and visual field discontinuity (i.e., screens are located in the same vertical plane, but spatially separated, which makes them appear as distinct displays instead of one visually contiguous screen). The movies displayed by each screen lasted 1 min each and were prepared in advance. The audio was turned off for all videos to isolate the effects of the visual information on attention. The FaceLab eye-tracking device<sup>1</sup> was employed during the experiment by following the practices of the visual attention community [15, 33] and those of previous experiment designs investigating visual attention for television watching [9, 20, 24].

<sup>1</sup> The FaceLab system is a head-free eye tracker, <http://www.seeingmachines.com/product/faceLab/>. We calibrated the tracker with a 9-dot grid and used it to record eye gaze at a rate of 60 Hz.

**Fig. 7** Eye gaze heatmaps for the  $3 \times 3$  tiled layout used in our preliminary experiment. Note how the *center* and *upper* screens attracted more of the participants' visual attention



#### 4.4 Design

The experiment was a within-participants design with two independent factors:

1. TV-COUNT, the number of distinct TV screens, with three values: 2, 3, and 4 screens, as informed by our preliminary study.
2. LAYOUT, representing the spatial arrangement of the TV screens and their sizes. For this factor, we designed three distinct conditions: TILED, PRIMARY, and ARBITRARY (see Fig. 8). In the TILED condition, all the screens have equal size and are arranged in a tiled, compact order. For PRIMARY, one screen acts as the main screen and is larger than all the rest, equally sized satellite screens. The ARBITRARY condition shows the screens in arbitrary sizes with random nonoverlapping locations.

We generated the TV layouts so that the total display area covered by constituent screens would be approximately the same for each layout, i.e., the size of the screens was larger for layouts with fewer screens. During the preliminary experiment, we displayed video sequences that were cut from the same movie scene. At that point, we adopted such an approach in order not to bias the visual attention of our participants to only some, presumably more captivating screens. However, we found that people were trying to put the individual movie pieces together in order to understand the full story, generating therefore a different visual attention pattern from what one would

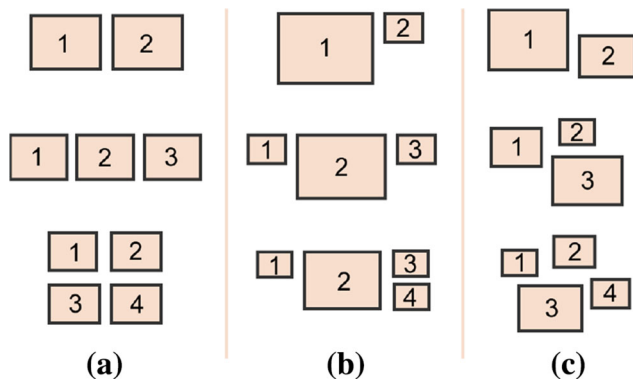
normally expect for general television watching. For the full experiment, all the TV screens displayed different content. However, we verified the content *a priori* by running a motion detector (employing frame-to-frame difference) to make sure that the motion level was roughly the same across all the screens of the same layout (the average motion level was 33 %, SD = 7 %).

#### 4.5 Task

Participants were asked to watch prerecorded movies for each combination of TV-COUNT and LAYOUT conditions resulting in a total number of nine experimental trials per participant and 9 min of watching TV (1 min per condition). Each participant watched the movies individually to eliminate any effect of social watching on visual attention. Participants were asked to watch the movies as if they were watching their TV set at home, and were told they had to answer a questionnaire after each trial in order to ensure a minimal level of attention. The order of the experimental conditions was randomized across participants. After each trial, participants were administered NASA TLX tests<sup>2</sup> using a computer version available online<sup>3</sup> to collect workload subjective ratings and they were handed questionnaires to evaluate their understanding of the content they had just watched. At the end of the experiment, participants filled a final questionnaire in which they reported

<sup>2</sup> <http://humansystems.arc.nasa.gov/groups/tlx/>.

<sup>3</sup> <http://www.keithv.com/software/nasatlx/>.



**Fig. 8** Screen layouts for the TILED (a), PRIMARY (b), and ARBITRARY (c) conditions with 2, 3, and 4 screens

the perceived comfortability (PC) of watching each layout. Participants were also asked to specify the maximum number of TV screens they would feel comfortable watching at the same time (MAX-TV), and to suggest their preferred layouts for 2, 3, and 4 screens by drawing them on paper.

#### 4.6 Hypotheses

We formulated the following hypotheses about viewers' distributions of visual attention in multi-screen TV scenarios to be verified in our experiment using our set of measures:

**Hypothesis A** Layouts with more screens will result in more user activity in terms of visual attention characterized by longer discovery times, more transitions, larger eye gaze travel distances, and longer switch times.

**Hypothesis B** Layout type will affect user activity in terms of visual attention characterized with the number of transitions between screens and the total distance traveled by viewers' eye gaze.

**Hypothesis C** Larger screens will receive more visual attention captured with screen watching times.

**Hypothesis D** There will be more transitions between screens during the discovery time than afterward, i.e., the discovery time acts as an accommodation period, during which viewers decide what to watch.

**Hypothesis E** More screens will result in higher cognitive load and lower perceived comfortability for viewers.

**Hypothesis F** Layout type will affect cognitive load and perceived comfortability of watching multi-screen television, e.g., some layouts may be more pleasant to look at with the effect of lowering cognitive effort as well.

**Hypothesis G** More screens and, implicitly, more content will result in lower rates of content understandability, e.g., there are more things to follow and

remember when more screens are available, but the short-term memory is nonetheless limited.

## 5 Results #1: Distribution of visual attention for multi-screen TV

### 5.1 Content discovery time

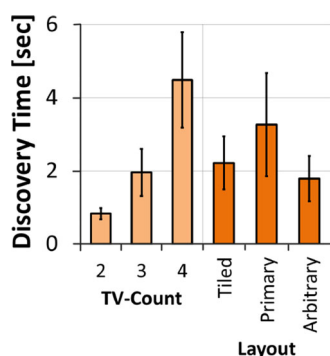
We found that our participants systematically attempted to discover content on each screen before committing to one screen to watch. In general, this process can be very fast as a single eye fixation is usually enough to roughly understand the topic being watched [29]. Discovery time values varied between 0.1 and 15.5 s for all our layout conditions with a mean time of 2.4 s ( $SD = 0.3$ ). We found a significant effect of TV-COUNT on discovery time ( $\chi^2(2) = 43.400$ ,  $p < .001$ ), showing that more time was needed by our participants to visually fixate more screens. For example, DT average values increased from 0.8 s for two screens up to 4.5 s for four screens (see Fig. 9). A second-degree polynomial showed a perfect fit with observed data ( $R^2 = 1.0$ ), suggesting that discovery time relates to the number of screens in a quadratic manner:

$$DT = 0.70 \times TV-COUNT^2 - 0.97 \times TV-COUNT + 1.10 \quad (9)$$

These results confirm Hypothesis A, according to which layouts with more screens will result in more user activity in terms of visual attention allocated to those screens, attention that we report from the perspective of the time needed to understand content displayed on all screens. We did not detect any significant effect of LAYOUT on discovery time ( $\chi^2(2) = 2.867$ ,  $n.s.$  at  $p = .01$ ).

### 5.2 Discovery sequences

The discovery sequence informs about the order in which individual screens are attended visually during the discovery time. For two screens, there are only two possible sequences, i.e., 1–2 and 2–1, and we found our participants preferring the former for all layouts, with preference counts of 8 out of 10 for the TILED layout, 10/10 for PRIMARY, and 9/10 for ARBITRARY (For convenience, screen numbers are shown in Fig. 8 for each layout.). For three screens, there are  $3! = 6$  possible sequences, out of which the permutation 2–3–1 occurred the most for TILED (5 out of 10), 2–1–3 for PRIMARY (7/10), and no majority preference could be identified for ARBITRARY. For four screens, there are  $4! = 24$  possible sequences, out of which 1–2–4–3 occurred the most for TILED (4/10), 2–1–3–4 for PRIMARY (4/10), and again, there was no majority preference for ARBITRARY,



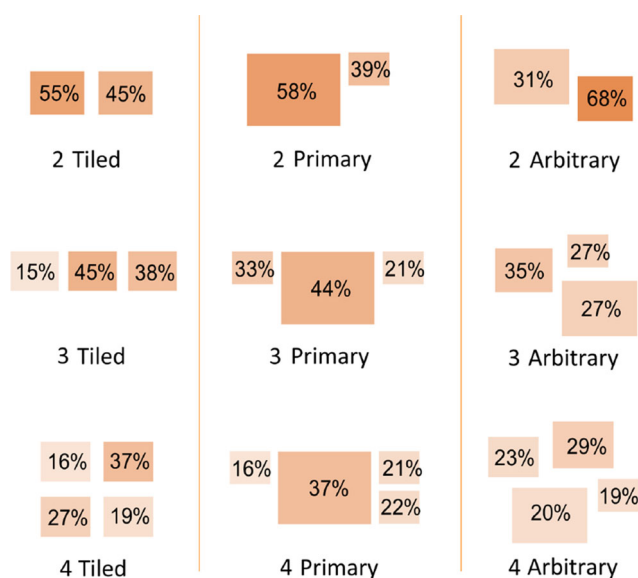
**Fig. 9** Average discovery times (in seconds) necessary for our participants to visually fixate all screens for all our TV-COUNT × LAYOUT experimental conditions. Note how discovery times increase with the number of screens (left), but remain roughly the same across different layout types (right). Note error bars show 95 % CI

for which eight different sequences were observed among the ten trials. These results show that viewers discover screens in order from left to right for 2-screen scenarios (i.e., the sequence 1–2); viewers are first attracted by the middle screen when three screens are present (sequences 2–3–1 and 2–1–3); and viewers follow a counterclockwise pattern (e.g., 1–2–4–3) in the absence of a primary screen to attract attention first (e.g., 2–1–3–4 for PRIMARY).

The discovery sequence points out the impact of the layout on attention. For example, a left-to-right model was adopted by our participants for screens of equal size, which corresponds to the reading order in Western culture (which was the case for our study). For the PRIMARY layouts, content discovery starts from the larger screen despite that it was not the leftmost screen. This finding shows that our participants immediately identified the largest screen as the main or primary one. The anticlockwise pattern is also interesting as it builds on the left-to-right model, but also exploits the shortest distance between screens. Consequently, it may represent an instance of the Z-shaped pattern observed during reading [32], but specific for multi-screen TV.

### 5.3 Distribution of screen watching time

The distribution of our participants’ visual attention between screens is illustrated in Fig. 10 using color codes, with darker colors denoting more visual attention, i.e., more eye gaze points falling within the boundaries of those screens. We found significant differences between the TILED and PRIMARY layouts with three and four screens, while only the ARBITRARY layout had a significant effect on SWT for two screens ( $p < .01$ ). We also found that screen watching time is related to the size of the screen, i.e., the large screen in all the PRIMARY conditions received more visual attention (see Fig. 10), a result that confirms our



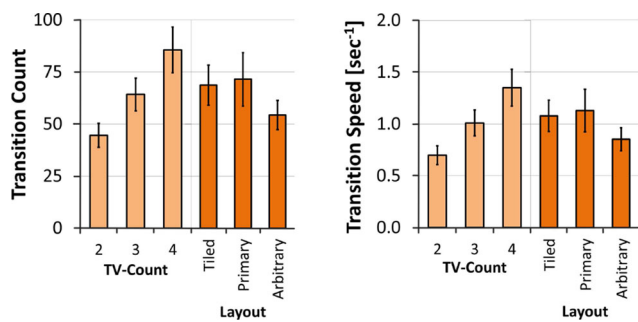
**Fig. 10** Screen watching times, shown in percentages, for all the TV-COUNT × LAYOUT conditions. Note: screen watching times do not always add to 100 % because there is some time lost while switching attention between screens, which is captured by our switch time measure (see next)

expectation from Hypothesis C. Our observations also revealed that screen watching time is also related with the type of content being displayed, as we found out by asking participants, e.g., participants’ visual attention was attracted more to the right screen of the 2-ARBITRARY condition that displayed a bicycle race rather than the first screen showing a news broadcast, resulting in 68 and 31 % levels of attention, respectively.

SWT can be further visualized as heatmaps (see Fig. 3 for eye gaze heatmaps for all our nine layouts) that use color codes to describe eye gaze density spatially along the screen. For layouts with screens with equal size, the eye gaze density reflects the SWT values exactly, e.g., low color density for the first screen, larger for the central, and moderate for the third in the 3-TILED condition (see both Figs. 3 and 10). However, the gaze density color of the largest screen has lower maximum in the PRIMARY condition, because gaze is distributed across a larger area, i.e., larger coverage.

### 5.4 Transition count and speed

The number of eye gaze transitions between screens varied from 11 to 200 for all our screen layouts during the monitoring interval of 1 min, with a mean value of 64.8 transitions (SD = 28.9) (see Fig. 11). We found significant effects for both TV-COUNT ( $\chi^2(2) = 41.667, p < .001$ ) and LAYOUT ( $\chi^2(2) = 10.237, p < .001$ ) with no significant difference between the TILED and PRIMARY conditions for



**Fig. 11** Average transition count and speed values for each TV-COUNT and LAYOUT experimental conditions. Note the increase in both transition time and speed for layouts with more screens. Note error bars show 95 % CI

LAYOUT). We also found a significant increase in eye gaze speed with more screens ( $\chi^2(2) = 41.667, p < .001$ ), as well as a significant effect of LAYOUT on transition speed ( $\chi^2(2) = 10.067, p < .01$ ). Figure 11, left reveals an expected yet strong positive correlation ( $R^2 = 1.0$ ) between TV-COUNT and TC: more screens determine more transitions during the first minute of television watching:

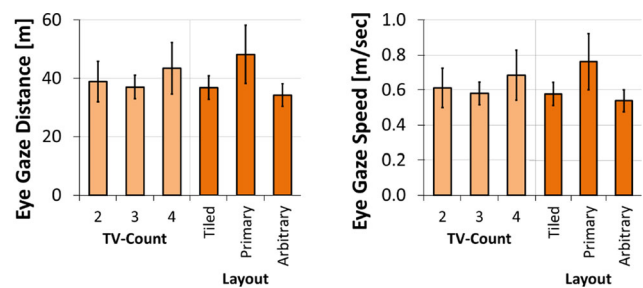
$$TC = 20.5 \times TV-COUNT + 23.9 \quad (10)$$

These results confirm the expectation of Hypothesis A that layouts with more screens result in more visual attention allocated to those screens, which is measured here from the perspective of the number of transitions performed by viewers' eye gaze between screens. Results also confirm Hypothesis B, according to which layout type affects viewers' visual activity, measured here in terms of the transition count measure.

We also found that the ARBITRARY layout led to significantly less transitions. This result may be explained by the fact that looking in the center of ARBITRARY layouts covers most of the screens that are close to this centered point of focus. Consequently, there is less need for the eye gaze to actually transition to other screens, as the peripheral information is already available to viewers' visual attention and is being accordingly processed, as explained by the zoom-lens model [12].

### 5.5 Eye gaze travel distance and speed

During the 1 min of each experimental trial, participants' eye gaze traveled on average 39.7 m (see Fig. 12, left) at an average speed of 0.63 m/s (Fig. 12, right). Interestingly, we did not detect any significant effect of the number of TV screens on EGTD ( $\chi^2(2) = 1.667, n.s.$  at  $p = .01$ ). This result partially invalidates Hypothesis A from the perspective of the EGTD measure, although significant more visual attention was detected by other measures. However, we found a significant effect of LAYOUT on EGTD

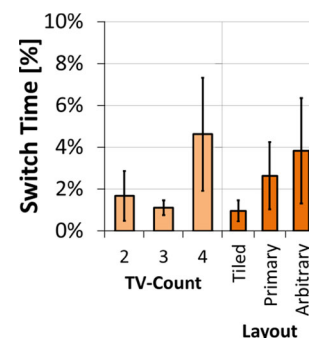


**Fig. 12** Average eye gaze travel distances and speeds for each TV-COUNT and LAYOUT experimental conditions. Note error bars show 95 % CI

( $\chi^2(2) = 8.867, p = .01$ ), which confirms Hypothesis B, according to which layout influences viewers' distribution of visual attention. Similar results were found for the eye gaze travel speed measure: no significant effect of TV-COUNT on EGTS ( $\chi^2(2) = 2.467, n.s.$ ), and a significant effect of LAYOUT on EGTS ( $\chi^2(2) = 9.800, p < .011$ ). Post hoc Wilcoxon tests (corrected at  $p = .05/3 = .017$ ) revealed significant differences only between the PRIMARY and ARBITRARY layouts, but not between the other two layout pairs. Participants seemed to have traveled roughly the same distance in terms of eye gaze for both TILED and PRIMARY, but the large screen of the PRIMARY condition led to larger eye gaze travel distances to reach the secondary screens when compared to the ARBITRARY layout.

### 5.6 Switch time

Switch time is the time required for eye gaze to travel between screens and is expressed as percentage of the total duration of the monitoring time interval. Overall, we found ST to vary up to 29 % of the total monitored time with an average of 2.5 % (SD = 5 %). We found a significant effect of TV-COUNT on ST ( $\chi^2(2) = 8.824, p = .01$ ) with post hoc tests showing significant differences only between two and four screens (see Fig. 13). These results confirm the



**Fig. 13** Average time to switch eye gaze between screens, reported as percentage for all TV-COUNT and LAYOUT experimental conditions. Note error bars show 95 % CI

**Table 1** The Pearson correlation coefficients between our objective measures ( $N = 90$ )

Measure	Measure	Pearson's $r$
Transition speed (TS)	Eye gaze travel distance (EGTD)	.650**
Transition speed (TS)	Eye gaze travel speed (EGTS)	.646**
Transition count (TC)	Eye gaze travel distance (EGTD)	.641**
Transition count (TC)	Eye gaze travel speed (EGTS)	.636**
Switch time (ST)	Eye gaze travel speed (EGTS)	.410**
Switch time (ST)	Eye gaze travel distance (EGTD)	.406**
Discovery time (DT)	Switch time (ST)	.349**

Only correlations with coefficient values greater than .315 are shown in this table (i.e., at least 10 % shared variance between the corresponding measures); correlations are presented in descending order of their Pearson's  $r$  values

\*\* Correlations are significant at the  $p = .01$  level

expectation of Hypothesis A that layouts with more screens will result in more user activity in terms of visual attention allocated to those screens, which is measured here from the perspective of the time needed to switch visual attention between different screens. We did not detect any significant effect of LAYOUT on switch time ( $\chi^2(2) = 2.867$ ,  $n.s.$  at  $p = .01$ ). Overall, this measure revealed that 4-screen layouts are less efficient in terms of actually fixating TV content, as they unnecessarily consume eye gaze for transitions between screens.

### 5.7 Correlations between visual attention measures

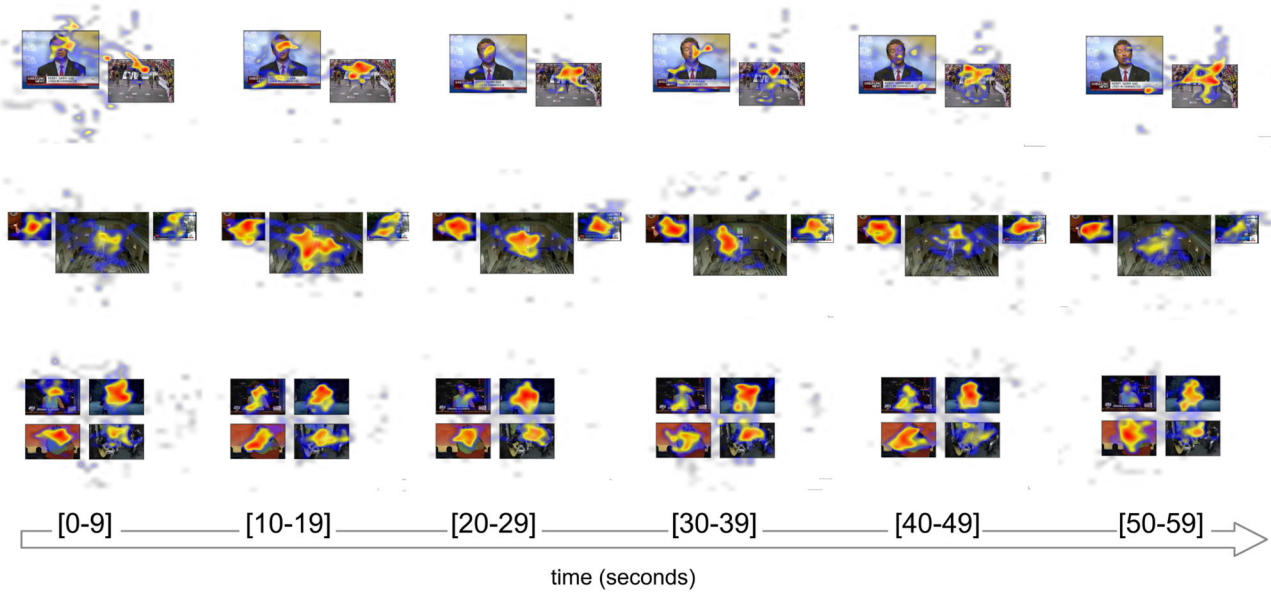
It is important to understand how our measures are correlated in order to evaluate their capacity to reflect and characterize different aspects of viewers' distributions of visual attention. Table 1 lists the values of Pearson's  $r$  correlation coefficients computed for our measures,<sup>4</sup> for which the magnitude of the correlation was at least .315 large (i.e., the corresponding measures share at least 10 % of their variance). We found high positive correlations between transition count (TC) and eye gaze travel distance (EGTD) ( $r = .641$ ), as well as between transition speed (TS) and eye gaze travel speed (EGTS) ( $r = .646$ ), significant at  $p = .01$ . These results are intuitive, confirming the fact that the more viewers transition between screens, the more their eye gaze travel distance will increase. Overall, transition and eye gaze travel measures share 40 % of their variance. However, these measures reflect different aspects of the interaction, i.e., how many times viewers switch screens versus how much their eye gaze travels both between and within screens. We also found high correlations between switch time (ST) and the two Eye Gaze Travel measures ( $r = .410$  with gaze speed and  $r = .406$  with gaze distance,  $p = .01$ ), showing that the

more eye gaze travels, the longer the switch time will also be (the shared variance between these measures is approximately 20 %). Finally, we found that discovery time (DT) correlated significantly and positively with switch time ( $r = .349$ ,  $p = .01$ ), suggesting a connection between the amount of time viewers take to understand content initially and the total time they spend afterward switching between content on the various screens of the layout; however, the shared variance between the two measures is small (12 % only).

Note that we do not show in Table 1 correlations between transition count and Speed (TC and TS), nor between eye gaze travel distance and speed (EGTD and EGTS), because they are .999 (i.e., perfect correlation), as we computed our speed measures by dividing transition count and travel distance by the duration of the time interval. However, speed measures are useful to complement the distance measures they were derived from by offering a different perspective for reporting results that are normalized across the time duration of the monitored intervals.

We also note that these correlation results are valid when the content of each screen remains the same over time (i.e., the same show) as in our experiment. Should that be the case, the discovery time is used by viewers to build a model of the kind of content broadcasted by each screen. The time-spent modeling helps viewers to be more efficient afterward in terms of distributing their visual attention across the entire layout of multiple TV screens. Should the composition of screen content type change (e.g., a new show starts on one of the screens), a new discovery period will probably be needed again for viewers to update their model about the content displayed by that screen. According to the classification of television looks reported in Hawkins et al. [19], the discovery period would consist mostly of monitoring looks (i.e., looks that last up to 1.5 s) and, possibly, yet less likely, of orienting looks (i.e., looks that last beyond 1.5 s and up to 5 s).

<sup>4</sup> Except for the screen watching time and discovery sequence measures, which return more than one value and, consequently, cannot be included in this analysis.



**Fig. 14** Evolution of viewers’ distributions of visual attention over multiple consecutive 10-s time intervals for three layouts: 2-ARBITRARY (*top*), 3-PRIMARY (*middle*), and 4-TILED (*bottom*)

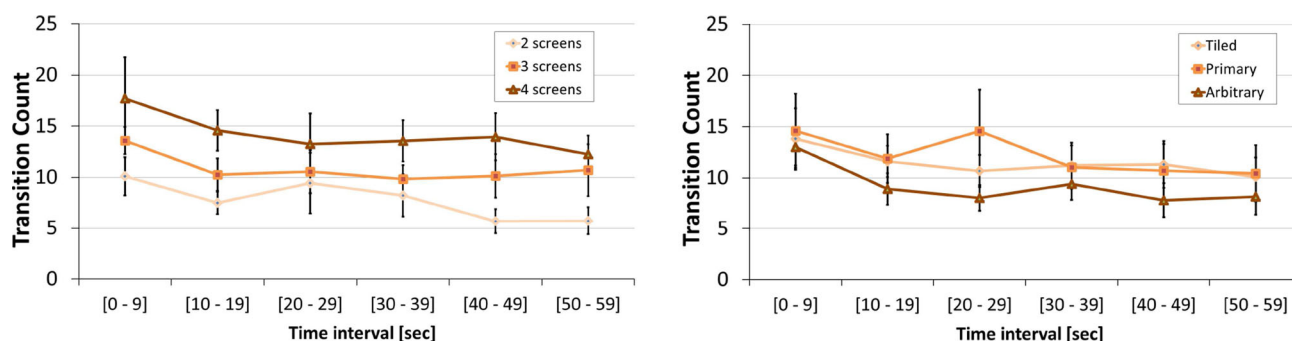
## 6 Results #2: Distribution of viewers’ visual attention over time

The previous section investigated viewers’ distributions of visual attention from the perspective of their average performance during the entire monitored time interval. In this section, we break down our previous analysis by looking at how visual attention unfolds over time. For example, Fig. 14 shows an overview of viewers’ average distributions of visual attention represented as eye gaze heatmaps computed for consecutive time intervals. As mentioned previously in the paper, heatmaps show the spatial distribution of eye gaze and, consequently, are able to reflect an overall trend, e.g., our participants seemed to have looked more at the primary screen of the 3-PRIMARY condition (see Fig. 14, middle) for the first 30 s of the experiment, after which their attention was drawn more to the two smaller screens located around the primary display. However, eye gaze heatmaps are less informative when it comes to analyze quantitatively the various aspects of viewers’ visual behaviors, such as those captured by our set of measures. In this section, we look at how the values reported by our measures change over time and we draw conclusions about viewers’ visual attention from the time distribution of our measures. In the following, we report data for transition count, eye gaze travel distance, and switch time that we select as representative measures from our set and for which we compute their values for time intervals of 10 s long. Since each experimental trial took 60 s to complete, we report and work with six distinct values for each measure.

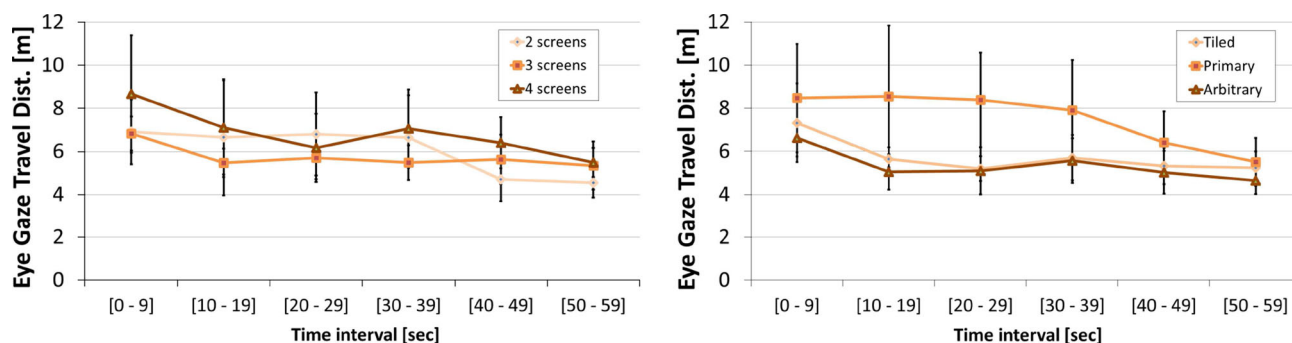
We found that the number of eye gaze transitions between screens remained relatively constant for each 10-s time interval under both TV-COUNT and LAYOUT factors (see Fig. 15). However, we detected a significant effect of the time interval on TC ( $\chi^2(5) = 44.330, p < .001$ ), but post hoc Wilcoxon signed-rank tests between consecutive intervals showed that only the first interval was significantly different from the second (at  $p = .05/5 = .01$ ). In average, our participants performed more eye gaze transitions during the first 10 s of watching (TC = 13.8), after which TC became relatively stable at approximately one transition per second. This result confirms our expectation set out in Hypothesis D, according to which the discovery time will generate more transitions than the other subintervals, because it is during that period that viewers decide what is interesting to watch.

We found a significant effect of time interval on the distance traveled by viewers’ eye gaze ( $\chi^2(5) = 40.590, p < .001$ ) (see Fig. 16). Paired post hoc comparisons (conducted with the Wilcoxon signed-rank test, Bonferroni corrected at  $p = .05/5 = .01$ ) showed significant differences between the first and second intervals only ( $Z = -4.364, p < .01$ ). During the first 10 s, eye gaze traveled in averaged 7.5 m, after which it became relatively stable at 5.9 m.

We also found a significant effect of time interval on the time consumed for switching attention between screens ( $\chi^2(5) = 36.287, p < .001$ ) (see Fig. 17). Post hoc Wilcoxon signed-rank tests (Bonferroni corrected at  $p = .05/5 = .01$ ) only showed a significant difference between the first and the second intervals ( $Z = -5.143, p < .01$ ). After



**Fig. 15** Evolution of the average number of transitions between screens over consecutive 10-s time intervals for each TV-COUNT (left) and LAYOUT (right) experimental conditions



**Fig. 16** Evolution of average eye gaze travel distance over multiple consecutive 10-s time intervals for TV-COUNT (left) and LAYOUT (right) experimental conditions

the first 10 s, for which the switch time was 3.8 %, switching screens lowered to 2.3 %.

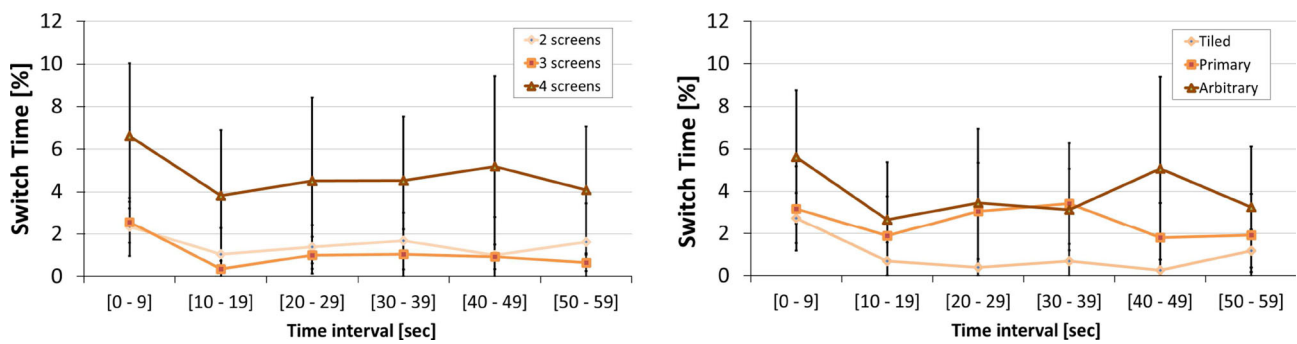
These results are in agreement with the findings reported in the previous section. Viewers tend to build a model of the content displayed on each screen during the discovery period, which is characterized by longer times for screen switching, longer eye gaze travel distances, and more transitions between screens. This sort of scene modeling is performed during the first few seconds. After that, viewers are more efficient in terms of distributing their visual attention, which we observed with a decrease in the average values of our reported measures. Some variation still remains, but we suspect it is related to changes in content on some of the screens.

## 7 Results #3: Viewers' cognitive load during multi-screen television

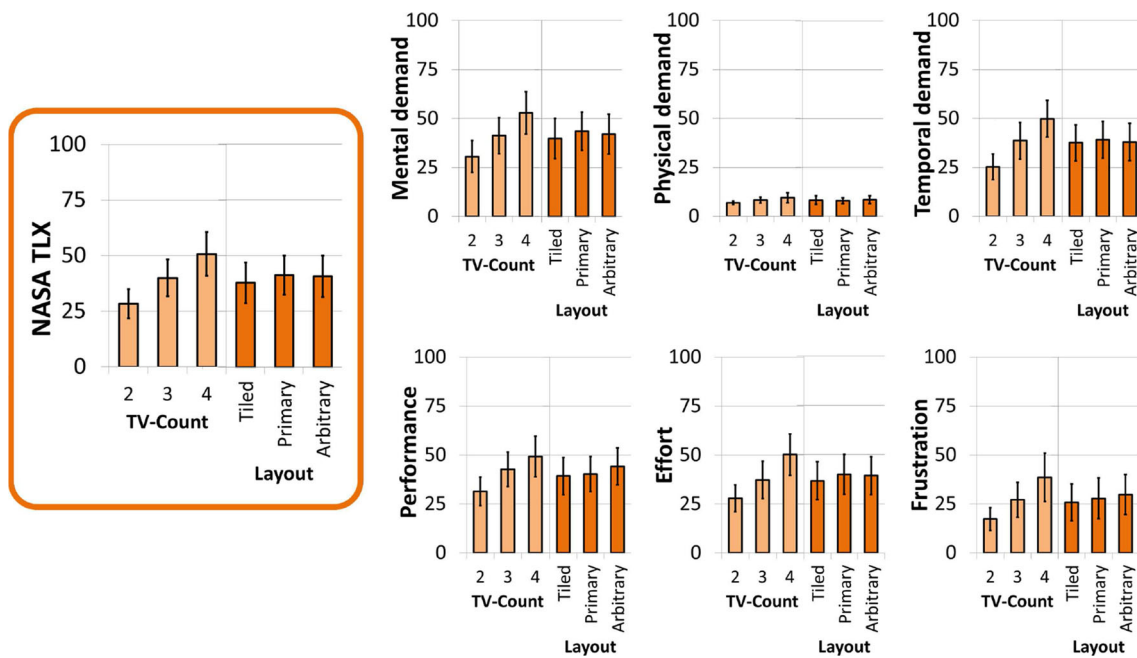
### 7.1 Cognitive load

After each trial, participants were administered NASA TLX tests to collect their subjective ratings of perceived

workload on a scale from 1 (low) to 100 (high workload). We found that the TLX value increased with the number of TV screens from 28.4 for two screens to 39.9 and 50.7 for three and four screens, respectively (see Fig. 18, left). More TV screens were perceived more difficult to follow, as shown by a Friedman test ( $\chi^2(2) = 27.214, p < .001$ ). Post hoc Wilcoxon signed-rank tests revealed significant differences (at  $p = .05/2 = .025$ ) between two and three, and three and four screens, with medium-to-large effect sizes ( $r = .42$  and  $r = .50$ , respectively). Significant effects of TV-COUNT were found for each dimension of the NASA TLX test (Fig. 18, right). These results confirm Hypothesis E, according to which we expected more screens to increase the workload perceived by viewers. We did not detect any significant effect of LAYOUT on the perceived task load measured by TLX ( $\chi^2(2) = 4.206, n.s.$  at  $p = .01$ ), nor on any of the six dimensions employed by the NASA TLX test. These results invalidate Hypothesis F in which we proposed that some layout presentations may look visually pleasing to viewers and, therefore, result in lower perceived work load. However, that was not the case with our participants and TV layouts.



**Fig. 17** Evolution of switch time between screens over multiple consecutive 10-s time intervals for each TV-COUNT (left) and LAYOUT (right) experimental conditions



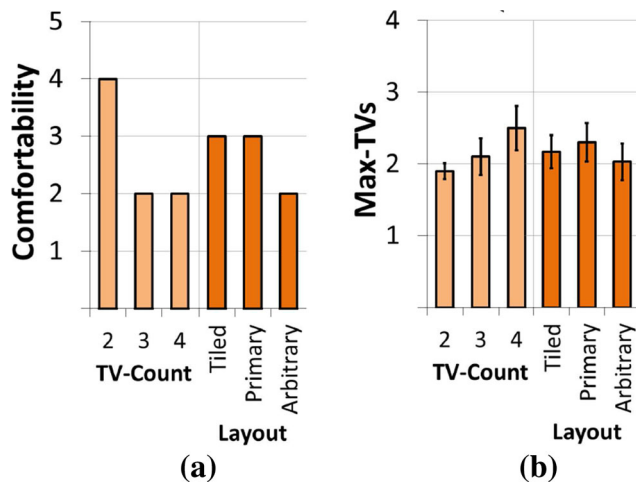
**Fig. 18** Participants' average workload ratings measured with the NASA task load test for each TV-COUNT and LAYOUT experimental conditions (left). Notes The NASA TLX test employs six dimensions (right) to measure workload in the range [0, 100] corresponding to the

subjective perceptions of Low/High (e.g., high mental demand, low physical effort) and Poor/Good for performance. Error bars show 95 % CI

## 7.2 Perceived comfortability

At the end of the experiment, we asked participants to rate the perceived comfort (PC) of watching each screen layout on a 5-point Likert scale. They were also asked to specify the maximum number of screens they felt could be watched comfortably at the same time (MAX-TV). The median PC rating over all trials was 2.5, in between “uncomfortable” and “neutral” (Fig. 19a). Maximum comfortability was perceived for layouts with two screens (4, “comfortable”), while perceived comfortability was 2 (“uncomfortable”) for layouts with more than two screens. We found a significant effect of TV-COUNT on PC ( $\chi^2(2) = 39.244, p < .001$ ) that was further confirmed by post hoc Wilcoxon

signed-rank tests (corrected at  $p = .05/3 = .017$ ) for paired TV-COUNT conditions (2,3) and (2,4) with large effect sizes ( $r = .51$  and  $r = .57$ , respectively). These results confirm Hypothesis E, according to which more screens would negatively affect perceived comfortability. However, no significant effect was detected between three and four screens, which suggests that the perceived comfortability may drop to a stable low level beyond three or four screens. We also found a significant effect of LAYOUT on PC ( $\chi^2(2) = 23.275, p < .001$ ). Post hoc Wilcoxon signed-rank tests revealed significant differences (at  $p = .05/3 = .017$ ) between PRIMARY and ARBITRARY ( $r = .44$ ), and TILED and ARBITRARY ( $r = .47$ ), but not between PRIMARY and TILED. These results make plausible



**Fig. 19** Median ratings of perceived comfort (a) and the average of the maximum number of TV screens (MAX-TV) our participants felt could be comfortably watched at the same time (b)

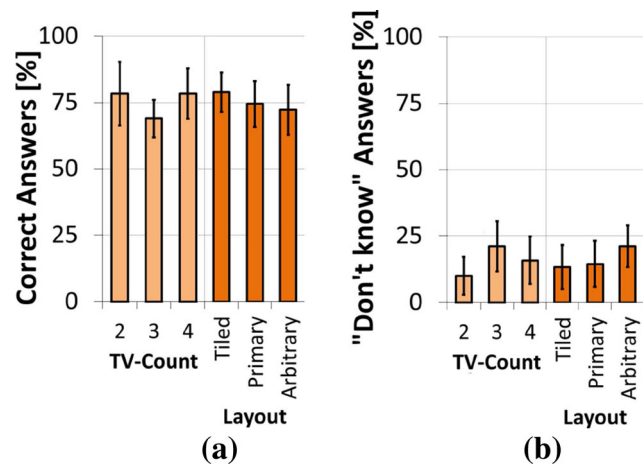
Hypothesis F (despite that perceived cognitive load was not affected by LAYOUT, see the previous section) and suggest that esthetic screen layouts could potentially be designed to compensate for the negative effects that more screens impose on work load.

The median value of MAX-TV was two screens (Fig. 19b). We found a significant effect of TV-COUNT on MAX-TV ( $\chi^2(2) = 13.565$ ,  $p < .001$ ), no significant difference between two and three screens, but significant differences between the pairs (3,4) and (2,4) with effect sizes  $r = .36$  and  $r = .40$ , respectively. We did not detect any significant effect of LAYOUT on MAX-TV ( $\chi^2(2) = 5.261$ , *n.s.* at  $p = .01$ ).

## 8 Results #4: Viewers' capacity to understand content and the perceived screen watching time

### 8.1 Content understandability

After each trial, participants were administered multiple-choice questions regarding the content displayed on each screen, with one question per screen. Each question had four possible choices with only one being correct. The last choice was always "I don't know the answer". We counted the number of correct answers as well as the number of "don't know" answers. We found that participants were able to remember content with an average accuracy of 75.2 %, while the percentage of "don't know" answers was 16.3 % (Fig. 20). We found no significant effects of TV-COUNT or LAYOUT on the mean number of correct answers ( $\chi^2(2) = 4.000$  and  $\chi^2(2) = 0.970$ , respectively, *n.s.* at  $p = .01$ ), but we found a marginally significant effect of TV-COUNT on the number of "don't know"



**Fig. 20** Participants' understandability of content measured as the accuracy of their answers (a) and the percentage of the time participants did not remember content (b)

answers ( $\chi^2(2) = 6.645$ ,  $p = .036$ ). These results invalidate Hypothesis G, according to which we anticipated that more screens would result in fewer correct answers about content. Instead, it seems that four screens are not too many for viewers to lose track of what they watch. Note that the hypothesis might hold true for more than four screens (as suggested by our marginally significant effect on "don't know" answers) but, for the moment, we leave more detailed explorations for future work.

### 8.2 Perceived watching time

After each trial, participants estimated in percentages how much they thought they watched each screen. When we correlated their answers with ground truth screen watching time computed from the eye tracker equipment, we found an overall Pearson's correlation coefficient of  $r = .763$ , significant at  $p = .01$ . This result shows a surprisingly good capacity of our participants to estimate what they were actually watching and for how long. Correlation coefficients computed for each experimental condition are shown in Table 2, with a maximum of  $r = .892$  for the 2-ARBITRARY layout. These high correlations suggest that the PSWT measure could be used to evaluate viewers' multi-screen television watching experience reliably, even in the absence of an eye-gaze-tracking device.

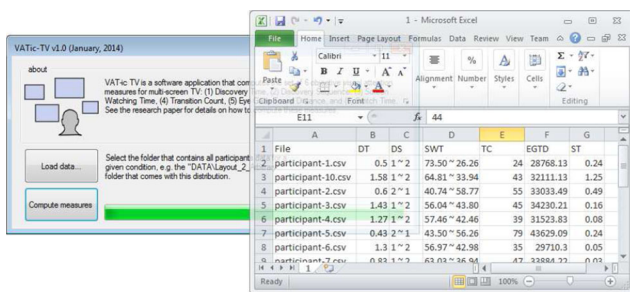
## 9 Toolkit for visual attention measures

We release our set of measures in the form of a software update for the Visual Attention Toolkit for TV (VATic-TV, see Fig. 21 for a snapshot) originally introduced by Vatavu and Mancas [46], which has now reached version v2 with the new measures from this work. We make the toolkit

**Table 2** Pearson's correlation coefficients ( $N = 40$ ) computed between objective screen watching time (SWT) and subjective perceived screen watching time (PSWT)

Condition (TV-COUNT $\times$ LAYOUT)	Pearson's $r$
2-Tiled	.848**
2-Primary	.755**
2-Arbitrary	.892**
3-Tiled	.676**
3-Primary	.519**
3-Arbitrary	<i>n.s.</i>
4-Tiled	.819**
4-Primary	.727**
4-Arbitrary	.686**

\*\* Correlations are significant at  $p = .01$



**Fig. 21** Snapshot of VATic-TV, the Visual Attention Toolkit for TV, originally introduced by Vatavu and Mancas [46], which has reached version v2 in this work. The toolkit computes the set of eight objective measures for evaluating visual attention from eye gaze text data files

available to the community as open-source software companion to this paper contribution. VATic-TV v2 can be downloaded for free from the web page <http://www.eed.usv.ro/~vatavu>. VATic-TV v2 computes our set of eight objective measures from eye gaze data encoded as .txt files and reports results in the .csv format, ready to be used by researchers and practitioners for their own data reporting and analysis. We also release all the data files that were collected and used in this study, i.e., the nine movies of the experiment and the ten eye gaze data logs, which constitute into a multi-screen eye gaze dataset to allow easy replication of our results and encourage further investigation of visual attention phenomena for multi-screen television scenarios and applications in the community.

## 10 Conclusion

We proposed in this work a set of 12 general and reusable measures to characterize viewers' visual attention patterns for multi-screen TV, out of which eight measures can be

computed automatically with the software toolkit accompanying the paper. We applied our measures to evaluate multi-screen TV layouts and showed how the number of screens and their arrangement in space affect viewers' distributions of visual attention and cognitive load. We also used our measures to characterize the way viewers' visual attention unfolds in time. We also reported and analyzed viewers' subjective perceptions and preferences for multi-screen TV layouts.

We believe that this initial investigation on designing new measures to evaluate and understand multi-screen TV scenarios will foster further developments in the interactive TV community toward a sound methodological framework for evaluating user and system performance for new multi-screen television. Such a methodology will enable examination of many interactive multi-screen scenarios. For instance, interesting lines of work are to evaluate viewers' distributions of attention for multi-screen systems composed of physical TV screens that may be distributed at different depths in the room, hybrid TV systems that employ one or more smart devices (e.g., smartphones, tablets, watches, etc.) connected to one or more Smart TVs, and evaluating the effect of the physicality of displayed content on visual attention for augmented reality environments in which both physical and projected screens coexist [41, 44]. Moreover, as the goal of our experimental examination in this paper was mainly to demonstrate how our set of measures can be applied in practice (for which we considered a minimal number of participants), other fruitful lines of work are new experimental designs to evaluate more conditions (i.e., more screens and different screen layouts) with more participants, as well as *in situ* studies conducted with various ethnographic groups. From this perspective, we look forward to see how our measures will be employed by the community to understand more about viewers' visual attention patterns for emerging multi-screen TV applications.

**Acknowledgments** This work was partially supported from the project "Integrated Center for research, development and innovation in Advanced Materials, Nanotechnologies, and Distributed Systems for fabrication and control", Contract No. 671/09.04.2015, Sectoral Operational Program for Increase of the Economic Competitiveness co-funded from the European Regional Development Fund.

## References

- Anderson J (2004) Cognitive psychology and its implications. Worth Publishers, New York
- Basapur S, Harboe G, Mandalia H, Novak A, Vuong V, Metcalf C (2011) Field trial of a dual device user experience for iTV. In: Proceedings of the 9th international interactive conference on interactive television, ACM, New York, NY, USA, EuroITV '11, pp 127–136. doi:10.1145/2000119.2000145

3. Basapur S, Mandalia H, Chaysinh S, Lee Y, Venkitaraman N, Metcalf C (2012) FANFEEDS: evaluation of socially generated information feed on second screen as a TV show companion. In: Proceedings of the 10th European conference on interactive TV and video, ACM, New York, NY, USA, EuroITV '12, pp 87–96. doi:[10.1145/2325616.2325636](https://doi.org/10.1145/2325616.2325636)
4. Bi X, Bae SH, Balakrishnan R (2010) Effects of interior bezels of tiled-monitor large displays on visual search, tunnel steering, and target selection. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, New York, NY, USA, CHI '10, pp 65–74. doi:[10.1145/1753326.1753337](https://doi.org/10.1145/1753326.1753337)
5. Borji A, Itti L (2013) State-of-the-art in visual attention modeling. *IEEE Trans Pattern Anal Mach Intell* 35(1):185–207
6. Cesar P, Bulterman DC, Jansen AJ (2008) Usages of the secondary screen in an interactive television environment: control, enrich, share, and transfer television content. In: Proceedings of the 6th European conference on changing television environments, Springer, Berlin, EuroITV '08, pp 168–177. doi:[10.1007/978-3-540-69478-6\\_22](https://doi.org/10.1007/978-3-540-69478-6_22)
7. Chorianopoulos K, Fernández FJB, Salcines EG, de Castro Lozano C (2010) Delegating the visual interface between a tablet and a TV. In: Proceedings of the international conference on advanced visual interfaces, ACM, New York, NY, USA, AVI '10, pp 418–418. doi:[10.1145/1842993.1843096](https://doi.org/10.1145/1842993.1843096)
8. Courtois C, D'heer E (2012) Second screen applications and tablet users: constellation, awareness, experience, and interest. In: Proceedings of the 10th European conference on interactive TV and video, ACM, New York, NY, USA, EuroITV '12, pp 153–156. doi:[10.1145/2325616.2325646](https://doi.org/10.1145/2325616.2325646)
9. Cummins R, Tirumala L, Lellis J (2011) Viewer attention to espns mosaic screen: an eye-tracking investigation. *J Sports Media* 6(1):23–54
10. Dezfuli N, Khalilbeigi M, Huber J, Müller F, Mühlhäuser M (2012) Palmrc: imaginary palm-based remote control for eyes-free television interaction. In: Proceedings of the 10th European conference on interactive TV and video, ACM, New York, NY, USA, EuroITV '12, pp 27–34. doi:[10.1145/2325616.2325623](https://doi.org/10.1145/2325616.2325623)
11. Eriksen CW, Hoffman JE (1972) Temporal and spatial characteristics of selective encoding from visual displays. *Percept Psychophys* 12(2):201–204. doi:[10.3758/BF03212870](https://doi.org/10.3758/BF03212870)
12. Eriksen CW, St James JD (1986) Visual attention within and around the field of focal attention: a zoom lens model. *Percept Psychophys* 40(4):225–240. doi:[10.3758/BF03211502](https://doi.org/10.3758/BF03211502)
13. Fitts PM (1954) The information capacity of the human motor system in controlling the amplitude of movement. *J Exper Psychol* 47(6):381–391. doi:[10.1037/h0055392](https://doi.org/10.1037/h0055392)
14. Forlines C, Shen C, Wigdor D, Balakrishnan R (2006) Exploring the effects of group size and display configuration on visual search. In: Proceedings of the 2006 20th anniversary conference on computer supported cooperative work, ACM, New York, NY, USA, CSCW '06, pp 11–20. doi:[10.1145/1180875.1180878](https://doi.org/10.1145/1180875.1180878)
15. Frintrop S, Rome E, Christensen HI (2010) Computational visual attention systems and their cognitive foundations: a survey. *ACM Trans Appl Percept* 7(1):6:1–6:39. doi:[10.1145/1658349.1658355](https://doi.org/10.1145/1658349.1658355)
16. Geerts D, Cesar P, Bulterman D (2008) The implications of program genres for the design of social television systems. In: Proceedings of the 1st international conference on designing interactive user experiences for TV and video, ACM, New York, NY, USA, UXTV '08, pp 71–80. doi:[10.1145/1453805.1453822](https://doi.org/10.1145/1453805.1453822)
17. Geerts D, Leenheer R, De Grooff D, Negenman J, Heijstraten S (2014) In front of and behind the second screen: viewer and producer perspectives on a companion app. In: Proceedings of the 2014 ACM international conference on interactive experiences for TV and online video, ACM, New York, NY, USA, TVX '14, pp 95–102. doi:[10.1145/2602299.2602312](https://doi.org/10.1145/2602299.2602312)
18. Grüniger J, Krüger J (2013) The impact of display bezels on stereoscopic vision for tiled displays. In: Proceedings of the 19th ACM symposium on virtual reality software and technology, ACM, New York, NY, USA, VRST '13, pp 241–250. doi:[10.1145/2503713.2503717](https://doi.org/10.1145/2503713.2503717)
19. Hawkins R, Pingree S, Hitchon J, Radler B, Gorham B, Kahlor L, Gilligan E, Serlin R, Schmidt T, Kannaovakun P, Kolbeins G (2005) What produces television attention and attention style? Genre, situation, and individual differences as predictors. *Human Commun Res* 31(1):162–187
20. Holmes ME, Josephson S, Carney RE (2012) Visual attention to television programs with a second-screen application. In: Proceedings of the symposium on eye tracking research and applications, ACM, New York, NY, USA, ETRA '12, pp 397–400. doi:[10.1145/2168556.2168646](https://doi.org/10.1145/2168556.2168646)
21. Hutchings D (2012) An investigation of Fitts' law in a multiple-display environment. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, New York, NY, USA, CHI '12, pp 3181–3184. doi:[10.1145/2207676.2208736](https://doi.org/10.1145/2207676.2208736)
22. Jones B, Sodhi R, Murdock M, Mehra R, Benko H, Wilson A, Ofek E, MacIntyre B, Raghuvanshi N, Shapira L (2014) Roomalive: magical experiences enabled by scalable, adaptive projector-camera units. In: Proceedings of the 27th annual ACM symposium on user interface software and technology, ACM, New York, NY, USA, UIST '14, pp 637–644. doi:[10.1145/2642918.2647383](https://doi.org/10.1145/2642918.2647383)
23. Jones BR, Benko H, Ofek E, Wilson AD (2013) Illumiroom: peripheral projected illusions for interactive experiences. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, New York, NY, USA, CHI '13, pp 869–878. doi:[10.1145/2470654.2466112](https://doi.org/10.1145/2470654.2466112)
24. Kallenbach J, Narhi S, Oittinen P (2007) Effects of extra information on TV viewers' visual attention, message processing ability, and cognitive workload. *Comput Entertain* 5(2). doi:[10.1145/1279540.1279548](https://doi.org/10.1145/1279540.1279548)
25. Miller G (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol Rev* 63:81–97
26. Nielsen (2014) <http://www.nielsen.com/us/en/insights/reports/2014/the-us-digital-consumer-report.html>
27. Posner M (1980) Orienting of attention. *Q J Exper Psychol* 32:3–25
28. Posner M, Cohen Y (1984) Components of visual orienting. *Atten Perform X Control Lang Process* 32:531–556
29. Potter M (1976) Short-term conceptual memory for pictures. *J Exper Psychol* 2:509–522
30. Rashid U, Nacenta MA, Quigley A (2012) The cost of display switching: a comparison of mobile, large display and hybrid UI configurations. In: Proceedings of the international working conference on advanced visual interfaces, ACM, New York, NY, USA, AVI '12, pp 99–106. doi:[10.1145/2254556.2254577](https://doi.org/10.1145/2254556.2254577)
31. Rashid U, Nacenta MA, Quigley A (2012) Factors influencing visual attention switch in multi-display user interfaces: a survey. In: Proceedings of the 2012 international symposium on pervasive displays, ACM, New York, NY, USA, PerDis '12, pp 1:1–1:6. doi:[10.1145/2307798.2307799](https://doi.org/10.1145/2307798.2307799)
32. Reichle E, Rayner K, Pollatsek A (2004) The e-z reader model of eye-movement control in reading: comparisons to other models. *J Behav Brain Sci* 26:445–476
33. Riche N, Mancas M, Culibrk D, Crnojevic V, Gosselin B, Dutoit T (2013) Dynamic saliency models and human attention: a comparative study on videos. In: Proceedings of the 11th Asian conference on computer vision—volume part III, Springer, Berlin, ACCV '12, pp 586–598. doi:[10.1007/978-3-642-37431-9\\_45](https://doi.org/10.1007/978-3-642-37431-9_45)
34. Schwarz J, Klionsky D, Harrison C, Dietz P, Wilson A (2012) Phone as a pixel: enabling ad-hoc, large-scale displays using

- mobile devices. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, New York, NY, USA, CHI '12, pp 2235–2238. doi:[10.1145/2207676.2208378](https://doi.org/10.1145/2207676.2208378)
35. Sohlberg M, Mateer C (1989) Introduction to cognitive rehabilitation: theory and practice. Guilford Press, New York
  36. Tan DS, Czerwinski M (2003) Effects of visual separation and physical discontinuities when distributing information across multiple displays. In: Proceedings of the IFIP TC.13 international conference on human–computer interaction, INTERACT '03, pp 252–255
  37. Theeuwes J (1991) Exogenous and endogenous control of attention: the effect of visual onsets and offsets. *Percept Psychophys* 49(1):83–90. doi:[10.3758/BF03211619](https://doi.org/10.3758/BF03211619)
  38. Thomas BH (2012) A survey of visual, mixed, and augmented reality gaming. *Comput Entertain* 10(3):3:1–3:33. doi:[10.1145/2381876.2381879](https://doi.org/10.1145/2381876.2381879)
  39. Toet A (2011) Computational versus psychophysical bottom–up image saliency: a comparative evaluation study. *IEEE Trans Pattern Anal Mach Intell* 33(11):2131–2146
  40. Treisman AM, Gelade G (1980) A feature-integration theory of attention. *Cogn Psychol* 12(1):97–136
  41. Vatavu RD (2012) Point & click mediated interactions for large home entertainment displays. *Multimed Tools Appl* 59(1):113–128. doi:[10.1007/s11042-010-0698-5](https://doi.org/10.1007/s11042-010-0698-5)
  42. Vatavu RD (2012) User-defined gestures for free-hand TV control. In: Proceedings of the 10th European conference on interactive TV and video, ACM, New York, NY, USA, EuroITV '12, pp 45–48. doi:[10.1145/2325616.2325626](https://doi.org/10.1145/2325616.2325626)
  43. Vatavu RD (2013) A comparative study of user-defined handheld vs. freehand gestures for home entertainment environments. *J Ambient Intell Smart Environ* 5(2):187–211. doi:[10.3233/AIS-130200](https://doi.org/10.3233/AIS-130200)
  44. Vatavu RD (2013) There's a world outside your TV: exploring interactions beyond the physical TV screen. In: Proceedings of the 11th European conference on interactive TV and video, ACM, New York, NY, USA, EuroITV '13, pp 143–152. doi:[10.1145/2465958.2465972](https://doi.org/10.1145/2465958.2465972)
  45. Vatavu RD, Mancas M (2013) Interactive TV potpourris: an overview of designing multi-screen TV installations for home entertainment. In: Mancas M, d'Alessandro N, Siebert X, Goselin B, Valderrama C, Dutoit T (eds) *Intelligent technologies for interactive entertainment*. Lecture notes of the institute for computer sciences, social informatics and telecommunications engineering, vol 124. Springer International Publishing, pp 49–54. doi:[10.1007/978-3-319-03892-6\\_6](https://doi.org/10.1007/978-3-319-03892-6_6)
  46. Vatavu RD, Mancas M (2014) Visual attention measures for multi-screen TV. In: Proceedings of the 2014 ACM international conference on interactive experiences for TV and online video, ACM, New York, NY, USA, TVX '14, pp 111–118. doi:[10.1145/2602299.2602305](https://doi.org/10.1145/2602299.2602305)
  47. Vatavu RD, Pentti SG (2008) Interactive coffee tables: interfacing TV within an intuitive, fun and shared experience. In: Proceedings of the 6th European interactive TV conference, Springer, EuroITV '08, pp 183–187
  48. Vatavu RD, Wobbrock JO (2015) Formalizing agreement analysis for elicitation studies: new measures, significance test, and toolkit. In: Proceedings of the 33rd annual ACM conference on human factors in computing systems, ACM, New York, NY, USA, CHI '15, pp 1325–1334. doi:[10.1145/2702123.2702223](https://doi.org/10.1145/2702123.2702223)
  49. Vatavu RD, Zaiti IA (2014) Leap gestures for TV: insights from an elicitation study. In: Proceedings of the 2014 ACM international conference on interactive experiences for TV and online video, ACM, New York, NY, USA, TVX '14, pp 131–138. doi:[10.1145/2602299.2602316](https://doi.org/10.1145/2602299.2602316)
  50. Wagner J, Nancel M, Gustafson SG, Huot S, Mackay WE (2013) Body-centric design space for multi-surface interaction. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, New York, NY, USA, CHI '13, pp 1299–1308. doi:[10.1145/2470654.2466170](https://doi.org/10.1145/2470654.2466170)
  51. Wallace JR, Vogel D, Lank E (2014) Effect of bezel presence and width on visual search. In: Proceedings of the international symposium on pervasive displays, ACM, New York, NY, USA, PerDis '14, pp 118:118–118:123. doi:[10.1145/2611009.2611019](https://doi.org/10.1145/2611009.2611019)
  52. Wallace JR, Vogel D, Lank E (2014) The effect of interior bezel presence and width on magnitude judgement. In: Proceedings of the 2014 graphics interface conference, Canadian Information Processing Society, Toronto, Ont., Canada, Canada, GI '14, pp 175–182. <http://dl.acm.org/citation.cfm?id=2619648.2619678>
  53. Wobbrock JO, Aung HH, Rothrock B, Myers BA (2005) Maximizing the guessability of symbolic input. In: CHI '05 extended abstracts on human factors in computing systems, ACM, New York, NY, USA, CHI EA '05, pp 1869–1872. doi:[10.1145/1056808.1057043](https://doi.org/10.1145/1056808.1057043)