Visual Attention Measures for Multi-Screen TV

Radu-Daniel Vatavu
University Stefan cel Mare of Suceava
Suceava 720229, Romania
vatavu@eed.usv.ro

Matei Mancaș
University of Mons
20, Place du Parc, 7000 Mons, Belgium
matei.mancas@umons.ac.be

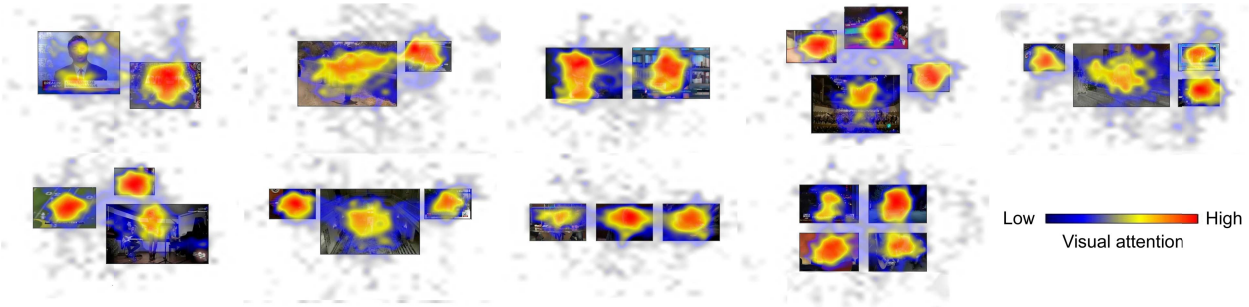

Figure 1. Visual attention heat maps for all nine multi-screen TV layouts evaluated in this study.

ABSTRACT
We introduce a set of nine measures to characterize viewers' visual attention patterns for multi-screen TV. We apply our measures during an experiment involving nine screen layouts with two, three, and four TV screens, for which we report new findings on visual attention. For example, we found that viewers need an average discovery time up to 4.5 seconds to visually fixate four screens, and their perceptions of how long they watched each screen are substantially accurate, *i.e.*, we report Pearson correlations up to .892 with measured eye tracking data. We hope our set of new measures (and the companion toolkit to compute them automatically) will benefit the community as a first step toward understanding visual attention for emerging multi-screen TV applications.

Author Keywords
Visual attention; interactive TV; multi-screen TV; multi-display; eye gaze; eye tracking; measures; evaluation; TLX.

ACM Classification Keywords
H.5.1 Multimedia Information Systems: Evaluation / methodology; Video. H.5.2 User Interfaces: Evaluation / methodology.

INTRODUCTION
Multi-screen systems that customize visual output to more than one screen are able to deliver more content and more control as well as new ways to enrich, share, and transfer content [2]. Due to such particular attractiveness, the multi-screen scenario has been investigated by the HCI community in terms of technical design [20,25] and performance evaluation [7,16,17,22]. In the context of the interactive TV, today's common implementation of the multi-screen concept is the secondary screen, with tablets used in conjunction with the TV set [2,4]. In a larger context, Vatavu and Mancaș referred to multi-screen TV systems as "TV potpourris", as they represent hybrids of screens with different form factors, layouts, and broadcasted programs and genres [24].

However, beside obvious advantages, more screens also demand higher cognitive load for viewers to understand what they watch, and increased visual attention distributed across displays. Therefore, it is likely for multi-screen TV to increase viewers' visual and cognitive attention load up to the point where the TV experience is no longer pleasant. Unfortunately, such important aspects have not been thoroughly addressed by the TV community up to now. In fact, in a recent work investigating visual attention in multi-display interfaces, Rashid et al. note that "further research is needed to investigate the influence of different categories of content coordination on attention switching and task performance" [17] (p. 4). In this work, we make one step further toward understanding viewers' visual attention patterns for multi-screen TV and, in doing so, we provide the community with a set of general and reusable measures to characterize visual attention for these scenarios.

Our contributions are as follows: (1) we introduce a set of measures that we compute from eye tracking data to characterize viewers' visual attention patterns for multi-screen TV; (2) we use these measures to report new findings on visual attention, such as viewers' subjectively-perceived watching time per screen is substantially accurate; and (3) we provide a toolkit to compute our measures automatically. In the end, we hope this work will benefit researchers and practitioners of the interactive TV community who will employ our measures to further investigate viewers' visual attention behavior for emerging multi-screen TV applications.

TVX '14, June 25 - 27, 2014, Newcastle Upon Tyne, United Kingdom

ACM 978-1-4503-2838-8/14/06…$15.00.
http://dx.doi.org/10.1145/2602299.2602305

RELATED WORK

The simplest form to emulate a multi-screen environment is to define individual screens as part of a video projection, with all screens controlled by the same computer [25]. Alternatively, physical screens can be put together to create multi-screen environments by using platforms that control the distribution of content. Phone as a Pixel is one such platform that can scale up to hundreds of displays [20].

More screens deliver more content and offer more control to viewers. Conversely, they may also have side effects on visual attention and task performance. In this section, we review previous work that showed interest in visual attention in general, but also in conjunction with the TV set.

Multi-screen environments and task performance

Multi-screen environments have been investigated in terms of attention demands [16,17] and their effects on task performance [7,22]. For example, Rashid et al. [16] 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 occur when shifting eye gaze between the two screens. Tan and Czerwinski [22] addressed the effect of visual separation between displays and physical discontinuities, such as monitor bezels. They found that discontinuities do not affect users' performance, but displaying content on screens positioned at different depths has small yet detrimental effects on task performance. Forlines et al. [7] observed participants performing worse during a visual search task when the information was displayed at different rotation angles on four vertical screens than when presented on a single screen. The authors of that study concluded that scanning of multiple views added to the length of the task, but not to its accuracy.

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). By definition, attention allows people to focus on a single task at one given time. Sohlberg and Mateer [21] identified five levels of attention, which are focused, sustained, selective, alternating, and divided attention.

Visual attention has been modeled by cognitive psychologists with the spotlight [5] and zoom-lens models [6]. The spotlight model describes 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 (the cut-off of the visual attention area). The zoom-lens model [6] upgraded the spotlight model by making it adaptable in size, and thus explained the trade-off in efficiency of processing visual information, *e.g.*, larger the focus, slower the processing will be.

Researchers have also modeled the way the brain attends to stimuli and processes information in what is known as bottom-up and top-down processing [23]. For example, some stimuli attract attention because of their stringent nature (*e.g.*, a quick motion or a telephone ring), which makes our brain process information at a preconscious level. On the other hand, top-down processing represents the act of individuals controlling their attention toward achieving a specific goal. Finally, attention is known to be overt (*i.e.*, when eye gaze attends to some region in space) and covert (mental focus can shift without necessarily moving the eyes) [14]. Overt attention is sequential by using eye saccades (*e.g.*, ballistic movements) and fixations (*e.g.*, the eye gaze stops at some spatio-temporal stable area). In contrast, covert attention can process several stimuli in parallel. Humans are known to be able to simultaneously attend to 7 ± 2 stimuli at once [13].

Visual attention and TV

Researchers have found that individual looks at the TV vary in length and people develop different watching strategies to follow content on TV. For example, people may look at the TV only at the right times, just enough to be aware of what is happening, while being engaged in some other activity. When investigating such phenomena, Geerts et al. [9] found that the genre of TV content correlates with how much people talk during watching TV, and that the plot structure influences talking during social television watching. Such findings reveal the importance of top-down attention during the everyday TV watching experience.

Surprisingly, most TV looks are very short, *e.g.,* 2 seconds, and can be described as mere glances [10]. This fact can be characterized with the "hazard look" function that gives the probability that looks persisting a given length will terminate in the next half second. Once a look begins, it is likely to terminate in the first second, with a hazard peak at 1−1.5 seconds. Hawkins et al. [10] investigated this phenomenon and identified monitoring looks less than 1.5 seconds, orienting looks up to 5 seconds, engaged looks between 6 and 15 seconds, and staring after 15 seconds.

To characterize visual attention, researchers have employed eye tracking devices that accurately follow viewers' eye gaze. For example, Kallenbach et al. [12] used an eye tracker and 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. [11] 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 a multi-screen sports study, Cummins et al. [3] found visual attention to vary function of screen size, game play (*i.e.*, action), and repeated exposure. They also reported that viewers had to adopt screen watching strategies to cope with the many pictures displayed simultaneously. Finally, Rashid et al. [17] identified five factors that affect visual attention patterns for multi-display user interfaces, namely display contiguity, angular coverage, content coordination, input directness, and input-display correspondence.

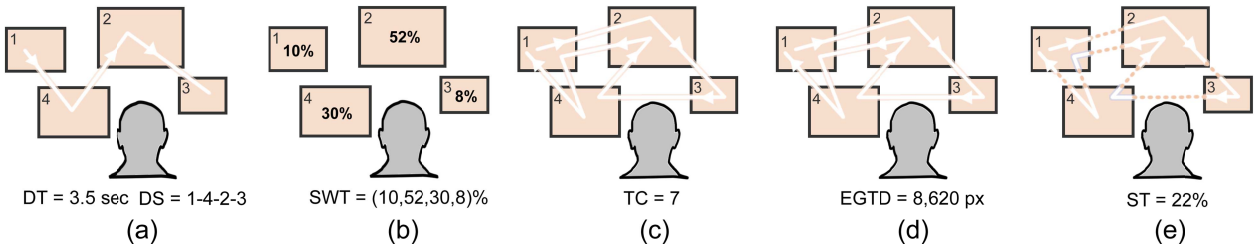

Figure 2. Illustrations of our six *objective* visual attention measures.

VISUAL ATTENTION MEASURES

By following the results of previous work in terms of experimental findings and modeling of visual attention, we define six *objective* measures to characterize viewers' visual attention behavior in the context of multi-screen TV:

1. **Discovery Time (DT)** is defined as the time required for the viewer to make a pass over all TV screens so that each screen has been fixated visually at least once (Figure 2a). DT is the minimum time imperative to understand what is happening on all screens in order to inform what to watch. The discovery time may depend on several factors, such as the number of screens, their layout and form factors, and displayed content.
2. **Discovery Sequence (DS)** is defined as the sequence of screens that was traversed by the viewer's eye gaze during the discovery time (Figure 2a).
3. **Screen Watching Time (SWT)** is the percentage of visual attention devoted by viewers to each TV screen during the monitoring time interval. For example, the SWT distribution for a 3-screen layout may be uniform, with ≈33% of time devoted to each screen, but it may also be non-uniform, with one screen capturing the viewer's visual attention more, such as 60%, 40%, and 10%. SWT may depend on the form factors, screen layout, and the content displayed by the screens, *e.g.*, the larger the screen or more attractive its content, more time will likely be devoted to consume that content. SWT can also be visualized as heat maps (see Figure 1).
4. **Transition Count (TC)** is defined as the number of eye gaze transitions between consecutively fixated TV screens that occurred during the monitoring time interval (Figure 2c). Larger TC values reflect more distributed attention, but may also show design flaws in the layout.
5. **Eye Gaze Travel Distance (EGTD)** represents the total distance travelled by the viewer's eye gaze during the monitoring time interval, expressed in the units reported by the eye tracking equipment, such as pixels (Figure 2d). Different EGTD values may reflect different visual attention patterns, different preferences for some channels over time, and they may possibly correlate with watching fatigue. Real-world distance units, such as centimeters, should be used when measuring EGTD for screens with different pixel pitch densities.
6. **Switch Time (ST)** is defined as the percentage of time during which eye gaze travels between screens (Figure 2e). Large ST values may show flaws in layout design, *e.g.,* larger the ST value, further apart the screens are located one from the other in that particular layout.

We introduce these six new measures to characterize viewers' TV watching behavior in more nuanced ways, not easily accessible with generic eye gaze heat maps and scan paths. For example, heat maps (as those shown in Figure 1) are generally used to describe the spatial distribution of eye gaze, which is useful for investigating specific elements that attract attention *within* the same screen. However, our SWT measure reflects viewers' allocated watching time for the entire screen with one single value integrating the heat map spatially-distributed data, 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 behavior occurring at specific moments, *e.g.*, during 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 introduced below) are expressed in this work using percentages that normalize these measures with respect to the entire monitoring 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 time values of their viewers' TV watching behavior.

We also define and employ three *subjective* measures:

1. **Perceived Screen Watching Time (PSWT)** is defined as the percentage of the visual attention devoted to each screen during the monitoring interval, as it was perceived by viewers themselves. We show later how subjective PSWT correlates with measured SWT.
2. **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, from 1 to 5: very uncomfortable, uncomfortable, neutral, comfortable, and very comfortable.
3. **Content Understandability (CU)** represents the capacity of the viewers to understand content delivered by the multiple screens of some layout. CU is assessed by asking viewers questions about the content they just watched, and is measured as the percentage of correct answers. For example, in our experiment we asked one question of moderate difficulty for each TV screen.

EXPERIMENT

We conducted an experiment to understand the effect of multiple TV screens on visual attention and to validate our new visual attention measures. To inform the design of our experiment, we first ran a preliminary study.

Preliminary study

Previous work showed that the number of screens, their form factors, and displayed content affect viewers' visual attention patterns. 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. Four participants were presented with five TV layouts composed of 2, 4, 6, 9, and 12 screens of equal size arranged in matrix-like configurations. To prevent participants from visually privileging some screens over the others, all the screens displayed non-overlapping sequences extracted from the same movie scene. All video sequences had one minute in length. The audio was turned off. Each participant watched the movies separately (there was no social TV watching).

We found participants generally looking in the center of the matrix layouts trying to cover most of the information within their visual field. While the eye repartition remained well distributed in the case of two and four screens, 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 [5] for our specific matrix-like screen layouts. Also, participants witnessed that more than two/three screens was too much to follow because they were trying to make sense of the various sequences of the same movie, *i.e.*, putting the pieces together. However, they were interested in multiple screens that would convey *complementary* information to a single, main screen. Therefore, findings revealed that the concept of a primary screen with an easily-identifiable form factor (*i.e.*, larger than all the other screens) is important to understand the layout. These preliminary findings informed the design of our experiment for which we investigated in detail layouts composed of two, three and four TV screens.

Participants

Ten volunteers (one female) participated in the experiment (mean age was 27.9 years, $SD = 3.7$ years). Participants' self-reported daily average time for watching television was 1.5 hours ($SD = 2.1$ hours). All participants had normal or corrected to normal vision.

Apparatus

TV screens were part of a large image (1.30 × 0.87 meters) that was projected on a wall with a standard projector (24.5 dpi). Participants were seated comfortably in a chair at a distance of 2.30 meters from the projection. Given the fact that the projected screens did not fill the entire projected image, the maximum visual angle was 25° on the x 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 [17], our screen scenario 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 are spatially separated, which makes them appear as distinct displays instead of a visual contiguous screen). The movies displayed by each screen lasted one minute each and were prepared in advance. The audio was turned off for all videos in order to isolate the effects of the visual information alone on attention. The FaceLab eye-tracking device[1] was employed during the experiment by following the practices of the visual attention community [8,19] and those of previous experiment designs investigating visual attention for TV [3,11,12].

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.

(2) LAYOUT, representing the space arrangement of the TV screens and their sizes. For this factor we designed three distinct conditions: TILED, PRIMARY, and ARBITRARY (Figure 3). In the TILED condition all the screens have equal size and are arranged in a 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 a random layout.

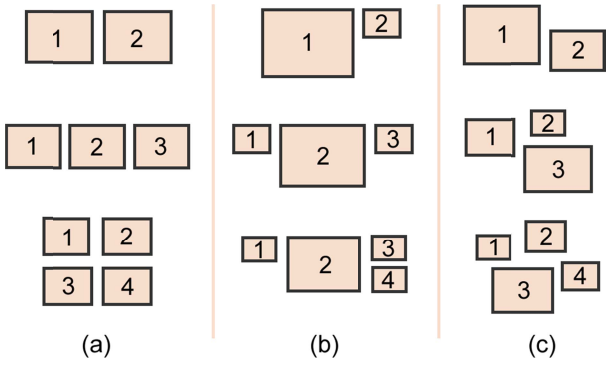

Figure 3. TV layouts for the TILED (a), PRIMARY (b), and ARBITRARY (c) conditions with two, three, and four screens.

TV layouts were created so that the total display area covered by constituent screens would be approximately constant for each layout (*i.e.,* the size of the TV screens was larger for layouts with fewer screens). In the preliminary experiment we displayed movie 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 when watching television.

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

For the full experiment, the TV screens displayed different content. However, we verified the content *a priori* by running a motion detector (frame-to-frame difference) to make sure that the motion level was roughly the same across screens of the same layout.

Task

Participants were asked to watch prerecorded movies for each combination of TV-COUNT and LAYOUT conditions, resulting in a total number of 9 experimental trials per participant corresponding to 9 minutes of watching TV (one minute per condition). Each participant watched the movies separately to eliminate any effect of social watching on attention. Participants were asked to watch the movies as if they were watching their TV at home, and were told they had to answer a questionnaire after each trial in order to ensure a minimal level of attention. Condition order was randomized across participants. After each trial, participants were administered NASA TLX tests[2] (using the computer version available on-line at[3]) to collect workload subjective ratings, and 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 the perceived comfortability (PC) of watching each layout on a 5-point Likert scale, with 1 being very uncomfortable and 5 very comfortable. Participants were also asked to specify the maximum number of TV screens they would feel comfortable watching at the same time (MAX-TV).

RESULTS #1: DISTRIBUTION OF VISUAL ATTENTION FOR MULTI-SCREEN TV

Discovery Time

Our participants systematically discovered all screens before committing to one screen to watch. In general, this process can be very fast and a single eye fixation is usually enough to roughly understanding the topic being watched [15]. Discovery time varied between 0.1 and 15.5 seconds, with a mean time of 2.4 seconds ($SD = 0.3$). 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 participants to visually fixate more screens until all have been discovered (Figure 4a). DT values ranged from 0.8 seconds for two screens up to 4.5 seconds for four screens. A second degree polynomial showed a perfect fit with observed data ($R^2 = 1$), suggesting that discovery time relates to the number of screens in a quadratic manner ($\mathrm{DT} = 0.70 \cdot \text{TV-COUNT}^2 - 0.97 \cdot \text{TV-COUNT} + 1.10$). There was no significant effect of LAYOUT on DT ($\chi^2(2) = 2.867, n.s.$).

Discovery Sequence

The discovery sequence informs about the order in which screens are visually attended 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 Figure 3.) For three screens, there are $3! = 6$ possible sequences, out of which 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, no majority preference for ARBITRARY (for which 8 different sequences were found among the 10 recordings). These results show that viewers discover screens from left to right for the two screen condition (sequence 1,2), are first attracted by the middle screen when three screens are present (sequences 2,3,1 and 2,1,3), and follow a counter-clockwise pattern (*e.g.*, 1,2,4,3) in the absence of a primary screen to attract attention first (2,1,3,4 for PRIMARY).

The discovery sequence mainly shows the impact of layout. For screens of equal size, a left-to-right model was adopted by our participants, which corresponds to the reading order in Western culture (which was the case for our participants). For the PRIMARY layouts, discovery begins with the larger screen despite it not being the left-most screen. This finding shows that participants immediately identified the largest screen as the main or primary one. The anti-clockwise pattern is also interesting, as it builds on the observed 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 [18], but specific for multi-screen TV.

Screen Watching Time

The average percentages of visual attention shared between screens are illustrated in Figure 4e using color codes, with darker values showing more visual attention. We found significant differences for the TILED and PRIMARY layouts and three and four screens, while only the ARBITRARY layout had a significant effect on SWT for two screens. Results show 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), but also with content, as we later found by asking participants (*e.g.*, participants' visual attention was more attracted by the right screen of the 2-ARBITRARY condition that displayed a bicycle race, instead by the first screen that showed news, resulting in 68% and 31% devoted attention, see Figure 4e).

SWT can be further visualized as heat maps (Figure 1) that use color codes to describe gaze density spatially along the screen area size. When all screens are of equal size, the 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 Figures 1 and 4e). However, in the PRIMARY condition, the gaze density color of the largest screen has lower maxima, as gaze is distributed across a larger area (larger coverage).

Transition Count

The number of eye gaze transitions between screens varied from 11 to 200 for the entire watching time of one minute,

[2] http://humansystems.arc.nasa.gov/groups/tlx/

[3] http://www.keithv.com/software/nasatlx/

with a mean value of 64.8 transitions ($SD = 28.9$). We found significant effects (at $p < .001$) for both TV-COUNT ($\chi^2(2) = 41.667$) and LAYOUT ($\chi^2(2) = 10.237$, with no significant difference between TILED and PRIMARY for LAYOUT). Figure 4b reveals an expected yet strong positive correlation ($R^2 = 1$) between TV-COUNT and TC: more screens determine more transitions during the first minute of watching ($TC = 20.5 \cdot \text{TV-COUNT} + 23.9$). We also found that the ARBITRARY layout led to significantly less transitions. This 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 to actually transitioning to other screens, as the peripheral information is available to viewers' visual attention and is processed accordingly, as explained by the zoom-lens model [6].

Eye Gaze Travel Distance

During the one minute of each trial, participants' eye gaze travelled in average 39.7 meters (Figure 4c). Interestingly, the number of TV screens had no significant effect on EGTD ($\chi^2(2) = 1.667, n.s.$), but the way the screens were arranged in space did ($\chi^2(2) = 8.867, p = .01$). 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 travelled the same amount of 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.

Switch Time

The switch time (Figure 4d) is the time required for eye gaze to travel between screens, and we found it to vary up to 29% of the total watching time, with an average of 2.5% ($SD = 5\%$). We found a significant effect of TV-COUNT on Switch Time ($\chi^2(2) = 8.824, p = .01$) with post-hoc tests showing significant differences only between two and four screens. There was no significant effect of LAYOUT on Switch Time ($\chi^2(2) = 2.867, n.s.$). Overall, this measure revealed that four-screen layouts are less efficient in terms of actually fixating TV content, as they unnecessarily consume eye gaze for transitions in-between screens.

RESULTS #2: COGNITIVE LOAD AND COMFORTABILITY FOR MULTI-SCREEN TV

Cognitive load

After each trial, participants were administered NASA TLX tests to collect their subjective ratings of the workload on a scale from 1 (low) to 100 (high). 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 (Figure 5, 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 $.50$). Significant effects of TV-COUNT were found for each dimension of the NASA TLX test (Figure 5, right). At the same time, there was no significant effect of LAYOUT on the perceived task load measured by TLX ($\chi^2(2) = 4.206, n.s.$), nor on any of the six dimensions employed by the NASA TLX test.

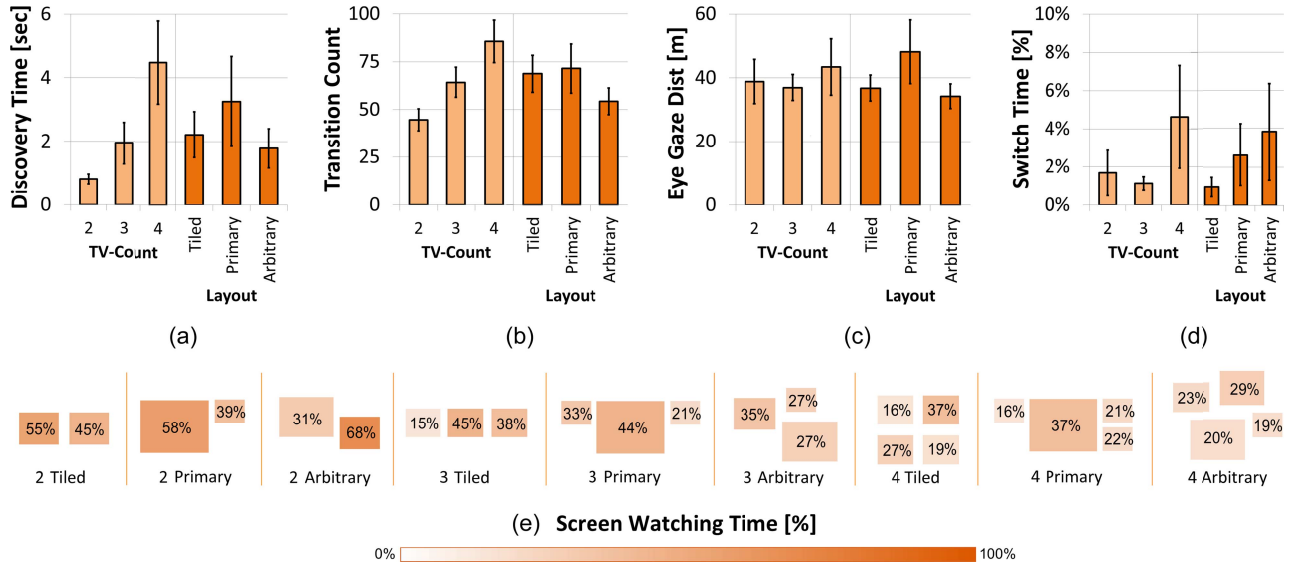

Figure 4. Visual attention measures for multi-screen TV: Discovery Time (a), Transition Count (b), Eye Gaze Distance (c), Switch Time (d), and Screen Watching Times (e). (Note that SWTs (e) do not always add to 100%; the remainder is the Switch Time.)

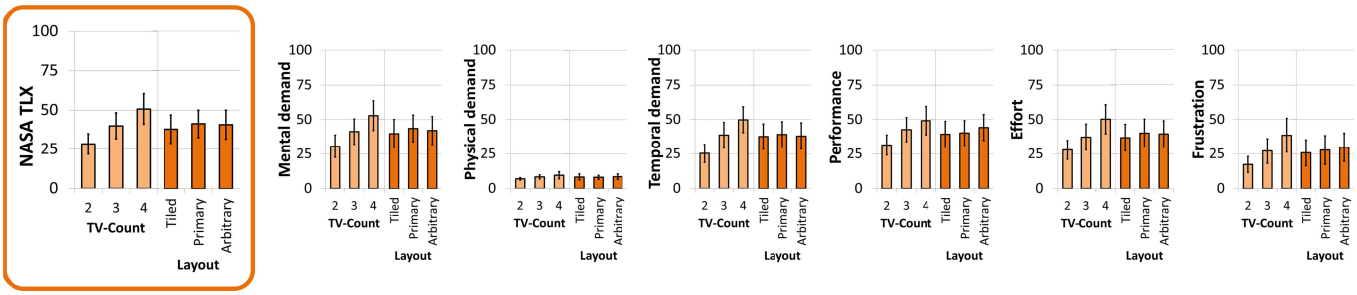

Figure 5. Participants' average workload ratings measured with the NASA task load test for each TV-COUNT and LAYOUT (left). NOTE: The NASA TLX test employs six dimensions (right) to measure workload in the range $[0, 100]$ corresponding to the subjective perceptions of Low/High (*e.g.*, high mental demand, low physical effort), and Poor/Good for Performance.

Comfortability

At the end of the experiment, participants were asked 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 of uncomfortable and neutral (Figure 6a). Maximum of comfortability was perceived for layouts with two screens (4, comfortable), and 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 (at $p = .05/3 = .017$) for paired conditions (2,3) and (2,4) with large effect sizes ($r = .51$ and $.57$ respectively). No significant effect was detected between three and four screens. We also found a significant effect of LAYOUT on perceived comfortability ($\chi^2(2) = 23.275, p < .001$). 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.

The median value of MAX-TV was 2 screens (Figure 6b). We found a significant effect of TV-COUNT on MAX-TV ($\chi^2(2) = 13.565, p < .001$), no significant difference between 2 and 3 screens, but significant between (3,4) and (2,4) with effect sizes $.36$ and $.40$. There was no significant effect of LAYOUT on MAX-TV ($\chi^2(2) = 5.261,\ n.s.$).

RESULTS #3: CAPACITY TO UNDERSTAND CONTENT AND PERCEIVED SCREEN WATCHING TIME

Content understandability

After each trial, participants were administered multiple-choice questions about the content displayed by each screen (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% (Figure 7). There were 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 "don't know" answers ($\chi^2(2) = 6.645,\ p = .036$).

Perceived Watching Time

After each trial, participants estimated in percentages how much they watched each screen. When we correlated this perceived SWT with measured SWT, we found a Pearson 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 how much. Correlation coefficients computed for each condition are shown in Figure 8, with a maximum of .892 for 2-ARBITRARY.

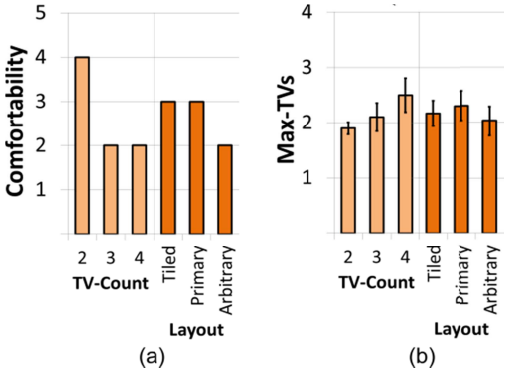

Figure 6. Participants' median ratings of Perceived Comfort (a) and the average of the maximum number of TV screens they felt could be watched comfortably at the same time (b).

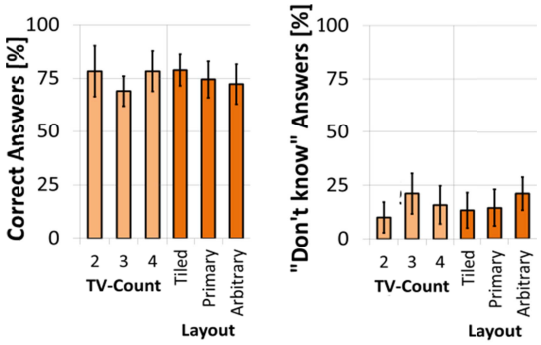

Figure 7. Participants' Content Understandability (CU).

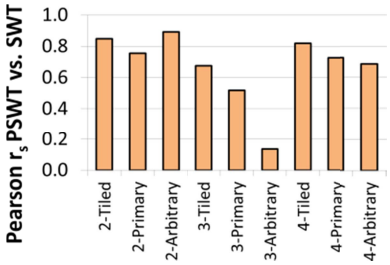

Figure 8. Correlations between perceived and measured SWT. (All significant at $p = 0.01$, except for 3-ARBITRARY.)

VISUAL ATTENTION TOOLKIT

We release our set of measures in the form of the Visual Attention Toolkit for TV (VATic-TV), which we make available to the community as open source software companion to this paper contribution (VATic-TV can be downloaded from `http://www.eed.usv.ro/~vatavu`). We also release all the data files (*i.e.,* movies of the experiment and eye gaze tracking data logs) collected and used in this study as a multi-screen eye gaze dataset to allow easy replication of results and encourage further investigation of visual attention phenomena for multi-screen TV applications.

CONCLUSION

We proposed in this work a set of nine general and reusable measures to characterize viewers' visual attention patterns for multi-screen TV, out of which the six objective measures can be computed automatically with the toolkit accompanying the paper. We applied our measures to evaluate multi-screen TV layouts and showed how the number of screens and their structure affect viewers' visual attention and cognitive load. 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.

ACKNOWLEDGEMENTS

The paper and research effort have been co-founded by the European Commission under project "NUBOMEDIA - An elastic Platform as a Service (PaaS) cloud for interactive social multimedia", reference FP7-ICT-2013.1.6 GA-610576.

REFERENCES

1. Anderson, J.R. 2004. *Cognitive psychology and its implications*. Worth Publishers, New York, USA
2. Cesar, P, Bulterman, D.C., Jansen, A.J. 2008. Usages of the Secondary Screen in an Interactive Television Environment: Control, Enrich, Share, and Transfer Television Content. *Proc. of EuroITV '08*, 168-177
3. Cummins, R.G., Tirumala, L.N., Lellis, J.M. 2011. Viewer Attention to ESPN's Mosaic Screen: An Eye-Tracking Investigation. *Journ. Sports Media* 6(1), 23-54
4. Courtois, C., D'heer, E. 2012. Second screen applications and tablet users: constellation, awareness, experience, and interest. *Proc. of EuroITV '12*, 153-156
5. Eriksen, C, Hoffman, J. 1972. Temporal and spatial characteristics of selective encoding from visual displays. *Perception & Psychophysics* 12 (2B), 201-204
6. Eriksen, C., St James, J. 1986. Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics* 40 (4), 225–240
7. Forlines, C., Shen, C., Wigdor, D., Balakrishnan, R. 2006. Exploring the effects of group size and display configuration on visual search. *Proc. of CSCW '06*, 11-20
8. Frintrop, S., Rome, E., Christensen, H.I. 2010. Computational visual attention systems and their cognitive foundations: A survey. *ACM TAP* 7(1), 39 pp.
9. Geerts, D., Cesar, P., Bulterman, D. 2008. The implications of program genres for the design of social television systems. In *Proc. of UXTV '08*. ACM, 71-80
10. Hawkins, R.P., Pingree, S., Hitchon, J., Radler, B., Gorham, B.W., Kahlor, L., Gilligan, E., Serlin, R.C., Schmidt, T., Kannaovakun, P., Kolbeins, G.H. 2005. What Produces Television Attention and Attention Style? Genre, Situation, and Individual Differences as Predictors. *Human Commun. Research* 31(1), 162–187
11. Holmes, M.E., Josephson, S., Carney, R.E. 2012. Visual attention to television programs with a second-screen application. In *Proc. of ETRA '12*. ACM, 397-400
12. Kallenbach, J., Narhi, S., Oittinen, P. 2007. Effects of extra information on TV viewers' visual attention, message processing ability, and cognitive workload. *Computers in Entertainment* 5(2). ACM, NY, USA
13. Miller, G. 1956. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for rocessing Information. *Psychological Review* 63, 81-97
14. Posner, M.I. 1980. Orienting of attention. *Quarterly Journal of Experimental Psychology* 32, 3-25
15. Potter, M.C. 1976. Short-term conceptual memory for pictures. *Journal of Exp. Psychol.* 2, 509-522
16. Rashid, U., Nacenta, M.A., Quigley, A. 2012. The cost of display switching: a comparison of mobile, large display and hybrid UI configurations. *Proc. AVI '12*, 99-106
17. Rashid, U., Nacenta, M.A., Quigley, A. 2012. Factors influencing visual attention switch in multi-display user interfaces: a survey. *Proc. of PerDis '12*. ACM, 6 pages
18. Reichle, E.D., Rayner, K., Pollatsek, A. 2004. The E-Z reader model of eye-movement control in reading: Comparisons to other models. *Behav. Brain Sci.* 26, 445-476
19. Riche, N., Mancas, M., Culibrk, D., Crnojevic, V., Gosselin, B., Dutoit, T. 2012. Dynamic saliency models and human attention: a comparative study on videos. *Proc. of ACCV'12*, 586-598
20. Schwarz, J., Klionsky, D., Harrison, C., Dietz, P., Wilson, A. 2012. Phone as a pixel: enabling ad-hoc, large-scale displays using mobile devices. *CHI '12*, 2235-2238
21. Sohlberg, M.M., Mateer, C.A. 1989. *Introduction to cognitive rehabilitation: theory and practice*. Guilford Press
22. Tan, D.S., Czerwinski, M. 2003. Effects of Visual Separation and Physical Discontinuities when Distributing Information across Multiple Displays. *OZCHI'03*, 184-191
23. Theeuwes, J. 1991. Exogenous and endogenous control of attention - the effect of visual onsets and offsets. *Perception & Psychophysics* 49(1), 83–90
24. Vatavu, R.D., Mancaș, M. 2013. Interactive TV Potpourris: An Overview of Designing Multi-Screen TV Installations for Home Entertainment. *INTETAIN'13*, 49-54
25. Vatavu, R.D. 2013. There's a world outside your TV: exploring interactions beyond the physical TV screen. In *Proc. of EuroITV '13*. ACM, NY, USA, 143-152