Discerning Ambient/Focal Attention with Coefficient $\mathcal{K}$

KRZYSZTOF KREJTZ, SWPS University of Social Sciences and Humanities, Warsaw, Poland
ANDREW DUCHOWSKI, Clemson University, Clemson, SC
IZABELA KREJTZ, SWPS University of Social Sciences and Humanities, Warsaw, Poland
AGNIESZKA SZARKOWSKA, University of Warsaw, Warsaw, Poland
AGATA KOPACZ, National Information Processing Institute, Warsaw, Poland

We introduce coefficient $\mathcal{K}$, defined on a novel parametric scale, derived from processing a traditionally eye-tracked time course of eye movements. Positive and negative ordinates of $\mathcal{K}$ indicate *focal* or *ambient* viewing, respectively, while the abscissa serves to indicate time, so that $\mathcal{K}$ acts as a dynamic indicator of fluctuation between ambient/focal visual behavior. The coefficient indicates the difference between fixation duration and its subsequent saccade amplitude expressed in standard deviation units, facilitating parametric statistical testing. To validate $\mathcal{K}$ empirically, we test its utility by capturing ambient and focal attention during serial and parallel visual search tasks (Study 1). We then show how $\mathcal{K}$ quantitatively depicts the difference in scanning behaviors when attention is guided by audio description during perception of art (Study 2).

Categories and Subject Descriptors: J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Psychology*

General Terms: Human Factors

Additional Key Words and Phrases: Ambient-focal attention, visual attention dynamics, serial versus parallel search

ACM Reference Format:

Krzysztof Krejtz, Andrew Duchowski, Izabela Krejtz, Agnieszka Szarkowska, and Agata Kopacz. 2016. Discerning ambient/focal attention with coefficient $\mathcal{K}$. ACM Trans. Appl. Percept. 13, 3, Article 11 (May 2016), 20 pages.
DOI: http://dx.doi.org/10.1145/2896452

1. INTRODUCTION

There is an increasing demand for characterization of viewer behavior through analysis of eye movements. Efforts are under way to surpass traditional categorization of the captured eye gaze sequence (x_i, y_i, t_i) as fixations and saccades into higher-level descriptors of visual behavior. For example, Bednarik et al. [2012] explored eye movement features that could best describe the differences in gaze behavior during intentional and nonintentional interaction, for example, deciding if the user

This work has been partly supported by research grant "Audio description in education" awarded by the Faculty of Applied Languages, University of Warsaw.

Authors' addresses: K. Krejtz and I. Krejtz, Department of Psychology, SWPS University of Social Sciences and Humanities, Warsaw, Poland, ul. Chodakowska 19/31, 03-815 Warsaw, Poland; emails: {kkrejtz, ikrejtz}@swps.edu.pl; A. Duchowski, School of Computing, Clemson Universiy, Clemson, SC, 100 McAdams Hall, Clemson, S.C. 29634; email: duchowski@clemson.edu; A. Szarkowska, Faculty of Applied Linguistics, University of Warsaw, Warsaw, Poland, ul. Dobra 55, 00-312 Warszawa, Poland; email: a.szarkowska@gmail.com; A. Kopacz, Interactive Technology Laboratory, Information Processing Institute, Warsaw, Poland, Al. Niepodleglosci 188b, 00-608 Warszawa; email: akopacz@opi.org.pl.

DOI: http://dx.doi.org/10.1145/2896452

is about to issue a command. They used a Support Vector Machine (SVM) approach to differentiate pupil diameter from a baseline recording for this purpose. Bulling et al. [2013] attempted to classify continuous Electro-Oculography (EOG) signals into a vector of binary descriptors of everyday life situations, that is, whether or not the user is interacting socially, concentrating on a mental task, engaging in a physical activity, or is inside or outside. Meanwhile, Toker et al. [2013] considered various eye-tracking features as potential indicators of perceptual speed and visual memory. They found that some features such as fixation measures and gaze transitions between Areas Of Interest (AOIs) could serve as potential indicators of perceptual speed, given visual inspection of two types of graphs: bar graphs and radar graphs. Whether the eye movement features themselves could serve as indicators of users' attentional patterns or whether results were largely due to the difference in stimulus (the graphs) was not entirely clear, however. More recently, Krejtz et al. [2015] modeled gaze transitions between AOIs as Markov chains and compared individuals' viewing behavior via statistical entropy of the computed transition matrices.

In this article, instead of pupil diameter or EOG, we analyze the traditionally eye-tracked time course of eye movements over AOIs and introduce coefficient $\mathcal{K}$, defined on a novel parametric scale where positive values on the ordinate indicate *focal* patterns, while negative values suggest *ambient* visual scanning. The abscissa serves to indicate time, so that $\mathcal{K}$ acts as a dynamic indicator of fluctuation between ambient/focal visual scanning. We derive coefficient $\mathcal{K}$ by applying Velichkovsky et al.'s [2005] original characterization of the two ambient and focal modes of attention but also consider the time course of an individual's eye movement record. Continuous attentional signal switching between focal and ambient scanning is derived from plotting $\mathcal{K}$ over time. First, we show how coefficient $\mathcal{K}$ captures ambient and focal attention during serial and parallel visual search tasks. Second, we demonstrate how $\mathcal{K}$ can be used to compare and contrast the time course of visual attention of groups of viewers, in this instance when viewing artwork. Evidence suggests that $\mathcal{K}$ depicts dynamic maintenance of focal attention when it is cued exogenously.

2. BACKGROUND

From a traditional eye-tracking perspective, the process of scene viewing, visual search, or reading, is a dynamic interplay between fixations and saccades. Saccades are rapid eye movements reaching velocities as high as 500° per second [Rayner 1998, 2009]. In silent reading, saccade amplitude is usually 2° visual angle, while during scene perception it increases on average to 4° [Rayner 1984]. Fixations can simply be defined as periods of time between two consecutive saccades. Previous research states that the eyes usually stop for about 200–300ms, however, during scene watching, fixations from under 100ms to over 500ms may also be observed [Rayner 1998].

The combination of fixation duration and saccadic amplitude immediately following the fixation has been proposed for the analysis of static and dynamic scene viewing [Velichkovsky et al. 2005; Krejtz et al. 2012b]. Specifically, short duration fixations followed by long saccades are characteristic of ambient processing, while longer duration fixations followed by shorter saccades are indicative of focal processing [Unema et al. 2005].

The interplay between focal and ambient modes of visual information processing changes dynamically. At early stages of scene perception, shorter fixations and longer saccades appear to govern initial scene exploration. Once a target has been identified, longer fixations are followed by shorter saccades suggesting a change to a focal mode of processing [Irwin and Zelinsky 2002; Velichkovsky et al. 2005].

The dynamic pattern of visual attention can be attributed to two modes of acquiring information: exploration and inspection. Pannasch et al. [2008] showed a systematic increase in the durations of fixations and a decrease of saccadic amplitudes over the time course of scene perception. This relationship was very stable across the variety of studied conditions, including repeated presentation of similar stimuli, object density, emotional stimuli, and mood induction. In their work, the time courses of fixation durations and saccadic amplitudes were considered as two independent streams of data. We combine both streams into a single dynamic stream defined on a novel parametric scale capturing explicitly the interplay of ambient and focal modes.

In a recent study during free viewing of paintings with gaze-contingent distractors and non-gaze-related, time-locked changes, Pannasch et al. [2011] demonstrated that any visual change prolongs the current fixation, strongly suggesting that fixations classified as focal are under the direct control of the stimulus information. These results support direct control theories of visual attention, which state that decisions about fixation termination are made during the ongoing fixation, as opposed to indirect control theories, which posit that prolongation of the current fixation is determined by other factors.

Holmqvist et al. [2011b] review several means of operationalization of ambient/focal viewing: thresholding on the ratio of fixation duration to saccade amplitude, and computation of a saccade/fixation ratio. None of these approaches, however, explicitly consider dynamics of how the fixation duration/saccade amplitude ratio changes over time.

Originally, Unema et al. [2005] averaged all fixations within 20ms bins and plotted saccade amplitudes against binned fixation duration. While this gives an indication of the ratio between fixation duration and saccade amplitude, it is a static snapshot and does not give an indication of how this ratio changes over time. Both fixation durations and saccade amplitudes (sa) were described by a simple exponential growth curve of total activity, that is, $sa = b \cdot e^{(-a/t)}$ with b an asymptotic value, a the acceleration, and t the time elapsed since the beginning of a fixation. No discussion was given of how a function combining saccadic amplitude and fixation duration could change over time.

Similarly, Follet et al. [2011] note that, during picture viewing, long saccades are performed first (probably to explore visual content quickly) and saccade amplitudes decrease over time. They also plot saccade amplitude versus fixation duration, and do not consider how the interplay changes over time. They consider probability distribution functions of focal and ambient fixations occurring over time, but the probability distribution pertains to either ambient or focal fixations and does not indicate the interplay between them. More recently, Holmqvist et al. [2011a] discussed a method for quantifying focused versus overview behavior, but this technique was based more on AOI transitions rather than on the relation of fixation durations and consecutive saccadic amplitudes.

To the best of our knowledge, we are not aware of a straightforward metric to distinguish ambient from focal viewing modes via an automatic approach that depicts the interplay between these modes over time, allowing subsequent statistical analyses. We develop such a metric and test its implementation in R, the language for statistical computing [R Development Core Team 2011], in two independent studies.

3. TECHNICAL DEVELOPMENT: AMBIENT/FOCAL ATTENTION COEFFICIENT $\mathcal{K}$

We introduce the $\mathcal{K}$ ambient/focal coefficient as an additional means of analyses of eye movements, which, in turn, are referred to as *process measures*, complementing traditional *performance measures*, for example, speed and accuracy, often used in the measurement of human behavior.

Velichkovsky et al. [2005] originally suggested characterization of fixations as ambient or focal based on their durations and the amplitude of successive saccades. Unfortunately, their presentation prevented collation and eventual comparison of means that is useful in inferential statistical analyses. Krejtz et al. [2012b] used an ambient/focal attention coefficient, defined as the relation between the current fixation duration and the subsequent saccade amplitude, but did not provide its derivation (see also Biele et al. [2013]). We formally define the ambient/focal attention coefficient proposed by Krejtz et al. [2012b] as $\mathcal{K}$, which permits statistical comparison between individuals and between groups.

In the calculation of $\mathcal{K}$, both fixation durations and saccade amplitudes are transformed into a standard score (z-score), allowing computation of an ambient/focal attentional coefficient per individual scanpath. The transformation of raw data from saccade amplitude and fixation duration into standard scores, representing the distance between the raw score and the mean in units of the standard deviation, allows for direct mathematical comparison of both measures.

We derive $\mathcal{K}$, calculated for each participant, as the mean difference between standardized values (z-scores) of each saccade amplitude (a_{i+1}) and its preceding i^{th} fixation duration (d_i):

$$\mathcal{K}_i = \frac{d_i - \mu_d}{\sigma_d} - \frac{a_{i+1} - \mu_a}{\sigma_a}, \quad \text{such that} \quad \mathcal{K} = \frac{1}{n}\sum_n \mathcal{K}_i, \tag{1}$$

where μ_d, μ_a are the mean fixation duration and saccade amplitude, respectively, and σ_d, σ_a are the fixation duration and saccade amplitude standard deviations, respectively, computed over all n fixations and hence n $\mathcal{K}_i$ coefficients (i.e., over the entire duration of stimuli presentation) [Krejtz et al. 2012b].

Note that computed means (μ_d and μ_a) and standard deviations (σ_d and σ_a) in (1) refer to sample statistics and are computed over all conditions and for all participants. This allows the reduction of bias of $\mathcal{K}$ rooted in the experimental manipulation. At the same, this procedure allows for capturing differences between experimental conditions where they actually exist.

Values of $\mathcal{K}$ that are close to zero indicate relative similarity between fixation durations and saccade amplitudes (in terms of their distance from their respective means). Positive values of $\mathcal{K}_i$ show that relatively long fixations were followed by short saccade amplitudes, indicating focal processing. Analogously, negative values of $\mathcal{K}_i$ refer to the situation when relatively short fixations were followed by relatively long saccades, suggesting ambient processing.

Values of $\mathcal{K}_i$ are expressed in Standard Deviation scores (SD) making them easy to interpret. For example, $\mathcal{K}_i = 1$ indicates that the duration of the current fixation is beyond 1SD longer than the subsequent saccade amplitude, whereas $\mathcal{K}_i = -1$ indicates a saccade more than 1SD longer than the preceding fixation duration. Finally, $\mathcal{K}_i = 0$ means that the fixation length and subsequent saccade amplitude are statistically equivalent. It is worth noting that $\mathcal{K}_i = 0$ refers to the rare and ambiguous situation when a person is exhibiting long saccades preceded by long fixations or is making short fixations followed by short saccades.

We are unaware of any previous formulation of such a coefficient expressing the dynamics between ambient and focal processing. Below we validate its effectiveness during serial and parallel visual search and then demonstrate its utility in distinguishing visual behavior when viewing artwork, in the presence or absence of exogenous cues guiding attention.

4. EMPIRICAL EVALUATION OF $\mathcal{K}$

To validate ambient/focal characterization through $\mathcal{K}$ and to demonstrate its utility, we describe results of two eye-tracking studies. The first was focused on validation of the $\mathcal{K}$ coefficient and shows that the coefficient successfully captures ambient and focal eye movement patterns when they are expected during parallel and serial visual search tasks. The second study demonstrates applied usage of the $\mathcal{K}$ coefficient in distinguishing the dynamics of ambient/focal attention when viewing artwork under the effects of auditory cuing (audio description). $\mathcal{K}$ is shown as an effective complement to analytical process measures accompanying traditional measures of performance, for example, time to completion.

5. STUDY 1: VALIDATION OF $\mathcal{K}$ DURING VISUAL SEARCH

The main motivation of the first study was to validate the $\mathcal{K}$ coefficient during visual behavior that is considered exemplary of the combination of ambient and focal visual scanning, namely, visual search.

In particular, we decided to replicate the empirical procedure put forth by Nothdurft [1999]. In his study, Nothdurft showed that serial visual search involves sequential shifts of focal attention, whereas no such shifts occur when performing parallel search. Nothdurft also concluded that during search the focus of attention moves to the target (and away from nontargets) both in serial and parallel search.

We hypothesized that during serial search participants would exhibit more focal attention compared to parallel search. That is, due to the pop-out effect during parallel search, one expects fast localization of the target evidenced by a long saccade (large amplitude) directed to the target. Reaction times reported by Nothdurft bear this out. By its definition, therefore, the $\mathcal{K}$ coefficient should characterize these eye movements to the pop-out target as ambient. Conversely, if serial search is composed of longer fixations interspersed by shorter saccades, $\mathcal{K}$ should characterize serial search as focal.

We also expected that target presence would stimulate more focal eye movements compared to its absence. Thus, we expected $\mathcal{K}$ in serial search to be largely positive, while in parallel search it should largely be negative.

5.1 Methodology

To replicate Nothdurft's [1999] study, we followed a within-subjects 2×2 factorial design. The two independent variables were search condition (serial versus parallel) and target presence (hit versus rejection). We considered the following main dependent variables: task accuracy (measured with proportion of correct answers), task speed (measured with time to task completion in milliseconds), and $\mathcal{K}$ obtained from eye-tracking data.

Each participant's task was to search for a vertically oriented Gabor patch and to indicate, as quickly and accurately as possible, whether the target was present or not.

5.1.1 *Participants.* Thirty-three university students took part in the experiment from which data of six were removed due to poor eye-tracker calibration or other technical problems resulting in a final sample of $N = 27$ (19 female and 8 male subjects). The average age of participants in the final sample was 23.33 (SD = 4.08). All participants were right-handed. Students participated voluntarily. They were rewarded with small gifts (sweets) and a study record point. Before starting the experiment all participants signed a consent form.

5.1.2 *Stimuli.* Participants were shown a set of ten circular Gabor patches (sinusoidal gratings at frequency of 3 Hz with linear interpolation and Gaussian distortion) sized at 2°. All stimuli were located at 8° from center. The target stimulus consisted of a vertically oriented Gabor patch (only one such target was displayed in the target-present conditions). In the parallel search task all nontarget stimuli consisted of horizontally oriented Gabor patches. In the serial search task all nontarget stimuli consisted of randomly oriented Gabor patches with angle of rotation between 20° and 340° (see Figure 1).

5.1.3 *Apparatus.* Eye movements were recorded at 60 Hz with a GazePoint2 eye tracker. Participants were seated in front of a laptop computer monitor (1600×900 resolution; 15″ LCD).

Python PsychoPy libraries [Peirce 2009] were used to prepare the experimental procedure and control the stimuli presentation synchronized with recorded eye movements. For detection of fixations and saccades we used custom software written in Python. The Python gaze analytics pipeline processed raw gaze points (x_i, y_i, t_i) with a second-order Butterworth filter set to 60Hz sampling and 6.15Hz cutoff frequencies, respectively, followed by differentiation by a six-tap Savitzky-Golay filter of degree 2. Assuming 60cm viewing distance, output of the Savitzky-Golay filter (signal velocity) was thresholded at $\pm 20^\circ$/s to identify saccades. Data not identified as saccades were treated as fixations (see Duchowski et al. [2014] for details).

(a) Serial search: "hit" (b) Serial search: "reject" (c) Parallel search: "hit" (d) Parallel search: "reject"

Fig. 1. Examples of visual stimuli in the four experimental conditions in Study 1. Following the study by Nothdurft [1999], starting by fixing their gaze on a central crosshair, subjects were asked to search for a vertically oriented Gabor patch (Nothdurft used lines) and to respond as soon as they detected the target, or to indicate if the target was absent. Search patterns contained 10 items at 8° eccentricity equally spaced on a circle.

5.1.4 *Procedure.* After completion of the consent form and familiarization with equipment and the experimental procedure, a five-point calibration of the eye-tracking system was performed. Each participant then started the experimental procedure preceded by a short training component. In both components, the participant's task was to determine whether the target stimulus was present or absent. Participants indicated their decision by pressing the "z" and "m" keyboard keys, respectively, as quickly and as accurately as possible.

In the training component, each participant completed four trials (two hit and two reject) for each of the two tasks (serial and parallel). Participants were given simple feedback ("Correct!" or "Wrong!") depending on their performance.

In the main component of the procedure, participants completed 10 hit and 10 reject trials in each of the serial and parallel task blocks.

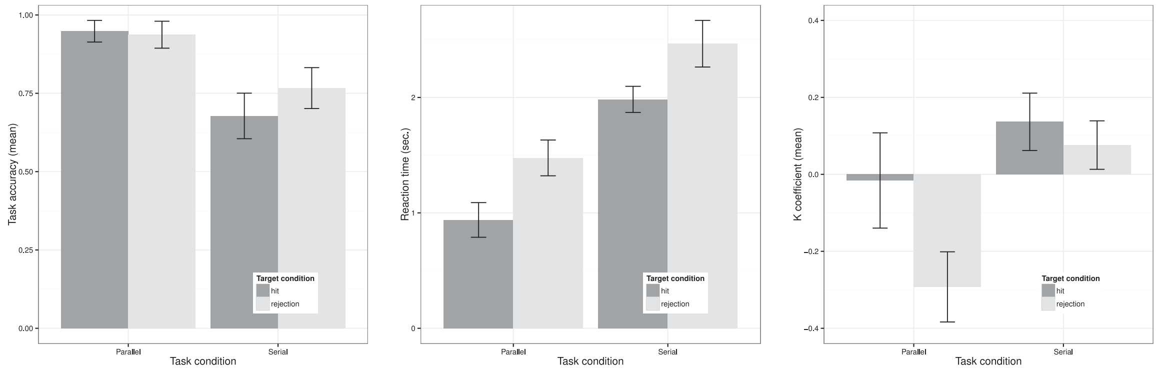
In both training and main components of the experimental procedure, trial blocks were presented in fully randomized order. In the target-present (hit) conditions, the position of the target stimulus was randomized with no repetitions. In the main part of the experimental procedure no feedback was provided. After completing the procedure, participants were thanked and rewarded.

5.2 Results

All statistical analyses were computed in *R*. For most of the analyses we used Analysis of Variance (ANOVA), with Bonferroni post hoc comparisons when called for.

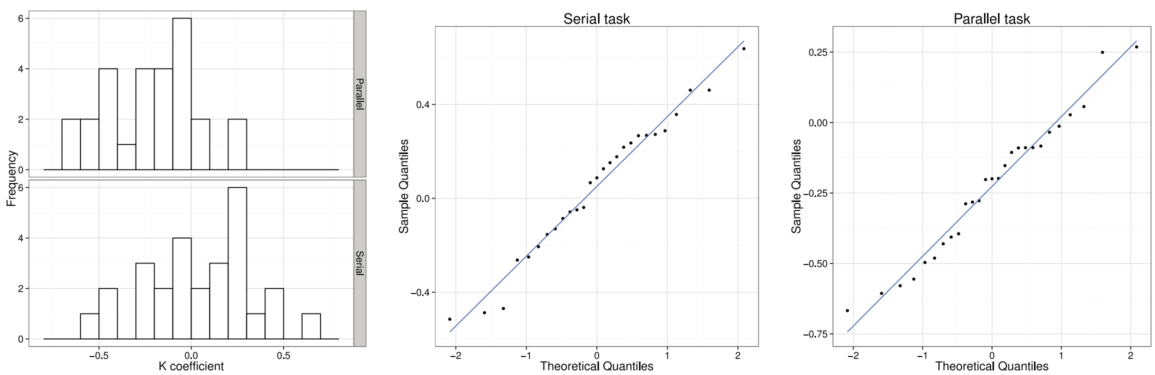
5.2.1 *Accuracy and Reaction Time.* We start with analysis of accuracy and reaction time in the visual search task. We conducted two-way within-subject ANOVAs separately for accuracy and reaction time as dependent variables. Similar to the results reported by Nothdurft [1999], the first analysis revealed the main effect of visual search type, $F(1, 26) = 46.30$, $p < 0.001$, $\eta^2 = 0.273$. Accuracy was significantly higher in parallel search ($\mathrm{M} = 0.95$, $\mathrm{SE} = 0.02$) than in serial search ($\mathrm{M} = 0.72$, $\mathrm{SE} = 0.03$). The main effect of target presence did not reach statistical significance, $F(1, 26) = 1.91$, $p > 0.1$, $\eta^2 = 0.012$. However, the interaction effect of visual search type and target presence reached statistical tendency, $F(91, 26) = 3.90$, $p = 0.059$, $\eta^2 = 0.019$ (see Figure 2(a)).

A similar analysis was conducted for task completion time as the dependent variable. Analysis yielded two significant main effects (see Figure 2(b)). As expected, the main effect of search type indicated that reaction times were significantly shorter in parallel search ($\mathrm{M} = 1.21$, $\mathrm{SE} = 0.07$) than in serial search ($\mathrm{M} = 2.22$, $\mathrm{SE} = 0.07$), $F(1, 26) = 81.42$, $p < 0.001$, $\eta^2 = 0.466$. The second main effect



(a) *Accuracy depending on visual search.* (b) *Reaction time depending on visual search and target presence (two main effects).* (c) *Interaction effect of visual search and target presence on $\mathcal{K}$ coefficient.*

Fig. 2. Differences in accuracy, reaction time, and $\mathcal{K}$ coefficient depending on experimental conditions in Study 1 (whiskers represent 0.95 confidence interval for the means).



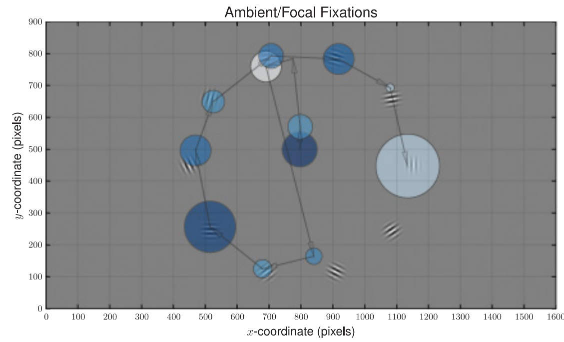
(a) *Histograms of $\mathcal{K}$ for both exp. conditions* (b) *QQ plot of $\mathcal{K}$ for serial search condition* (c) *QQ plot of $\mathcal{K}$ for parallel search condition*

Fig. 3. Distributions of $\mathcal{K}$ coefficient for serial and parallel experimental conditions in Study 1.

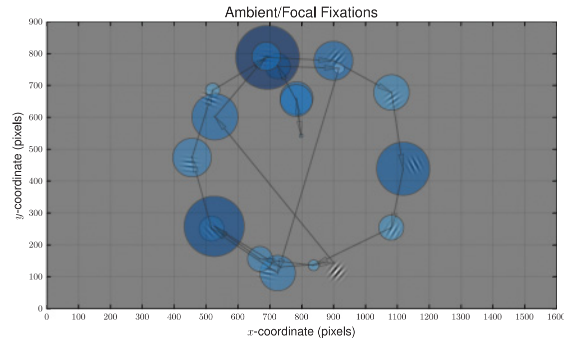
showed that participants made their decisions faster when the target was present (M = 1.46, SE = 0.12) than when it was absent (M = 1.97, SE = 0.13), $F(1, 26) = 65.46$, $p < 0.001$, $\eta^2 = 0.018$.

5.2.2 *Ambient/Focal Attention ($\mathcal{K}$ Coefficient Validation).* Prior to the main analysis of variance of $\mathcal{K}$ between experimental conditions, we conducted normality tests on $\mathcal{K}$ as the dependent variable to verify assumptions of normality distribution, using a series of Shapiro-Wilk tests. The analyses were run separately for each of serial and parallel search conditions. Results of these tests revealed that for both conditions $\mathcal{K}$ was normally distributed about the mean. For serial search, $W = 0.973$, $p = 0.69$ (see Figure 3(b)), and for parallel search, $W = 971$, $p = 0.62$ (see Figure 3(c)). Histograms of $\mathcal{K}$ distributions for both experimental conditions are shown in Figure 3(a).

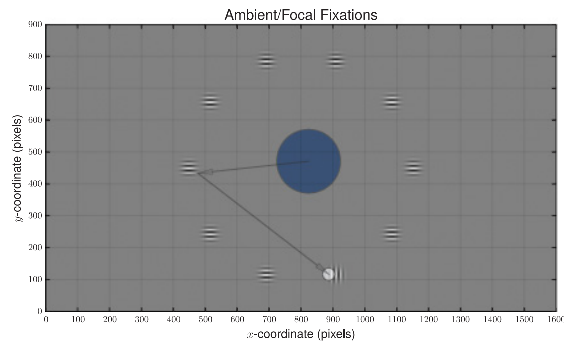
A two-way ANOVA of $\mathcal{K}$ with search type (serial vs. parallel) and target presence (hit versus reject) as independent within-subject factors revealed a significant main effect of visual search, $F(1, 26) = 12.37$, $p < 0.001$, $\eta^2 = 0.126$. Consistent with the hypothesis, descriptive statistics showed that in parallel search, $\mathcal{K}$ showed eye movements to be significantly more ambient (M = −0.19, SE = 0.04) than in



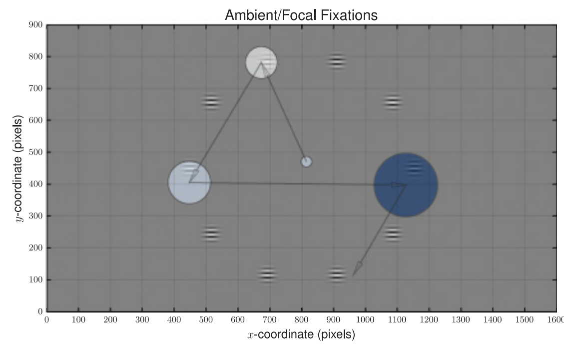
(a) Example of scanpath in serial search, hit condition



(b) Example of scanpath in serial search, reject condition



(c) Example of scanpath in parallel search, hit condition



(d) Example of scanpath in parallel search, reject condition

Fig. 4. Examples of scanpaths in the four experimental conditions in Study 1: fixations with darker shade indicate more focal viewing; fixations with lighter shade indicate ambient viewing (see Duchowski and Krejtz [2015]).

serial search (M = 0.10, SE = 0.03). The mean value of $\mathcal{K}$ for serial search suggests focal eye movements (see Figures 4(a) and 4(b), which depict scanpaths during serial search, color-coded to indicate focal fixations).

Also consistent with expectations, the effect of target presence was also statistically significant, $F(1, 26) = 23.87$, $p < 0.001$, $\eta^2 = 0.029$. Descriptive statistics showed that $\mathcal{K}$ showed eye movements in the rejection trials to be significantly more ambient (M = −0.06, SE = 0.03) than in the hit trials (M = 0.09, SE = 0.04) (see Figures 4(c) and 4(d), which depict scanpaths during parallel search, color-coded to indicate ambient fixations).

The interaction effect of visual search and target presence was marginally significant, $F(1, 26) = 4.09$, $p = 0.054$, $\eta^2 = 0.02$. As seen in Figure 2(c), the results suggest the most ambient eye movements were exhibited in parallel search reject conditions (M = −0.29, SE = 0.05), while most focal eye movements were observed in the serial search hit condition (M = 0.14, SE = 0.04).

5.3 Brief Discussion

The present study replicated the results reported by Nothdurft [1999]. When induced to perform parallel search, participants decided on target presence much faster and with greater accuracy than when induced to perform serial search. The presence of the target (*hit* condition) made the decision easier resulting in faster reaction times than in the *rejection* condition, in which the search target was absent.

These results suggest that one may expect ambient attention to dominate in parallel search, while focal attention is required for serial search, corroborating Nothdurft.

More importantly, the present study validates $\mathcal{K}$ in its ability to successfully distinguish between ambient and focal eye movement patterns. Ambient attention was expected in parallel search accompanied by lower $\mathcal{K}$ values than in serial search. When the task is made especially difficult for viewers (i.e., target-absent *rejection* condition), the difference in $\mathcal{K}$ is clear.

The utility of the $\mathcal{K}$ coefficient is seen despite very short stimulus presentation times. In this instance, stimuli presentation time was self-paced and averaged at about 1.5s. The short duration of this study precludes deeper analysis of $\mathcal{K}$ dynamics. To do so, longer presentation times are needed to allow examination of shifts between ambient and focal attention, taking advantage of $\mathcal{K}$'s dynamic expressiveness. The desire to explore ambient/focal dynamics was a partial motivation of Study 2.

6. STUDY 2: DYNAMICS OF VISUAL ATTENTION DURING ART VIEWING

Visual perception is guided by two types of processes: exogenous cues (stimuli characteristics) or endogenous control of attention (individual intentions or characteristics) [Posner 1980; Klein 2004].

An example of exogenous cuing was developed by Bailey et al. [2012] who showed that introducing peripheral motion to scene perception is an effective way of modifying scanpaths. They combined eye tracking with subtle image-space modulations (peripheral blinking) to guide the viewer's gaze about a scene. This type of exogenous manipulation significantly improved the accuracy of the target count and spatial location recall.

Similarly, Audio Description (AD) can be considered an exogenous cue, although an acoustic and not a visual one. AD is a form of verbal narration accompanying artwork [De Coster and Muhleis 2007], film [Orero 2007; Krejtz et al. 2012a, 2012b], theater performance [Holland 2008], opera [Cabeza-Cáceres 2010], and museum exhibitions [Matamala and Orero 2007]. AD of an art piece usually contains: (a) historical information (the name of the artist, the title of the painting/artwork, date, dimensions (height/width), technique); and, more importantly, (b) a verbal description of the visual aspects of the work.

Exogenous, audio-described cues to the visual elements of the artwork appear to elicit endogenous control of visual attention, focusing the observer's attention and facilitating memory of the artwork (see Nummenmaa et al. [2006]). Relatively few studies have examined the influence of endogenous control on attentional processes. Endogenous, intentional exploration of artwork may take control from exogenous cues, but generally requires some background knowledge [Hekkert and van Wieringen 1996], personal interest, or external task requirements [Yarbus 1967; Rothkopf et al. 2007]. For example, Zangemeister et al. [1995] modified viewers' scan patterns of artwork by suggesting different instructions such as memorization of content, or focus on the artistic aspects of the artwork. We use the exogenous AD cue to show how coefficient $\mathcal{K}$ can quantitatively tease apart subsequent, we presume endogenously controlled, spatial distribution of visual attention.

In visual processing of a scene, there is interplay between both types of attentional shifts: attention is reflexively drawn to spatial locations either by external cues or by goal-driven intent, for example, see Zoeat and Donk [2004]. Similarly, visual exploration of an art piece is the result of the interplay between endogenous attentional control and exogenous cues guiding the observer's attention [Massaro et al. 2012; Cupchik et al. 2009]. Research investigating these two aspects of attentional guidance has mainly focused on common (e.g., salient) exogenous factors [Castelhano and Henderson 2007].

Naive art viewers view artwork mostly relying on exogenous cues, which automatically direct attention toward pop-out elements of a scene [Mackworth and Morandi 1967; Santella and DeCarlo 2004]. Massaro et al. [2012] enumerates exogenous cues that modify observers' viewing patterns: luminance contrast [Ramachandran and Hirstein 1999], balance [Boselie and Leeuwenber 1985], symmetry [Jacobsen and Hofel 2002], color [Martindale and Moore 1988], and orientation [Le Meur and Chevet 2010]. Examining the dynamics of eye movements, one may see an analogy between exogenous/endogenous processes and ambient/focal attention shifts. At early stages of viewing time, viewers exhibit ambient eye movements, allowing them to get the general sense (gist) of the entire scene [Velichkovsky et al. 2005]. During this time, attention is likely guided by stimuli characteristics, then slowly shifting toward more focal processing. During the latter phase the focus of the attention is likely motivated by endogenous factors [Velichkovsky et al. 2005; Krejtz et al. 2012b].

Recording of eye movements provides clear manifestation of the overt orientation of visual attention. When monitoring gaze, we assume that direction of gaze corresponds to visual attention. As such, eye tracking has become an important method with which to study human attention and the correspondence between attention and behavior [Duchowski 2002]. Recording of eye movements when viewing a painting also provides useful insight into the interaction between endogenous attentional shifts and exogenous cues [Locher 2006].

The use of audio description as an exogenous factor appears to influence the interplay between focal and ambient visual processing. We hypothesize that listening to AD while visually scanning a painting modifies the distribution of attention compared to the control condition. Thus, AD, in guiding attention, induces endogenous control of attention beyond its response to exogenous visual factors such as luminance or color and keeps the attention focused for a longer period when viewing the art piece.

A number of studies have indicated beneficial effects of audio description on comprehension among the blind [Schmeidler and Kirchner 2001; Peli et al. 1996] and sighted viewers [Krejtz et al. 2012a; Vilaró et al. 2012]. Schmeidler and Kirchner [2001] verified the impact of watching scientific programs with and without audio description. Results confirmed that blind respondents retained more details from audio-described programs.

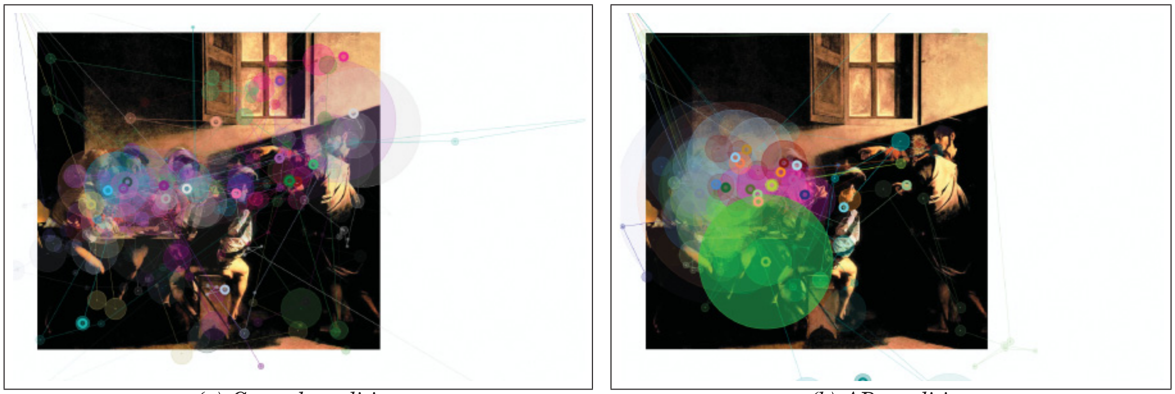
Vilaró et al. [2012] confirmed that an audio track influences visual perception and understanding of audiovisual texts. In line with these findings, Kruger [2010] demonstrated that visually peripheral elements may gain particular narrative importance and compete with more exogenous cues. Moreover, Krejtz et al. [2012a] showed that AD helped to focus children's visual attention on relevant elements of the scene. The AD script also provided children with a specialized vocabulary, which fostered their understanding of the presented material.

In the present study, in order to evaluate the effect of audio description on patterns of visual attention during perception of art, as well as its effect on content memorization, still images were presented accompanied by audio description. Eye movements were tracked while high school students, naive to art critique, viewed two paintings in two experimental conditions: with and without AD. Analysis of results tests the effectiveness of coefficient $\mathcal{K}$ in distinguishing the dynamics of viewing patterns in the presence of AD as the external source of attentional cuing when visually exploring classic artwork.

6.1 Methodology

The present eye-tracking experiment followed a 2×4 mixed factorial design with the following *independent variables*:

(1) Experimental condition: the paintings were presented with AD or without audio description (CONTROL).
(2) Description segments (time intervals): due to the unequal duration of stimulus presentation, we decided to cut each presentation time into four equal segments. This recoded the data into a four-level factor, for which each level represents the beginning, two middle segments, and the ending segment of the stimulus presentation (viewing of the painting). This construction of the time interval factor allowed us to include it within the full factorial analyses together with the three



(a) Control condition (b) AD condition

Fig. 5. Scanpaths on the stimulus in control and audio description over 120 seconds.

remaining experimental conditions. We hypothesized that during the first segments, we would observe ambient attention becoming more focal in the latter segments. We also expected the AD condition would modify this dynamic of ambient/focal attention in a way when focal viewing would be reached faster and sustained longer than in the CONTROL condition.

6.1.1 *Participants.* Thirty seven junior high school students took part in the study. Data from one of them were excluded from the analyses due to poor eye-tracker calibration resulting in a final sample of N = 36 with ages between 14 and 16 years old and gender equally distributed (19 male and 19 female). Participants were randomly assigned to the experimental condition resulting in 18 participants in the AD group and 20 in the CONTROL group. Students participated voluntarily and were rewarded with small gifts. Before starting the experimental procedure, participants signed a consent form and were obliged to provide consent forms approved by their parents.

6.1.2 *Stimuli.* Participants were presented with two paintings: “The Calling of St. Matthew” by Caravaggio and “The Umbrellas” by Pierre-August Renoir, in random order. Before presentation of each picture, textual information about the painter was given, for example,

> *Caravaggio was an Italian painter, one of the most prominent artists of the early Baroque. He is credited with the invention of Tenebrism, a technique using pronounced chiaroscuro, producing stark contrast between light and dark, resulting in figures and elements emerging from the dark background.*

In the CONTROL condition the painting presentation was without any additional description, whereas in the AD condition audio description accompanied the picture (see Figure 5). The audio description was presented aurally by a recording of a voice talent performer. The audio description consisted of two parts, each lasting 2 minutes. The first part was general information about the painting, for example,

> *Caravaggio’s The Calling of St. Matthew. Oil on canvas, nearly 11 feet × 11 feet 5 inches. The painting was completed in 1599-1600 in the Contarelli Chapel in the church of San Luigi dei Francesi in Rome, where it remains today. It was the first serious commission that allowed Caravaggio to become an independent artist.*

The second part was a detailed description of the painting's key elements, for example,

> *The light, whose source is unknown, falls like a wedge from the top right-hand corner, from outside of the image, over the heads of Christ and his disciple. The beam of light diffuses and falls on the men at the table. Farthest at left, at the head of the table, sits a young man painted in profile.*

6.1.3 *Apparatus.* Eye movements were recorded at 250Hz with an SMI eye-tracking system. Participants were seated in front of a computer monitor (1024 × 768 resolution; 22″ LCD, 60Hz refresh rate). The eye tracker's remote mode of tracking was used to promote participants' comfort. The spatial resolution of the eye-tracking system was reported by the vendor as 0.03° visual angle with gaze position accuracy of 0.4° visual angle. SMI's Experiment Center software was used to present stimuli synchronized with recorded eye movements. For detection of fixations and saccades following raw data cleaning SMI's BeGaze software was employed. We used default BeGaze settings for detecting fixations and saccades with their High Speed Event Detection method. With this method saccades are treated as primary events and fixations are derived from them. In order to capture saccades the method uses a velocity-based algorithm. With this algorithm the peak velocity threshold is set to 40°/s and the minimum saccade duration is set to 22ms, while the minimum fixation duration is set to 50ms. Participants in the AD condition wore headphones in which the AD was played.

6.1.4 *Procedure.* After completion of the consent form and familiarization with equipment and the experimental procedure, a nine-point calibration of the eye-tracking system was performed. The following procedure consisted of two steps:

(1) Learning phase. The presentation of paintings by Caravaggio or Renoir in random order, on the computer screen, with eye movement recording. Depending on the experimental condition, participants either heard, in their headphones, the recorded voice description (AD condition), or were presented with only the painting (CONTROL condition).
(2) Memory task. After viewing each painting, participants were asked to solve an interactive 25-piece jigsaw puzzle prepared from the Caravaggio or Renoir painting consistent with step 1 of the procedure.

The presentation time in the CONTROL condition was shorter (2 minutes) than in the AD condition to prevent boredom or fatigue setting in, in the absence of additional materials.

6.1.5 *Dependent Measures.* Results were analyzed in terms of eye movement characteristics during the learning phase of the experiment and behavioral measures during the memory task. Eye movement analyses focused on classical eye movement process measures, the measure of ambient/focal attention ($\mathcal{K}$ coefficient), and accuracy of the memory test. The following *dependent variables* were derived:

(1) Fixation Frequency: defined as the number of fixations per second. We expected that in AD condition participants would make fewer fixations per second than in the CONTROL condition.
(2) Fixation Duration: measured in milliseconds, summing all fixation durations divided by the number of fixations. We hypothesized that the average fixation duration would be higher in the AD condition compared to the CONTROL condition.
(3) Saccade Amplitude: measured in degrees, summing all saccade amplitudes divided by the number of saccades. We assumed shorter saccades in the AD condition compared to the CONTROL condition due to more focused eye movement patterns.
(4) Ambient/Focal $\mathcal{K}$ Coefficient: we expected differences between ambient and focal viewing with more focal processing by the AD group, especially during the description of the painting's elements. We

also expected differences in the temporal pattern of $\mathcal{K}$ between experimental conditions: AD would cause faster switching toward focal attention as well as stronger sustainability of focal attention.

(5) Completion Time of Jigsaw Puzzle: we examined how long it took participants to reconstruct the visually presented paintings using their jigsaw puzzles. The prediction was that audio description would strengthen memory of the painting, and therefore result in shorter completion times by the AD group rather than by the CONTROL group.

6.2 Results

All statistical analyses were computed with R. For most of the analyses we used ANOVA tests. For analyses of memory test results we used Pearson's Chi-squared test of independence and Pearson's (first-order) correlation test.

6.2.1 *Eye Movement Process Measures.* Using traditional process measures (fixation frequency, fixation duration, and saccade amplitude), we tested the hypothesis that AD would draw and sustain attention over the paintings, resulting in a more deliberate visual investigation. If so, longer and less frequent fixations would be expected along with smaller saccade amplitudes. A series of three one-way between-subjects ANOVAs was carried out with the experimental condition as the fixed factor.

Analysis of fixation duration supported the hypothesis, with significant group differences, $F(1, 36) = 13.05$, $p < 0.001$, $\eta^2 = 0.266$. Descriptive statistics confirmed that AD the condition fixations were significantly longer (M = 344.73, SE = 14.31) compared to the CONTROL condition (M = 260.76, SE = 10.12).

ANOVA of fixation frequency as the dependent measure also showed a significant main effect of experimental condition, $F(1, 36) = 8.41$, $p < 0.001$, $\eta^2 = 0.189$. In line with the hypothesis, participants in the AD condition produced significantly fewer fixations (M = 2.4, SE = 0.07) than those in the CONTROL condition (M = 2.78, SE = 0.07).

Finally, analysis of saccade amplitude revealed a main effect of experimental condition that was only marginally significant, $F(1, 36) = 3.38$, $p = 0.074$, $\eta^2 = 0.086$. Saccade amplitudes were shorter in the AD condition (M = 3.02, SE = 0.16) compared to the CONTROL condition (M = 3.74, SE = 0.23).

6.2.2 *Ambient/Focal Attention.* Prior to analysis of ambient/focal attention per condition with the $\mathcal{K}$ coefficient, we wanted to examine the static relationship between saccade amplitude and fixation duration as per Unema et al. [2005]. Figure 6 shows the nonlinear distribution of saccade amplitude as a function of fixation duration. As per Unema et al., saccade amplitudes were averaged over all fixations within 20ms bins. Figure 6 shows the average subsequent saccade amplitude for all fixations up to 600ms. Following Unema et al., longer fixation durations were left out of this analysis. As is clearly visible, a subset of fixations of $\leq$180 ms are accompanied, in general, by larger than average subsequent saccades. As with Unema et al., this appears to be omnipresent over two experimental conditions (AD and control group). In terms of the static relationship between saccade amplitude and fixation duration, our results replicate those of Unema et al., although the relation is less clear. The remainder of our analysis is a departure from this prior work, extending the analysis to show what happens to the two modes of visual attention over time as expressed by the $\mathcal{K}$ coefficient.

Prior to hypotheses testing with the use of $\mathcal{K}$, we decided to check the assumption of normality of frequency distributions. Similar to analyses described in Section 5, two Shapiro-Wilk normality tests were conducted separately for each experimental condition. Results for both analyses revealed that $\mathcal{K}$ frequency distributions do not deviate significantly from the normal distribution, for the AD condition: $W = 0.93$, $p = 0.17$ (see Figure 7(a)); and for the CONTROL condition: $W = 0.98$, $p = 0.86$ (see Figure 7(b)). Plotting quantiles of observed data and theoretical normal distributions against each other (see Figure 7) shows that data points lie close to the line indicating normal distribution of $\mathcal{K}$ in both conditions.

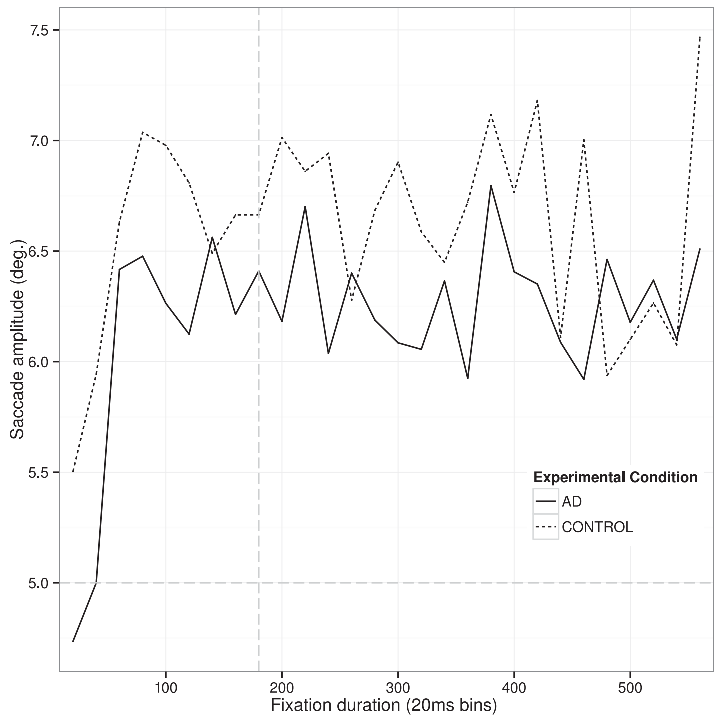
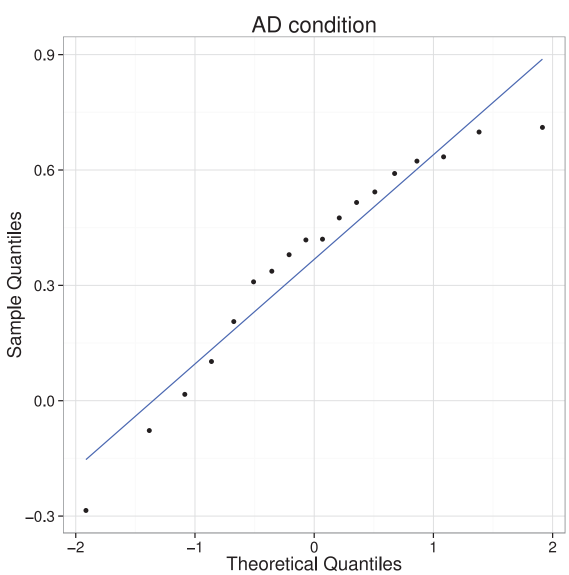
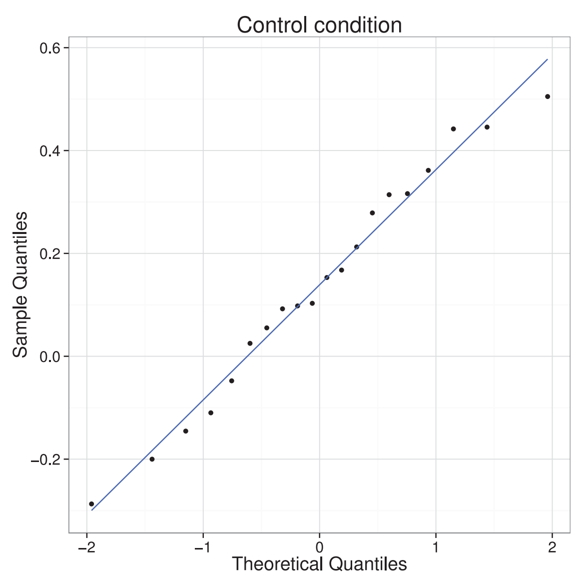


Fig. 6. Distribution of saccade amplitude as a function of fixation duration, following Unema et al. [2005].

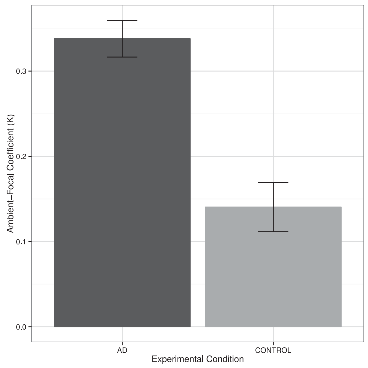


(a) $\mathcal{K}$'s QQ-plot in AD condition

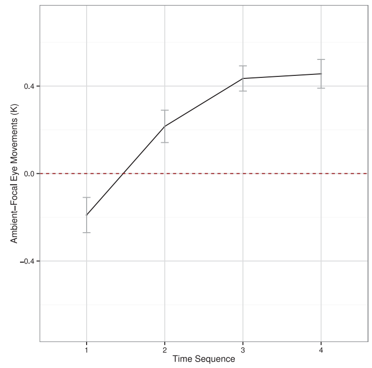


(b) $\mathcal{K}$'s QQ-plot in control condition

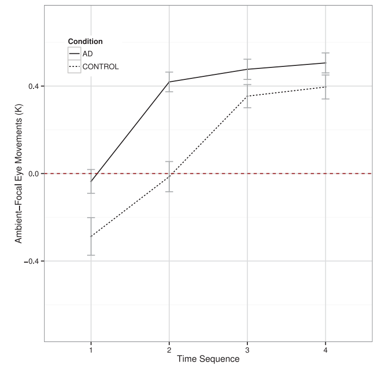
Fig. 7. Quantile-Quantile plots of $\mathcal{K}$ coefficient distributions in the experimental conditions.



(a) *Ambient/Focal attention depending on exp. condition.*



(b) *Dynamics of ambient/focal attention.*



(c) *Dynamics of ambient/focal depending on exp. condition.*

Fig. 8. Ambient/focal attention effects (whiskers represent 0.95 confidence interval for the means).

We hypothesized that AD accompanying the artwork would elicit and sustain more deliberate visual investigation, manifested by more focused viewing lasting for a longer period of time. To verify this prediction, we used $\mathcal{K}$ as a dependent measure in a mixed-design (2×4) ANOVA. The experimental condition (AD versus CONTROL) was treated as the between-subject factor and the time interval as the within-subject factor. Stimulus presentation time in each condition was split into four equal segments.

As hypothesized, analysis revealed a significant main effect of condition, $F(1, 36) = 10.25$, $p < 0.01$, $\eta^2 = 0.164$. Participants in the AD condition exhibited, on average, significantly more focal eye movements ($M = 0.37$, $SE = 0.07$) compared to the CONTROL condition ($M = 0.14$, $SE = 0.05$).

We also obtained, as expected, a significant main effect of time, $F(3, 108) = 81.93$, $p < 0.001$, $\eta^2 = 0.41$. Descriptive statistics showed that during the first quarter of viewing, the eye movement pattern was generally ambient ($M = -0.19$, $SE = 0.24$), becoming focal in the second ($M = 0.22$, $SE = 0.04$), third ($M = 0.43$, $SE = 0.03$), and fourth quarters ($M = 0.46$, $SE = 0.03$) (see Figure 8(b)). *Post hoc* tests showed that the observed differences between the first and all subsequent quarters were statistically significant ($p < 0.01$). Also, the differences between the second versus third and fourth time periods reached statistical significance ($p < 0.05$). The mean $\mathcal{K}$ in the third and fourth time periods was not statistically different ($p = 1$). These results are again in line with Unema et al. [2005].

Consistently with the hypothesis, the interaction of time and experimental condition was also statistically significant, $F(3, 108) = 5.04$, $p < 0.01$, $\eta^2 = 0.042$, suggesting that changes of the ambient and focal attention coefficient are moderated by experimental condition (see Figure 8(c)).

Decomposing the interaction effect by experimental condition levels we see the significant simple main effect of time for the AD condition, $F(3, 51) = 56.02$, $p < 0.001$, $\eta^2 = 0.361$. The following pairwise comparisons with Bonferroni correction showed the significant differences between first and all later time periods ($p < 0.001$) and no significant differences between second, third, and fourth periods. Besides the first quarter, where slightly ambient eye movements were recorded ($M = -0.02$, $SE = 0.05$), in the remaining three quarters viewing patterns were focal: second quarter ($M = 0.45$, $SE = 0.03$), third quarter ($M = 0.52$, $SE = 0.03$), and fourth quarter ($M = 0.53$, $SE = 0.02$). Note that in the second quarter participants reached focal viewing with $\mathcal{K}$ remaining stable until the end of the experimental task.

The simple main effect for the CONTROL condition also reached significance, $F(3, 57) = 41.35$, $p < 0.001$, $\eta^2 = 0.482$. Pairwise comparisons with Bonferroni correction showed that in the first quarter, the viewing pattern was significantly ($p < 0.01$) more ambient than in all other time periods. But there was also significantly ($p < 0.01$) more focal attention in the third and fourth periods compared with the second. Descriptive stats for $\mathcal{K}$ in all time periods are as follows: for the first time period (M = −0.35, SE = 0.06), for the second (M = 0.002, SE = 0.05), for the third (M = 0.36, SE = 0.04), and for the fourth (M = 0.39, SE = 0.05). One may clearly notice that, unlike the AD condition, the stabilization of focal viewing was reached later in the third and fourth quarters.

6.2.3 *Memory Effects of Audio Description.* As a consequence of more focal viewing, we hypothesized that viewers exhibiting greater focal attention would better recollect the paintings they viewed. In order to examine group differences in the recollection of the paintings, the time to complete the jigsaw puzzles was examined via analysis of variance.

Generally, most participants managed to finish the jigsaw puzzles: 65% (72% in the AD condition and 58% in the CONTROL group) of the sample completed the puzzles for both paintings, 22% did so for only one painting (22% in the AD and 21% in the CONTROL condition), and 14% did not complete either (6% in the AD condition and 21% in the CONTROL group). To check the differences between experimental conditions in the rate of jigsaw puzzle task completion we performed a Pearson's Chi-squared test. The result of the test was not statistically significant, $\chi^2(2) = 1.94$, $p = 0.38$, indicating that there is no relation between experimental condition and jigsaw puzzle completion time.

Furthermore, we analyzed the time to completion of the jigsaw puzzles among those participants who completed the task. One-way ANOVA was performed with condition as a between-subjects factor. In line with the hypothesis, results revealed a significant main effect, $F(1, 30) = 4.95$, $p < 0.05$, $\eta^2 = 0.142$. The following descriptive statistics showed that participants in the AD group completed the jigsaw puzzle task faster (M = 92.71, SE = 3.86) compared to the CONTROL condition (M = 106.53, SE = 4.97).

Bearing in mind previous results showing that the the AD condition elicited more focal viewing patterns on average, one may claim that the focal pattern of visual stimuli inspection fosters better performance in the memory test. In order to justify such a claim we conducted a direct simple Pearson's correlation of ambient/focal $\mathcal{K}$ coefficient with jigsaw puzzle completion time. The analysis was conducted again on only those participants who successfully completed the jigsaw puzzle task. The result of this analysis partially confirms our prediction (the correlation test turned out to be marginally significant), $r = -0.35$, $t(29) = -2.03$, $p = 0.051$, with the memory task completed faster given a more focal viewing pattern during the learning phase.

6.3 Brief Discussion

Study 2 demonstrates the utility of the ambient/focal coefficient $\mathcal{K}$ in discerning the difference between participants' eye movements subject to different exogenous visual attention cues. Plotting $\mathcal{K}$ over the time course of the stimuli presentation allows depiction of dynamic patterns defined by switching between ambient ($\mathcal{K} < 0$) and focal ($\mathcal{K} > 0$) modes of visual attention.

The main hypothesis stated that audio description would modify and align the participants' distribution of visual attention in the AD condition compared to the CONTROL condition and also change the dynamic pattern of ambient/focal attention switching (keeping viewers' attention more focused for a longer period of time). The analyses employed classical process measures of eye movements (fixation duration, fixation frequency, and saccade amplitude) and the novel ambient/focal ($\mathcal{K}$) measure.

First, we demonstrated that in the AD condition participants exhibited fewer but longer fixations than in the CONTROL group. This suggested more deliberate viewing of the presented artwork.

Second, comparison of averaged $\mathcal{K}$ between experimental conditions corroborated the preceding results and provided more evidence for interpretation of more deliberate viewing while visually inspecting the audio-described picture.

Third, the interaction effect of time and experimental condition on $\mathcal{K}$ supports our claim that audio description stimulates quicker achievement of focal viewing compared to the same stimuli presented with no additional description. In other words, $\mathcal{K}$ differentiates dynamic patterns of ambient/focal attention shifts between experimental conditions. It is clear that audio description elicited faster focal viewing and its sustainability at that level. What is more, this conclusion would be hard to support with analysis of only classical eye movement process measures.

Finally, participants completed the memory task (jigsaw puzzle) faster when the picture was accompanied by audio description compared to the CONTROL group. This finding supports the notion that audio description strengthens memory of the artwork. The analysis of $\mathcal{K}$ suggests that one of the mechanisms is fostering and sustaining more deliberate focal visual inspection.

According to the Dual Coding Theory [Sadoski and Paivio 2004], humans process information in parallel through auditory and visual channels, therefore information is stored and retrieved by both, in nonverbal and verbal form. The dual coding hypothesis in the AD condition explains facilitation of memory retrieval because the same information was accessible via auditory and visual inputs. The temporal coincidence of audio description and visual perception prompted participants' visual attention toward spatial locations described by the audio. Consequently, perception of the art followed the "what and where" verbal direction embedded within the audio description. This type of audio/video integration forms one of the basic rules of construction of multimedia learning, namely, the temporal contiguity principle, which states that audio should be synchronized in time with presented visual information [Mayer 2002]. The present results suggest that AD can serve designers of multimedia learning environments in facilitating adhesion to Mayer's principle when constructing complex visual stimuli such as art pieces or film [Krejtz et al. 2012a].

There are several research examples showing that what we hear can change what and how we see. For example, when one observes a computer screen with two identical objects moving toward each other, overlapping, and then moving apart, two different interpretations can be made. Either the objects pass each other, or collide and reverse course. Sekuler et al. [1997] showed that presentation of sound synchronized with the overlap increased the number of interpretations of collision. This simple example supports conclusions made in the present study. Synchronizing audio description with visual perception of a painting significantly influenced how participants saw the art pieces, and how they encoded representation of the images in memory. The coefficient $\mathcal{K}$ was useful in visualizing and quantitatively capturing the groups' dynamics.

Future studies should consider effects previously described by Hollingworth and Henderson [2002], who suggest that relatively detailed visual information is retained in memory from previously attended objects in natural scene viewing. Our present results are in line with this hypothesis and can be explained by their model of scene perception and long-term memory: facilitation of focal attention and its guidance toward important regions of art piece fosters its memorization. However, since there is a relatively strong relation between fixation time spent on complex visual stimuli and memory [Hollingworth and Henderson 2002; Melcher and Kowler 2001], more careful separation of focal attention process and overall fixation time is needed in future studies.

7. GENERAL DISCUSSION AND FUTURE DEVELOPMENTS

We have demonstrated the utility of the coefficient $\mathcal{K}$ to capture the dynamics of ambient/focal attention in the perception of complex visual stimuli and how its analysis can fit in among traditional analyses of performance and process metrics.

We note that the coefficient may need further development focused on two aspects, its standardization and its application to recording of gaze behavior under differing conditions. First, the coefficient's standardization should be aimed at developing clear boundaries (lowest and highest limits) within which the coefficient is to be used. This would facilitate the comparisons of $\mathcal{K}$ between different experimental conditions as well as between different studies. One may expect that, for different stimuli or different viewing situations, thresholds should be established for clearer interpretation of ambient and focal modes of attention. The categorization of eye movement behavior into one of these two modes would clearly broaden the interpretation possibilities of analysis based on $\mathcal{K}$. Second, future work should also focus on evaluation of $\mathcal{K}$ for different types of stimuli. Dynamic stimuli, that is, motion pictures, would be a good example. A key aspect that requires exploration is the temporal granularity of $\mathcal{K}$'s computation. Time segments that are too short or too long are likely to result in smaller variability of $\mathcal{K}$, which in turn would affect its sensitivity toward detecting significant differences between experimental conditions. Appropriate use of the coefficient for dynamical stimuli needs extensive future research. Nevertheless, we believe $\mathcal{K}$ opens new perspectives and possibilities for investigation of the dynamics of visual attention as reflected in the time course of recorded eye movements.

Finally, we plan to develop a clearer interpretation of the null value of $\mathcal{K}$, as it currently is ambiguous. It can mean long fixations followed by long saccades, or short fixations followed by short saccades. Although both seem to be rare, the distinction and understanding of underlying attentional processes is needed.

8. CONCLUSIONS

We introduced a new coefficient, $\mathcal{K}$, which allows for capturing the dynamics of visual attention switching along an ambient/focal parametric scale. The coefficient relates the standardized measure of fixation duration and subsequent saccade amplitude. Its validity was empirically demonstrated in capturing ambient/focal attention shifts during parallel and serial visual search. Its utility was then shown in its ability to discern viewing behavior cued by audio description when viewing complex static visual stimuli, namely, artwork. The results of the study provided support for the significant influence of simultaneous auditory and visual stimuli when perceiving and retaining artwork. The integration of audio description and visual information increased participants' ability to localize and memorize elements of the artwork.

ACKNOWLEDGMENTS

We thank Ms. Hanna Stachera and Regina Lewkowicz from *Staszic* High School in Warsaw, Poland, for their help in organizing Study 2. We also thank Marek Młodożeniec, Ewa Domaradzka, Dr. Rafał Albiński, and Anna Niedzielska for their help in conducting the studies.

REFERENCES

R. Bailey, A. McNamara, A. Costello, S. Sridharan, and C. Grimm. 2012. Impact of subtle gaze direction on short-term spatial information recall. In *Proceedings of the 2012 Symposium on Eye Tracking Research & Applications (ETRA'12)*. ACM, New York, NY, 67–74. DOI:http://dx.doi.org/10.1145/2168556.2168567

R. Bednarik, H. Vrzakova, and M. Hradis. 2012. What do you want to do next: A novel approach for intent prediction in gaze-based interaction. In *Proceedings of the 2012 Symposium on Eye Tracking Research & Applications (ETRA'12)*. ACM, New York, NY, 83–90. DOI:http://dx.doi.org/10.1145/2168556.2168569

C. Biele, A. Kopacz, and K. Krejtz. 2013. Shall we care about the user's feelings? Influence of affect and engagement on visual attention. In *Proceedings of the International Conference on Multimedia, Interaction, Design and Innovation (MIDI'13)*. ACM, New York, NY, Article 7, 8 pages. DOI:http://dx.doi.org/10.1145/2500342.2500349

F. Boselie and E. Leeuwenber. 1985. Birkhoff revisited: Beauty as a function of effect and means. *The American Journal of Psychology* 98, 1 (1985), 1–39.

A. Bulling, C. Weichel, and H. Gellersen. 2013. EyeContext: Recognition of high-level contextual cues from human visual behaviour. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'13)*. ACM, New York, NY, 305–308. DOI:http://dx.doi.org/10.1145/2470654.2470697

C. Cabeza-Cáceres. 2010. Opera audio description at Barcelona's Liceu theatre. In *Media for All 2: New Insights into Audiovisual Translation and Media Accessibility*, Jorge Díaz Cintas, Anna Matamala, and Josélia Neves (Eds.). Rodopi, Amsterdam, Holland, 227–237.

M. S. Castelhano and J. M. Henderson. 2007. Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance* 33, 4 (2007), 753–763.

G. C. Cupchik, O. Vartanian, A. Crawley, and D. J. Mikulis. 2009. Viewing artworks: Contributions of cognitive control and perceptual facilitation to aesthetic experience. *Brain and Cognition* 70, 1 (2009), 84–91.

K. De Coster and V. Muhleis. 2007. Intersensorial translation: Visual art made up by words. In *Media for All: Subtitling for the Deaf, Audio Description and Sign Language*, Jorge Díaz Cintas, Pilar Orero, and Aline Remael (Eds.). Rodopi, Amsterdam, Holland, 175–187.

A. T. Duchowski. 2002. A breadth-first survey of eye tracking applications. *Behavior Research Methods, Instruments, and Computers (BRMIC)* 34, 4 (2002), 455–470.

A. T. Duchowski, S. V. Babu, J.f Bertrand, and K. Krejtz. 2014. Gaze analytics pipeline for unity 3D integration: Signal filtering and analysis. In *Proceedings of the 2nd International Workshop on Eye Tracking for Spatial Research (ET4S)*.

A. T. Duchowski and K. Krejtz. 2015. Visualizing dynamic ambient/focal attention with coefficient $\mathcal{K}$. In *Proceedings of the 1st Workshop on Eye Tracking and Visualization (ETVIS)*.

B. Follet, O. Le Meur, and T. Baccino. 2011. New insights on ambient and focal visual fixations using an automatic classification algorithm. *i-Perception* 2, 6 (2011), 592–610.

P. Hekkert and P. C. W. van Wieringen. 1996. The impact of level of expertise on the evaluation of original and altered versions of post-impressionistic paintings. *Acta Psychologica* 94 (1996), 117–131.

A. Holland. 2008. Audio description in the theatre and the visual arts: Images into words. In *Audiovisual Translation: Language Transfer on Screen*, J. Díaz Cintas and G. Anderman (Eds.). Palgrave Macmillan, Basingstoke, UK.

A. Hollingworth and J. M. Henderson. 2002. Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance* 28, 1 (2002), 113–136.

K. Holmqvist, P. Lindström, and F. Ferrara. 2011a. A method for quantifying focused versus overview behavior in AOI sequences. *Behavior Research* 43 (2011), 987–998.

K. Holmqvist, M. Nyström, Ri. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer. 2011b. *Eye Tracking: A Comprehensive Guide to Methods and Measures.* Oxford University Press.

D. E. Irwin and G. J. Zelinsky. 2002. Eye movements and scene perception: Memory for things observed. *Perception and Psychophysics* 64 (2002), 882–895.

T. Jacobsen and L. Hofel. 2002. Aesthetic judgments of novel graphic patterns: Analyses of individual judgments. *Perceptual and Motor Skills* 95 (2002), 755–766.

R. M. Klein. 2004. On the control of visual orienting. In *Cognitive Neuroscience of Attention*, M. I. Posner (Ed.). The Guilford Press, 29–44.

I. Krejtz, A. Szarkowska, K. Krejtz, A. Walczak, and A. Duchowski. 2012b. Audio description as an aural guide of children's visual attention: Evidence from an eye-tracking study. In *Proceedings of the 2012 Symposium on Eye Tracking Research & Applications (ETRA'12)*. ACM, New York, NY, 99–106. DOI:http://dx.doi.org/10.1145/2168556.2168572

K. Krejtz, A. Duchowski, T. Szmidt, I. Krejtz, F. González Perilli, A. Pires, A. Vialró, and N. Villalobos. 2015. Gaze transition entropy. *ACM Transactions on Applied Perception (TAP)* 13, 1 (2015), 4:1–4:20.

K. Krejtz, I. Krejtz, A. Duchowski, A. Szarkowska, and A. Walczak. 2012a. Multimodal learning with audio description: An eye tracking study of children's gaze during a visual recognition task. In *Proceedings of the ACM Symposium on Applied Perception (SAP'12)*. ACM, New York, NY, 83–90.

J.-L. Kruger. 2010. Audio narration: Re-narrativising film. *Perspectives* 18, 3 (2010), 231–249.

O. Le Meur and J. C. Chevet. 2010. Relevance of a feed-forward model of visual attention for goal-oriented and free-viewing tasks. *IEEE Transactions on Image Processing* 19 (2010), 2801–2813.

P. Locher. 2006. The usefulness of eye movement recordings to subject an aesthetic episode with visual art to empirical scrutiny. *Psychological Science* 48, 2 (2006), 106–114.

N. H. Mackworth and A. J. Morandi. 1967. The gaze selects informative details within pictures. *Attention, Perception, and Psychophysics* 2 (1967), 547–554.

C. Martindale and K. Moore. 1988. Priming, prototypicality, and preference. *Journal of Experimental Psychology: Human Perception and Performance* 14, 4 (1988), 661–670.

D. Massaro, F. Savazzi, C. Di Dio, D. Freedberg, and V. Gallese. 2012. When art moves the eyes: A behavioral and eye-tracking study. *PLoS ONE* 7, 5 (2012).

A. Matamala and P. Orero. 2007. Accessible opera in Catalan: Opera for all. In *Media for All. Subtitling for the Deaf*, J. Díaz Cintas, P. Orero, and A. Remael (Eds.). Rodopi, Amsterdam, 201–214.

R. E. Mayer. 2002. Multimedia learning. *Psychology of Learning and Motivation* 41 (2002), 85–139.

D. Melcher and E. Kowler. 2001. Visual scene memory and the guidance of saccadic eye movements. *Vision Research* 41 (2001), 3597–3611.

H.-C. Nothdurft. 1999. Focal attention in visual search. *Vision Research* 39 (1999), 2305–2310.

L. Nummenmaa, J. Hyöna, and M. G. Calvo. 2006. Eye movement assessment of selective attentional capture by emotional pictures. *Emotion* 6, 2 (2006), 257–268. DOI:http://dx.doi.org/10.1037/1528-3542.6.2.257

P. Orero. 2007. Sampling audio description in europe. In *Media for All: Subtitling for the Deaf, Audio Description and Sign Language*, J. Díaz Cintas, P. Orero, and A. Remael (Eds.). Rodopi, Amsterdam, Holland, 111–125.

S. Pannasch, J. R. Helmert, K. Roth, A.-K. Herbold, and H. Walter. 2008. Visual fixation durations and saccade amplitudes: Shifting relationship in a variety of conditions. *Journal of Eye Movement Research* 2, 2 (2008), 1–19.

S. Pannasch, J. Schulz, and B. M. Velichkovsky. 2011. On the control of visual fixation durations in free viewing of complex images. *Attention, Perception, and Psychophysics* 73, 4 (2011), 1120–1132.

J. W. Peirce. 2009. Generating stimuli for neuroscience using psychopy. *Frontiers in Neuroinformatics* 2, 10 (2009).

E. Peli, E. M. Fine, and A. T. Labianca. 1996. Evaluating information provided by audio description. *Journal of Visual Impairment and Blindness* 90 (1996), 378–385.

M. I. Posner. 1980. Orienting the attention. *Quaterly Journal of Experimental Psychology* 32 (1980), 3–25.

R Development Core Team. 2011. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

V. S. Ramachandran and W. Hirstein. 1999. The science of art: A neurological theory of aesthetic experience. *Journal of Consciousness Studies* 6, 6 (1999), 15–51.

K. Rayner. 1984. Visual selection in reading, picture perception and visual search: A tutorial review. In *Attention and Performance*, H. Bouma and D. Bouwhuis (Eds.). Erlbaum, Hillsdale, NJ.

K. Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 85 (1998), 618–660.

K. Rayner. 2009. Eye movements in reading: Models and data. *Journal of Eye Movement Research* 2, 5 (2009), 1–10.

C. A. Rothkopf, D. H. Ballard, and M. M. Hayhoe. 2007. Task and context determine where you look. *Journal of Vision* 7 (2007), 1610–1620.

M. Sadoski and A. Paivio. 2004. A dual coding theoretical model of reading. In *Theoretical Models and Processes of Reading*, R. B. Ruddell and N. J. Unrau (Eds.). Vol. 5. International Reading Association, 1329–1362.

A. Santella and D. DeCarlo. 2004. Robust clustering of eye movement recordings for quantification of visual interest. In *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications (ETRA'04)*. Vol. 23. 27–34.

E. Schmeidler and C. Kirchner. 2001. Adding audio description: Does it make a difference? *Journal of Visual Impairment and Blindness* 95, 4 (2001), 198–212.

R. Sekuler, A. B. Sekuler, and R. Lau. 1997. Sound alters visual motion perception. *Nature* 385 (1997), 308.

D. Toker, C. Conati, B. Steichen, and G. Carenini. 2013. Individual user characteristics and information visualization: Connecting the dots through eye tracking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'13)*. ACM, New York, NY, 295–304. DOI:http://dx.doi.org/10.1145/2470654.2470696

P. J. A. Unema, S. Pannasch, M. Joos, and B. Velichkovsky. 2005. Time course of information processing during scene perception. *Visual Cognition* 12, 3 (2005), 473–494.

B. M. Velichkovsky, M. Joos, J. R. Helmert, and S. Pannasch. 2005. Two visual systems and their eye movements: Evidence from static and dynamic scene perception. In *Proceedings of the XXVII Conference of the Cognitive Science Society (CogSci'05)*. 2283–2288.

A. Vilaró, P. Duchowski, A. T. Orero, T. J. Grindinger, S. Tetreault, and E. Di Giovanni. 2012. How sound is the pear tree story? Testing the effect of varying audio stimuli on visual attention distribution. *Perspectives: Studies in Translatology* 20, 1 (2012), 55–65. Special Issue on The Pear Stories.

A. L. Yarbus. 1967. *Eye Movements and Vision*. Plenum Press, New York.

W. H. Zangemeister, U. Oechsner, and C. Freksa. 1995. Short-term adaptive of eye movements in patients with visual hemifield defects indicates high level control of human scan path. *Optometry and Vision Science* 72 (1995), 467–478.

W. Zoeat and M. Donk. 2004. Bottom-up and top-down control in visual search. *Perception* 33 (2004), 927–937.

Received August 2014; revised October 2015; accepted November 2015