AGE-DEPENDENT SACCADIC MODELS FOR PREDICTING EYE MOVEMENTS

Olivier Le Meur[(a)], Antoine Coutrot[(b)], Adrien Le Roch[(a)], Andrea Helo[(c)(e)], Pia Rämä[(c)], Zhi Liu[(d)]*

[(a)] Univ. Rennes 1 (FR); [(b)] Univ. College London (UK);
[(c)] Laboratoire Psychologie de la Perception, Univ. of Paris Descartes, CNRS (UMR 8242) (FR);
[(d)] School of Communication and Information Engineering, Shanghai Univ. (CN);
[(e)] Departamento de Fonoaudiologa, Universidad de Chile, Santiago (CH)

ABSTRACT

How people look at visual information reveals fundamental information about themselves, their interests and their state of mind. While previous visual attention models output static 2-dimensional saliency maps, saccadic models predict not only what observers look at but also how they move their eyes to explore the scene. Here we demonstrate that saccadic models are a flexible framework that can be tailored to emulate the gaze patterns from childhood to adulthood. The proposed age-dependent saccadic model not only outputs human-like, i.e. age-specific visual scanpath, but also significantly outperforms other state-of-the-art saliency models.

Index Terms— age-dependent saccadic model, eye movement, visual scanpath, saliency

1. INTRODUCTION

Oculus animi index is an old Latin proverb that could be translated as *the eyes reflect our thoughts* [1]. Eye-movements, revealing how and where observers look within a scene, are mainly composed of fixations and saccades. Fixations aim to bring areas of interest onto the fovea where the visual acuity is maximum. Saccades are ballistic changes in eye position, allowing to jump from one position to another. Visual information extraction essentially takes place during the fixation period. The sequence of fixations and saccades an observer performs to sample the visual environment is called a visual scanpath [2].

Predicting where we look within a scene is of particular relevance for many computer vision applications such as computer graphics [3], quality assessment [4, 5] and compression [6] to name a few. Saccadic models aim to predict where we look within a scene by generating plausible visual scanpaths, i.e. the actual sequence of fixations and saccades an observer would do while viewing stimuli onscreen [7, 8, 9, 10, 11, 12]. By the term plausible, we mean that the predicted scanpaths should be as similar as possible to human scanpaths. They should exhibit similar characteristics, such as the same distributions of saccade amplitudes and saccade orientations. In addition a saccadic model has to predict where the observer looks, i.e. the salient areas.

In this paper, we tailor a saccadic model for different age groups, ranging from childhood to adulthood. In section 2, we present the eye tracking experiment which has involved four age groups, from infancy to adult ages. Section 3 presents an age-dependent saccadic model. In section 4, we elaborate on the model's performance. This consists in evaluating the relevance of scanpaths as well as saliency maps. Finally we draw some conclusions in Section 5.

2. EYE TRACKING EXPERIMENT

In this section, we analyze eye tracking data collected from observers of a wide range of ages and investigate whether the key component of Le Meur et al.'s saccadic model [11, 12], i.e. the joint distribution of saccade orientations and amplitudes, can be tailored to simulate the gaze biases of different age groups.

2.1. Maturation of eye-movements

The visual system at birth is limited but develops rapidly during the first years of life and continues to improve through adolescence [13, 14]. Helo et al. [15] give evidence of age-related differences in viewing patterns during free-viewing natural scene perception. Fixation durations decrease with age while saccades turn out to be shorter when comparing children with adults. Materials and methods of this eye-tracking experiment are briefly summarized below.

Participants. A total of 101 subjects participated in the experiments, including 23 adults and 78 children. These subjects were divided into 5 groups: 2 yo group, 4-6 yo group, 6-8 yo group, 8-10 yo group and adults group. Participants were instructed to explore the images. The 4-10 yo and the adults were instructed to perform a recognition test to determine whether an image segment presented at the center of the screen was part of the previous stimulus (more details on experimental design is available in [15]).

Stimuli. Thirty color pictures taken from children books, as illustrated in Figure 1 (a), are displayed for 10s. Children's

*This work was supported by the National Natural Science Foundation of China under Grant No. 61471230.

Fig. 1. From left to right: original stimulus; fixation maps (red crosses indicate fixation) for 2 yo and adult groups; actual human saliency maps for 2 yo and adults groups.

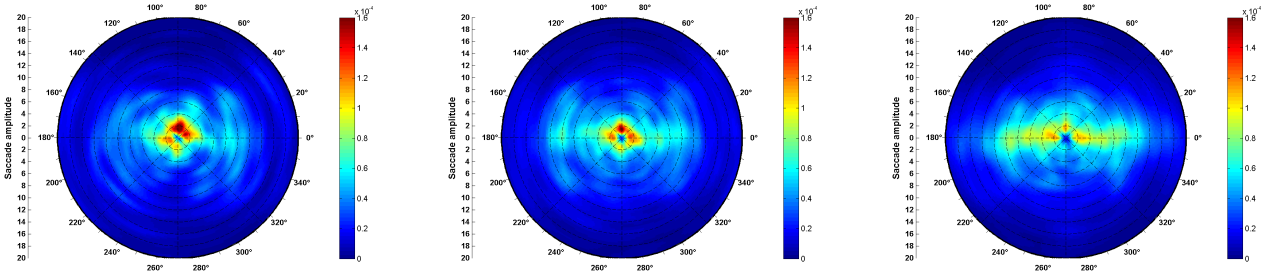

Fig. 2. Polar plots of joint distribution of saccade amplitudes and orientations for different age groups. From left to right: 2, 4-6 yo groups and adult group. The 6-8 and 8-10 yo distributions are not displayed for the sake of clarity.

books are used as stimuli in order to keep the youngest children as focussed as possible. A drift correction is performed before each stimulus. The viewing distance is 60 cm. One degree of visual angle represents 28 pixels. For all the results reported in this paper, the first fixation has been removed.

2.2. Joint distribution of saccade orientations and amplitudes

Following the method proposed in [11], we estimate the joint probability distribution of saccade amplitudes and orientations $p_B(d, \phi)$ for each age group. This nonparametric distribution is obtained by using a 2D Gaussian kernel density estimation. The two bandwidth parameters are chosen optimally based on the linear diffusion method proposed by [16]. The joint probability $p_B(d, \phi)$ is given by:

$$p_B(d, \phi) = \frac{1}{n} \sum_i K_h(d - d_i, \phi - \phi_i) \qquad (1)$$

where d_i and ϕ_i are the distance and the angle between each pair of successive fixations respectively. n is the total number of samples and K_h is the two-dimensional Gaussian kernel. Figure 2 shows the joint probability distributions of saccade amplitudes and orientations in a polar plot representation. Radial position indicates saccadic amplitudes expressed in degree of visual angle. A number of observations can be made: first, eye-movement patterns change with age. Saccade amplitudes are shorter in the 2 yo group than in adults group. Saccade amplitudes increase with age (see also Figure 3 top row). This first observation was consistent with the ones made in [15]. Regarding the saccade orientations, we observe a strong horizontal bias in the adult group which is also consistent with previous studies [17, 18]. This horizontal bias can be explained by several factors, such as biomechanical factors, physiological factors and the layout of our natural environment [19]. Figure 2 also shows that the distribution shape of the 2 yo group (a) is much more isotropic than the adults' one (d), but with a marked tendency for making upward vertical saccades.

A two-sample two-dimensional Kolmogorov-Smirnov test [20] is performed to test whether the difference between the joint distributions illustrated in Figure 2 is statistically significant. For two given distributions, we randomly draw 5000 samples and test whether both data sets are drawn from the same distribution. The tests show significant differences between 2 yo and 4-6 yo groups, and between 4-6 yo and 6-8 yo groups (all $p < .001$). There is no difference between 6-8 yo and 8-10 yo groups ($p = 0.2$). A significant difference is however observed between adults and 8-10 yo groups ($p = 0.0049$). We reduced the within-group variance by increasing the sample size and merging the 6-8 and 8-10 yo groups together. The resulting group is called the 6-10 yo group.

In summary, these results suggest that the joint distribution of saccade amplitudes and orientations is able to grasp gaze behavior differences across age, as well as to reflect important features of development on the visual deployment. By plugging these distributions into the saccadic model of [11, 12], we should be able to replicate the gaze behavior of a spe-

cific age group.

3. AGE-DEPENDENT SACCADIC MODEL

In this section we tailor Le Meur's saccadic model [11, 12] to the different age groups, namely, 2, 4-6, 6-10 yo and adults.

This model is based on a conditional probability $p\left(x|x_{t-1}\right)$ which is used for selecting the next fixation point x knowing the previous point x_{t-1}. This conditional probability is a combination of low-level salience p_{BU} (i.e. representing the saliency map), memory effects p_M and viewing biases $p_B(d, \phi)$ (i.e. joint probability density function of saccade amplitude d and orientation ϕ) (see [11, 12] for details). To define the next fixation point, N_c candidates are drawn from the conditional probability $p\left(x|x_{t-1}\right)$. The candidate having the highest saliency is eventually pick up as the next fixation point.

We perform two main modifications to adapt this model to the purpose of our study. The first modification consists in using a joint probability density function $p_B(d, \phi)$ that has been learned from eye tracking data collected from different age groups, as presented in section 2.2. This prior knowledge represents the viewing tendencies, expressed in this study in terms of saccade amplitudes and orientations, which are common across all observers of a given age. The use of such a prior is fundamental to constrain how we explore scenes and to generate saccade amplitudes and orientations that match those estimated from human eye behavior. As suggested in [12], we use a spatially-variant joint distribution. The image is then split into a non-overlapping 3×3 grid; for each cell in the grid, the joint distribution of saccade amplitudes and saccade orientations is estimated following the procedure detailed in section 2.2. This spatially-variant prior is more appropriate for catching important viewing tendencies. One of the most important priors is the central bias.

The second modification concerns the selection of the most appropriate candidate among the N_c candidates drawn from the conditional probability $p\left(x|x_{t-1}\right)$. In [11], the next fixation point is selected as being the candidate having the highest bottom-up salience. This selection rule is modified to take into account the probability $p_B(.,.)$, the bottom-up saliency $p_{BU}(.)$ and the distance d between the candidate and the previous fixation point. The next fixation point x^* is then selected as

$$x^* = \arg\max_{s \in \Theta} \frac{p_{BU}(s) \times p_B(d(s, x_{t-1}), \phi(s, x_{t-1}))}{d(s, x_{t-1})} \quad (2)$$

where, Θ is composed of the spatial coordinates of the N_c chosen candidates, x_{t-1} is the previous fixation point and d is the Euclidean distance between the candidate s and the previous fixation point. This new rule allows us to favor the candidates that are close to the previous fixation point and featured by both a high probability to be attended and high bottom-up salience.

4. PERFORMANCES

First, we evaluate the extent to which the predicted fixations fall within salient areas. Second, we test the plausibility of the generated scanpaths with respect to the actual scanpaths of the four age groups. To perform this evaluation, we proceed as follows: for each image of the dataset and for each age group, we generate 20 scanpaths, each composed of 15 fixations. The first fixation is randomly chosen. The input saliency map p_{BU} is computed by using either GBVS [21] or RARE2012 [22] model. From the generated scanpaths, a saliency map, called scanpath-based saliency map, is computed by following the classical procedure (see [23, 24]).

Table 1. Performance of GBVS, RARE2012 and saccadic model (using an input saliency map from GBVS or RARE2012). Best scores are in bold. ↑ means the higher the better, ↓ means the lower the better.

Metrics	CC ↑	SIM ↑	EMD ↓	AUC-Judd ↑	AUC-Borji ↑	NSS ↑
Adults						
GBVS	0.531	**0.731**	0.868	**0.644**	**0.634**	0.463
Our model	**0.636**	0.706	**0.759**	**0.644**	**0.639**	0.561
RARE2012	0.290	0.638	1.228	0.592	0.575	0.256
Our model	**0.566**	**0.701**	**0.787**	**0.640**	**0.630**	**0.492**
6-10 y.o.						
GBVS	0.589	**0.761**	0.776	**0.661**	**0.640**	0.478
Our model	**0.686**	0.732	**0.744**	0.659	**0.640**	**0.562**
RARE2012	0.328	0.661	1.183	0.607	0.578	0.266
Our model	**0.617**	**0.717**	**0.792**	**0.648**	**0.630**	**0.505**
4-6 y.o.						
GBVS	0.544	**0.691**	1.052	**0.690**	0.675	0.597
Our model	**0.673**	0.690	**0.806**	0.688	**0.681**	**0.745**
RARE2012	0.275	0.592	1.464	0.614	0.587	0.296
Our model	**0.602**	**0.683**	**0.900**	**0.678**	**0.672**	**0.661**
2 y.o.						
GBVS	0.501	**0.662**	1.071	**0.674**	0.667	0.570
Our model	**0.579**	0.659	**0.906**	**0.674**	**0.671**	**0.666**
RARE2012	0.264	0.578	1.431	0.601	0.579	0.292
Our model	**0.517**	**0.639**	**1.014**	**0.662**	**0.653**	**0.585**

4.1. Prediction of salient areas

Table 1 presents the similarity degree between scanpath-based saliency map and the ground truth (i.e. either human saliency map or eye tracking data). These performances are obtained for an optimal value of N_c. Six metrics, i.e. CC, SIM, EMD, AUC-Judd, AUC-Borji, and NSS, are used; they are described in [24, 25]. Our model significantly outperforms RARE2012 model, whatever age groups and metrics. Compared to GBVS model, it performs better according to 4 metrics. The results were analyzed using a three-way mixed ANOVA design when the RARE2012 model is used for computing the input saliency map p_{BU}. The proposed analysis uses age groups (adults, 6-10 yo, 4-6 yo, or 2 yo) as the between-subjects variable, type of saliency model (RARE2012 or RARE2012-based saccadic model) and type of metric (CC, SIM, EMD, AUC-Judd, AUC-Borji, or NSS) as the within-subjects variables. The three-way ANOVA yielded a significant main effect of age ($F(3, 95) = 6.81$, $p < .001$), model ($F(1, 95) =$

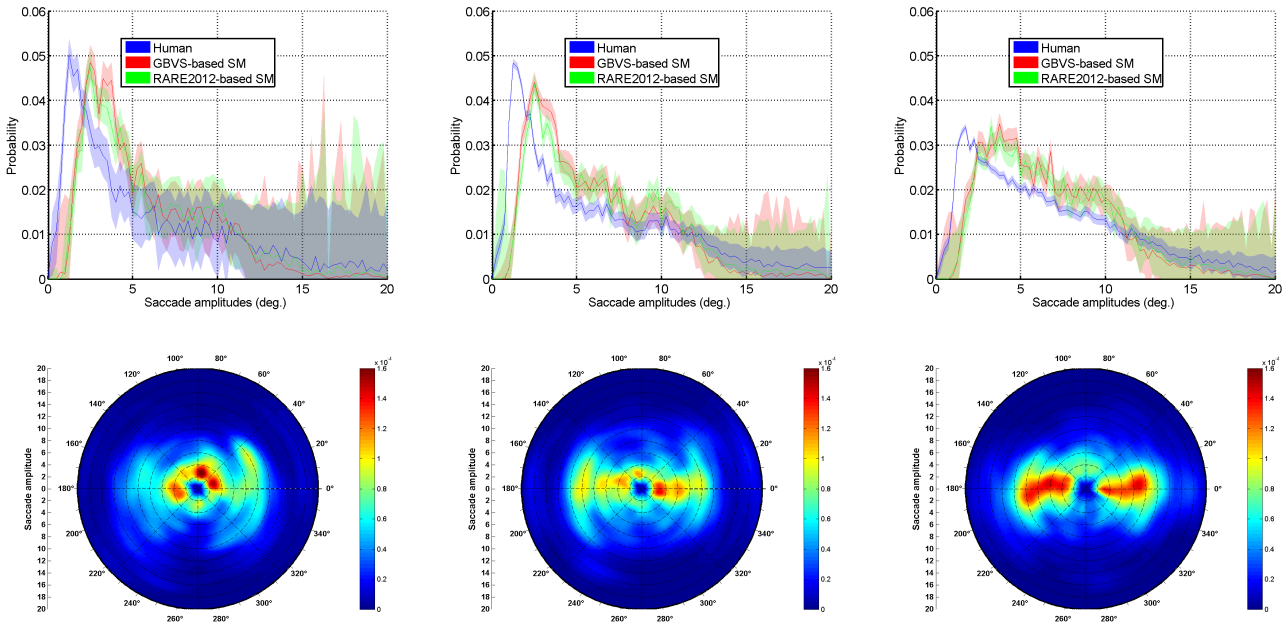

Fig. 3. Features of the predicted scanpaths. Top row: the actual and the predicted distributions of saccade amplitudes are plotted for $N_c = 5$, $N_c = 4$ and $N_c = 4$ corresponding to 2 yo (left), 4-6 yo (middle) and adults groups (right), respectively. Bottom row: joint distributions of saccade amplitudes and saccade orientations computed from RARE2012-based saccadic model for 2 yo (left), 4-6 yo (middle) and adults groups (right), respectively.

$67.92, p < 0.001$) and metric ($F(5, 91) = 262.45, p < .001$). The $metric \times age$ interaction is significant ($F(15, 279) = 7.43, p < .001$), as well as the $model \times metric$ interaction ($F(5, 91) = 96.143, p < .001$). The $model \times age$ interaction is not significant ($F(3, 95) = 0.62, p = 0.60$). Post-hoc Bonferroni comparisons show significant differences between adults and 4-6 yo ($p < .001$), marginal differences between adults and 2 yo ($p = 0.086$) and no difference between adults and 6-10 yo ($p = 1$). There is a significant difference between 6-10 yo and 4-6 yo ($p = 0.01$) but not between 6-10 yo and 2 yo ($p = 0.57$). There is no significant difference between 2 yo and 4-6 yo ($p = 1$). Note that we observe similar results when GBVS model is used to compute the input saliency map.

4.2. Are visual scanpaths plausible?

Saccadic models predict both salient areas and plausible scanpaths. From the predicted scanpaths, we compute, for each age group and for both saliency models (i.e. GBVS and RARE2012), the 1D distribution of saccade amplitudes and the 2D joint distribution of saccade amplitudes and saccade orientations. Figure 3 shows these distributions for three age groups. We observe that the distributions of saccade amplitudes computed with the proposed saccadic model have a shape similar to the distribution of human saccade amplitudes. We note, however, that the proposed model tends to generate larger saccades. The main peak of the predicted distributions is between 2 and 3 degrees of visual angle, whereas the main peak of actual distributions is about 2 degrees of visual angle. This discrepancy might be due to the computational modelling of the inhibition-of-return mechanism which does not entirely reflect the reality. A strong similarity is also observed between the actual (see Figure 2) and predicted joint distributions (bottom row of Figure 3).

5. CONCLUSION

In this paper, we show that saccadic models can be tailored for different age groups. This model combines low-level salience, memory effects and viewing biases. The latter provides fundamental information about how observers explore a visual scene. We show that these viewing biases evolve with the maturation of the visual system. These differences can be captured with joint distributions of saccade amplitudes and orientations. By using this age-based visual signature, we propose the first age-dependent saliency model. This model outperforms not only GBVS and RARE2012 saliency models but succeeds in generating scanpaths that match actual eye tracking data. Thanks to this new age-dependent model, we can adapt saliency-based applications in order to match a target age. Obviously, this approach cannot fully account for the complex nature of overt visual attention and it would be required to incorporate other known properties of gaze behavior.

6. REFERENCES

[1] Roger PG Van Gompel, *Eye movements: A window on mind and brain*, Elsevier, 2007.

[2] David Noton and Lawrence Stark, "Scanpaths in eye movements during pattern perception," *Science*, vol. 171, no. 3968, pp. 308–311, 1971.

[3] P. Longhurst, K. Debattista, and A. Chalmers, "A gpu based saliency map for high-fidelity selective rendering," in *Proceedings of the 4th international conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*. ACM, 2006, pp. 21–29.

[4] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba, "Does where you gaze on an image affect your perception of quality? applying visual attention to image quality metric," in *2007 IEEE International Conference on Image Processing*. IEEE, 2007, vol. 2, pp. II–169.

[5] H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: based on eye-tracking data," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 7, pp. 971–982, 2011.

[6] Z. Li, S. Qin, and L. Itti, "Visual attention guided bit allocation in video compression," *Image and Vision Computing*, vol. 29, no. 1, pp. 1–14, 2011.

[7] D. Brockmann and T. Geisel, "The ecology of gaze shifts," *Neurocomputing*, vol. 32, no. 1, pp. 643–650, 2000.

[8] G. Boccignone and M. Ferraro, "Modelling gaze shift as a constrained random walk," *Physica A: Statistical Mechanics and its Applications*, vol. 331, no. 1, pp. 207–218, 2004.

[9] H. R. Tavakoli, E. Rahtu, and J. Heikkilä, "Stochastic bottom–up fixation prediction and saccade generation," *Image and Vision Computing*, vol. 31, no. 9, pp. 686–693, 2013.

[10] R. Engbert, H. A. Trukenbrod, S. Barthelmé, and F. A. Wichmann, "Spatial statistics and attentional dynamics in scene viewing," *Journal of vision*, vol. 15, no. 1, pp. 14–14, 2015.

[11] O. Le Meur and Z. Liu, "Saccadic model of eye movements for free-viewing condition," *Vision research*, vol. 116, pp. 152–164, 2015.

[12] O. Le Meur and A. Coutrot, "Introducing context-dependent and spatially-variant viewing biases in saccadic models," *Vision Research*, vol. 121, pp. 72–84, 2016.

[13] B. Luna, K. Velanova, and C. F. Geier, "Development of eye-movement control," *Brain and cognition*, vol. 68, no. 3, pp. 293–308, 2008.

[14] Eva Aring, Marita Andersson Grönlund, Ann Hellström, and Jan Ygge, "Visual fixation development in children," *Graefe's Archive for Clinical and Experimental Ophthalmology*, vol. 245, no. 11, pp. 1659–1665, 2007.

[15] A. Helo, S. Pannasch, L. Sirri, and P. Rämä, "The maturation of eye movement behavior: Scene viewing characteristics in children and adults," *Vision research*, vol. 103, pp. 83–91, 2014.

[16] Z.I. Botev, J.F. Grotowski, and D. P. Kroese, "Kernel density estimation via diffusion," *The annals of Statistics*, vol. 38, no. 8, pp. 2916–2957, 2010.

[17] T. Foulsham, A. Kingstone, and G. Underwood, "Turning the world around: Patterns in saccade direction vary with picture orientation," *Vision Research*, vol. 48, pp. 1777–1790, 2008.

[18] B.W. Tatler and B.T. Vincent, "Systematic tendencies in scene viewing," *Journal of Eye Movement Research*, vol. 2, pp. 1–18, 2008.

[19] D.R. Van Renswoude, S.P. Johnson, M.E.J. Raijmakers, and I. Visser, "Do infants have the horizontal bias?," *Infant Behavior and Development*, vol. 44, pp. 38–48, 2016.

[20] J.A. Peacock, "Two-dimensional goodness-of-fit testing in astronomy," *Monthly Notices of the Royal Astronomical Society*, vol. 202, no. 3, pp. 615–627, 1983.

[21] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proceedings of Neural Information Processing Systems (NIPS)*. 2006, MIT Press.

[22] N. Riche, M. Mancas, M. Duvinage, M. Mibulumukini, B. Gosselin, and T. Dutoit, "Rare2012: A multi-scale rarity-based saliency detection with its comparative statistical analysis," *Signal Processing: Image Communication*, vol. 28, no. 6, pp. 642 – 658, 2013.

[23] D.S. Wooding, "Fixation maps: quantifying eye-movement traces," in *Proceedings of the 2002 symposium on Eye Tracking Research & Applications*. ACM, 2002, pp. 31–36.

[24] O. Le Meur and T. Baccino, "Methods for comparing scanpaths and saliency maps: strengths and weaknesses," *Behavior Research Method*, vol. 45, no. 1, pp. 251–266, 2013.

[25] Nicolas Riche, Matthieu Duvinage, Matei Mancas, Bernard Gosselin, and Thierry Dutoit, "Saliency and human fixations: State-of-the-art and study of comparison metrics," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 1153–1160.