Are You Really Looking? Finding the Answer through Fixation Patterns and EEG

Anne-Marie Brouwer[1], Maarten A. Hogervorst[1], Pawel Herman[2], and Frank Kooi[1]

[1] TNO Human Factors, Kampweg 5, 3769 ZG Soesterberg, The Netherlands
[2] Donders Institute for Brain, Cognition and Behaviour, The Netherlands
`{anne-marie.brouwer,maarten.hogervorst,frank.kooi}@tno.nl,`
`pawel.herman@donders.ru.nl`

Abstract. Eye movement recordings do not tell us whether observers are 'really looking' or whether they are paying attention to something else than the visual environment. We want to determine whether an observer's main current occupation is visual or not by investigating fixation patterns and EEG. Subjects were presented with auditory and visual stimuli. In some conditions, they focused on the auditory information whereas in others they searched or judged the visual stimuli. Observers made more fixations that are less cluttered in the visual compared to the auditory tasks, and they were less variable in their average fixation location. Fixated features revealed which target the observers were looking for. Gaze was not attracted more by salient features when performing the auditory task. 8-12 Hz EEG oscillations recorded over the parieto-occipital regions were stronger during the auditory task than during visual search. Our results are directly relevant for monitoring surveillance workers.

1 Introduction

Imagine a man sitting behind a desk, looking at several monitors that display the images produced by surveillance cameras. His eyes are open and the radio is turned on. How do we know whether he is actually paying attention to what happens on the monitors rather than listening attentively to the radio or thinking about the groceries? We want to see whether we can distinguish between these kinds of situations - in particular, whether an observer's main current occupation is visual or not. We propose that fixation patterns and EEG could contribute to this distinction.

With respect to fixation patterns, we will test three different hypotheses.

Firstly, we expect that fixations will be guided more by bottom-up, low level visual features when observers are not occupied with a visual task compared to when they are. The idea is that if there is no explicit top-down, high level visual goal, low-level visual features will play a more important role. In order to test this hypothesis, we will compare fixation locations to saliency maps [1, 2]. Saliency maps indicate how saliency varies across the visual environment, based on low-level visual features such as, in our case, orientation, intensity and color. Rather than absolute values, center-surround values are used such that it is the contrast within a feature space that counts.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 329–338, 2009.

Secondly, we predict that in a visual search task fixations will be spread more than in an auditory task since in the first case, subjects have to scan an area whereas in the second case, they can choose to view one particular interesting area. We will both look at the standard deviation of fixation location and at measures defined by using Voronoi diagrams, following [3]. A Voronoi diagram is a division of an area into cells, with the borders of each cell surrounding a point of interest, in our case, a fixation. Every point within the cell is closer to the cell's fixation than to any other fixation. Small cell sizes go together with a densely fixated region whereas large cell sizes indicate occasional fixations. Fixation clutter produces a large difference in cell sizes, while evenly distributed fixations produce more similar cell sizes. Thus, fixation clutter can be quantified by the standard deviation of cell size or the skewness of the cell size distribution. [3] determined skewness in a free viewing condition, visual search in structured images and visual search in homogeneous images. Fixation clutter as determined by skewness decreased across these three conditions.

Thirdly, we expect that image patches around fixations will reveal features of a searched target (only when observers are performing a visual search task).

With respect to EEG correlates, we expect low alpha power over the parieto-occipital (visual) cortex when a visual search task is performed. Alpha refers to 8 to 12 Hz oscillatory EEG activity. It has been shown to be negatively correlated with visual (and not auditory) attention [4, 5] which is in agreement with the alpha inhibition hypothesis [6].

In order to determine whether a person is engaged in a visual task or not, the proposed fixation- and EEG cues could be combined. The particular (non)visual tasks included in this study are a visual search task, a visual judgment task and an auditory attention or short term memory task.

2 Methods

2.1 Equipment and Stimuli

The setup involved four computers. One was used to control the eye recordings, one to record EEG data, and one to present stimulus images. These computer units were synchronized. The fourth computer was used to present auditory stimuli.

Subjects were sitting in a shielded room (Faraday cage) with their head in a chinrest and their right index and middle fingers on computer mouse buttons. Their eyes were approximately 65 cm away and at the height of the center of a 20 inch LCD monitor. Auditory stimuli were presented using speakers in the room. A Tobii eyetracker sampling eye position at 50 Hz, was positioned under the monitor.

EEG activity was recorded at the Fz, Cz, Pz, Oz, C3, C4, TP7, TP8, P3, P4, O1 and O2 electrode sites. A ground electrode was attached to the forehead and the reference was positioned on the left mastoid. The impedances between relevant pairs of electrodes were below 5 kΩ. The sampling rate was 256 Hz and the signal was bandpass filtered in the range of 0.1 - 100 Hz. Also, a 50 Hz notch filter was applied.

We used both natural and artificial images as visual stimuli (see for examples Figure 1). The natural images were pictures of landscapes, sometimes containing a (rather inconspicuous) military vehicle. These images were chosen from the Search_2 database [7] and from a larger collection of images of which Search_2 is a subset. The artificial images were generated by custom software, and consisted of a collection of colored simple geometric shapes. 20% of the images contained a green diamond and a blue pentagon.

The landscape images subtended the complete width of the monitor but not the complete height. The shapes images subtended the whole screen. Each image was presented for 5 s. In between images, a black screen was presented for 1 s.

The auditory stimuli consisted of the spoken letters a, b, c, x, y and z. Every 2.0 ± 0.2 s a letter randomly chosen from this sequence was presented for the entire duration of the experimental block.

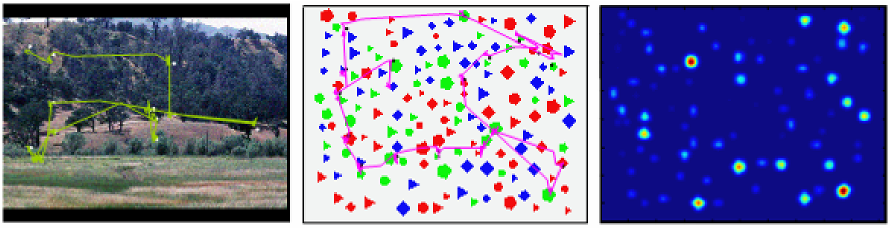

Fig. 1. On the left are examples of a landscape and a shapes image with accompanying gaze tracks and fixations as recorded during a trial of visual search. On the right is the saliency map from the shapes image example as computed by the Saliency Toolbox 2.1 [2].

2.2 Tasks

In the visual search task, subjects were asked to look for a target: a military vehicle in the landscape images or, either a green diamond (5 subjects) or a blue pentagon (6 subjects) in the shapes images. Subjects were requested to press the right mouse button when they had found the target and the left button if they could not find it. They were told to ignore the auditory presented letters.

In the visual judgment task, subjects were asked to judge the spread of the trees in the landscape images, and the spread of the red symbols in the shapes images. If subjects assessed these to be grouped, they pressed the left button, if they thought they were evenly spread, they pressed the right button. Subjects were told to ignore the auditory presented letters and advised to fixate their gaze at the center of the screen.

The auditory task was an auditory version of the Continuous Memory Task [8]. Subjects had to keep track of the letters 'a' and 'z'. Each time that one of these letters was presented for the first time, they were asked to right-click. When they were presented for the second time, subjects were supposed to left-click. After the second presentation subjects received feedback (a spoken 'correct' or 'false'). After the feedback, the counting started anew for the given letter. We refer to this task as an auditory task, but one could also consider it to be a short term memory or an auditory attention task. Although subjects were asked to watch the presented pictures, we stressed that keeping track of the letters was their main task.

2.3 Design and Procedure

10 naïve subjects plus the first author voluntarily participated in the study. Subjects were between 21 and 32 years old. Prior to the experiment they received written and spoken instructions.

The task order was random. For every task, a block of landscape images and a block of shapes images was presented (or vice versa). Each experimental block consisted of 40 images, shown in random order. Each block was preceded by a practice block of 5 images. Different images were used for each block. Before the start of each experimental block, a 9-point eye calibration session (ClearView 2.7) was performed to minimize fixation errors.

The first six subjects only performed the visual search and the auditory task; the last five subjects additionally performed the visual judgment task. Thus, the first group performed a total of 2 (tasks) * 2 (stimulus type) * 40 (number of repetitions) = 160 experimental trials, and the second group 3 (tasks) * 2 (stimulus type) * 40 (number of repetitions) = 240 experimental trials.

2.4 Analysis

We used the ClearView 2.7 default settings to define fixations. Examples of fixations plotted on the accompanying gaze track are presented in Figure 1. For each trial, we determined the number of fixations and the average fixation location. We also calculated the standard deviation of average fixation location between subjects. For trials with more than one fixation, the standard deviations in the horizontal and vertical direction were computed. In addition, we created Voronoi diagrams with every cell containing a fixation. The boundaries of the diagram were defined as the boundaries of the image. For each Voronoi diagram we normalized the cell sizes such that the average cell size was 1. Then we computed the standard deviation. A large standard deviation indicates more fixation clutter. Further, we computed the skewness of the distribution of cell sizes [3]. Higher values indicate more clutter.

For every stimulus image, we computed intensity, orientation, color and overall saliency maps according to [1] using the SaliencyToolbox 2.1 [2]. Figure 1 shows an example saliency map. For each fixation, we multiplied the saliency maps with a Gaussian window (standard deviation of 3 pixels) centered on the fixation. The average pixel value was determined. The results for all fixations within a trial were averaged. This value will be high when subjects fixate salient locations. To be able to compare between different images, the fixation saliency measure was defined as the ratio between this value and the average pixel value of the saliency map multiplied by the inverse of the Gaussian fixation windows (per trial, not per fixation).

For the shapes images, we wanted to identify the looked-for target from the fixations. Voronoi diagrams were calculated such that the center of each shape was the point around which cells were created. Next, we determined for each fixation location the color and the type of shape that was in the same cell as the fixation.

Repeated measures ANOVAs and paired t-tests were used to evaluate the results. For most variables, two ANOVAs were performed: one on the data of all 11 subjects with factors task ('visual search' and 'auditory') and stimulus type ('landscape' and 'shapes'), and one on the data of the last 5 subjects where the factor task had the levels 'visual judgment' and 'auditory'. Note that for the auditory task, the data in the second type of ANOVA (the grey symbols in the graphs) are a subset of the first (the black symbols). The significance level was set to 0.05. All significant effects are mentioned.

We visually inspected EEG signals in order to reject trials with excessive noise and artefacts. Then we performed power spectral analysis using the short-time Fourier transform (STFT) with Hanning window. We compared the average 8-12 Hz alpha power over the parieto-occipital region (Pz, P3, P4, Oz, O1 and O2) between the visual search and the auditory task, separately for shapes and landscapes images. The hypothesis that alpha power was larger in the auditory than in the visual task was verified using a one-sided Monte-Carlo testing procedure with dependent t-statistics. A Bonferroni correction was applied to correct for multiple comparisons (i.e. the significance level for each pair was set to 0.025). The entire EEG analysis was carried out using the Matlab package Fieldtrip (http://www.ru.nl/fcdonders/fieldtrip).

3 Results

3.1 Fixation Number

Figure 2 shows the average number of fixations per trial for each condition. In the visual task there are more fixations than in the auditory task, both when the task is visual search (effect of task p<0.01) and visual judgment (p=0.02). For visual search and the auditory task, there is an interaction between task and stimulus (p<0.01), such that the number of fixations is higher for the shape than the landscape stimuli in the search task, whereas there is no difference between the two in the auditory task.

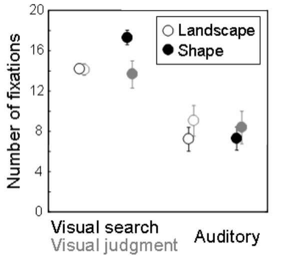

Fig. 2. Number of fixations. In all figures, error bars indicate standard errors of the mean. Black symbols represent data of all eleven subjects, grey symbols represent the subset of five subjects who also performed the visual judgment task.

3.2 Fixation Location

On average, subjects fixate around the center of the screen for all tasks and all stimuli. However, the average fixation locations are more variable between subjects in the auditory task compared to the visual tasks. The standard deviation of the average horizontal fixation location is 23 pixels (≈0.80 deg) for visual search, 28 pixels for visual judgment and 174 for the auditory task (all collapsed over stimulus type). For the vertical fixation location, these numbers are 22, 17 and 133, respectively. This suggests that in the auditory task, each individual subject had his/her own preferred fixation location in space.

3.3 Fixation Clutter

To investigate the spread of the fixation locations within trials, we computed the standard deviation of the horizontal and vertical fixation location for each trial that contained more than one fixation. Fixations were clearly more spread in the visual search condition as compared to the auditory condition (Figure 3, effect of task as indicated by p-values <0.01 for both the horizontal (3A, black symbols) and vertical direction (3B, black symbols)). There were also interactions with stimulus type (p=0.02 for the horizontal and p<0.01 for the vertical direction), indicating that the fixations were especially spread out when subjects were searching for vehicles compared to shapes, whereas the effect of stimulus type was the other way around during the auditory task. This indicates that the subjects scan more of the scene when looking for the target among shapes than in landscapes. For the subjects who performed the visual judgment task and the auditory task (Figure 3, grey values), there is only an effect of task (p=0.03 for the horizontal direction and p=0.02 for vertical), again indicating that the fixations are more spread in the visual condition compared to the auditory condition.

We also quantified clutter of fixations using Voronoi diagrams. Figure 4 displays the results of the two Voronoi measures, the standard deviation of the normalized cell size (A) and the skewness of the cell size distribution (B). The pattern of results is very similar for both measures: comparing visual search to the auditory task (black symbols) results in a main effect of task (p<0.01 for both standard deviation and skewness) and an interaction between task and stimulus type (p=0.04 for standard deviation and p<0.01 for skewness). The main effect of task indicates more fixation clutter in the auditory task than in visual search. The interaction suggests a trend toward more fixation clutter for the landscapes than for the shapes in the auditory condition, whereas it tends to be the other way around in the visual search condition. The ANOVAs performed on the data of the subjects who performed the visual judgment task did not result in any significant effect, neither for the standard deviation of the normalized cell size nor for the skewness.

Note that stimulus type can have an effect on fixation clutter in the auditory task. Paired t-tests comparing the two stimulus types in the auditory condition result in significant effects for the standard deviation of fixations in both directions, and for the standard deviation of the normalized Voronoi cell size (p-values <0.05). For skewness significance is approached (p=0.06).

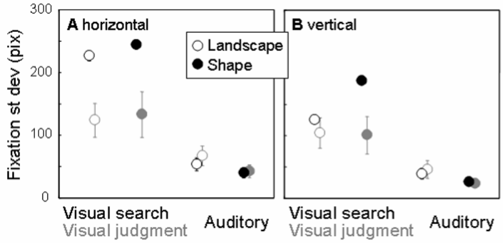

Fig. 3. Standard deviation of horizontal (A) and vertical (B) fixation locations

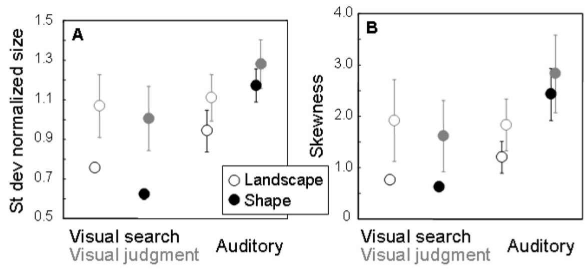

Fig. 4. Vonoroi measures of clutter: standard deviation of normalized cell size (A) and the skewness of cell size distribution (B)

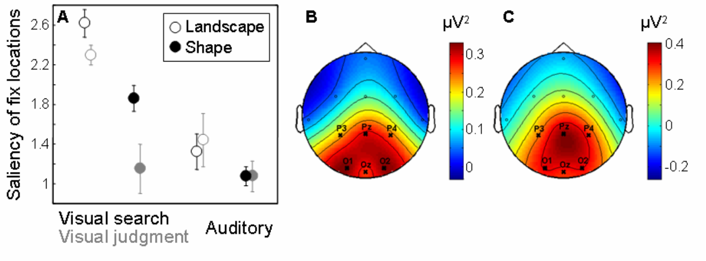

Fig. 5. Saliency of fixated locations (A) and distribution of the difference in alpha EEG power between the auditory and visual search tasks (B: landscapes, C: shapes)

3.4 Fixation Saliency

Figure 5 shows to what extent subjects fixate salient regions in the different conditions. When comparing the visual search task to the auditory task, the overall fixation saliency value is affected by task ($p<0.01$) and stimulus ($p<0.01$). Contrary to what we predicted, the task effect indicates that subjects fixate salient features more in the visual search task than in the auditory task. This is especially the case for the landscape stimuli. When comparing the visual search task to the auditory task, there is no significant effect of task, but an effect of stimulus type ($p=0.02$) and an interaction ($p=0.02$). When judging landscape rather than shape stimuli, subjects fixate highly salient features.

3.5 Fixation Target

For the shape stimuli, we had a closer look at the specific shapes that were fixated in the different conditions. Figure 6A shows the proportion of fixations close to blue, green and red shapes in the different tasks. In 6B, the data is split up by the form of the shape that subjects fixated. The data of the auditory and visual judgment task overlap in both graphs. This is in accordance with the saliency data where we did not observe any difference between the two tasks when the shapes were presented. One might have expected to find relatively many fixations close to red shapes in the visual judgment task (where the spread of red symbols had to be judged) but this is not

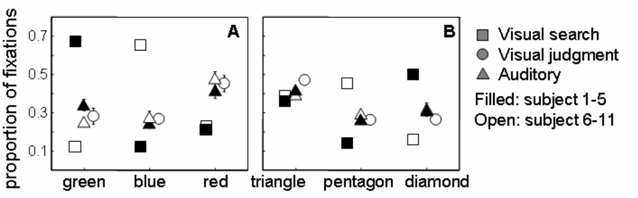

Fig. 6. Proportions of fixations close to differently colored (A) and formed (B) shapes. Note that all data points are together; except for the squares representing the search condition. The open squares indicate subject 1-5 who searched for the green diamond, the filled squares subject 6-11 who looked for the blue pentagon.

apparent. In general, red shapes (6A) and triangles (6B) seem to draw gaze. Our main objective was to determine the looked-for target from fixated features. Indeed, almost 70% of the fixations is on shapes of the same color as the target. Also, the form corresponding to that of the target is fixated most. Note that the target itself is hardly ever fixated because most trials did not contain a target, and if they did, those trials usually contained many more fixations on the other shapes.

3.6 EEG

As hypothesized, the average power of the parieto-occipital EEG alpha activity was higher during the auditory task than during visual search (p=0.015 for the landscape images with an average increase of alpha power of 52%, and p=0.022 for the shapes with an increase of 55%). Figure 5BC shows the distribution of the alpha power difference over the skull.

4 Discussion

Our experiment yielded several important findings.

Subjects make more fixations in both of the visual tasks compared to the auditory task. This is the case even though we asked subjects to view the pictures during the auditory task, and we tried to minimize the number of eye movements in the visual judgment task by instructing subjects to look at the center of the image. Note that the number of eye movements in visually judging the landscape images is the same as in searching them.

In the auditory task, subjects seem to have their own preferred fixation location in space. This suggests that fixation locations during non-visual tasks are not completely determined by low-level visual features. Although all subjects used a chin rest and an adjustable chair such that their eyes were approximately directed to the center of the screen, the preferred fixation location may have been influenced by the exact sitting and 'straight ahead' posture of the subject.

Fixations are more widely spread in both visual tasks than in the auditory task. The standard deviations of fixation locations indicate that when searching the shape stimuli, fixations were more spread than when searching the landscapes, while in the auditory task, fixations were less spread when viewing shape stimuli than landscapes.

Perhaps, the difficult shape stimuli forced subjects to scan a larger area when searching for the target, whereas, when there was no visual task to be done, subjects preferred scanning the landscape stimuli over the shape stimuli.

Computationally, the standard deviation does not depend on the number of data points. However, in fixation studies, the number of fixations could correlate positively with the standard deviation of fixation location: if subjects want to scan a larger area they will probably make more fixations. Still, this relation is not very strong as seen when Figures 2 and 3 are compared. This means that besides the number of fixations, fixation spread as indicated by the standard deviation may still add information when determining which task is being performed.

Fixation spread or clutter as measured using Voronoi diagrams provided similar results as when it was measured by the standard deviation. Fixations are less cluttered in the visual search task than in the auditory task, with an interaction effect such that landscapes cause more clutter in visual search than shape stimuli (consistent with [3]). The opposite effect of image type is observed in the auditory task. In contrast to the standard deviation, the Voronoi measures did not indicate any significant effect when comparing visual judgment to the auditory task. This is probably due to the larger variability in the Voronoi measures - the trend of the results is similar.

Interestingly, we found that the type of visual stimulus affects fixation spread in the auditory task. Thus, while the subject specific preferred fixation location in space during the auditory task indicates that fixation locations during non-visual tasks are not only determined by low-level visual features, visual stimuli do have an effect. This appears to be a higher level effect, such as an increased tendency to look around when natural, more meaningful stimuli are presented.

Our prediction that gaze is especially attracted by salient stimuli when subjects do not have a specific visual task did not bear out. On the contrary, especially for the landscape stimuli, subjects look more at salient features during visual tasks than during the auditory task. In the visual search task, this could be explained by the fact that some aspects of the target are salient. The findings suggest once again that during non-visual tasks low-level visual features do not determine fixation. We do not expect that these results can be easily reversed by using another definition of saliency since the significant effects of our variables indicate that the saliency maps as defined here are indeed meaningful.

We can deduce from fixations which target subjects are looking for, even if they hardly ever fixate the actual target. We showed this now for simple geometrical shapes, but if the eye tracking is accurate enough, it could work for more complex images as well. This could be done by applying different filters to the image patch around fixation and comparing the averaged results to randomly chosen patches.

The power of EEG alpha activity monitored over the parieto-occipital area is higher for the auditory task than visual search. This finding is completely in accordance with the alpha inhibition hypothesis [4, 5]. The emerging concept goes beyond the common notion of alpha as an ‘idling’ rhythm [9].

To sum up, our results suggest the following markers for a visual task: many fixations, a large fixation spread, and an average fixation location close to the center of the region of interest. Depending on the specific relevant items in the visual field, fixated features tend to be salient and can reveal which target the observer is looking for. The power of EEG alpha is low. Markers identified for a non-visual task are the following: few and cluttered fixations, average fixation locations far from the center of the region of interest, fixations that are not specifically directed at salient features and a relatively high EEG alpha power.

In order to effectively exploit these findings in monitoring surveillance workers, these different cues should be incorporated in a collective model. One way to address this challenge is to use classification techniques. This would entail training a classification model on features extracted from experimental data with the aim of enhancing the classifier's capability to generalize and categorize unseen data examples (in the given context, visual versus non-visual task). Pilot investigations into the robustness of this approach using solely the EEG data are promising.

Acknowledgements. BrainGain Smart Mix Programme of the Netherlands Ministry of Economic Affairs and the Netherlands Ministry of Education, Culture and Science. We also thank Sjaak Kriekaard, Antoon Wennemers, Carien Caljouw, Lex Toet and Jan van Erp.

References

1. Itti, L., Koch, C.: A saliency-based search mechanism for overt and covert shifts of visual attention. Vis. Res. 40, 1489–1506 (2000)
2. Walther, D., Koch, C.: Modeling attention to salient proto-objects. Neural Networks 19, 1395–1407 (2006)
3. Over, E.A.B., Hooge, I.T.C., Erkelens, C.J.: A quantitative measure for the uniformity of fixation density: The Voronoi method. Behav. Res. Meth. 38(2), 251–261 (2006)
4. van Dijk, H., Schoffelen, J.M., Oostenveld, R., Jensen, O.: Pre-stimulus oscillatory activity in the alpha band predicts visual discrimination ability. J. Neurosci. 28, 1816–1823 (2008)
5. Foxe, J.J., Simpson, G.V., Ahlfors, S.P.: Parieto-occipital 10 Hz activity reflects anticipatory state of visual attention mechanisms. Neuroreport 9, 3929–3933 (1998)
6. Klimesch, W., Sauseng, P., Hanslmayr, S.: EEG alpha oscillations: the inhibition-timing hypothesis. Brain Res. Rev. 53, 63–88 (2007)
7. Toet, A., Bijl, P., Valeton, J.M.: Image dataset for testing search and detection models. Optical Engineering 40(9), 1760–1767 (2001)
8. Veltman, J.A., Gaillard, A.W.K.: Physiological workload reactions to increasing levels of task difficulty. Ergonomics 5, 656–669 (1998)
9. Pfurtscheller, G., Stancak Jr., A., Neuper, C.: Event-related synchronization (ERS) in the alpha band: an electrophysiological correlate of cortical idling (review). Int. J. Psychophysiol. 24, 39–46 (1996)