Title
Individual Differences in Pupil Dilation during Naming Task

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 38(0)

Authors
Loo, Kaidi
Rij, Jacolien Van
Jarvikivi, Juhani
et al.

Publication Date
2016

Peer reviewed

Individual Differences in Pupil Dilation during Naming Task

Kaidi Lõo (kloo@ualberta.ca)
Department of Linguistics, University of Alberta
4-32 Assiniboia Hall, Edmonton, AB T6G 2E7, Canada

Jacolien van Rij (vanrij.jacolien@gmail.com)
Center of Language and Cognition, University of Groningen
P.O.box 716, 9700AS Groningen, The Netherlands

Juhani Järvikivi (jarvikiv@ualberta.ca)
Department of Linguistics, University of Alberta
4-55 Assiniboia Hall, Edmonton, AB T6G 2E7, Canada

Harald Baayen (harald.baayen@uni-tuebingen.de)
Department of Linguistics, University of Tübingen
Wilhelmstrae 19 D-72074 Tübingen, Germany

Abstract

The present study investigates individual differences in pupil dilation during standard word naming. We looked at (i) how individual subjects' pupil size changes over the course of time and (ii) how well pupil size is predicted by the frequency of the stimuli. The time course of the pupil size was analysed with generalized additive modeling. The results show large individual variations in the pupil response pattern in this very simple task. Although, we see a pupil response to both stimulus onset and articulation onset and offset, both the amplitude of change and the direction of change differ substantially between subjects. This raises the question of what makes the pupil response functions so diverse, and one factor indicated by the frequency effect or the lack thereof might be shallow reading versus reading for content.

Keywords: pupillometry; word naming task; individual differences; word frequency; lexical processing

Introduction

The eye's pupil diameter size reflects changes in luminance, emotional state, but also cognitive processes (Ahern & Beatty, 1979; Hess & Polt, 1960; Kahnemann & Beatty, 1966; Young & Biersdorf, 1954). For example, Kahnemann and Beatty (1966) conducted a recall task to show that pupil size can be used to measure mental effort. The recall of more complex number strings triggered a stronger pupil response compared to the recall of less complex strings. Pupil dilation has also been found to reflect *linguistic* processing, such as the complexity of a linguistic task (Hyönä, Tommola, & Alaja, 1995), sentence intelligibility (Zekveld, Kramer, & Festen, 2010), sentence complexity (Ben-Nun, 1986; Just & Carpenter, 1993; Schluroff et al., 1986) and spoken language comprehension (Engelhardt, Ferreira, & Patsenko, 2010).

Finally, pupil diameter has been used to investigate lexical processing costs (Geller, Still, & Morris, 2015; Kuchinke, Võ, Hofmann, & Jacobs, 2007; Papesh & Goldinger, 2012). Kuchinke et al. (2007) measured pupil dilation in high and low frequency words and found that high frequency words receive higher peak pupil dilation compared to low frequency words. Papesh and Goldinger (2012) replicated the effect, using delayed naming paradigm. They delayed the time of the naming response up to 2000 ms and observed that frequency affects pupil size, even after the naming responses were issued. Geller et al. (2015) reported for a masked priming study that higher-frequency words elicited an earlier dilation of the pupil compared to low-frequency words. They also observed that dilation increased when words had more lexical competitors, both for low- and high-frequency words.

The present study investigates changes in pupil size while participants read out loud words presented to them one by one on a computer screen, using an inter-stimuli interval of around 4000 ms in order to allow the pupil to contract to baseline after each trial. The present study addresses the question of the extent to which the response to cognitive load in lexical processing varies across individual participants. The response of the pupil to a cognitive task has been described mathematically by Hoeks and Levelt (1993) as single smoothly increasing and then slowly decreasing function of time. Wierda, van Rijn, Taatgen, and Martens (2012) argued that a time series of pupil dilation values can be a superposition of several such functions, arising as a consequence of several cognitive events taking place within the window of time under investigation.

If in a controlled task such as word naming the pupil indeed responds in a fixed and consistent way to the demands of the task (reading the word, preparing for articulation, controlling the articulation itself, and preparation for the next trial), then one would expect the same pupil dilation function for all participants, possibly with minor variations in shape, similar to minor variations in intercept and slope that one may detect for linear response functions with the help of (generalized) linear mixed models. However, when reading styles differ substantially across participants, a general response function may not be appropriate. In what follows, we address these questions using generalized additive mixed modeling (Wood (2006); see also Baayen, van Rij, de Cat, & Wood, 2016). First, however, we introduce the details of our naming experiment.

Word Naming Experiment

Participants

Thirty-three native speakers of Estonian (18 women; 22-69 years; mean age 38), with normal or corrected to normal vision and no diagnosed speech impairments, participated in the experiment. The participant's dominant eye was tracked (30 right eye, 3 left eye). Participants received 15 euros for their participation.

Materials and Design

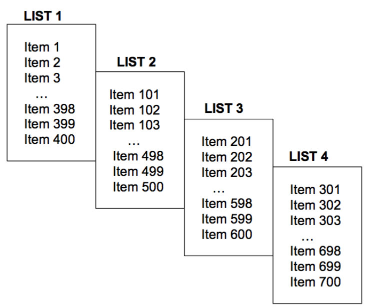

Figure 1: Outline of the experimental design.

A total of 2800 case-inflected nouns were randomly selected from the Balanced Corpus of Estonian. The frequency distribution resembled the distribution of the complete corpus, and ranged between 1 and 1000 per million (mean 14.27). The length of stimuli varied between 2 and 19 characters (mean 7.88; sd 2.62 characters). Twenty-eight experimental lists were created from the randomized set of 2800 items, each with 400 words. To maximize the number of items in the experiment, an overlapping design was used (see Figure 1), with a 300-word overlap between successive lists. A given word occurred four times in the experiment, once in each of four lists.

Apparatus

The experiment was conducted in a medium illuminated sound-attenuated booth. Eye movements and pupil size were recorded using the head-mounted EyeLink II eye tracker by SR Research Ltd. EyeLink II is a video-based tracking system with a resolution of 500 Hz, ca 3.0 ms delay, and an average spacial accuracy of approximately 0.5 degrees of arc.

The naming data was recorded separately from the stimulus presentation program ExperimentBuilder by SR Research with a Marantz PMD670 digital recorder, using a supercarioid condenser table top microphone by Beyerdynamic, placed approximately 10 cm from the participant's mouth.

Procedure

Participants were tested individually. First, they were familiarized with the procedure: reading aloud, as naturally as possible, words presented one by one on the computer screen. They were asked to start speaking as soon as the word appeared on the screen. As participants were wearing a head-mounted eye tracker, they were instructed to move their head as little as possible.

Participants were seated in front of the computer screen at a distance of approximately 60 cm. The experimenter placed the headband of the eye tracker over participant's head and adjusted it such that the eye position could be correctly tracked on the computer screen. Further, the eye tracking system was calibrated. Adjustments were made until the spacial accuracy of the eye location measurement was smaller than 0.5 degrees of arc. The stimuli were presented on a 21-inch Dell grey background computer screen in black lower case 26-point Courier New Bold font. The screen resolution was 1024x768 pixels.

Each trial started with a drift correction on the left of the screen, after which the target appeared in the center. The target stayed on the screen for 1500 ms and was then replaced by a fixation cross that remained on the screen for 2500 ms. Thus, in total each trial lasted 4000 ms. We extended the length of each trial to ensure that the pupil size was recorded with enough time delay for the pupil to contract to the baseline. The experiment started with ten practice trials, which were followed by the 400 experimental trials. Every 100th trial was followed by a short break. At the end of the experiment participants filled out a language background questionnaire. The whole procedure lasted approximately 90 minutes.

Data Preparation and Analysis

Naming latency and articulation duration were calculated directly from the audio recordings using Matlab (version 8.5.0, (MATLAB, 2015). Prior to the analysis, data for two participants were removed due to technical problems during tracking. Thus, the analysis was conducted on the data from 31 subjects.

Trials with misspoken stimuli, eye movement spikes and saccades due to eye-blinks were removed from the analysis using R (version 3.2.1, R Development Core Team, 2015, 6.3% of the trials).

The statistical analysis was performed with R (version 3.2.1, (R Development Core Team, 2015), using the *mgcv* package, version 1.8.6 of 2015, for generalized additive mixed regression (GAMM) modeling (Wood, 2006); see also Baayen et al., 2016, for visualization, we made use of the *itsadug* package (version 1.0.1, van Rij, Baayen, Wieling, and van Rijn (2015).

The dependent variable of interest was log-transformed *Pupil size*, measured in the standard (arbitrary) units delivered by the eye tracking system. The main predictors were *Time* in milliseconds and *Frequency*. Frequency was transformed to normality using the Johnson transformation (version 1.3, R package *Johnson*, Fernandez (2014)). *Gaze coordinates* (x- and y-axis position on the screen in pixels) were added to account for changes in measured pupil size due to the location on the screen.

For each participant, we fitted a separate GAMM to the 400 time series of pupil dilation values (resulting in 31 models). In addition to a general smooth for a subject's pupil dilation curve, we also included, for each word, a nonlinear random effect curve in time, using shrunk factor smooths. The X and Y coordinates of the fixation position were included as controls using a tensor product smooth, and as we anticipated the effect of frequency to vary over time, we also included a tensor product smooth for frequency and time. The model was checked for autocorrelation in the residual error, and an AR (1) autocorrelation parameter was then added to the model to remove, as far as possible, autocorrelative structure from the residuals (see Baayen et al. (2016) for a detailed discussion).

In what follows, we first discuss the estimated pupil dilation functions for the ensemble of 31 subjects. We then provide more details on three subject groups resulting from a clustering analysis.

Individual Patterns over Time

Figure 2 presents the estimated pupil dilation functions provided by GAMMs that focused on the main effect of time (leaving out frequency as covariate). The x-axis represents time and the y-axis presents pupil size. The stimulus was presented at time 0 ms, and a trial ended after 4000 ms. The first black dotted vertical line in a panel represents the median onset of articulation and the second dotted line the median offset of articulation.

Figure 2 shows that the relation between the Pupil size and the Time differs substantially between subjects. The different dilation functions fall into five groups, obtained with a divisive hierarchical clustering method using Manhattan distance applied to the first three principal components of a principal component analysis of the correlation matrix of the empirical first derivative of the subject time smooths, and indicated by different color coding.

Group 1. The first group is the largest (12 subjects: s02, s04, s05, s07, s08, s09, s18, s19, s21, s26, s30 and s31). For these subjects, the pupil dilation function shows a first peak shortly after stimulus onset and a second peak at or shortly after the onset of articulation. However, some subjects show a slightly different pattern from this general trend. For subject s02, s21, s26 and s30 there is no clear first peak, and for subject s04 and s09 the second peak much later after the onset of articulation.

Group 2. The second group includes eight subjects (s03, s11, s13, s15, s22, s25, s32, s33). Here only one clear pupil response is visible, which occurs slightly after speech onset and peaks after speech offset. Somewhat different are subject s11 and s15 who also show a slight peak after stimulus onset. However, the main difference is a well-differentiated initial peak in the pupil dilation function which is present in group 1, but absent in group 2.

Group 3. The third group also includes eight subjects (s06, s12, s14, s23, s24, s27, s28 and s29). Compared to the first two groups, this group is more heterogeneous. Also here, only one pupil peak is visible and it occurs after speech onset. However, unlike in the second group, in this group pupil size is declining instead of increasing from stimulus onset. Some subjects again deviate from the general group pattern. For example, for subject s14 there is no peak after speech onset, but the pupil stays at high plateau even after the articulation offset. Furthermore, pupil fluctuation patterns of subject s27 and s29 are somewhat similar to the first group as also here two clear pupil responses are present. However, for these subjects the second peak is much smaller than the second peak in the first group. Finally, unlike the first two groups, the relative heights of the pupil maxima are quite diverse.

Other dilation curves. The last two groups obtained by the hierarchical clustering technique clearly stand out from the rest (see the last row in Figure 2: subject s10, s16 and s17). Subject s10 and s16 start with a high pupil size, but it declines significantly after the stimulus onset. The pupil size increases also only a little after the articulation onset and stays relatively constant without a clear second peak. Finally, the last cluster only included subject s17. This subject seems to be similar to the first group and also has an initial pupil peak. However, like subject s10 and s16, this subject has no clear second peak. Because the two last groups include only three subjects, we excluded them from further group analysis.

In the present experiment, there are three key events to which the pupil appears to be sensitive: the presentation of the stimulus, the onset of articulation, and the offset of articulation. The first three groups show an increase in pupil size around the onset of articulation. The first group also shows an increase in pupil size after the word is presented, whereas it is less so for the rest of the four groups. The difference between the second and third is group is that although for both a second peak is present, subjects in the second group have an increasing pupil size and subjects in the third group decreasing pupil size. Finally, subject s10, s16 and s17 (group 4 and 5) start with a decline in pupil size from stimulus onset and show only a small increase at articulation onset.

The subjects-specific pupil dilation functions indicate that even though participants engage in exactly the same task, with exactly the same procedure, they apparently engage in this task in cognitively significantly different ways. Subjects who show little or no dilation following stimulus onset might be experienced readers, and subjects who show a decreasing pupil dilation function might be highly skilled talkers.

Next, we included a smooth for frequency and a tensor product smooth for frequency and time to the GAMM model. We compared the frequency effect differences between the three largest subject groups.

Frequency Effects in Subject Groups

Figure 3 presents the main partial effect of frequency (top panels) and the way frequency over time modifies the pupil dilation function for the three main subject groups (bottom panels). As in Figure 2, the first vertical dotted line represents median articulation onset and the second line median articulation offset. The frequency measure, normalized (and thus centered), ranges from -2 to +3; pupil size was log-transformed.

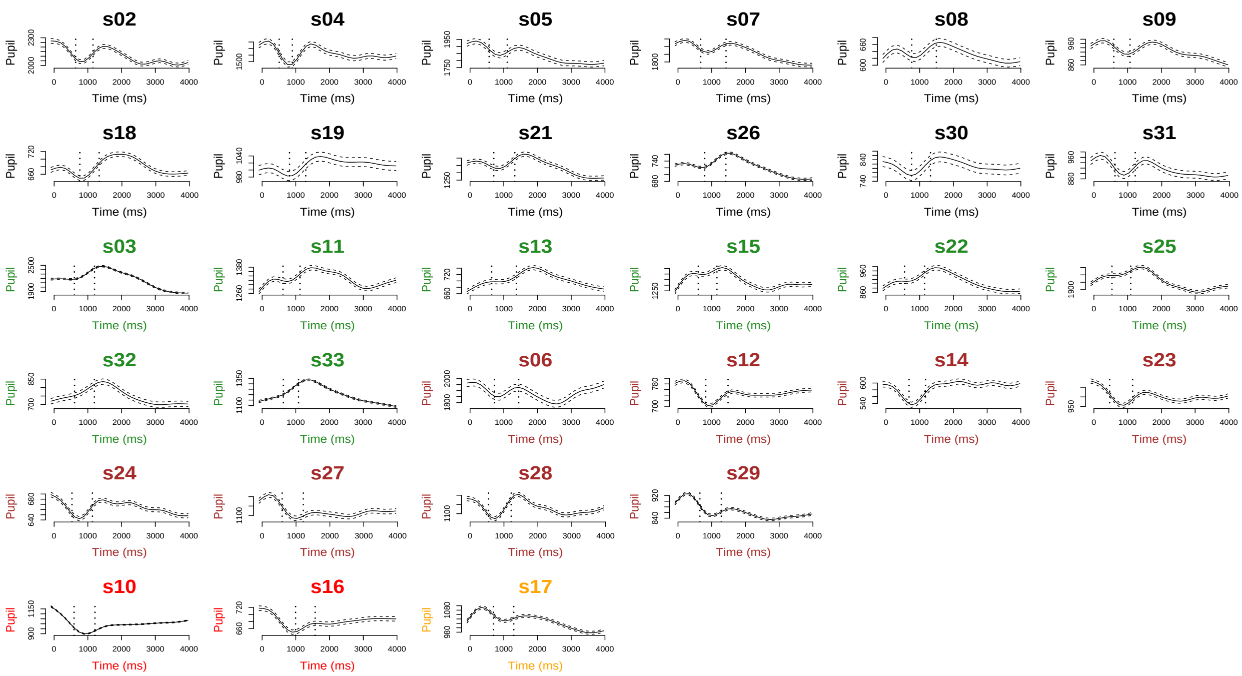

Figure 2: The fitted effects of time smooths for 31 subjects, five groups are indicated by different color-coding. The first vertical dotted line indicates the median articulation onset and the second dotted line the median articulation offset.

Group 1. The summary of the statistical model for the first group indicated a significant main effect of Frequency (t(898846)=-4.93; p-value < 0.0001). Pupil dilation decreases linearly with increasing frequency, which is consistent with high-frequency words being cognitively less demanding, typically affording shorter responses latencies in the word naming task (Forster & Chambers, 1973).

Frequency entered into a significant nonlinear interaction with time (F(14.02, 896425.4)=8.72; p-value < 0.0001). The way in which frequency modulates the pupil dilation is presented in the bottom-left panel of Figure 3. The contour plot can be read like a topographic map with peaks and valleys, yellow indicates the highest and blue the lowest elevation. The contour lines show the slope of the Pupil size as a function of Time and Frequency. The lines that are closely spaced represent steep slope and contour lines further apart gentler slopes, i.e., slower changes in Pupil size.

When the value on the y-axis is kept constant at zero, we see a similar pattern to the simple time smooths for members of the first group in Figure 2. The pupil starts off with a slight peak in pupil size shortly after the stimulus onset, contracts before the articulation, dilates again during articulation and finally, contracts at the end of the trial. This is color-coded by changes from green to yellow, from yellow to green, from green back to yellow and from yellow back to green and then blue. However, as we can also see, the pupil dilation changes differently over time dependent on the frequency of the word. The pupil size increases earlier in time and more with lower frequency words than higher frequency words (see the peak around 0 ms and the yellow peak around 1500 ms in the bottom-left panel of Figure 3).

It is noteworthy that the effect of frequency on pupil size is the largest for this group *after* speech offset. This can be seen by considering the gradient before and after articulation. Before articulation, we find three contour lines, after articulation, we find five. Even around 3000 ms after stimulus onset, frequency still has a strong effect. This suggests the effect of frequency is not restricted to the early stages of information uptake, and is likely to have a strong semantic component.

Group 2. As the middle panels of Figure 3 show, the second group has no main effect of Frequency (F(1, 578743.7)=1.17; p-value=0.28. However, the frequency and time interaction is significant (F(15.034, 578743.7)= 9.90; p-value < 0.0001). The contour lines in the bottom-middle panel are all fairly vertical, indicating an effect of time and hardly any modulation by frequency, except a weak effect before speech onset at high frequency range.

This pattern of results suggests that the second group might not be semantically engaged. It is well known that experienced readers can read out text while thinking about other issues. We suspect that the members of the second group are rather 'mechanically' performing the task, but are not deeply engaged in interpretation, possibly because isolated words in a word naming experiment are out of context and have no communicative value.

Group 3. This group also shows no significant main effect of Frequency (F(1.255, 591182.6)=8.95; p-value=0.42), but a non-linear interaction (F(14.388, 591182.6)=7.52.; p-

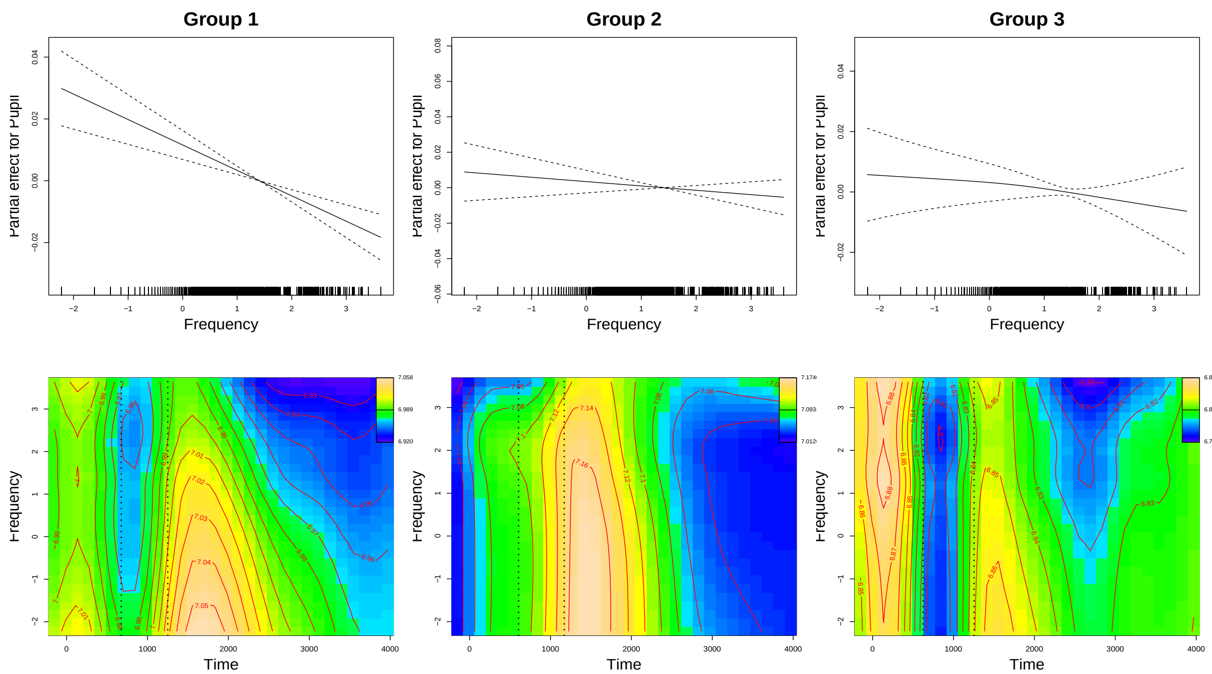

Figure 3: The partial effects of Frequency (the upper panels) and the tensor product smooth for Frequency and Time for three subject groups without random effects. The first vertical dotted line is the median articulation onset, the second vertical dotted line is the median articulation offset.

value <0.0001). The interaction is presented in the bottom-left panel of Figure 3.

The interaction effect of the third group is somewhat similar to the first group. However, compared to the first group, the effect is more shallow and gradient in particular during and after articulation. At stimulus onset, there is no frequency effect, but there is a gradual decline in pupil size across frequency span until the speech onset. Further, compared to the first group, we see fewer contour lines and a more gradual change from yellow to green after the end of the articulation. Also for this group the pupil dilation peak is weaker after speech offset. The results suggests more engagement with words compared to the second but less compared to the first group.

General Discussion

The results of this study can be summarized as follows. Inspection of the pupil dilation function for 31 participants revealed substantial variation in not only the magnitude of dilation in response to stimulus onset, speech onset, and articulation offset, but also in the direction of change, with some subjects showing contraction and others dilation. In the light of such substantial inter-subject variability, it makes little sense to try to extract a 'population dilation curve' from this kind of data. Such a curve would not come close to characterizing the pupil response function for many of the participants. This seems to apply in particular to pupillometry data, but also cognitive research in general (see e.g., Roehm, Bornkessel-Schlesewsky, Rösler, & Schlesewsky, 2007, for similar conclusions in EEG data). What this shows is either that different subjects engage in exactly the same task in very different ways or that subjects are all engaged in exactly the same way, but their engagement is manifested differently in their pupil dilations. However, based on the group differences in frequency, we argue for the first option.

Given substantial variability in subjects' reading abilities (see e.g., Kuperman & Van Dyke, 2011), loquaciousness, amount of education, social status and responsibilities, as well as differences in age, gender, and motivation for participating in a psycholinguistic experiment, these differences are perhaps unsurprising. But these differences clearly indicate that general statements about loci of processing effects in the 'population' based on pupillometry data are potentially hazardous, if the present pattern of results, revealed by detailed investigation with generalized additive mixed models, turn out to be replicable in future experiments.

This conclusion is supported by an examination of the frequency effect of three subject groups. Two of these subject groups (group 1 and 3) showed a frequency effect that was the strongest after the offset of articulation. One of the groups (group 1) also showed a weak frequency effect immediately after stimulus onset. The second group showed only a weak effect of frequency, even though the members of this group, just as the others, read the words and produced them correctly. The weak frequency effect, in combination with a relatively shallow pupil response, may be indicative of short but semantically shallow lexical processing (e.g., Baayen and Milin (2010), who observed the absence of a word frequency effect for fast readers, and the strongest effect of frequency for the slowest readers in self-paced reading of continuous text).

We think that for a proper understanding of lexical processing, in all its currently bewildering variability, it will be essential to consider in much more detail the vast differences in experience, motivation, socio-cultural background, as well as differences in brain morphology, that subjects bring into an experiment.

Acknowledgments

We would like to thank Dr. Fabian Tomaschek for his help with extracting the naming latencies. This research was supported by the national scholarship program Kristjan Jaak, funded and managed by Archimedes Foundation in collaboration with the Estonian Ministry of Education and Research and by an Alexander von Humboldt research award to the first and last author, respectively.

References

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with scholastic aptitude test scores. *Science*, *205*(4412), 1289–1292.

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*, 12–28.

Baayen, R. H., van Rij, J., de Cat, C., & Wood, S. N. (2016). Autocorrelated errors in experimental data in the language sciences: Some solutions offered by generalized additive mixed models. *arXiv preprint arXiv:1601.02043*.

Ben-Nun, Y. (1986). The use of pupillometry in the study of on-line verbal processing: Evidence for depths of processing. *Brain and Language*, *28*(1), 1–11.

Engelhardt, P. E., Ferreira, F., & Patsenko, E. G. (2010). Pupillometry reveals processing load during spoken language comprehension. *The Quarterly Journal of Experimental Psychology*, *63*(4), 639–645.

Fernandez, E. S. (2014). Harrell miscellaneous [Computer software manual]. (R package version 1.4)

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of verbal learning and verbal behavior*, *12*(6), 627–635.

Geller, J., Still, M. L., & Morris, A. L. (2015). Eyes wide open: Pupil size as a proxy for inhibition in the masked-priming paradigm. *Memory and Cognition*, *15*, 1-11.

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science*, *132*(3423), 349–350.

Hoeks, B., & Levelt, W. J. (1993). Pupillary dilation as a measure of attention: A quantitative system analysis. *Behavior Research Methods, Instruments, & Computers*, *25*(1), 16–26.

Hyönä, J., Tommola, J., & Alaja, A.-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology*, *48*(3), 598–612.

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *47*(2), 310.

Kahnemann, D., & Beatty, J. (1966). Pupil Diameter and Load on Memory. *Science*, 1583–1585.

Kuchinke, L., Võ, M. L. H., Hofmann, M., & Jacobs, A. M. (2007). Pupillary responses during lexical decisions vary with word frequency but not emotional valence. *International Journal of Psychophysiology*, *65*(2), 132–140.

Kuperman, V., & Van Dyke, J. A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of memory and language*, *65*(1), 42–73.

MATLAB. (2015). *version 8.5.0 (r2015a)*. Natick, Massachusetts: The MathWorks Inc.

Papesh, M. H., & Goldinger, S. D. (2012). Pupil-blah-metry: Cognitive effort in speech planning reflected by pupil dilation. *Attention, Perception, & Psychophysics*, *74*(4), 754–765.

R Development Core Team. (2015). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/`

Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., & Schlesewsky, M. (2007). To predict or not to predict: Influences of task and strategy on the processing of semantic relations. *Journal of Cognitive Neuroscience*, *19*(8), 1259–1274.

Schluroff, M., Zimmermann, T. E., Freeman, R., Hofmeister, K., Lorscheid, T., & Weber, A. (1986). Pupillary responses to syntactic ambiguity of sentences. *Brain and language*, *27*(2), 322–344.

van Rij, J., Baayen, R. H., Wieling, M., & van Rijn, H. (2015). *itsadug: Interpreting time series, autocorrelated data using GAMMs.* (R package version 1.0.1)

Wierda, S. M., van Rijn, H., Taatgen, N. A., & Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences*, *109*(22), 8456–8460.

Wood, S. N. (2006). *Generalized additive models*. New York: Chapman & Hall/CRC.

Young, F. A., & Biersdorf, W. R. (1954). Pupillary contraction and dilation in light and darkness. *Journal of comparative and physiological psychology*, *47*(3), 264.

Zekveld, A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: the influence of sentence intelligibility. *Ear and hearing*, *31*(4), 480–490.