When a dog is a cat and how it changes your pupil size:
Pupil dilation in response to information mismatch

Lena F. Renner, Marcin Włodarczak

Department of Linguistics
Stockholm University
Stockholm, Sweden
{lena.renner|wlodarczak}@ling.su.se

Abstract

In the present study, we investigate pupil dilation as a measure of lexical retrieval. We captured pupil size changes in reaction to a match or a mismatch between a picture and an auditorily presented word in 120 trials presented to ten native speakers of Swedish. In each trial a picture was displayed for six seconds, and 2.5 seconds into the trial the word was played through loudspeakers. The picture and the word were matching in half of the trials, and all stimuli were common high-frequency monosyllabic Swedish words. The difference in pupil diameter trajectories across the two conditions was analyzed with Functional Data Analysis. In line with the expectations, the results indicate greater dilation in the mismatch condition starting from around 800 ms after the stimulus onset. Given that similar processes were observed in brain imaging studies, pupil dilation measurements seem to provide an appropriate tool to reveal lexical retrieval. The results suggest that pupillometry could be a viable alternative to existing methods in the field of speech and language processing, for instance across different ages and clinical groups.

Index Terms: pupil dilation, speech and language processing, pupillometry

1. Introduction

Cognitive processes related to language and speech perception are observable with neuroimaging methods that assess brain responses. One of those methods is electroencephalography (EEG), which uses electrodes placed on the scalp to measure electrical properties of neurons. Neural responses occurring at different latencies with respect to a stimulus onset (called event-related potentials, or ERPs) have been found to correspond to specific cognitive processes, several of which are directly linked to speech and language processing. For instance, a negative reaction occurring about 200 ms after the stimulus (also known as mismatch negativity, or MMN) has been found to correspond to registering an auditory change, a negative reaction with a 400-ms latency (N400) has been linked to lexical retrieval and a positive reaction 600 ms (P6) after the stimulus has been found to correspond to syntactic processing (see [1] for an overview).

While EEG has repeatedly proven useful for studying neural processes involved in language perception and understanding, the method is relatively invasive. By contrast, eye-tracking provides a cheap and easy possibility to examine cognitive functions [2]. Indeed, there exists a sizable body of work using pupillometry as a method for investigating cognitive processes. For instance, systematic changes in pupil size have been observed in participants solving simple mathematical problems such as number multiplication [3], where pupil size increased

with the difficulty of the operation. Pupil dilation also reflects mental effort in linguistic tasks: listening to two sentences at the same time is harder than listening to only one sentence, which is reflected in both larger pupil dilation and a greater peak latency when paying attention to two sentences [4]. Another study investigated cognitive effort related to semantic ambiguity in pronoun processing in Dutch and found that the size of the pupil increased more when a pronominal object followed a subject pronoun rather than a full noun phrase [5].

Using pupillometry is particularly promising in those fields where neuroimaging techniques are difficult to use, such as studies with infants and toddlers. In one such study, changes in infants' pupil size were observed to reflect a reaction to an unexpected event, such as a train changing the color after passing through a tunnel [6]. Another study with 30-month-old toddlers demonstrated changes in pupil sizes in response to mispronunciations of a word [7]. Similarly, pupillometry offers an attractive alternative to EEG in speech perception studies. For instance, it could be used to detect perception of speech intelligibility, where participants' reaction times in a behavioral task are a potentially confounding variable [8].

In the present study we investigate whether pupillometry is a viable alternative to EEG in linguistic studies of lexical retrieval. We adapt the picture-word matching paradigm used in EEG studies by Friedrich and Friederici [9, 10] to investigate robustness of pupil dilation responses in adults. This paradigm has been previously used to study semantic congruity in infants and adults [10]. The procedure consists of presenting a picture alongside an auditory stimulus, which is either the name of the object in the picture (match) or of another, mismatched entity. The response to a mismatched word is known to elicit an N400 effect, which, as indicated above, is linked to lexical retrieval. Specifically, we expect greater pupil dilation when the visual and the auditory information do not match than when they do.

A similar attempt has been previously made by Kuipers and Thierry [11], who investigated lexical retrieval within a semantic priming task where ERPs and pupil size were recorded simultaneously. They showed that pupil size increases when a prime word precedes an unrelated picture but no effect was observed when the word followed the picture. In addition, a negative correlation between pupil diameter and ERP amplitude (whereby larger N400 negativity was accompanied by smaller pupil dilation) indicated a functional link between the two measures. The authors subsequently repeated the experiment with mono- and bilingual toddlers [12]. They only found the effect for the latter group, suggesting that, similar to adults, bilingual toddlers are more sensitive to unexpected visual stimuli.

Here, we revisit the question of visually primed mismatch but look for an effect beyond the first 800 ms after the stimulus

onset, which was the time-window of interest in [11]. Moreover, as pupillary data are inherently functional (i.e. they describe variation of pupil diameter over time), we analyze the data using Functional Data Analysis, which, to our knowledge, with the exception of [6] has not been used with pupillary data in linguistic studies so far. Most importantly, by treating the data as continuous-time signals, we avoid reducing it to a handful of arbitrarily selected features (such as amplitude, slope, etc). Instead, we use Principal Component Analysis (PCA) to identify the main sources of variability in our data. While PCA is widely used in the analysis of EEG signals, e.g. [13], here we use its functional analogue (fPCA), which yields more easily interpretable results. Moreover, by employing functional inferential methods we are able to compare the time-course of the pupillary data without the need for repeated testing over a series of sliding windows, thereby preserving statistical power.

2. Methods

2.1. Data collection

We measured pupil size change in reaction to a match or a mismatch between a picture and an auditorily presented word. We presented 120 trials to ten native speakers of Swedish (seven females, three males; mean age: 38.5, range: 25-63) in a windowless test room. The luminance of the testing room was constant within each participant. In each trial a picture was displayed for six seconds, and 2.5 seconds into the trial a word was played through loudspeakers. The picture and the word were matching in half of the trials, and all stimuli were common high-frequency monosyllabic Swedish words. The frequency was controlled with the Korp tool comprising 401 Swedish corpora [14]. The frequencies of the words included in this study ranged from 107,648 to 1,992,285 occurrences (with an average of 416,938). The participants were seated 60 cm from a Tobii T120 eye-tracker. The trials were randomized and presented by E-prime 2.8 and the gaze data was recorded using Tobii Studio 3.2. All pictures were of equal size and the rest of the screen was uniformly set to gray (RGB value: 179, 179, 179) for all stimuli, so that the majority of the screen had identical luminance across the trials. The baseline period of 2.5 s provided the time for the participants' eyes to adjust to change in ambient light as pupil dilation latencies in response to light are typically between 150 and 400 ms [15].

2.2. Data pre-processing and analysis

Pupil diameter data from both eyes were averaged and trials with more than 10 per cent of missing data points were excluded from the analysis. Subsequently, the data were smoothed using a 20-point Hampel filter to get rid of outliers and tracking artifacts. Missing data points (for instance due to blinks) were interpolated with a cubic spline and the resulting contours were aligned at the sound onset (that is 2.5 seconds into the trial) and scaled by individual participant's standard deviation.

The data were then submitted to Functional Data Analysis (FDA) [16]. FDA is a method for representing and analysing data in which one quantity (e.g. pupil diameter) varies *as a function* of another quantity (e.g. time). Most notably, FDA preserves the dynamic character of the relationship by representing each data point as a smoothed contour. Unlike the standard analysis methods, which require reduction of continuous functional dependencies to an arbitrary set of *scalar* features (e.g. amplitude or slope), it provides methods of identifying those dimensions which explain most of the variance in the data (functional Prin-
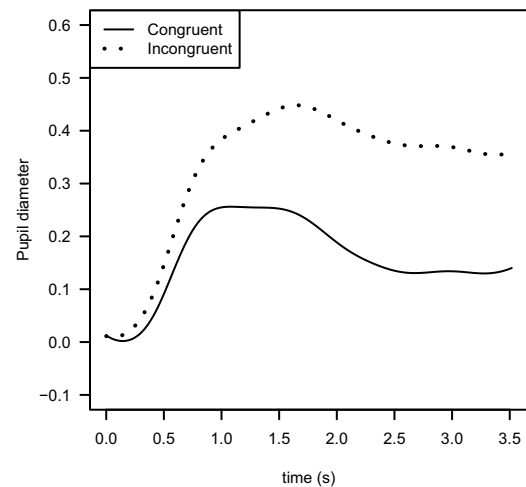

Figure 1: *Mean pupil diameter trajectories in the congruent and incongruent conditions.*

cipal Component Analysis or fPCA). It also offers inferential techniques which go beyond time-step significance testing, thus avoiding the pitfalls of carrying out repeated comparisons with its detrimental effect on statistical power.

Here, we followed the procedure outlined in [17] by adapting the R code accompanying the paper.[1] Specifically, the pupil trajectories were smoothed by fitting B-spline cubic polynomials ($\lambda = -4$, number of knots = 28) and condition means as well as functional principal components were obtained. Since the signals were already time-aligned, registration of landmarks was not necessary. Finally, congruent and incongruent conditions were compared by means of a functional t-test.

3. Results

Figure 1 plots mean trajectories of pupil diameter in the congruent (C) and the incongruent conditions (N). As expected, incongruent stimuli result in greater pupil dilation. In addition, small temporal differences in peak location are discernible: the pupil in the congruent condition reaches its maximum earlier, producing a wide plateau between 0.8 and 1.8 second after the stimulus onset, while in the incongruent condition the pupil continues to increase its size up to the end of that region.

To establish to what extend each of these parameters contributes to the distinction between the condition as well as to identify other possible sources of variability, the data were treated with functional Principal Component Analysis (fPCA), as explained above. The results are illustrated graphically in Fig. 2. Notably, since FDA describes functional relationships (in this case change in pupil diameter over time), each principal component (PC) is in itself a function (not shown here for lack of space). These functions are then added to the mean contour with a specific score *s* to reconstruct the original trajectories. Shown in the top panel of Fig. 2 are PCA functions for the first two components which jointly account for 90 per cent of variation. Specifically, each plot shows the mean pupil trajectory (in solid line) along with contours resulting from adding or subtracting

---

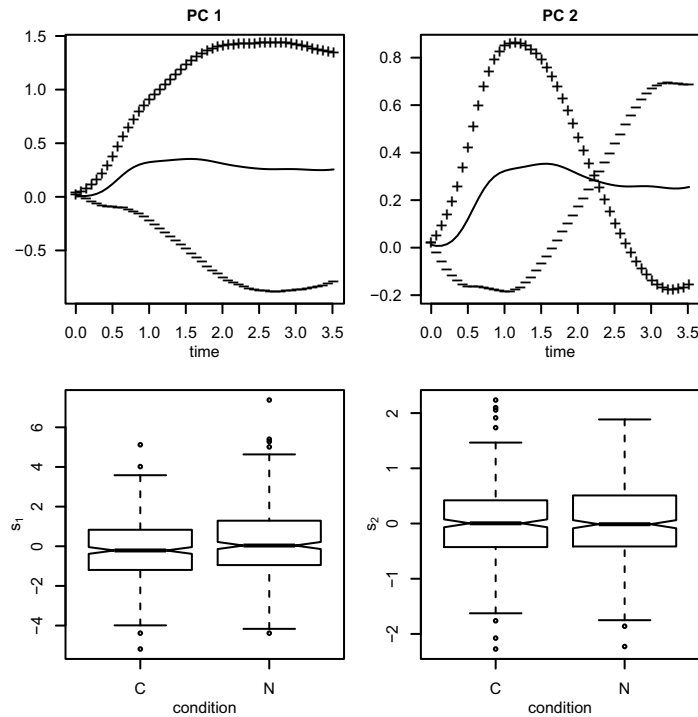[1]https://github.com/uasolo/FDA-DH, accessed 7 Feb. 2017.

Figure 2: *The first two principal component functions (top row) and the distribution of each component's scores across the experimental conditions (bottom row; C: congruent condition, N: incongruent condition). The principal component functions are visualized by the degree of excursion from the mean (solid line) when the PC is added (+'s) or subtracted (−'s) from the mean.*

some amount of that PC. From the plot it can be appreciated that the first PC, which explains 78 per cent of the variance, corresponds mainly to the magnitude of pupillary response, with positive values of s_1 resulting in increasing dilation and negative values corresponding to increasing constriction of the pupil. By contrast, PC2 (accounting for 12 per cent of the variance) determines mainly timing of the peak with large values of s_2 corresponding to earlier peak. The third PC is not shown here but it only accounts for 5 per cent of variance.

The bottom row in Fig. 2 shows the distribution of the scores for the first two components across the experimental conditions. While the distribution of s_2 does not seem to differ between the congruent and incongruent stimuli (means and standard deviations equalled -0.007 ± 0.66 and 0.006 ± 0.65 for C and N, respectively), there is a tendency for higher PC1 score values in case of a mismatch between visual and auditory stimuli (C: -0.19 ± 1.56; N: 0.17 ± 1.74). Given that higher values of s_1 correspond to increasing magnitude of pupil dilation, the distribution of s_1 indicates that, in line with the visual comparison of the trajectories in Fig. 1, the mismatch between visual and auditory stimuli does in fact result in increased pupil size.

The scores from the first two PCs were subsequently analyzed with logistic regression to test whether these components significantly predict the experimental condition. The models were build hierarchically by including in turn each of the PCs as well the interaction and using reduction of $-2 \times$ log-likelihood as a criterion for model selection. Since inclusion of PC1 resulted in significant reduction of log-likelihood ($p = 0.003$) and inclusion of further predictors did not improve the fit, the model with PC1 as the sole term was chosen. Analysis of the model revealed a significant ($p = 0.004$) effect of the first com-

ponent. Namely, an increase of one unit (i.e. corresponding to participant's one standard deviation) increases the odds of the incongruent condition by 0.13. However, the discriminative power of the model is relatively low, as evidenced in the degree of overlap between the two conditions (see Fig. 2) and the pseudo-R^2 value (0.015).

Finally, in order to pinpoint the region where the pupillary response differs between the conditions, a functional variant of the t-test was used. In addition to the overall p-value ($p = 0.02$), the test produces point estimates of the observed and critical values or the t-statistic (plotted in Fig. 3). The test revealed statistically significant differences between the conditions starting around 800 ms after the sound stimulus onset.

4. Discussion and future work

The results support our expectation of a greater pupil dilation effect in the incongruent condition. Principal component analysis revealed significant differences in the degree of maximal pupil size but no differences in its timing. The time window for this difference occurred 800 ms after stimulus onset. The delay is thus substantially longer than described by Kuipers and Thierry [11], who observed the differences between 366 and 800 ms, and is in fact closer to that reported for bilingual toddlers (566-766 ms, [12]). It should be borne in mind, however, that Kuipers and Thierry only observed an effect when the prime was auditory. While no significant differences were found for visual priming, the authors acknowledge the possibility of a later response. Indeed, the delay of 800 ms observed in the present study coincides precisely with the upper bound of their time window of interest. In addition, Kuipers and Thierry used a

slightly different experimental setup, where the picture appears a few seconds after the sound onset. By contrast, we kept the picture on the screen throughout the duration of the trial in order to keep the screen luminance constant.

The results of the present study thus support and extend on the results of [11, 12], which indicate that pupil dilation measurements are related to (and can be used as an index of) the processes of lexical retrieval. Moreover, we have used Functional Data Analysis, which is better suited to the dynamic character of the pupillary data and helps to avoid the pitfalls of repeated significance testing over a series of sliding time windows.

In spite of the by now strong evidence for the link between pupillary response and lexical retrieval, several questions remain open. For instance, it is not clear whether the observed effects are specific to lexical retrieval or whether they could be attributed to semantic integration in general.[2] If the latter were true, hearing the sound associated with the object (e.g. a dog barking) should be equivalent to hearing the word itself. We leave these question to future research.

More broadly, pupillometry is an attractive and non-intrusive technique for studying language processing phenomena also beyond lexical access. In particular, the present study indicates that speech stimuli can evoke pupillary response, which in turn opens the way for many potential applications in the field of speech perception. For instance, it could be used to study perception and the degree of mispronunciation without the need for eliciting any explicit behavioral reaction (such as clicking a button, cf. [8]). From that point of view, pupillometry could be thought of not merely as an alternative for investigating a particular ERP (say, N400 as we did here) but could be useful whenever a deviation from an expected speech pattern is produced, as in the widely-used oddball-paradigm and the mismatch negativity (MMN). We are planning to explore these possibilities in the future.

Additionally, while our study used adult subjects, the non-invasive character and the relatively low cost (compared to a standard EEG system) of the eye-tracking technology makes it particularly suitable for studies involving infants, clinical groups or field recordings. In particular, the experimental setup adapted for the purpose of the present experiment from EEG literature [9], can be easily applied to infants to investigate how children acquire lexical representations without the complications of an EEG setup. Indeed, some of these possibilities have been already explored in a recent study [7]. In the experiment, Tamási and colleagues measured pupil size changes in 30-month-old toddlers in response to picture/word matches in which the auditory stimuli consisted of feature manipulations of the target word. The results revealed a significant differences between the different phonological manipulations, thereby demonstrating that pupil dilation is a suitable tool for investigating phonological development in toddlers and younger children.

In general, our results demonstrate suitability of pupillometry for linguistic research. By using FDA, which does justice to the functional character of pupillary data, our results corroborate and substantially strengthen earlier findings on pupil dilation in priming tasks. They indicate that pupil dilation is a promising alternative to EEG for examining acquisition of lexical representations and, more broadly, speech and language processing.

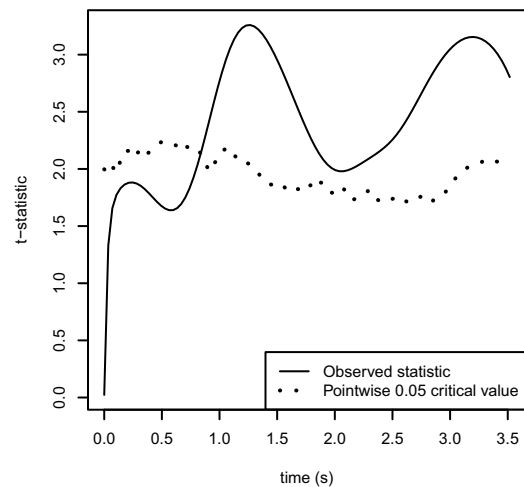[2]However, note that there is a similar debate pertaining to N400 effects (e.g. see [18]).

Figure 3: *Point-wise observed t-static and 0.05 critical values for the comparison of pupil dilation across the conditions.*

5. Acknowledgments

6. References

[1] S. J. Luck, *An Introduction to Event-Related Potential Techniques*. Cambridge: MIT Press, 2005.

[2] B. Laeng, S. Sirois, and G. Gredebäck, "Pupillometry: A window to the preconscious?" *Perspectives on Psychological Science*, vol. 7, no. 1, pp. 18–27, 2012.

[3] E. H. Hess and J. M. Polt, "Pupil size in relation to mental activity during simple problem-solving," *Science*, vol. 143, no. 3611, pp. 1190–1192, 1964.

[4] T. Koelewijn, B. G. Shinn-Cunningham, A. A. Zekveld, and S. E. Kramer, "The pupil response is sensitive to divided attention during speech processing," *Hearing Research*, vol. 312, pp. 114–120, 2014.

[5] M. Vogelzang, P. Hendriks, and H. van Rijn, "Pupillary responses reflect ambiguity resolution in pronoun processing," *Language, Cognition and Neuroscience*, pp. 1–10, 2016.

[6] I. Jackson and S. Sirois, "Infant cognition: Going full factorial with pupil dilation," *Developmental Science*, vol. 12, no. 4, pp. 670–679, 2009.

[7] K. Tamási, C. McKean, A. Gafos, T. Fritzsche, and B. Höhle, "Pupillometry registers toddlers? Sensitivity to degrees of mispronunciation," *Journal of Experimental Child Psychology*, vol. 153, pp. 140–148, 2017.

[8] S. Strömbergsson and C. Tånnander, "Correlates to intelligibility in deviant child speech-comparing clinical evaluations to audience response system-based evaluations by untrained listeners," in *Proceedings of Interspeech 2013*, Lyon, France, 2013, pp. 3717–3721.

[9] M. Friedrich and A. D. Friederici, "N400-like semantic incongruity effect in 19-month-olds: Processing known words in picture contexts," *Journal of Cognitive Neuroscience*, vol. 16, no. 8, pp. 1465–1477, 2004.

[10] ——, "Lexical priming and semantic integration reflected in the event-related potential of 14-month-olds," *NeuroReport*, vol. 16, no. 6, pp. 653–656, 2005.

[11] J. R. Kuipers and G. Thierry, "N400 amplitude reduction correlates with an increase in pupil size," *Frontiers in Human Neuroscience*, vol. 5, p. 61, 2011.

[12] ——, "ERP-pupil size correlations reveal how bilingualism enhances cognitive flexibility," *Cortex*, vol. 49, no. 10, pp. 2853–2860, 2013.

[13] J. Dien, "The ERP PCA toolkit: An open source program for advanced statistical analysis of event-related potential data," *Journal of Neuroscience Methods*, vol. 187, no. 1, pp. 138–145, 2010.

[14] L. Borin, M. Forsberg, and J. Roxendal, "Korp – the corpus infrastructure of Språkbanken," in *Proceedings of LREC 2012. Istanbul: ELRA*, 2012, pp. 474–478.

[15] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, J. Halszka, and J. van der Weijer, *Eye tracking A comprehensive guide to methods and measures*. New York: Oxford University Press, 2011.

[16] J. O. Ramsay and B. W. Silverman, *Functional data analysis*. New York: Springer-Verlag, 2005.

[17] M. Gubian, F. Torreira, and L. Boves, "Using functional data analysis for investigating multidimensional dynamic phonetic contrasts," *Journal of Phonetics*, vol. 49, pp. 16–40, 2015.

[18] M. Kutas and K. D. Federmeier, "Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP)," *Annual Review of Psychology*, vol. 62, no. 1, pp. 621–647, 2011.