Pupillometry: the Index of Cognitive Activity in a dual-task study

Vera Demberg (vera@coli.uni-saarland.de)
Cluster of Excellence, Saarland University,
Campus C7.4, 66123 Saarbrücken, Germany

Abstract

This paper reports experimental results on the index of cognitive activity (ICA), a recent micro-level measure in pupillometry, which relates processing load to the frequency of rapid small dilations of the pupil. We collected pupil size during a tracking task which was cast in a simulated driving context, as well as for a dual task of simultaneous tracking and language processing. The present results are the first to evaluate the ICA measure on these tasks. We find that the ICA is sensitive both to the simulated driving and the language task, and that it is more responsive to our driving task than overall pupil dilation. Overall, the use of the ICA as opposed to traditional pupillometry seems promising, as our data provide initial evidence that the ICA may be more responsive, and a more fine-grained measure of cognitive load than traditional macro-scale pupil dilation measures.

Keywords: Pupillometry, Index of Cognitive Activity, Dual Task, Language, Driving

Introduction

The size of the pupil has long been known to reflect arousal (Hess & Polt, 1960) and cognitive load in a variety of different tasks such as arithmetic problems (Hess & Polt, 1964), digit recall (Kahneman & Beatty, 1966), attention (Beatty, 1982) as well as language complexity (Schluroff, 1982; Just & Carpenter, 1993; Hyönä, Tommola, & Alaja, 1995; Zellin, Pannekamp, Toepel, & der Meer, 2011; Frank & Thompson, 2012), grammatical violations (Gutirrez & Shapiro, 2010) and context integration effects (Engelhardt, Ferreira, & Patsenko, 2010). All of these studies have looked at the macro-level effect of the overall dilation of the pupil as response to a stimulus. Recently, another micro-level measure of pupil dilation has been proposed, called the "Index of Cognitive Activity" or ICA (Marshall, 2000, 2002, 2007), which does not relate processing load to the overall changes in size of the pupil, but instead counts the frequency of rapid small dilation, which are usually discarded as pupillary hippus (Beatty & Lucero-Wagoner, 2000). The ICA has been argued to be robust to changes in ambient light and eye-movements, and can therefore be hoped to be more reliable and robust than overall pupil dilation. Furthermore, as it does not use the overall dilation of the pupil which can vary as a function of lighting and individual, the frequency of the rapid pupil dilations is argued to be more comparable across tasks and subjects.

If it reliably reflects processing load, the ICA would be a convenient method to assess processing load using an eye-tracker, in naturalistic environments, e.g. while driving a car, and could therefore usefully complement the range of experimental paradigms currently used.

To our knowledge, the present paper is the first to test its response to a tracking task, and to analyze properties of the Index of Cognitive Activity such as its response delay to a stimulus. The application of the method in a realistic scenario (measuring linguistically induced cognitive load during driving) also bears relevance for practical applications.

The Index of Cognitive Activity

The Index of Cognitive Activity is a patented measure of cognitive load which has previously only been evaluated on a small range of tasks (Marshall, 2000, 2002, 2007; Schwalm, 2008; Schwalm, Keinath, & Zimmer, 2008) including digit span tasks, and a simulated driving task. Using the ICA as a measure of processing load is motivated by the finding that pupil size can be affected by two different processes: lighting conditions and cognitive activity. In the overall pupil dilation, these two effects are confounded, even in stable lighting because there is a so-called "light reflex", meaning that the pupil oscillates irregularly and continually. Pupil dilation is controlled by two groups of muscles: circular muscles, which make the pupil contract and radial muscles, which make the pupil dilate. Furthermore, we know that the activation and inhibition patterns are different for reaction to light and reaction to cognitive activity (Marshall, 2000): dilations due to cognitive activity are very short and abrupt, while pupil size changes due to lighting are slower and larger. The ICA therefore tries to disentangle these patterns by performing a wavelet analysis on the pupil dilation record to remove all large oscillations and retain only the very short and rapid events (larger than a specified threshold), which are then attributed to the effect of cognitive activity.

The ICA events (rapid small dilations) per second are counted, divided by the number of expected ICA events per second (30), and the resulting number is then transformed using the hyperbolic tangent function, in order to obtain a number between zero and one[1]. To obtain a continuous measure, blinks are factored out by linear interpolation of adjacent events. When using the EyeTracking.Inc software, an ICA value per second is produced. To obtain finer granularity, we also calculated a per-100-msec ICA value from the ICA events (i.e. the rapid dilation events). Due to the short time span, we could not interpolate for blinks (which take about 100msecs) and therefore simply excluded from our analysis time all frames during which a blink or partial blink occurred.

Background on Pupillometry and the LC-NE area

It has been observed that pupil dilation is strongly correlated with activity in the locus caeruleus (LC) region of the brain.

[1]The method is patented, and the analysis program has to be licensed from EyeTracking, Inc., San Diego, CA. For details see (Marshall, 2000).

LC neurons is bilateral and emits the neuro-transmitter norepinephrine (NE) (Aston-Jones & Cohen, 2005; Laeng, Sirois, & Gredebäck, 2012). The LC-NE system is known to be activated by stress and is thought to also have a role in memory retrieval and memory consolidation. The activity of the LC-NE system as reflected in pupil dilations can therefore be a valuable method of inspecting cognitive load, and might be particularly useful also in multi-tasking settings.

Experimental Setup

We conducted an experiment with 24 subjects, during which participants had to simultaneously perform a tracking task as well as a language comprehension task. We also collected data for the tracking task in a single-task setting. Our tracking task was cast as a simulated driving task ("ConTRe task", (Mahr, Feld, Moniri, & Math, 2012)). The screen displays a moving road with two periodically moving bars at the horizon. One of the bars moves randomly across the screen ("reference bar"), while the other bar is controlled by the subject with a gaming steering wheel. The task of the participants is to cover the reference bar with their "steering bar", as exactly as possible. Difficulty of the ConTRe task was manipulated by changing the intervals at which the bar moves, as well as the speed at which it moves (the bar then always travels at a constant speed to a randomly determined destination on the horizon), to create an easy and a difficult driving setting [2].

The linguistic stimuli (loosely based on Bader & Meng, 1999; see Example (1)) consisted of 40 locally ambiguous subject and object relative clauses in German, where the relative pronoun *die* is ambiguous between nominative and accusative case. The following NP (*einige der Mieter*) is also ambiguous between these cases. Accordingly, the relative clause type (subject vs. object relative clause) is ambiguous until the disambiguating verb (hat vs. haben) is encountered.

(1) *Die Nachbarin, [**die**$_{sg,\ nom/acc}$ einige$_{pl,\ nom/acc}$ der Mieter auf Schadensersatz verklagt **hat**$_{sg}$/ **haben**$_{pl}$]$_{relative\ clause}$, traf sich gestern mit Angelika.*
"The neighbor, [whom some of the tenants sued for damages / who sued some of the tenants for damages]$_{relative\ clause}$, met Angelika yesterday."

The language stimuli were synthesized using the MARY text-to-speech system (Schröder & Trouvain, 2003). Synthesized stimuli were used to control the exact duration and timing of stimuli and pauses, so that we could more easily align our data for analysis. In particular, we made sure that the disambiguating region (*hat / haben*) was equally long in both conditions, by manipulating the duration of the pause after *hat/haben*. Furthermore, using synthesized speech avoids problems with large differences in intonation.

Our experiment was conducted in four phases, between which participants were offered to take a break. Each phase included 10 stimuli and 20 fillers, as well as 10 comprehension questions. The order of the stimuli was randomized. We recorded pupil dilations on both eyes using the head-mounted SR EyeLink II eyetracker at 250 Hz.

Data Analysis and Results

Methods All analyses reported below were done using the lme4 (Baayen, Davidson, & Bates, 2008) and mgcv (Wood, 2001) packages in R.

Distribution of the ICA Figure 1 shows the distribution of the ICA calculated per second (top plot) and calculated for a window of 100ms (bottom plot). While the aggregation is smooth for the 1s window, there are only few possible distinct events in a 100msec window (the bumps correspond to 0 events, 1 event up to 5 ICA events). Due to the tanh transformation of the ratio between observed and expected ICA events, the bulk of ICA values lies in a narrow range between 0.7 and 0.95 for the standard per-second aggregation.

The left and right ICA values are strongly correlated with each other (Spearman's rank correlation $\rho = 0.71$; $p < 0.0001$; per-second ICA), but clearly not identical.

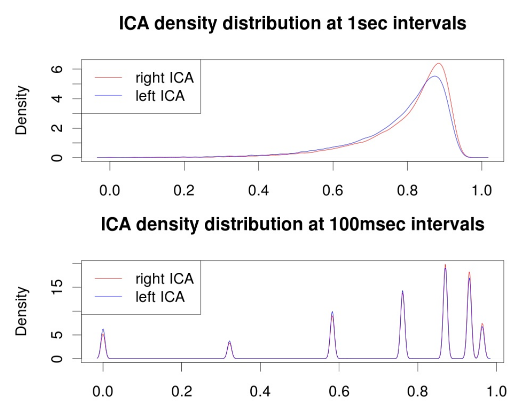


Figure 1: Understanding the distribution of ICA values: Density plot for the ICA for different aggregations.

Relationship between the ICA and pupil area Next, we inspect the relationship between the ICA and the overall pupil area. The correlation between these two measures are small (left eye per-second ICA: $\tau = 0.105; p < 0.0001$; and right eye per-second ICA: $\tau = 0.0146; p < 0.01;$). The autocorrelation plot in Figure 2 shows how dynamics of the two measures differ (Figure 2 only shows the left eye but the right eye looks very similar): while the ICA has little autocorrelation in the time-series analysis and changes dynamically, the overall pupil size has a high autocorrelation.

ICA and the ConTRe Driving Task

The reference bar moves periodically at a constant speed (ca. every 4 seconds for 1-3 seconds in easy driving and every 2.5 seconds for .5 to 1 seconds in difficult driving). This periodicity can also be seen in the autocorrelation plots shown in

[2] Driving speed was set to 40km/h in easy setting, 70km/h in the difficult driving setting; maximal speed setting for reference bar in easy setting was 1, and 2.5 in difficult setting; maximal speed setting for steering bar was 2 in easy setting and 4 in difficult setting.

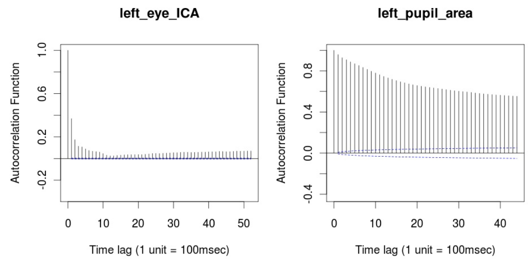


Figure 2: Auto-correlations for the ICA and pupil area.

Figure 4(a). More interestingly, we can also inspect the temporal relationship between the movement of the reference bar and any effect of this in the ICA or the overall pupil area, as shown in Figure 4(b). We can see that there is a time-shifted correlation between the movement of the reference bar and a reaction which we can measure in the ICA, starting at about 700msec after a movement in the reference bar and peaking at about 1.1 seconds after reference bar movement. This effect is more pronounced in the difficult driving conditions than in the easy driving conditions (these results hold both for the driving only and the driving plus language conditions). As Figure 4(c) shows, there is however almost no discernible effect of the reference bar movement on overall pupil dilation.

These time series analyses are interesting because there was previously no published information on how quickly to expect an effect on the ICA. We however also don't yet know enough about what we actually see in the ICA: is it related to the reference bar stimulus? or maybe rather an effect of the action taken by the participant in the task? In order to shed some light on this question, we also ran an autocorrelation analysis for the ICA and the subject controlled steering bar. As Figure 4(d) shows, the correlation between the ICA and the steering bar is stronger than the correlation between the ICA and the reference bar. As people moved the

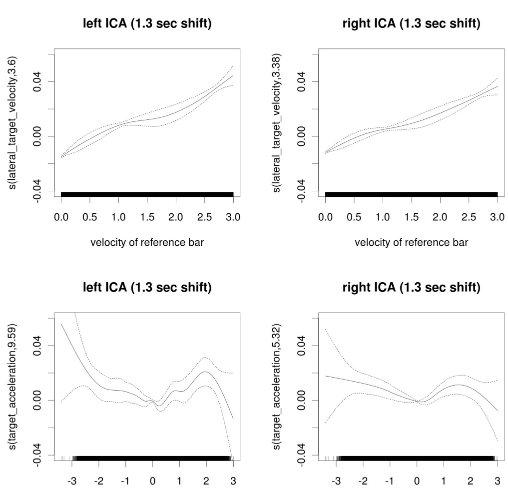
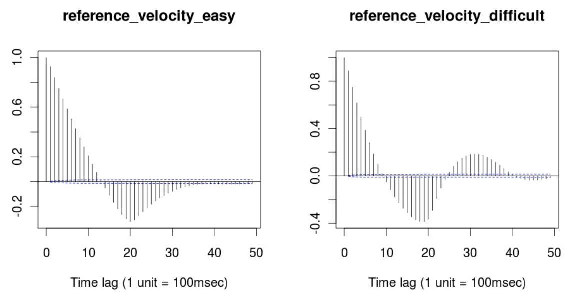
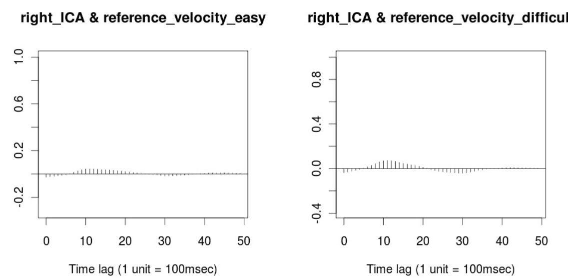


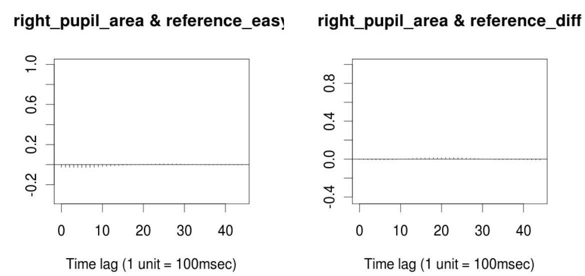
Figure 3: Spline plot (k=10) for reference bar velocity and acceleration in the same model fitting the ICA.



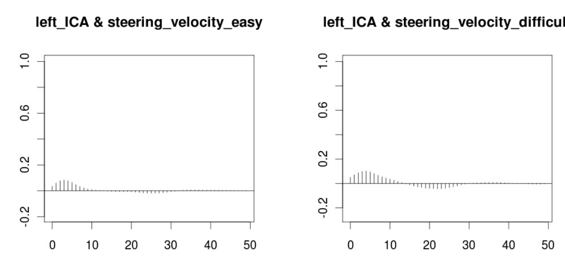
(a) Auto-correlation for the speed of the reference bar.



(b) Correlation of the right eye ICA with the speed of the reference bar at different time lags; (left eye looks the same).



(c) Almost no time-series correlation can be found between movement of the reference bar and overall pupil size.



(d) Time-series analysis: ICA and subject-controlled steering bar.

Figure 4: Time-series correlations left plots show easy driving, right plots show difficult driving.

steering bar as a reaction to the movement of the reference bar, the latency of the ICA with respect to the steering bar is also much smaller (starting right away and peaking at about 400msec). For further analysis, we re-aligned our measurements of the reference bar movement (shift by 1.3s) and steering bar movement (shift by 400msec) in order to align with the ICA.

Table 1: ICA estimates for the driving plus language phases.

	left ICA			right ICA		
	coef	t val	sig	coef	t val	sig
(Intercept)	0.704	49.30	***	0.730	50.49	***
sound file playing	0.034	9.18	***	0.033	8.99	***
easy driving	-0.008	-1.01		-0.012	-2.08	*

This adjusted alignment then allows us to enter these factors in regression and spline models. In a first analysis, we tested whether the ICA is explained only by the speed of the reference bar, or also by its acceleration. Figure 3 shows a spline plot for a model including both reference bar velocity and reference bar acceleration in fitting in turn left and right ICA. The patterns are independent of the driving condition (easy or difficult) and of the presence of language stimuli. We see a roughly linear relationship between reference bar speed and the ICA. The bottom plots of Figure 3 furthermore show a u-shaped correlation between the acceleration of the reference bar and the ICA, indicating that the ICA is larger when the reference bar starts moving or stops moving, and lower when it is not moving or moving at its constant top speed.

ICA and the language task

The Effect of Language Figure 5 shows how the left and right ICA and left and right overall pupil sizes evolve during the phases of the experiment, which consist of approximately two minutes of driving followed by four minutes of driving and listening to speech and answering yes-no questions. The speech signal consists of 10 blocks of one item, two fillers and a yes-no question. The blocks are separated by a pause of 2 seconds. It is very interesting to compare the pupil area plots and the ICA plots: pupil area is large at the beginning of a phase, but the pupil contracts soon afterwards. At the beginning of the language phase, pupil dilation increases again, which is what we expected, given the additional load of language processing. Interestingly, this is not the case in the ICA data: The ICA only goes down very little during the driving-only phase, and is overall *lower* in the dual-task section than in the single task section.

Another relevant observation is that we can observe 10 clear peaks in the ICA data, corresponding exactly to our 10 items. Such a relationship is not visible in the pupil area data (which also shows some periodicity but without a clear correspondence to stimuli). In a linear mixed effects model including only data from the driving plus language phase with the ICA as a response variable and two predictors (a flag whether a sound file is playing and a flag indicating whether the driving condition was easy or difficult), we find that the ICA is significantly higher when a sound file is being played than when it is not (i.e., between stimuli), see Table 1.

In regressions with *pupil area* as a response variable, whether the sound file is playing is a significant negative predictor on both the left eye (coef$= -0.058; t = -4.9; p > 0.001$) and the right eye (coef$= -0.067; t = -5.1; p > 0.001$), while the driving difficulty manipulation does not reach significance on either eye.

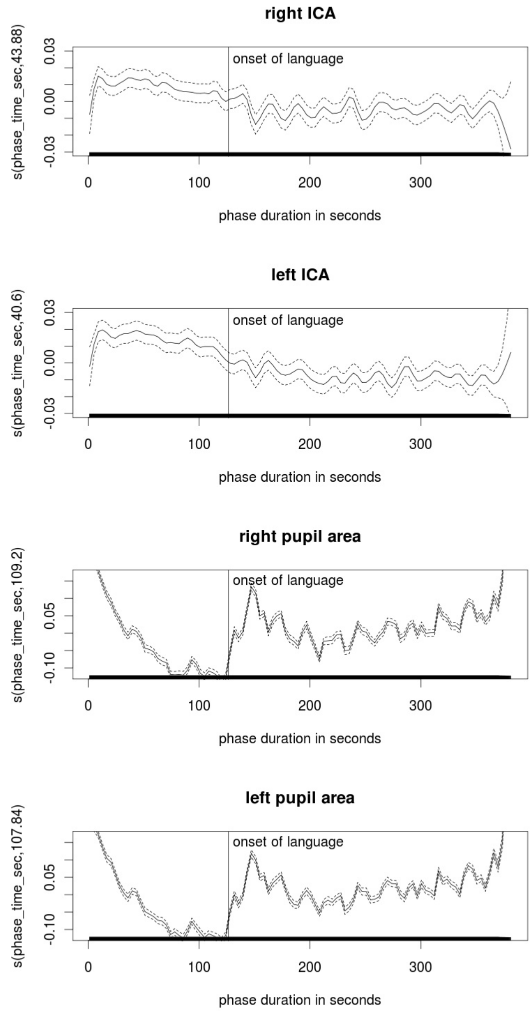
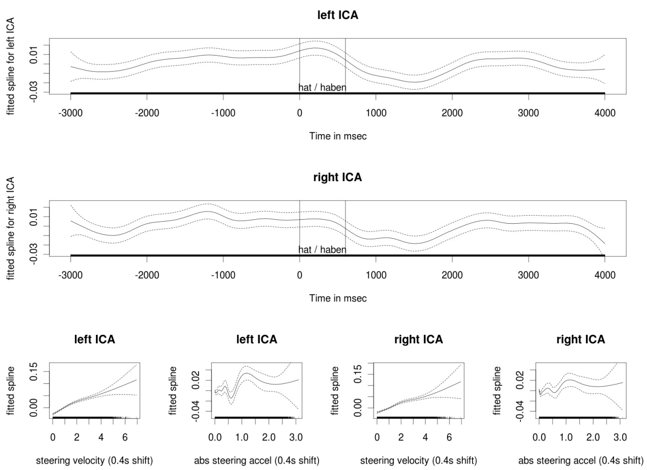


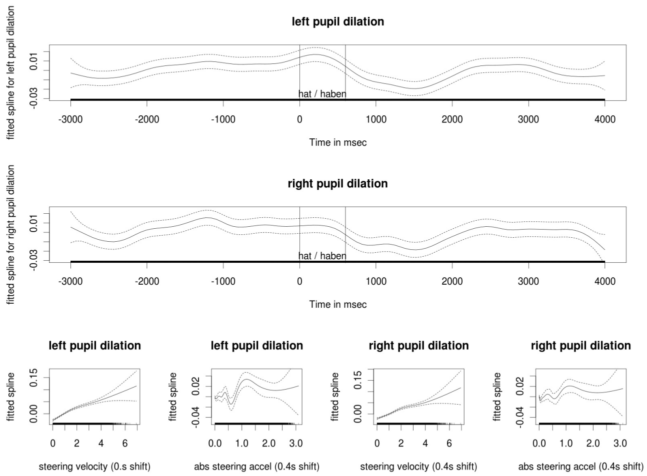
Figure 5: Spline plot (120 knots) for ICA and overall pupil dilation as a function of the duration of the driving only followed by driving with language task.

Ambiguous Region Next, we would like to see whether the ICA reflects in some way our critical region, i.e. whether we see an effect to the relative clause ambiguity. To this end, we run a spline model showing the development of the ICA during the duration of an item, with three predictor variables: time-shifted steering bar velocity, time-shifted steering bar acceleration and the distance from the critical region. Reference bar velocity does not explain any of the variance in the ICA data once steering bar velocity has been included as a predictor, therefore, our models include only the steering bar data. Figure 6 shows that the ICA is relatively high during the ambiguous region but starts falling right after disambiguation.

Disambiguating Region Note that the two relative clause conditions are collapsed in Figure 6– but can we measure a facilitation in the subject relative clause condition as opposed to the object relative clause? We ran a mixed effects regres-



(a) Left and right ICA.



(b) Overall pupil dilation of left and right eye.

Figure 6: Spline plots with confidence intervals for the ambiguous and critical region. Sentences are aligned for the onset and end of the disambiguating word "hat" / "haben".

sion model with right and left ICA (in turn) as response variables and (time shifted) reference bar velocity and acceleration, (shifted) steering bar velocity and acceleration, relative clause type, phase time (indicating how far into the phase the measurement was taken) and driving difficulty as explanatory variables. We also enter item and subject as random effects, as well as a random slopes for relative clause condition under item and subject.

The mixed effects models shown in Table 2 include data from the time window of 100 msec till 1800 msec after the onset of the critical region. Due to co-articulation, we expect that differences of *hat* vs. *haben* should be audible from about 100msec after the onset, and given our finding of the 1.3s lag between the reference bar movement and the ICA reaction, the window up to 1.8s after the onset of the critical region makes sure that we include the relevant part of the data in our model.

	left ICA			right ICA		
	Estimate	t val	sig	Estimate	t val	sig
(Intercept)	7.247e-01	39.24	***	0.718417	45.54	***
subject RC	-3.777e-02	-2.26	*			
phase time	-1.199e-07	-2.68	**			
steering velocity	2.541e-02	11.08	***	0.022656	10.34	***
steering accel.	1.094e-02	2.01	*			
SRC:phase time	1.411e-07	2.23	*			

Table 2: Mixed effects regression analysis with ICA as response variable, for region of 100 – 1800msec after the onset of critical region. (Duration of critical region: 0-600msec)

We found again that steering bar velocity is a better predictor of the ICA than reference bar velocity. For the left eye, we find that steering bar velocity, steering bar acceleration, phase time and our critical manipulation, the relative clause type, are significant predictors. In particular, we find that the left eye ICA is significantly lower when the item is a subject relative clause. We also find that the left ICA decrease as a function of when the item is presented within a phase (see also Figure 5). Additionally, we find a significant interaction between phase time and the relative clause condition, which indicates that the difference in ICA between the subject and object relative clauses gets weaker as the experiment proceeds – it is possible that this is a learning effect.

In the right ICA, we see similar tendencies, but, with the exception of steering bar velocity, the predictors fail to significantly improve model fit. It should be noted though that this finding replicates the finding of a language-only study using the same relative clause stimuli, which also found a significant effect of relative clause type on the left eye's ICA but not on the right eye (Demberg, Kiagia, & Sayeed, 2013).

While we cannot find a significant main effect of relative clause type in a regression with overall pupil size as a response variable, but we do find that the pupil size decreases significantly more quickly in the subject relative clause condition than in the object relative clause condition (this holds for both right eye ($p < 0.01$) and left eye ($p < 0.05$)).

Discussion and Conclusions

In this paper, we have reported our first experimental results on using the index of cognitive activity (ICA) as a measure of cognitive load in a dual-task scenario. Our analysis results show that the ICA and pupil dilation are rather different measures. They have a very low correlation to each other, and also behave differently: the ICA is very dynamic, while pupil dilation changes only slowly. This observation is particularly interesting, as it indicates that the ICA might be used at higher time resolution than overall pupil dilation.

The distribution of the ICA however also shows that there are some limits as to how incrementally it can be used: When calculating the ICA events at 100msec intervals, the distribution is not smooth, and there is little bandwidth of distinct events (in our data we observe between 0 and 6 such rapid movements per 100msec window).

The time series analyses reported here furthermore indicate that the ICA is reflecting the ConTRe task steering events, while no such effect is detectable in pupil area. The autocorrelation analyses also allowed us to understand more about the delay between stimulus and effect in the ICA: there is a lag of about 1.2s between the movement of the reference bar and an effect in the ICA, and a lag of about 400msec between the subject's steering action and the ICA. The fact that the correlation of the ICA and the steering bar is larger than the correlation with the reference bar indicates that the ICA might be related to the participants action execution (as opposed to their perception of the steering task). This is also confirmed by mixed effects regression models with the ICA as a response variable and re-aligned steering bar and reference bar velocity as predictors: the reference bar velocity predictor variable does not improve model fit over models which already include steering bar velocity.

Furthermore, we find that the ICA record reflects our secondary task, language comprehension. In a more detailed analysis, we find that the ICA is significantly higher within the dual task condition whenever the language stimulus is not playing, and that the ICA is high during the ambiguous region of our language stimulus and decreases following disambiguation. We also find a significant effect of our language manipulation, showing that the ICA of the left eye is significantly higher in the object relative clause condition than in the subject relative clause condition.

We also compare the ICA measure to traditional overall pupil dilation and find that our primary tracking task is not reflected in pupil dilation. For our language manipulation, the results in overall pupil dilation are consistent with our findings in the ICA: in the subject relative clause condition, the pupil contracts significantly faster than in the object relative clause condition.

References

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus norepinephrine function: adaptive gain and optimal performance. *Annual review of neuroscience*, *28*, 403–450.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*.

Bader, M., & Meng, M. (1999). Subject-object ambiguities in german embedded clauses: An across-the-board comparison. *Journal of Psycholinguistic Research*, *28*(2).

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin*, *91*(2), 276.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system.

Demberg, V., Kiagia, E., & Sayeed, A. (2013). The index of cognitive activity as a measure of linguistic processing. In *Proceedings of the 35th annual meeting of the cognitive science society (cogsci-13).*

Engelhardt, P. E., Ferreira, F., & Patsenko, E. G. (2010). Pupillometry reveals processing load during spoken language comprehension. *Quarterly journal of experimental psychology*, *63*, 639—645.

Frank, S., & Thompson, R. (2012). Early effects of word surprisal on pupil size during reading. In *Proc. 34th annu. conf. cognitive science society (eds n. miyake, d. peebles & rp cooper)* (pp. 1554–1559).

Gutirrez, R. S., & Shapiro, L. P. (2010). Measuring the timecourse of sentence processing with pupillometry. In *Cuny conference on human sentence processing.*

Hess, E., & Polt, J. (1960). Pupil size as related to interest value of visual stimuli. *Science*.

Hess, E., & Polt, J. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*.

Hyönä, J., Tommola, J., & Alaja, A. (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*, *47*.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*.

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: a window to the preconscious? *Perspectives on psychological science*, *7*(1), 18–27.

Mahr, A., Feld, M., Moniri, M., & Math, R. (2012). The contre (Continuous Tracking and Reaction) task: A flexible approach for assessing driver cognitive workload with high sensitivity. In *Cognitive load and in-vehicle human-machine interaction workshop at automotiveui.*

Marshall, S. (2000). *U.s. patent no. 6,090,051.*

Marshall, S. (2002). The index of cognitive activity: Measuring cognitive workload. In *proc. 7th conference on human factors and power plants* (pp. 7–5).

Marshall, S. (2007). Identifying cognitive state from eye metrics. *Aviation, space, and environmental medicine*, *78*(Supplement 1), B165–B175.

Schluroff, M. (1982). Pupil responses to grammatical complexity of sentences. *Brain and language*, *17*(1), 133–145.

Schröder, M., & Trouvain, J. (2003). The german text-to-speech synthesis system mary: A tool for research, development and teaching. *International Journal of Speech Technology*, *6*(4), 365–377.

Schwalm, M. (2008). *Pupillometrie als methode zur erfassung mentaler beanspruchungen im automotiven kontext*. Unpublished doctoral dissertation, Universitätsbibliothek.

Schwalm, M., Keinath, A., & Zimmer, H. (2008). Pupillometry as a method for measuring mental workload within a simulated driving task. *Human Factors for assistance and automation*, 1–13.

Wood, S. (2001). mgcv: Gams and generalized ridge regression for r. *R news*, *1*(2), 20–25.

Zellin, M., Pannekamp, A., Toepel, U., & der Meer, E. (2011). In the eye of the listener: Pupil dilation elucidates discourse processing. *Int. Journal of Psychophysiology*.