Language and cognitive load in a dual task environment

Nikolaos Engonopoulos (nikolaos.engonopoulos@uni-potsdam.de)
Potsdam University, Department of Linguistics
Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany

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

Abstract

We investigate the effect of linguistic complexity on cognitive load in a dual-task scenario, namely simultaneous driving and language use. To this end, we designed an experiment where participants use a driving simulator while listening to spoken stimuli and answering comprehension questions. Online physiological measures of cognitive load, including the recently established Index of Cognitive Activity, as well as measures of performance in both tasks have been collected with high temporal resolution. The resulting aligned data streams can be used to test a vast array of different hypotheses about the relationship between performance, difficulty, and cognitive load in dual tasks at various levels of temporal resolution and linguistic structure. We present results of the data analysis, including evidence that different linguistic structures may cause measurable changes in cognitive workload on a very fine temporal scale in cases of increased primary task difficulty.

Keywords: relative clause; dual task; cognitive load; pupillometry; skin conductance; tracking task; driving; multitasking

Introduction

Is there a relationship between psycholinguistic measures of language complexity and quantified cognitive workload in dual-task environments? To answer this question, we experimentally evaluate these measures of language processing in an environment where one task is language-related and the other not. Such language complexity measures have been shown in single-task studies to account for processing difficulty. This work represents a first step in which we investigate the effect of a grammatical structure (German locally ambiguous subject vs. object relative clauses) on a simplified, well-controlled non-linguistic task, a driving task.

Dual tasks are ubiquitous in everyday life, often in situations where attention and performance in the primary task is critical. An example is driving while engaging in dialogue, be it with a passenger, a dialogue-controlled interface, or remotely via mobile phone. Engaging in dialogue generally affects driving performance and safety (Just, Keller, & Cynkar, 2008; Young, Regan, & Hammer, 2007).

We manipulated the driving task difficulty and the structural complexity of the linguistic items. We also collected measurements of performance in both tasks and fine-grained physiological indicators of cognitive load, namely skin conductance levels and pupil sizes. We computed values from pupil size for the recent Index of Cognitive Activity (ICA). To our knowledge, this is the first study using the ICA measure in a setting with a language task.

Background and Related Work

There is a rich literature on language use while driving a car, largely showing that speaking on the telephone has a negative effect on driving performance (Just et al., 2008; Kubose et al., 2006). Further studies found that this is specific to conversations with remote speakers (independent of whether one uses a hand-held device or free speaking), but that conversations with an in-car passenger are less problematic (Strayer, Drews, & Johnston, 2003; Drews, Pasupathi, & Strayer, 2004). It appears that passengers adapt their conversation to the traffic situation, leaving the driver more resources to deal with demands of the driving task when driving becomes difficult (Drews, Pasupathi, & Strayer, 2008; Crundall, Bains, Chapman, & Underwood, 2005; Villing, 2009). By contrast, remote conversational partners cannot adapt their speech, so that the driver may reach the point of cognitive overload more easily and thus commit driving errors. However, these lines of research have not taken into account how the fine-grained details of linguistic complexity affect cognitive load and driving task performance.

On the other hand, there is a very rich literature on linguistic processing difficulty in single tasks using brain imaging, ERPs, and reading time studies, as well as a number of dual task experiments generally showing that performance on the linguistic task deteriorates with increased complexity of the other task, see for example King and Just (1991). Finally, multiple models explain the effect of cognitive load in one task on performance in another (Baddeley, 2003; Wickens, 2008; Just, Carpenter, & Miyake, 2003).

We see, however, unbroken ground in relating the effect of linguistic complexity on a realistic task (e.g., driving) and the size of the interference of linguistic processing with driving performance. This study takes a step in this direction in testing different methods for assessing cognitive load and the effect of one particular linguistic structure—incrementally ambiguous relative clauses—on driving performance in a simplified but controllable and continuous driving task.

The dual-task experiment

The ConTRe task

Our primary task was a tracking task (Jagacinski & Flach, 2003) presented as a car driving scenario and called the "Continuous Tracking and Reaction" (ConTRe) task (Mahr, Feld,

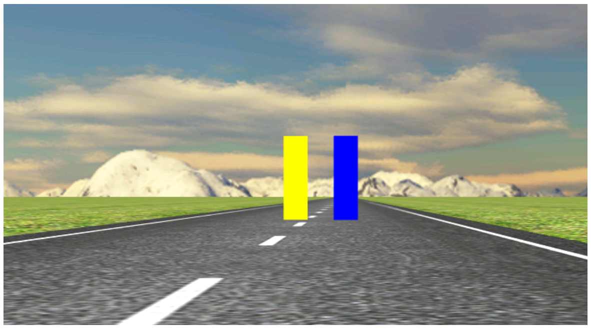

Figure 1: A screenshot of the ConTRe steering task.

Moniri, & Math, 2012). In this task, participants see a simulated 3-D road moving at a constant speed, intended to simulate a moving vehicle. Additionally, two bars of different color appear approximately 20m in front of the simulated vehicle. The two bars represent the vehicle's position and the target (reference) position. They move laterally across the screen. The reference bar's movement is pseudo-randomly generated by an algorithm, while the "vehicle" bar is controllable by the participant by means of a gaming steering wheel. Participants were instructed to track the reference bar's movements with the controllable bar as closely as possible. To reduce noise in our data, we removed all other elements of the original ConTRe environment (e.g., buildings along the side of the road, and traffic lights), except for the road and the moving bars. A screenshot of the simulated environment can be seen in fig. 1.

This task is a useful abstraction of driving, since it allows a precise and continuous performance measure for steering, essential to driving. We manipulated the difficulty of the ConTRe task by changing the speed of the reference and vehicle bars in order to create a "difficult driving" condition and an "easy driving" condition[1].

Language comprehension task

The spoken comprehension task consists in listening to a sentence containing a relative clause followed by two thematically related 'filler' sentences and a yes/no comprehension question. Questions were related to the relative clause (50% of the stimuli) or to the filler sentences. All sentences and questions are in German, inspired by Bader and Meng (1999). The stimuli are designed in pairs in such a way that the items in each pair are identical except for the form of the auxiliary of the relative clause (RC), which determines whether it is an object RC (ORC) or a subject RC (SRC). An example of such a relative clause pair is the following:

> *Die Lehrerin, die einige Eltern wegen einer solchen Kleinigkeit angerufen* ***[haben / hat]****, hat nun eine Elternversammlung einberufen.*
>
> "The teacher$_{FEM}$ [who called some parents / whom some parents called] because of such a trivial issue, has now called a parents' meeting."

The sentence is locally ambiguous between ORC and SRC until reaching the auxiliary; in previous experiments, increased reaction times in a speeded judgment task (Bader & Meng, 1999) have been observed when subjects read "haben" (ORC) compared to "hat" (SRC). This is evidence for an interpretive bias toward SRC. All items were synthesized prior to the experiment using MARY TTS (Schröder, Charfuelan, Pammi, & Türk, 2008) and pauses manipulated so that the critical region duration (*hat* / *haben*) is always identical.

Experimental setup

Each experiment is divided into 4 recording phases, each lasting about 6 minutes, with short pauses in-between. Each phase is composed of a driving-only phase of 2 minutes followed by a driving-with-language phase of approximately 4 minutes, during which 10 blocks, consisting each of one relative clause, two fillers and one question. Participants answer the question verbally and their response is coded by the experimenter. In the first and the third phase, the driving difficulty is set to "easy", while in the second and fourth phase it is set to "difficult". The order of presented items in the language condition was randomized, and we ensured that each person only saw one condition of each item.

Measures of cognitive workload

We have two principal sources of quantified cognitive workload data: physiological and task dependent. Our physiological measures are further divided into two subtypes: pupil area-based (pupillometry) and skin conductance-based, both of which have been widely used in cognitive workload studies, although principally on non-linguistic tasks. Our study is an opportunity to evaluate the relative efficacy of these data sources on linguistic tasks. We also take the opportunity to evaluate a novel form of pupillometric data processing: the Index of Cognitive Activity (ICA). To the best of our knowledge, ours is the first study to investigate the potential of the ICA as a measure of linguistically-induced cognitive load in a dual-task scenario.

Our task-dependent measure is driving performance in our simulated environment, which serves to confirm the "real-world" effect of variations in cognitive workload.

The Index of Cognitive Activity (ICA) Research in pupillometry (Just et al., 2003; Engelhardt, Ferreira, & Patsenko, 2010; Palinko, Kun, Shyrokov, & Heeman, 2010) has found that cognition-related changes in pupil size typically amount to a difference of 20% relative to the typical pupil size (Laeng, Sirois, & Gredebäck, 2010). However, light conditions also affect pupil sizes, with brightness-induced changes being much larger than cognitively induced ones (up to 120% of typical pupil size).

The *Index of Cognitive Activity* (ICA; Marshall (2002)) is a patented measure which applies signal processing techniques to filter out slow, large light-induced changes and identify

---

[1] Easy: reference bar maximum speed = 1m/s, controllable bar = 2m/s. Difficult: reference bar = 2.5 m/s, controllable bar = 4m/s.

the occurrence of short, abrupt changes in pupil size, held to be caused by cognitive load. The ICA measure is argued to be robust with respect to changes in light conditions and eye movement. It relates the frequency of rapid small changes in pupil size (also known as pupillary hippus) to cognitive load. The ICA measure has been used for measuring cognitive load in driving simulation tasks (Schwalm, Keinath, & Zimmer, 2008), simulated driving and visual search (Marshall, 2007), detecting different levels of surgical skill (Richstone et al., 2010), and for measuring linguistically induced cognitive load (Demberg, Kiagia, & Sayeed, 2013) among other uses. Demberg (2013) provides a more detailed analysis of the ICA measure in the dual task setting presented here.

ICA measurements have been shown to be relatively stable across several commonly used eye tracker models and sample rates ranging from 60 to 300 Hz (Bartels & Marshall, 2012). We used a head-mounted Eyelink II and sampled at 250Hz.

Skin conductance response Our second physiological proxy for measuring cognitive load is skin conductance response (SCR), which we calculate from skin conductance level (SCL). Changes in the electrical conductance of the skin are due to activity of the sweat glands, which are in turn controlled by the sympathetic nervous system. Skin conductance amplitude usually changes with respect to its "neutral" (*tonic*) level in response to unexpected, significant, or aversive stimuli. SCL has been previously used as a measure of cognitive load (Shi, Ruiz, Taib, Choi, & Chen, 2007). In a dual task experiment with simulated driving and a secondary cognitive task, B. Mehler, Reimer, Coughlin, and Dusek (2009) found that skin conductance levels peaked in cases of mental overload caused by incrementally increasing secondary task difficulty, which was followed by a deterioration in the performance of the primary task. Son and Park (2011) found skin conductance levels along with steering wheel reversals (used as a measure of task performance) to be good input features for an artificial neural network built to predict task difficulty.

We used the Ledalab software (Benedek & Kaernbach, 2010) to separate our raw skin conductance measurements into an estimate of the tonic component and the phasic component. The software also allows to calculate the number of skin conductance response events. SCR events are the "peaks" of the phasic component of skin conductance; both the number of such events per time unit and the amplitude of the peaks are used in the analysis below.

Driving performance We use performance on the ConTRe task as an additional measure of cognitive load. The task lets us define several measures of task success, including the distance between the reference bar and the controllable bar at each point in time and the speed and acceleration of the controllable bar.

Results

We ran our experiment with 24 German native speakers aged 20-34, with the total duration of the recorded samples summing up to about 12 hours. We performed our data analysis in R using linear mixed effects (LME) modeling with lme4 (Baayen, Davidson, & Bates, 2008) and mgcv (Wood, 2001).

Correlation between physiological measures

The first question we explored was whether our physiological measures are correlated with one another. While there is no significant correlation between the raw skin conductance levels and the ICA, we do find a significant positive correlation between the number of skin conductance events and the ICA (using Spearman's ρ; left ICA: $\rho = 0.06; p < 0.0001$; right ICA: $\rho = 0.09; p < 0.0001$). One important aspect to keep in mind is also possibly different latencies of the two measures in reaction to a stimulus.

We find a strong correlation between the ICA of the left and right eye ($cor = 0.74; p > 0.001$, Pearson's product-moment correlation coefficient).

Response to experimental phases

Driving performance The next hypothesis we tested was whether our task performance measure in the driving task, i.e., the steering deviation, is sensitive not only to the driving task difficulty, but also to the presence of language. In figure 2, we have plotted the mean deviation for each of the difficulty settings (easy and difficult driving), with and without the secondary linguistic task. Using linear mixed effects models with a random intercept and random slopes by subject, we found a large significant main effect of driving difficulty ($coef = 0.3; t = 20.33; p < 0.001$), showing that steering was less accurate when driving was more difficult. We also found a significant positive main effect of whether we are in a language phase ($coef = -0.05; t = -5.00; p < 0.001$; steering is worse when people are listening to language, see also figure 3), as well as a significant interaction between driving difficulty and the language phase, indicating that the effect of language was more burdensome in the difficult driving condition ($coef = -0.024; t = -6.98; p < 0.001$). To confirm whether the effect of language is significant in both driving conditions, we also split the data into two subsets, easy driving and difficult driving, and found that the effect of language was significant in both linear mixed effects models.

This figure illustrates an obvious difference between steering deviation in the easy and difficult driving conditions.

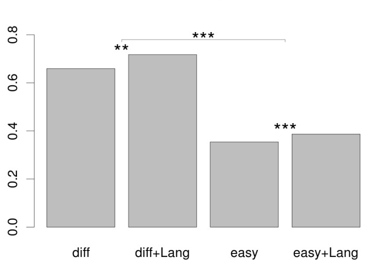

Figure 2: Driving condition/language vs. steering deviation.

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

	right ICA			left ICA		
	coef	t-value	sign	coef	t value	sign
(Intercept)	0.8116	123.40	***	0.7965	135.63	***
sound playing	0.0198	10.88	***	0.0186	10.10	***
easy driving	-0.0057	-2.44	*	-0.0004	-0.21	

Table 2: # of SCR events reduced during easy driving. (Random slope of driving condition by subject included.)

	Estimate	t value	signif.
(Intercept)	0.68626	12.550	***
difficulty=easy	-0.06495	-4.274	***

Pupillometry For the ICA, we find a main effect of driving difficulty in the ICA of the right eye, but not in the left eye (Table 1). Furthermore, we find significantly more blinks during the phases when language was playing. In-depth analysis of the pupillometric data reveals that overall dilation was larger when people were listening to language stimuli, but the number of ICA events was lower (Figure 3). If we look into the language phase, however, the ICA of both eyes went down significantly whenever language wasn't playing (e.g., between stimuli; Table 1: we factored out the effect of blinks or partial blinks on both the pupil area calculations and the ICA). This effect can also be seen in Figure 3, where the 10 ICA spikes in the language region coincide with our 10 blocks of language stimuli.

Skin conductance For skin conductance, we cannot easily compare the easy vs. difficult driving settings, as the skin conductance measuring device was removed between phases, and comparison of absolute values between phases is thus impossible. A measure that can be compared between driving conditions is however the number of *skin conductance events*. When running a linear mixed effects regression model with this measure as a response variable, we find that more such skin conductance events happened, as expected, in the difficult driving condition, see Table 2.

We do however not find any significant effect of the language vs. no language condition on this measure. Unexpectedly, we find that tonic skin conductance is *lower* in the driving plus language condition, see Figure 3.

Cognitive load and language processing difficulty

To this point, we find that the measures largely behave as expected. Thus we come to our main question: can they detect the effect of fine-grained language complexity? To this end, we analysed the data to see whether we can find a) a correlate for higher processing difficulty in the ambiguous region or right after the disambiguation at *hat*/*haben*, and b) whether ORCs lead to less cognitive load than SRCs.

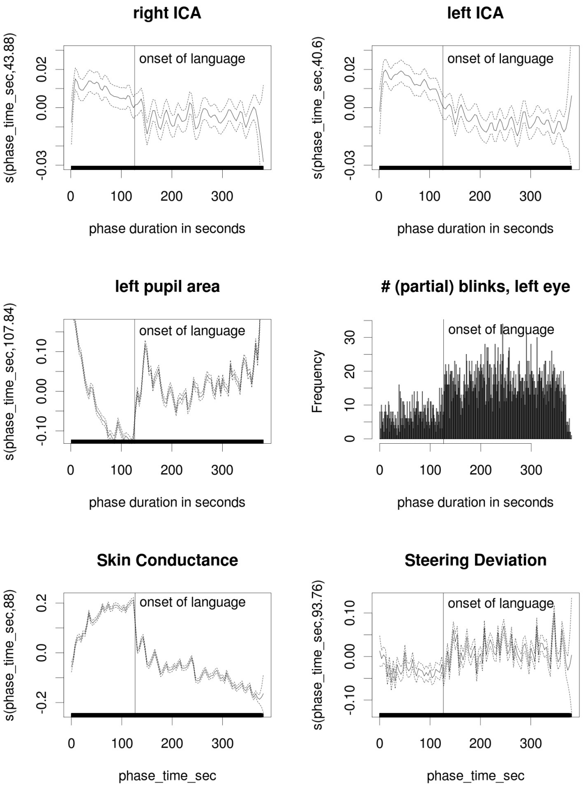

Figure 3: Spline plots (120 knots; with 0.95 conf intervals) showing the effect of language on physiological measures during an experimental phase (2 min driving only followed by 4 min of driving plus language).

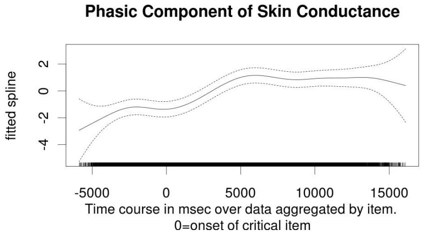

Figure 4: SCR during time that stimulus is spoken.

Disambiguating region Detailed analysis of the ambiguous region of the relative clause shows that the **Index of Cognitive Activity** is high during the ambiguous region of the relative clause (during the time span of -2000msec to 0msec), and that the ICA sharply falls right after disambiguation (see Table 4 which shows a significant reduction in ICA of both eyes following disambiguation, encoded as *time wrt. onset*). These effects hold over and above effects of the steering task, which have been mathematically accounted for by including the task difficulty as a factor in the model. These results indicate that subjects encounter processing difficulty due to the ambiguity. (This is possibly also something they learn during the experiment.)

For **skin conductance**, we know that effects can be expected 2-4 seconds after the stimulus. Figure 4 shows a significant rise in skin conductance during the five seconds after

Table 3: Mixed effects regression analysis with steering deviation as response variable, for region of 2s before the onset till 2s after end of the critical region.

	Estimate	t-value	
(Intercept)	3.562e-01	17.07	***
phase time	8.459e-08	3.44	***
target velocity	3.832e-01	205.08	***
critical region	1.396e-02	2.88	**
easy driving	-2.248e-01	-64.91	***
target acceleration	-2.680e-02	-5.90	***

Table 4: Mixed effects regression analysis with left and right ICA as response variable, 100–1800msec after critical region onset. (Critical region duration: 0-600msec)

	left ICA			right ICA		
	Estimate	t-value		Estimate	t-value	
(Intercept)	0.7504	35.71	***	0.736	37.82	***
subject RC	-0.0354	-2.12	*			
phase time	-1.16×10^{-7}	-2.59	*			
time wrt. onset	-2.78×10^{-5}	-6.38	***	-1.84×10^{-5}	-4.36	***
steering veloc	0.0257	5.37	***	0.0226	4.88	***
steering accel	0.0108	2.00	*			
SRC:phase time	1.34×10^{-7}	2.12	*			

the critical region, which would be consistent with an interpretation that the ambiguity causes higher cognitive load.

But can we see any effect of our linguistic stimuli on the **driving performance**? We compared steering accuracy at the time of the disambiguating region with steering accuracy during the two seconds before and after, and indeed found that deviation of the controllable bar from the reference bar was significantly larger during the disambiguating region than before or after; see the positive coefficient (Table 3) for the binary variable "critical region".

Subject vs. object relative clauses Finally, we test whether the ICA is sensitive to fine-grained linguistic complexity effects. We isolated the subset of the data which fell within the 1800msec following the onset of the critical region *hat / haben*. The duration of this critical region at *hat / haben* is 650 ms in both conditions, which we imposed by manipulating the duration of the phrase boundary pause during synthesis. On this subset of the data, we built two LME models (one for each eye) with the ICA measure as the response variable and the relative clause type as the fixed effect, while also introducing a random effect per participant.

The results of this analysis are shown in Table 4. We can see that there is a negative effect for the SRC type in both cases, although only the result for the right eye is significant. The interpretation of the coefficient is that SRCs tend to occur with smaller values of ICA than ORCs.

We did not find any significant effects of relative clause condition on skin conductance, overall pupil dilation or steering performance.

Table 5: LME model for answer accuracy.

	Estimate	t-value	Sig
INTERCEPT	2.663	5.72	***
RC-TYPE (OBJ)	0.445	1.17	
VOICE (PASSIVE)	-1.802	-3.11	**
DRIVINGDIFFICULTY (EASY)	-0.222	-1.18	

Performance in the language task

A last link that we wanted to investigate was the one between performance in the linguistic task (i.e., answer accuracy) and the difficulty of the driving and language tasks. We built a binomial LME model with the answer accuracy as the response factor and driving task difficulty, relative clause type, and the voice (passive vs. active) of the question as fixed effects with a random intercept per participant and subject and a random slope for relative clause type by item[2]. The resulting coefficients are presented in Table 5. While answer accuracy was lower for object relative clauses (74%) than for subject relative clauses (78%), and lower in difficult driving (75%) than in easy driving (77%), these differences did not reach significance. (NB: questions related to relative clauses were only asked after half of the items; i.e., this analysis is based on relatively little data.) The only significant negative effect on answer accuracy was found for passive voice questions, which means that there are significantly more wrong answers to passive voice questions than to active voice ones (this is not unexpected, as it has long been known that passives are more difficult to process than actives (J. Mehler, 1963)).

Discussion and conclusions

We designed the tasks in our experiment to require continuous attention. The language task clearly affects performance on the primary steering task: we see the effect of the secondary task in all of our measures. Furthermore, we find effects of linguistic ambiguity and complexity in our measures of cognitive load: during the ambiguous region in our stimuli, we see evidence for higher cognitive load in our pupillometric measure, which is also reflected in a slightly later galvanic skin response. During the disambiguating region, we observe significantly higher steering deviation, which indicates that people are allocating more mental resources to the linguistic task, hence impeding steering performance. We also found evidence for a measurable effect of linguistic complexity in our pupillometric measure ICA: the ICA was significantly higher during the disambiguating region and the following second for the ORC condition compared to SRC. This experiment provides early support for the ICA as a useful measure to assess language-induced cognitive load.

References

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects

---

[2] Best model according to standard comparison with AIC and χ^2.

and items. *Journal of memory and language*, *59*(4), 390–412.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews: Neuroscience*, *4*.

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

Bartels, M., & Marshall, S. (2012). Measuring cognitive workload across different eye tracking hardware platforms. In *Proceedings of the symposium on eye tracking research and applications* (pp. 161–164).

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of neuroscience methods*, *190*(1-5), 80.

Crundall, D., Bains, M., Chapman, P., & Underwood, G. (2005). Regulating conversation during driving: A problem for mobile telephones? *Transportation Research Part F*, *8(3)*, 197-211.

Demberg, V. (2013). Pupillometry: the index of cognitive activity in a dual-task study. In *Proceedings of the 35th annual meeting of the cognitive science society (cogsci-13).*

Demberg, V., Kiagia, E., & Sayeed, A. (2013). Language and cognitive load in a dual task environment. In *Proceedings of the 35th annual meeting of the cognitive science society (cogsci-13).*

Drews, F. A., Pasupathi, M., & Strayer, D. L. (2004). Passenger and cell-phone conversations in simulated driving. In *Proceedings of the human factors and ergonomics society 48th annual meeting.*

Drews, F. A., Pasupathi, M., & Strayer, D. L. (2008). Passenger and cell phone conversations in simulated driving. *Journal of Experimental Psychology: Applied*, *14*(4), 392.

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

Jagacinski, & Flach. (2003). *Control theory for humans: Quantitative approaches to modeling performance.* CRC.

Just, M. A., Carpenter, P. A., & Miyake, A. (2003). Neuroindices of cognitive workload: neuroimaging, pupillometric and event-related potential studies of brain work. *Theoretical Issues in Ergonomics Science*, *4*(1-2), 56–88.

Just, M. A., Keller, T. A., & Cynkar, J. (2008). A decrease in brain activation associated with driving when listening to someone speak. *Brain research*, *1205*, 70—80.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*(5), 580 - 602.

Kubose, T. T., Bock, K., Dell, G. S., Garney, S. M., Kramer, A. F., & Mayhugh, J. (2006). The effects of speech production and speech comprehension on simulated driving performance. *Applied cognitive psychology*, *20*(1), 43–63.

Laeng, B., Sirois, S., & Gredebäck, G. (2010). 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 *Automotive user interfaces and interactive vehicular applications* (pp. 88–91).

Marshall, S. (2002). The Index of Cognitive Activity: measuring cognitive workload. In *Proceedings of the 7th conference on human factors and power plants, 2002.*

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

Mehler, B., Reimer, B., Coughlin, J., & Dusek, J. (2009). Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record: Journal of the Transportation Research Board*, *2138*(-1), 6–12.

Mehler, J. (1963). Some effects of grammatical transformations on the recall of english sentences. *Journal of Verbal Learning and Verbal Behavior*, *2*(4), 346–351.

Palinko, O., Kun, A., Shyrokov, A., & Heeman, P. (2010). Estimating cognitive load using remote eye tracking in a driving simulator. In *ETRA.*

Richstone, L., Schwartz, M., Seideman, C., Cadeddu, J., Marshall, S., & Kavoussi, L. (2010). Eye metrics as an objective assessment of surgical skill. *Annals of surgery*, *252*(1).

Schröder, M., Charfuelan, M., Pammi, S., & Türk, O. (2008). The MARY TTS entry in the Blizzard Challenge 2008. In *Proc. blizzard challenge.*

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.

Shi, Y., Ruiz, N., Taib, R., Choi, E., & Chen, F. (2007). Galvanic skin response (gsr) as an index of cognitive load. In *CHI '07 extended abstracts on human factors in computing systems* (pp. 2651–2656).

Son, J., & Park, M. (2011). Estimating cognitive load complexity using performance and physiological data in a driving simulator. In *AutomotiveUI, 2011, adjunct proceedings.*

Strayer, D., Drews, F., & Johnston, W. (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of experimental psychology: Applied*, *9*(1), 23.

Villing, J. (2009). In-vehicle dialogue management - towards distinguishing between different types of workload. In *Proceedings of SimPE, fourth workshop on speech in mobile and pervasive environments.*

Wickens, C. D. (2008). Multiple resources and mental workload. *Human factors*, *50*(3), 449-455.

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

Young, K., Regan, M., & Hammer, M. (2007). Driver distraction: A review of the literature. *Distracted driving. Sydney, NSW: Australasian College of Road Safety*, 379–405.