Measuring cognitive load

Vera Demberg

Saarland University

ESSLLI, August 20th, 2014

Measuring Cognitive Load

Motivations:

- online measurement for online adaptation
- measurement during development of dialog system

Requirements:

- sensitive to linguistic manipulations
- allows us to measure the effect of a single difficult linguistic events, as opposed to a whole block of difficult events
- can separate the effect of difficulty in language and difficulty in driving tasks

Measuring Cognitive Load

Motivations:

- online measurement for online adaptation
- measurement during development of dialog system

Requirements:

- sensitive to linguistic manipulations
- allows us to measure the effect of a single difficult linguistic events, as opposed to a whole block of difficult events
- can separate the effect of difficulty in language and difficulty in driving tasks

! I want you to think about in how far the methods we will review are suitable to our requirements!

Language and realistic dual tasking

Measures of Cognitive Load:

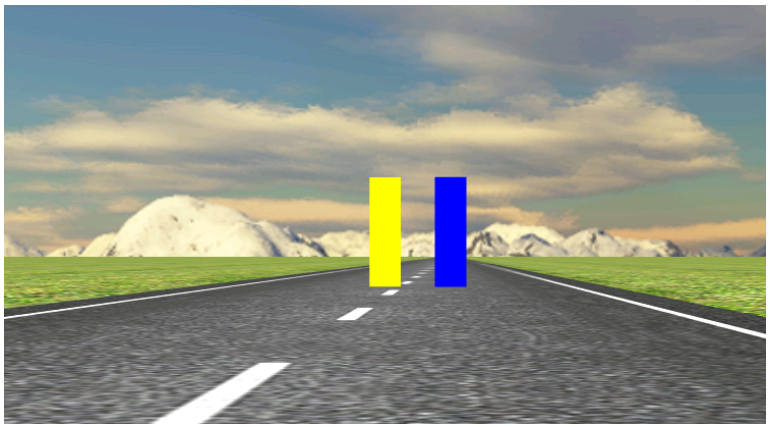
- Primary Task Performance (Driving)
- Steering Wheel Reversal
- Eye-Gaze, Blink Rate
- Skin Conductance, Heart Rate
- Pupillometry
- EEG / ERPs

Table of Contents

Driving: the ConTRe task

Driving task:

ConTRe (Mahr et al., 2012)

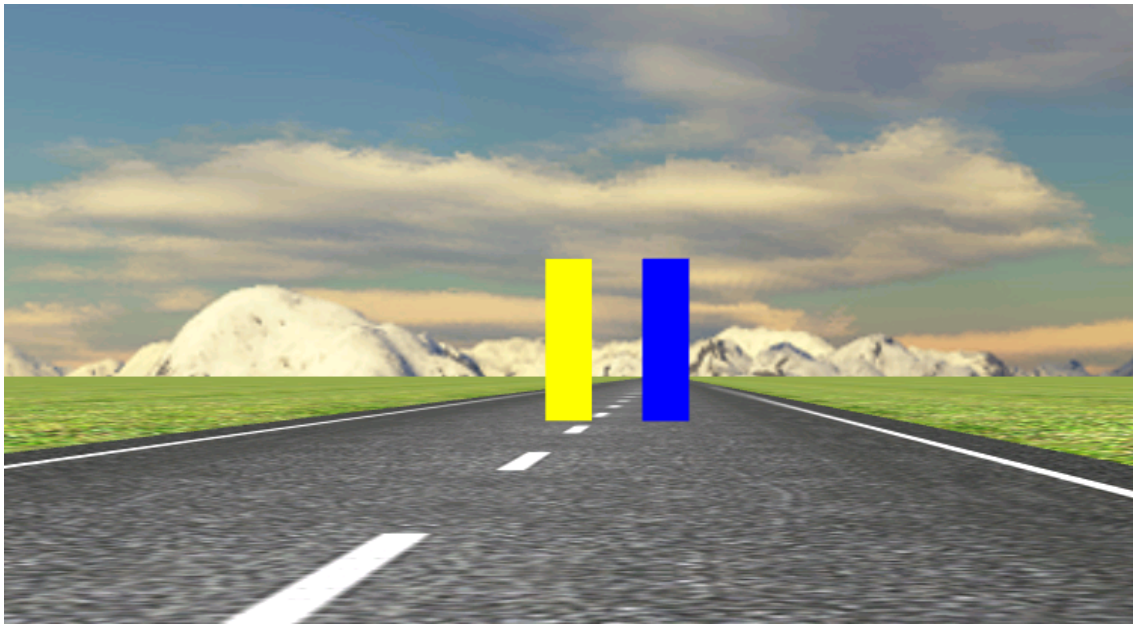


- 24 participants aged 20-34

Training Driving	2 min
Training Driving + language	1 min
Break	
Driving	2 min
Driving + language (10 items)	4 min
Break	
Driving	2 min
Driving + language (10 items)	4 min
Break	
Driving	2 min
Driving + language (10 items)	4 min
Break	
Driving	2 min
Driving + language (10 items)	4 min

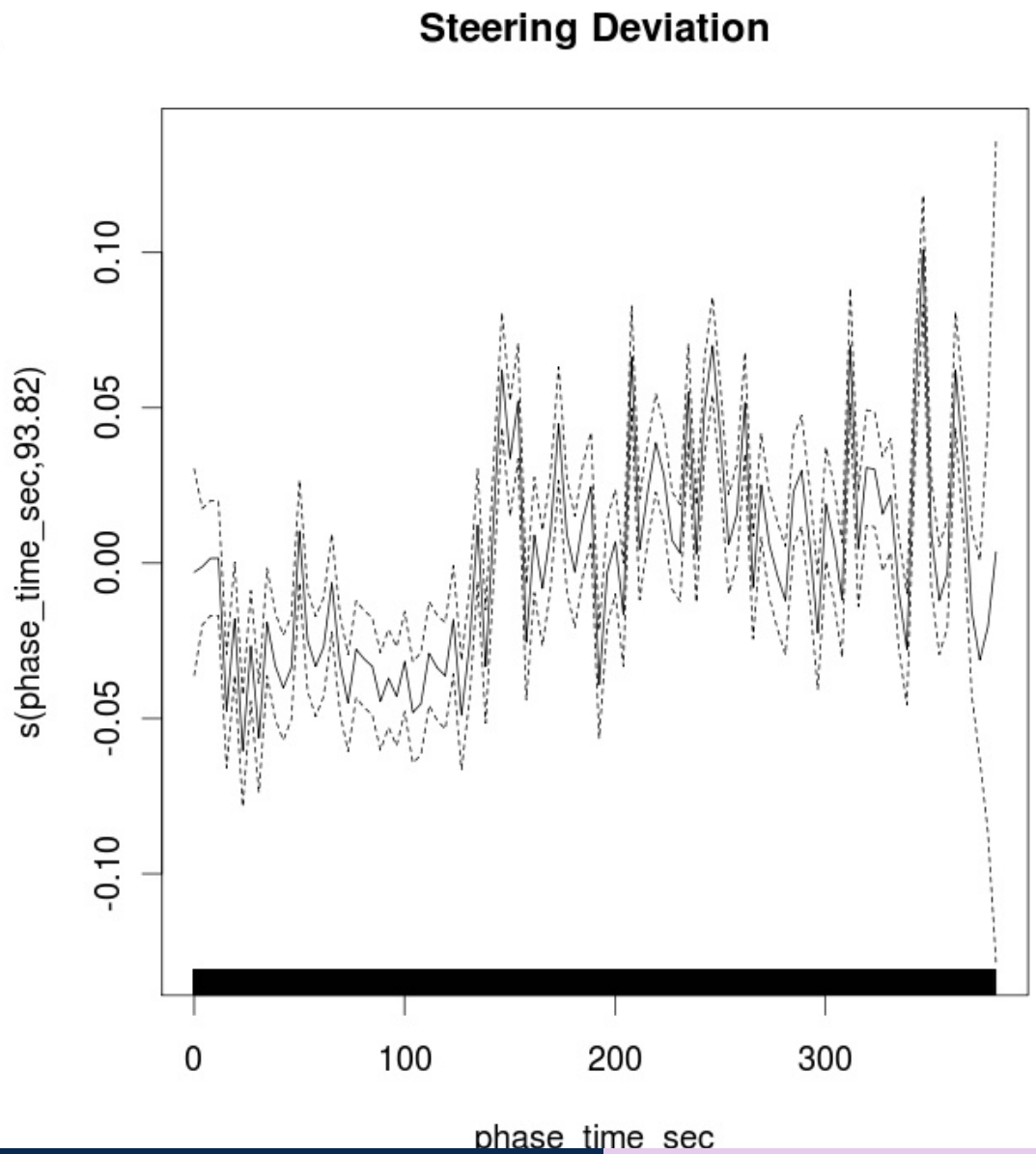
easy: target = 1m/s, controllable = 2m/s
difficult: target = 2.5m/s, controllable = 4m/s

Primary task performance



- measure how much performance on primary task degrades in presence of second task.
- task: keep both bars overlapped
- performance measure: distance between the two tasks at each moment in time.

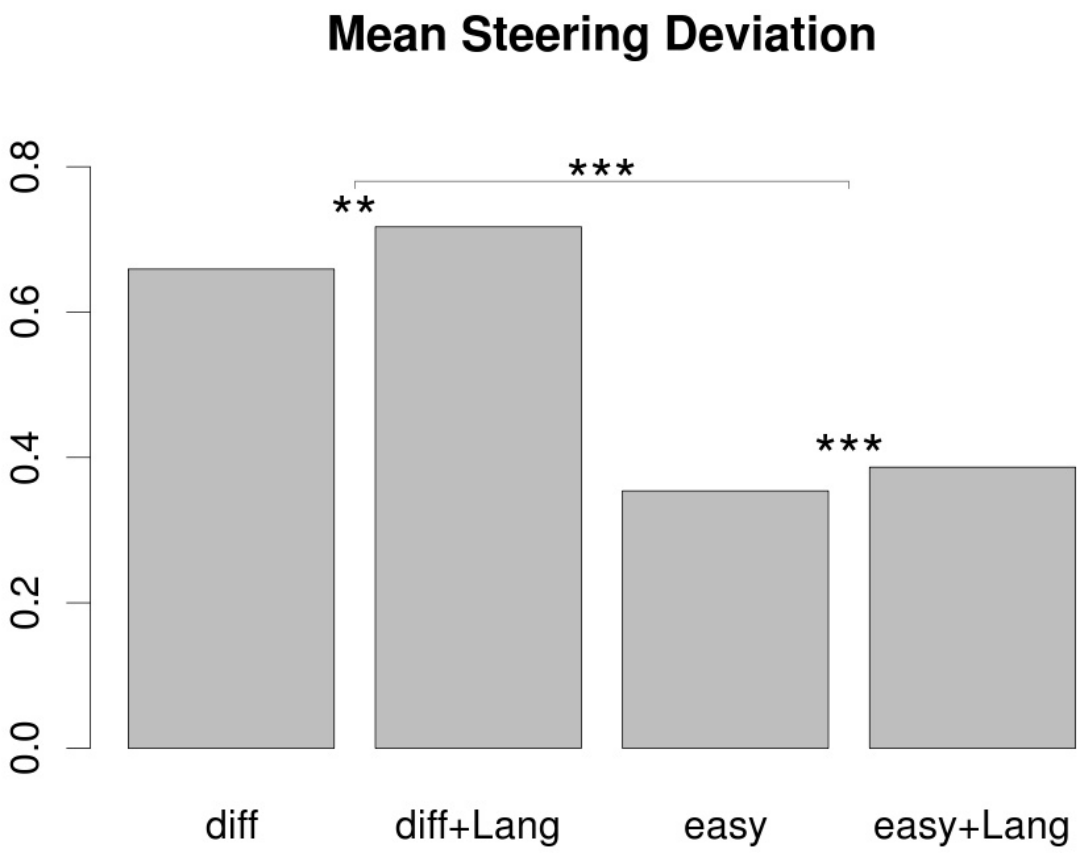
Steering Deviation



- Steering deviation (distance between the bars) much higher in dual task condition (after 120 sec).
- Variance also higher $\rightarrow$ possible sign of slower reactions

Steering Deviation

Effect of language is even stronger in difficult driving condition.



(Interaction also significant.)

Discussion Driving Performance

Driving performance as a measure of cognitive load:

- directly relevant to safety question
- sensitive to language task

Discussion Driving Performance

Driving performance as a measure of cognitive load:

- directly relevant to safety question
- sensitive to language task

Does it satisfy our requirements?

Discussion Driving Performance

Driving performance as a measure of cognitive load:

- directly relevant to safety question
- sensitive to language task

Does it satisfy our requirements?

Discussion: Realistic vs. controlled task?

- depends on what we're after
- advantages of controlled task:
  - continuous
  - fine-grained,
  - well-defined goals
  - easy to control (difficulty levels well-defined)
  - direct correspondence to well-understood standard tasks in psychology

Let's try it out!

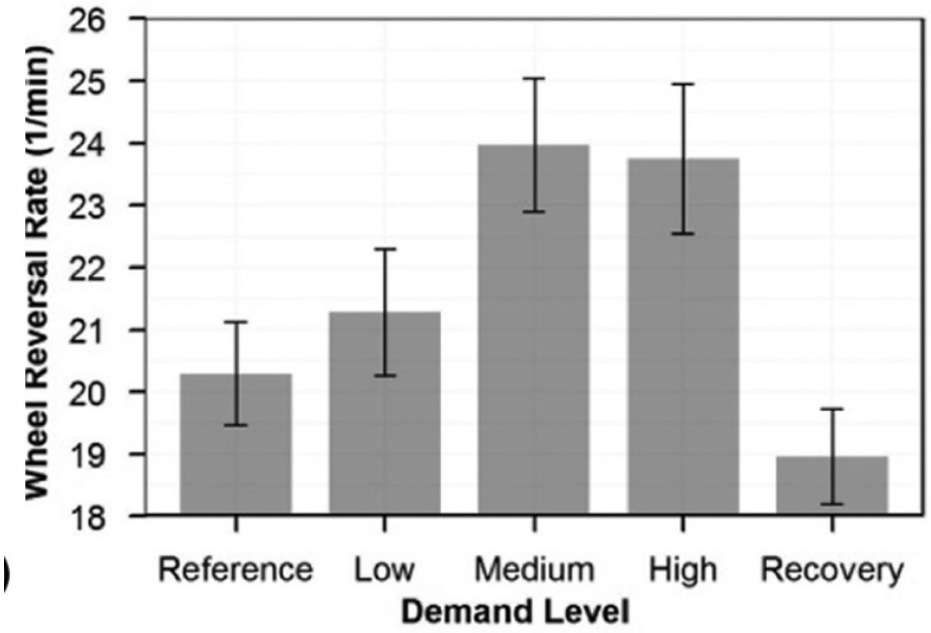
In groups of 2 people:

- each group needs one laptop that has the driving task installed.
- pick up a strip for the n-back task
- one person is the experimenter, the other one the driver
  - for each of the following, restart the driving simulation.
  - 2 min of getting used to the driving simulator (YOURNUMBER-train)
  - do the n-back task (YOURNUMBER-nback)
  - 2 min of driver telling the experimenter about their summer holidays (YOURNUMBER-story)
  - 2 min of driving only (YOURNUMBER-drive)
- then swap roles
- be back here in ca. 20 min.
- send the database files to vera@coli.uni-saarland.de

Table of Contents

Steering Wheel Reversal

- **how many steering movements** does the participant make?
- more steering movements = higher cognitive load
- well-established measure (MacDonal and Hoffmann, 1980), still used in automotive UI community today



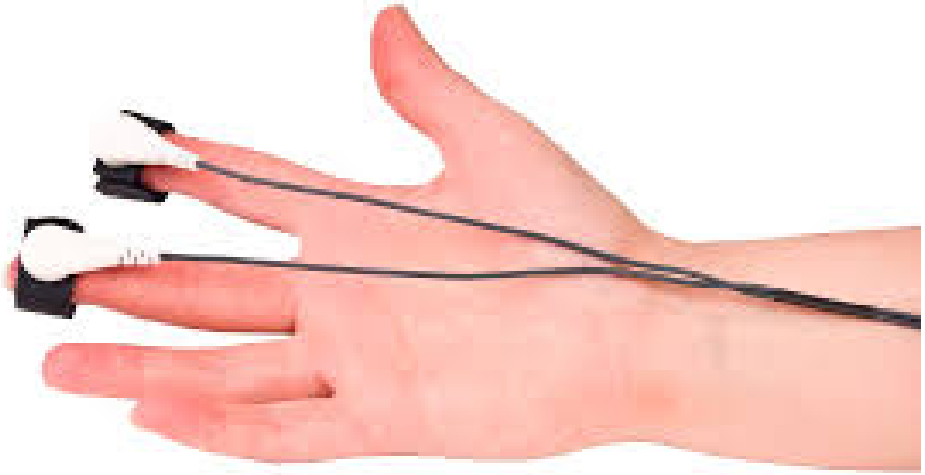
(graph from Reimer et al., 2012)

Suitability of Wheel Reversal

- good for block designs
- not well-suited for fine-grained differences such as subj vs. obj relative clause: too few events, not dynamic enough.

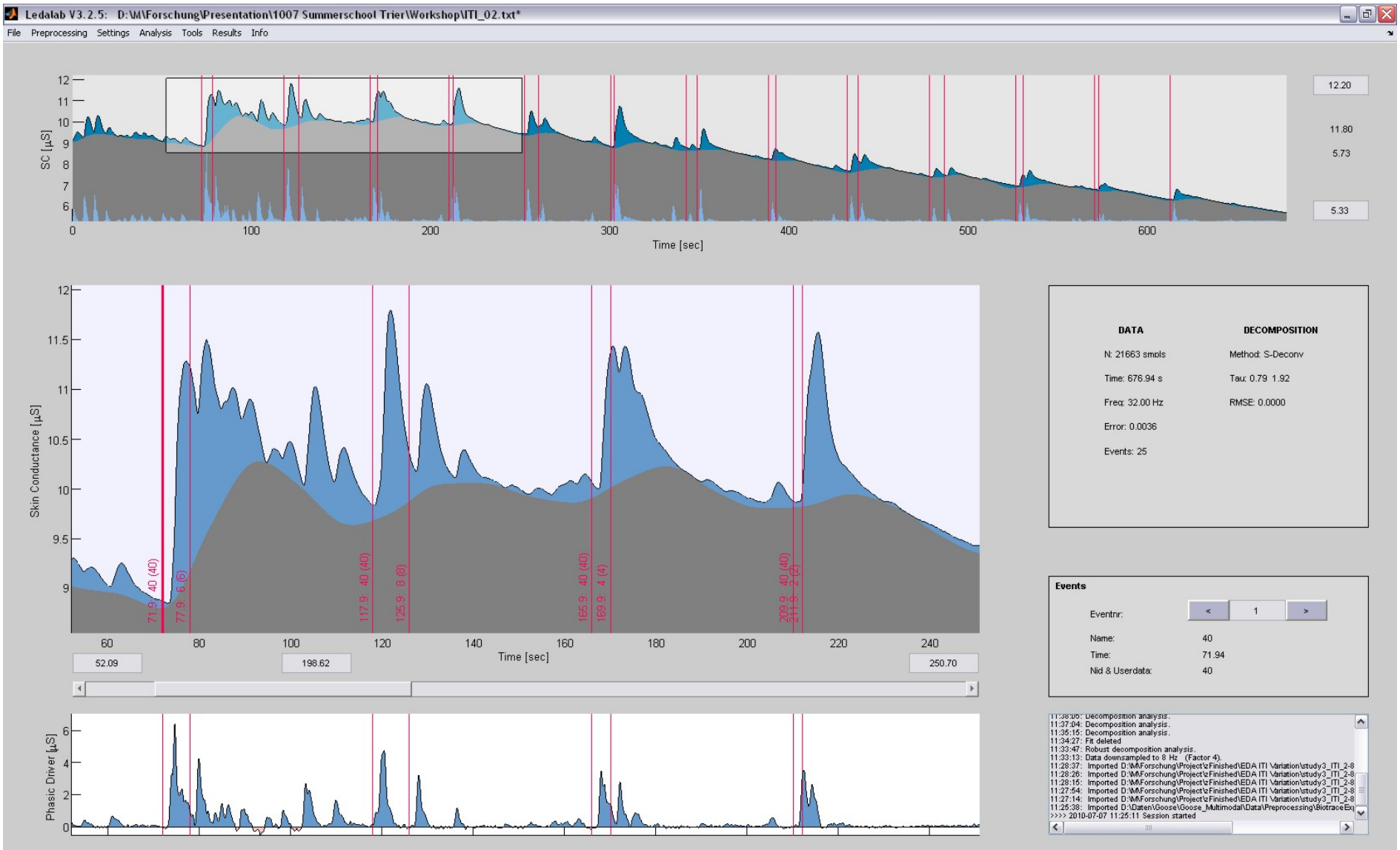
Table of Contents

Skin Conductance



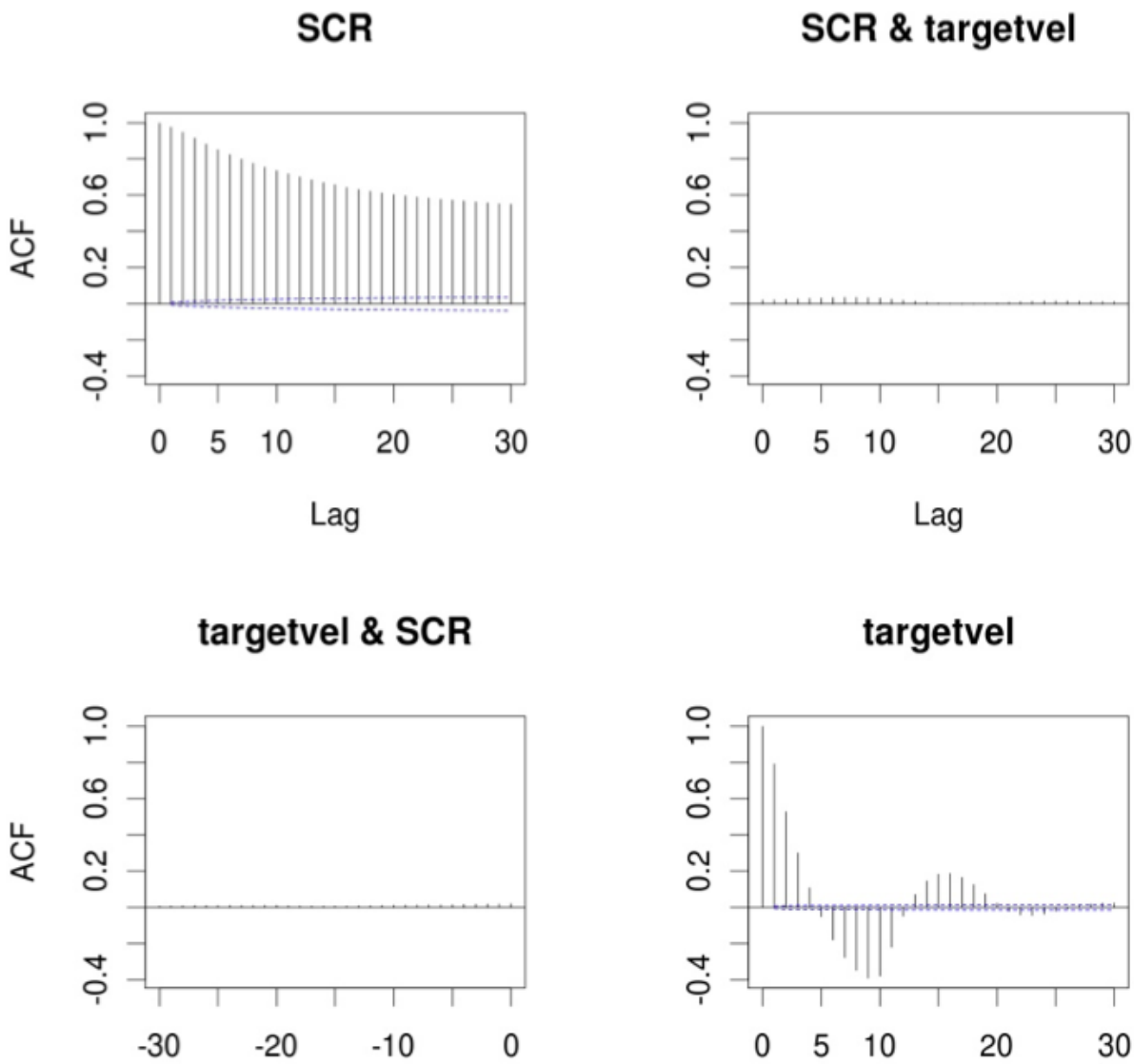
- measuring the electrical conductance of the skin
- electrical conductance varies depending on the amount of sweat-induced moisture on the skin
- sweat is controlled by the sympathetic nervous system
- $\rightarrow$ indication of psychological or physiological arousal

Skin Conductance Analysis



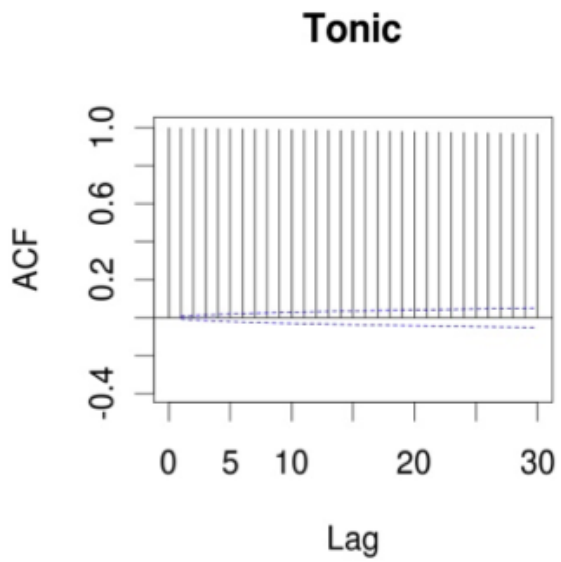
Results: phasic skin conductance vs. steering

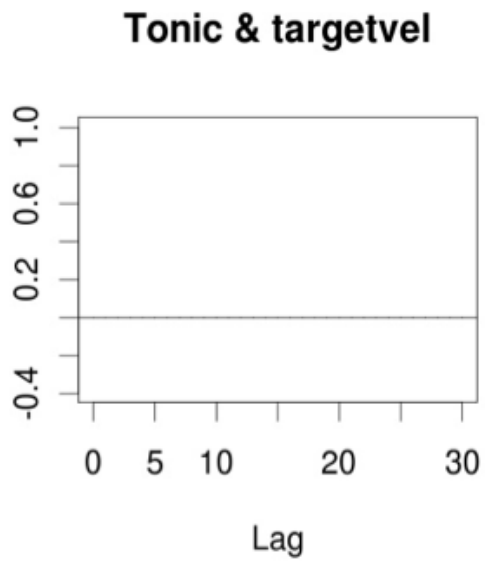
- high auto-correlation even for phasic skin conductance
- no significant cross-correlation to steering task

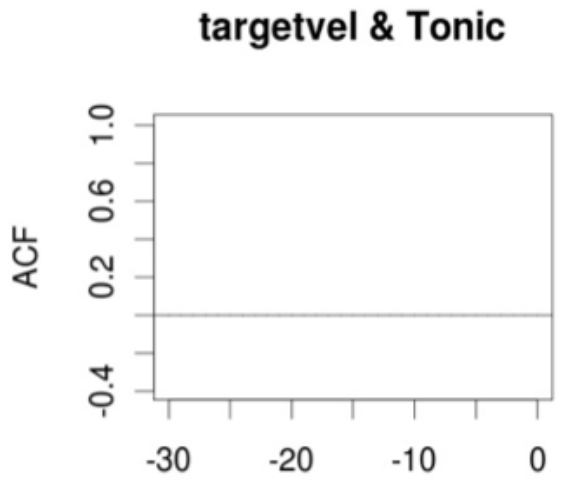


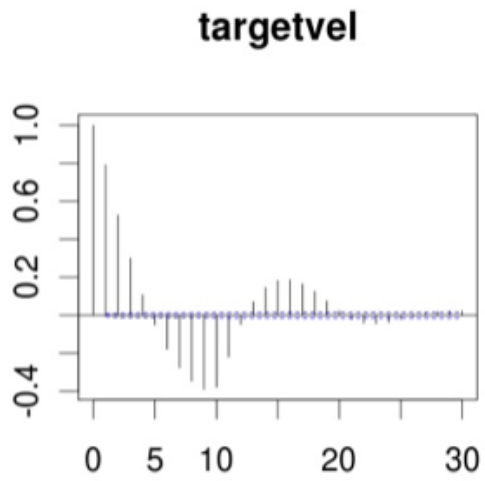
Results: SC tonic component vs. target bar movement

- almost perfect auto-correlation for tonic skin conductance
- no significant cross-correlation with steering task

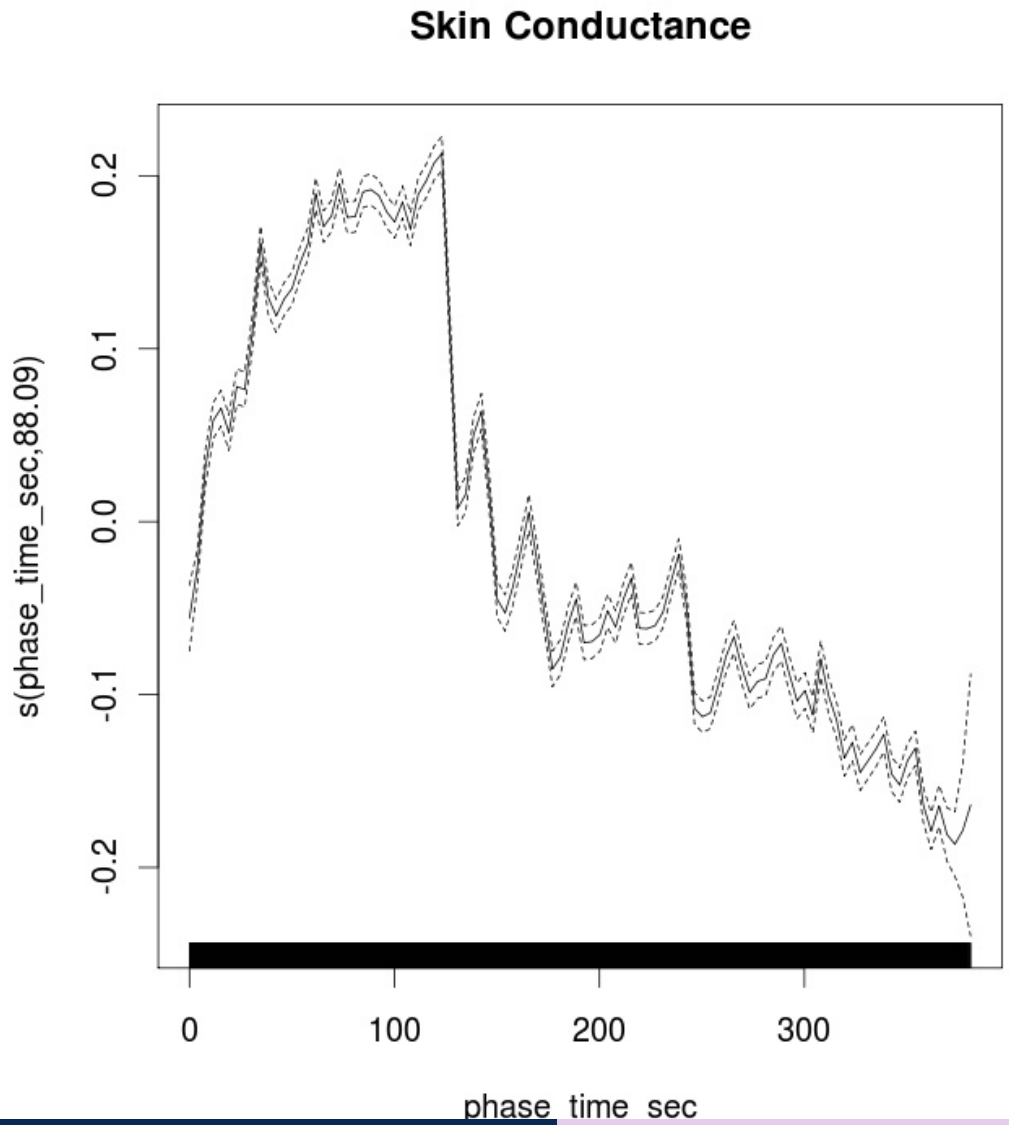








Skin Conductance



- skin conductance on finger LOWER in dual task condition (after 120sec) than in single task condition.
- opposite results if measured on neck?

Results: skin conductance during stimulus

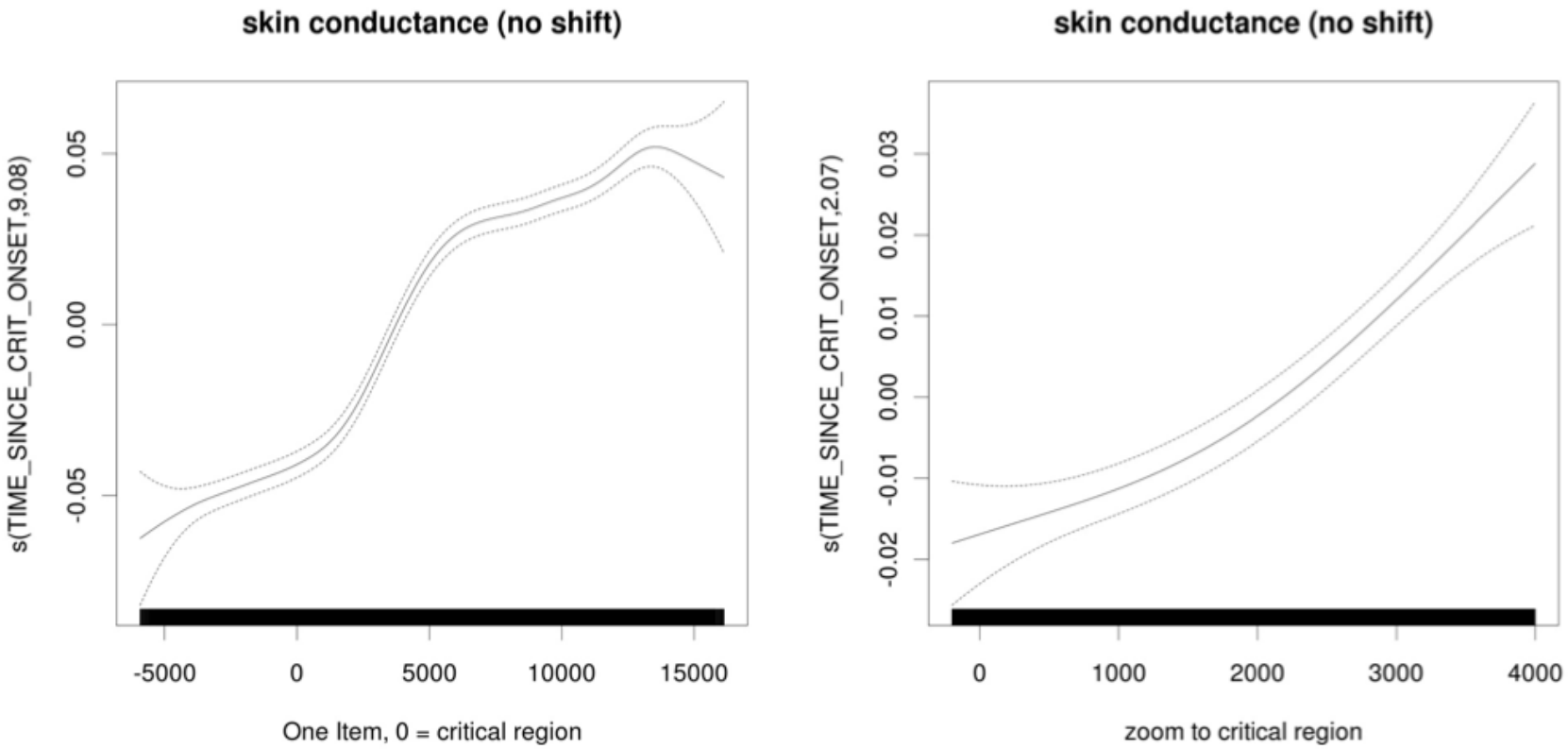


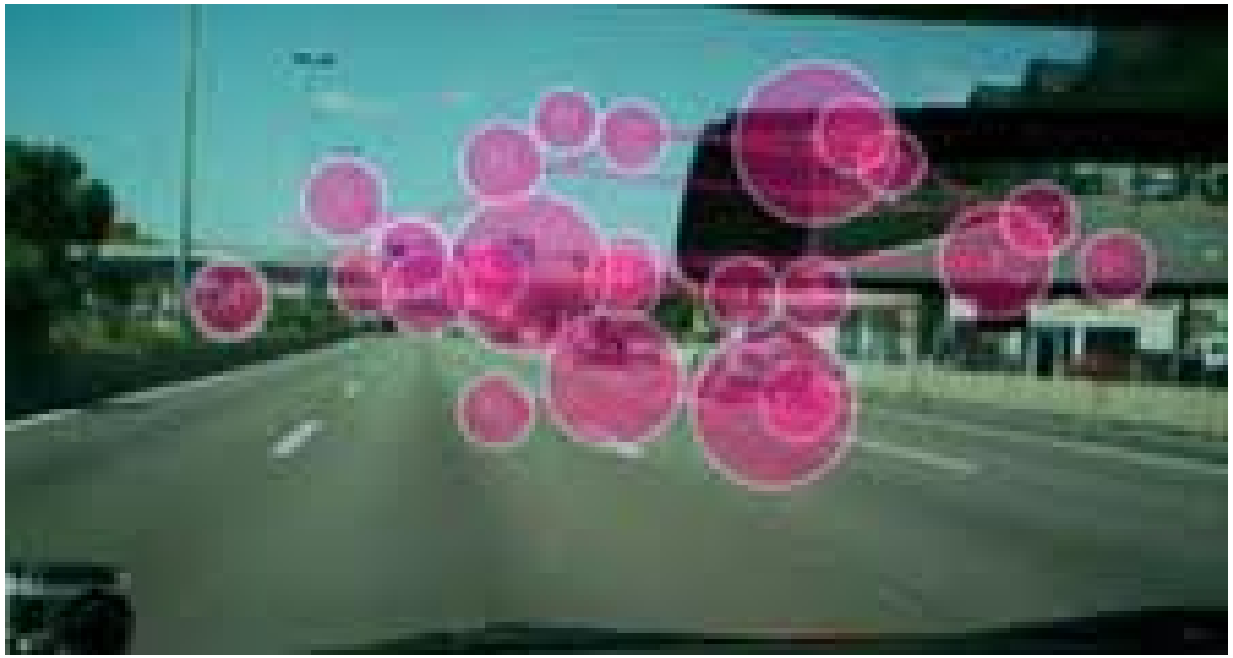
Figure : SCR during time that stimulus is spoken.

Conclusions

How well does skin conductance satisfy our requirements?

Table of Contents

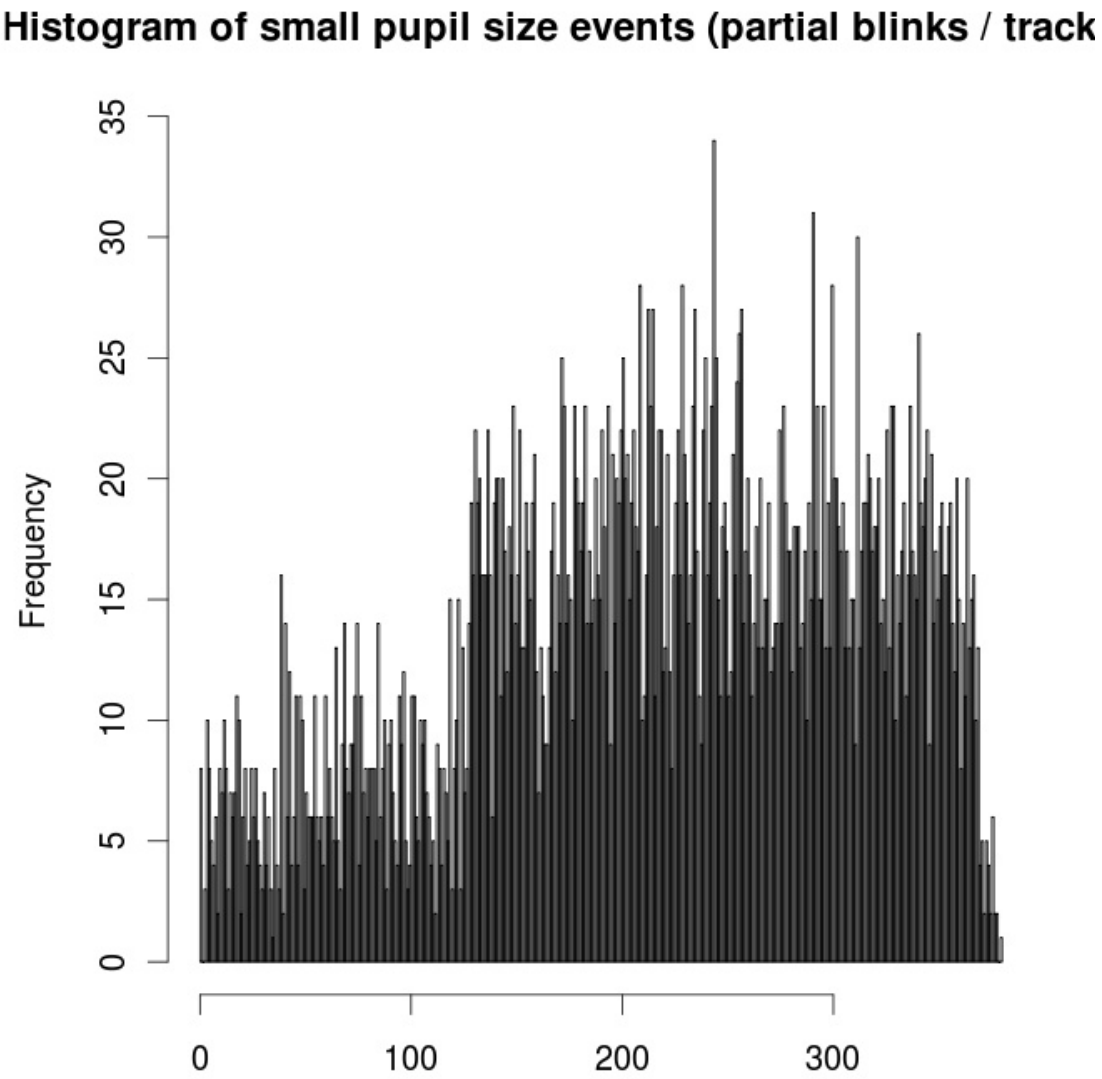
Eye-Gaze



Under cognitive load:

- less scanning
- later recognition (and less visual attention directed to relevant objects)

Blink Rate



- people blink more under high cognitive load!
- to the left: blink histogram from single task (first 120 sec) and dual task (120s till end).

Suitability of Eye-Gaze and Blink Rate?

Suitability of Eye-Gaze and Blink Rate?

as for Steering Wheel Reversal:

- good for block designs
- not well-suited for fine-grained differences such as subj vs. obj relative clause: too few events, not dynamic enough.

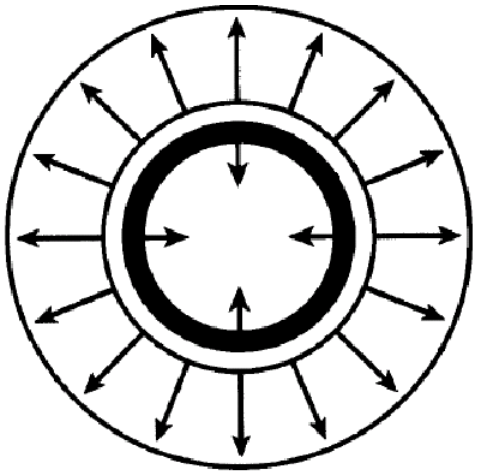
Table of Contents

Pupillometry – What is it and why do we want it?

- Pupillometry = measuring the change of size of the pupil
- Pupil size is not only influenced by luminance but also by emotion, arousal and **cognitive load**.
- Increased cognitive load $\rightarrow$ larger pupil.
- Non-invasive, can be installed in car context.
- First known observations in 1890s, measured since the 60s, recently some new interest also in context of language processing

How does it work?

- two muscles (Dilator Pupillae & Sphincter Pupillae)
- 1mm min – 9mm max diameter (Beatty & Lucero-Wagoner, 2000)
- response delay 200-300ms (so relatively short latency)
- peak after about 1200ms
- pupil dilation change when the muscles are activated or inhibited
- correlates with heart rate and skin conductance.



from: Beatty & Lucero-Wagoner 2000

Background: What does it mean?

- Pupil dilation is strongly correlated with activity in locus caeruleus (LC)
- LC neurons are bilateral and emit the neuro-transmitter norepinephrine (NE)
  (Aston-Jones and Cohen, 2005)
  (Laeng et al., 2012)
- LC-NE system activated by stress
- related to memory retrieval and memory consolidation

Relationship pupil – LC neuron

Relationship between tonic pupil diameter and baseline firing rate of an LC neuron in monkey.

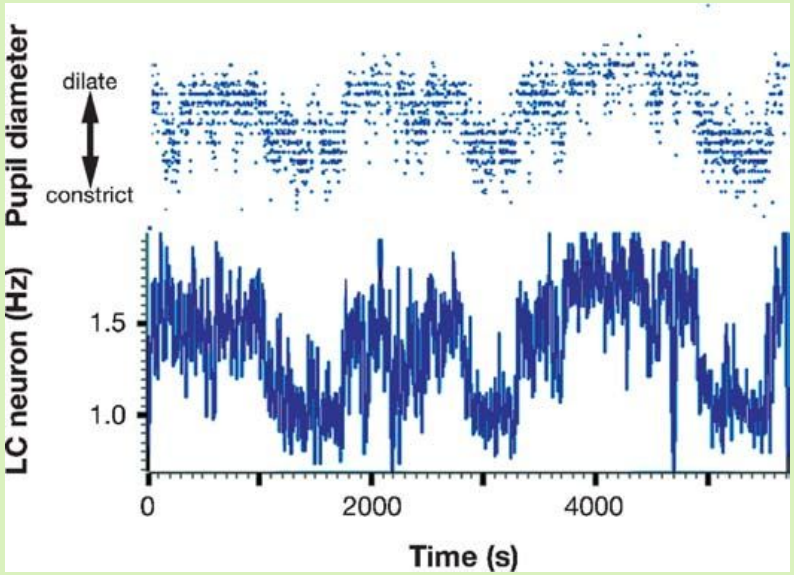
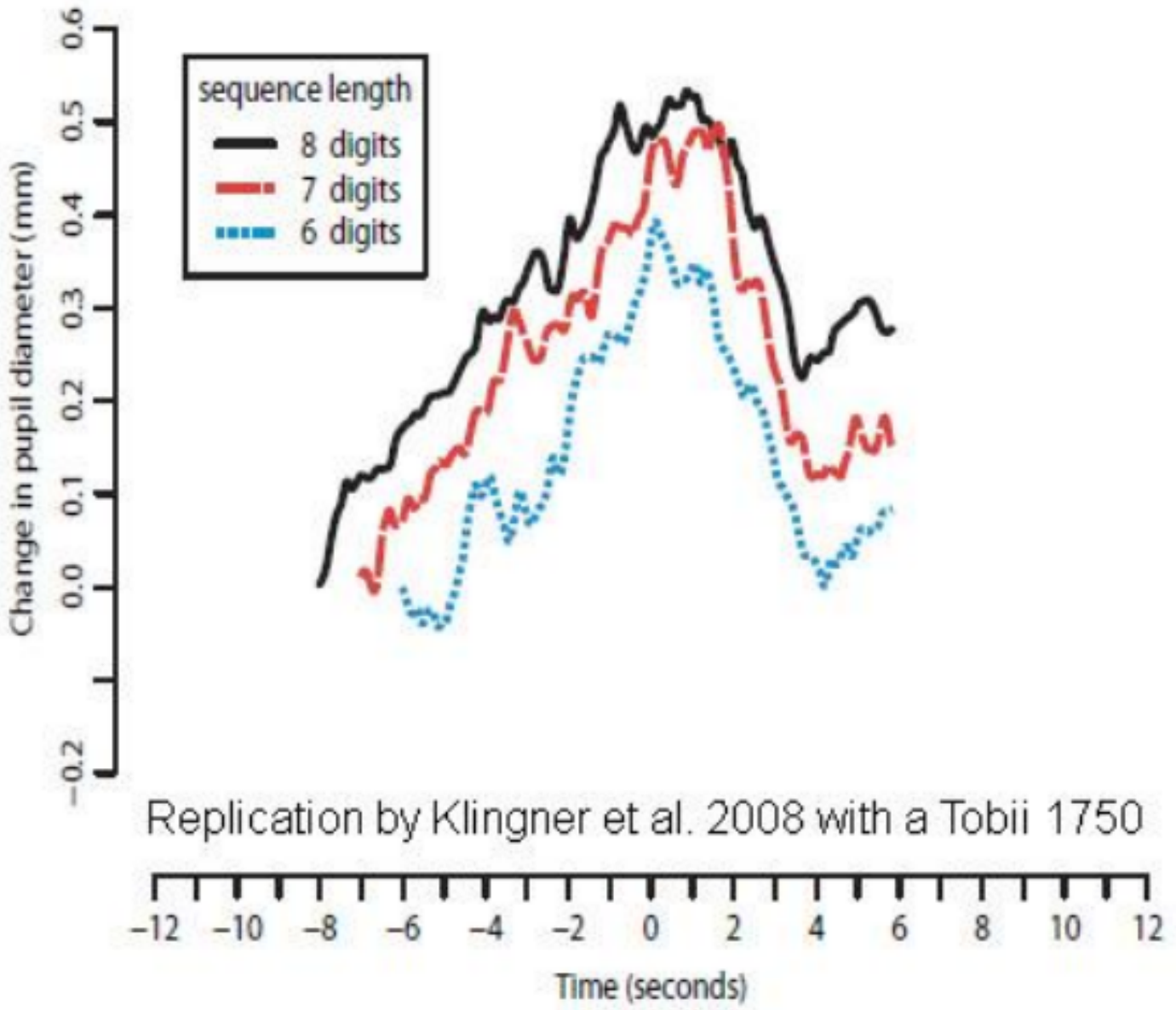


Figure taken from Aston-Jones and Cohen (2005).

Evaluated for different tasks

- arithmetic problems (Hess & Polt 1964)
- digit recall
- memory (Kahnemann & Beatty 1966)
- attention (Beatty, 1982)
- concentration
- inference
- language
  - syntactic complexity (SRC vs. ORC reading task; Just & Carpenter 1993)
  - translation (Hyönä, Tomola & Alaja, 1995)
  - grammaticality violations (Gutierrez & Shapiro 2010)
  - context integration (Engelhardt et al. 2010)

Classic Example: digit recall

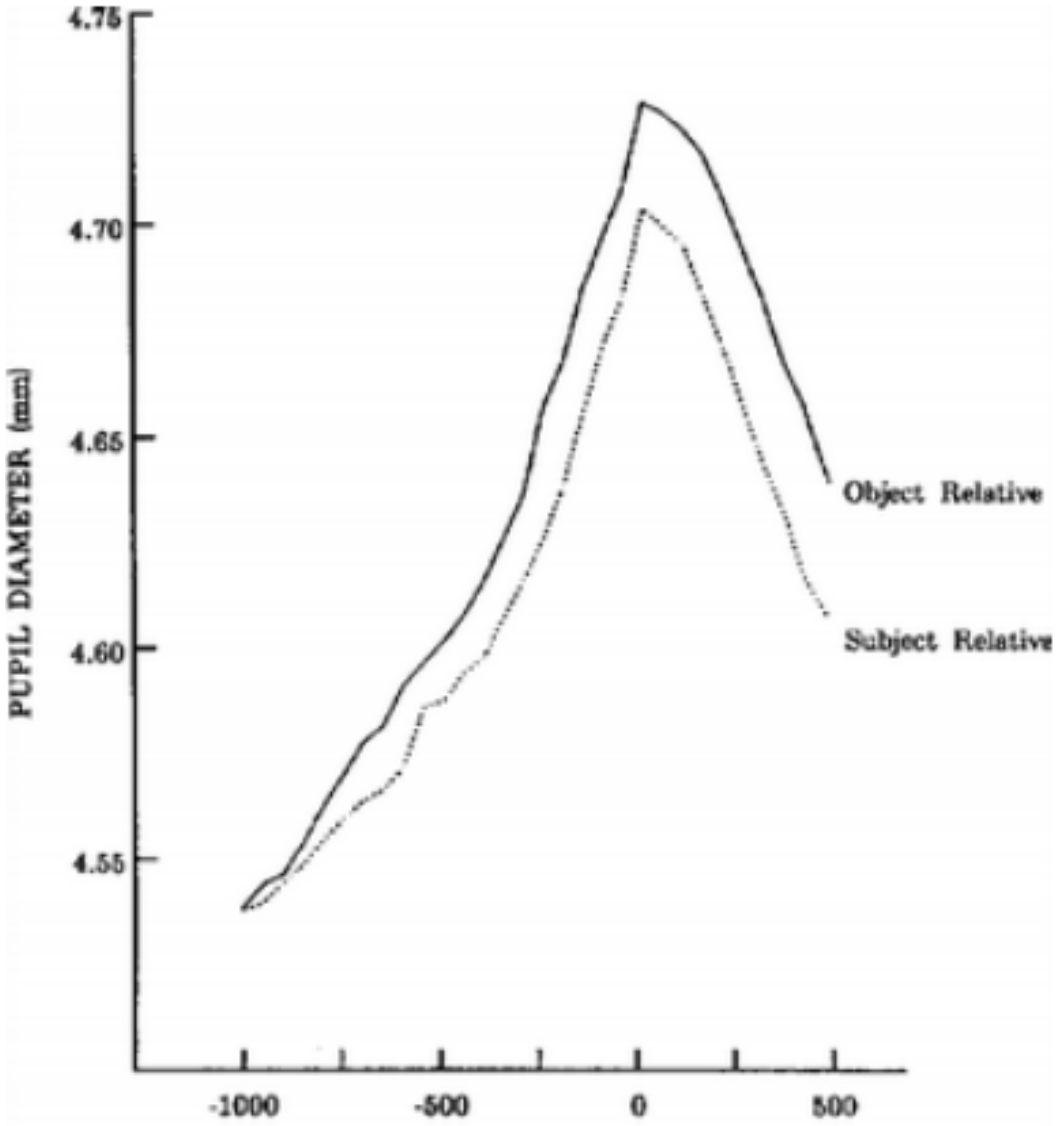


memorization
$\rightarrow$ dilation

recall
$\rightarrow$ constriction

Linguistic Example (Just & Carpenter '93)

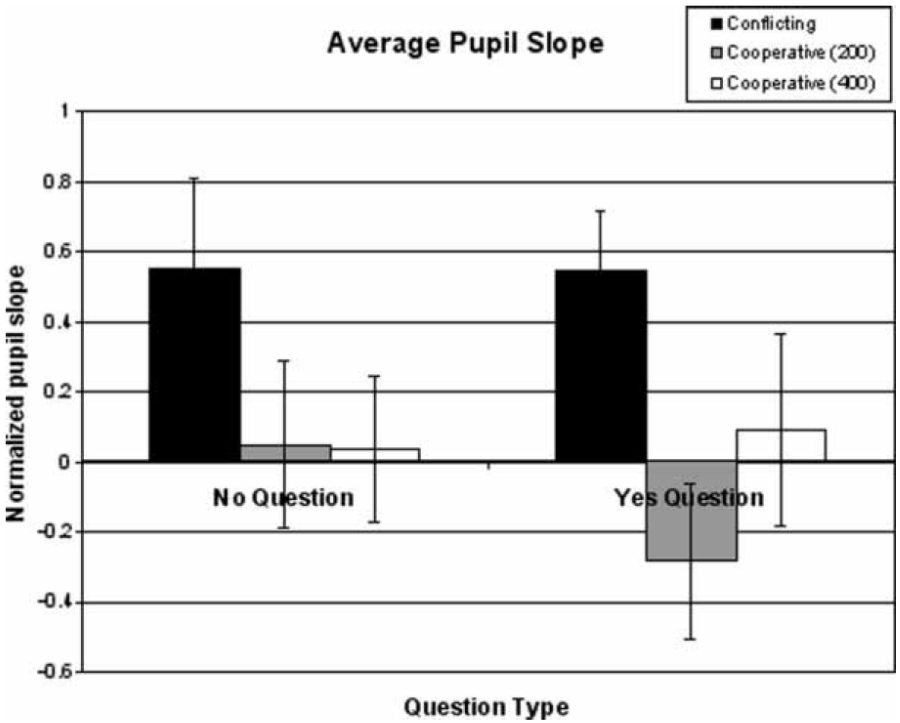
- object relative clauses more difficult than subject relative clauses
- confirmed by pupil dilation: pupil opens wider in ORC



Engelhardt & Ferreira 2010

Experiment about conflicting prosodic / visual context.

- time window of 1.2 sec, 200ms after word onset



"While the woman cleaned the dog that was big and brown stood in the yard."

Conclusions Pupil Size

Conclusions pupil size:

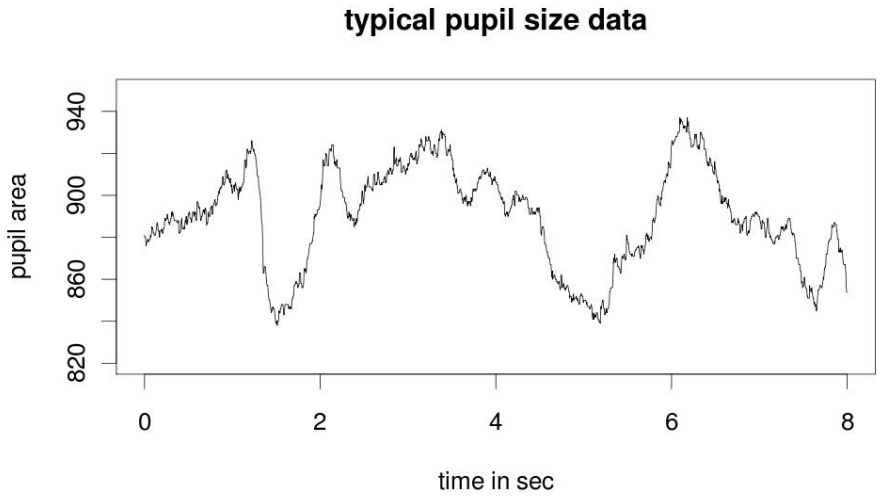
- sensitive to linguistically induced cognitive load
- but: researchers have sometimes not found effects in language studies
- for some tasks not fast enough

In order to work with overall change in pupil size need

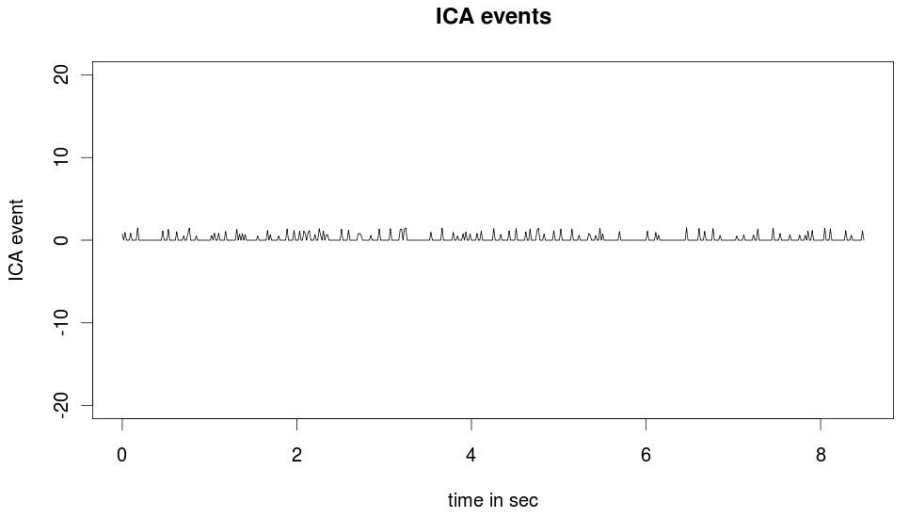
- constant lighting of room
- control for luminance of stimuli (difficult)
- normalize wrt. pupil size
- relatively slow measure compared to EEG, gaze, SPR

Table of Contents

Index of Cognitive Activity (ICA; Marshall, 2000)



↓ *wavelet analysis* ↓



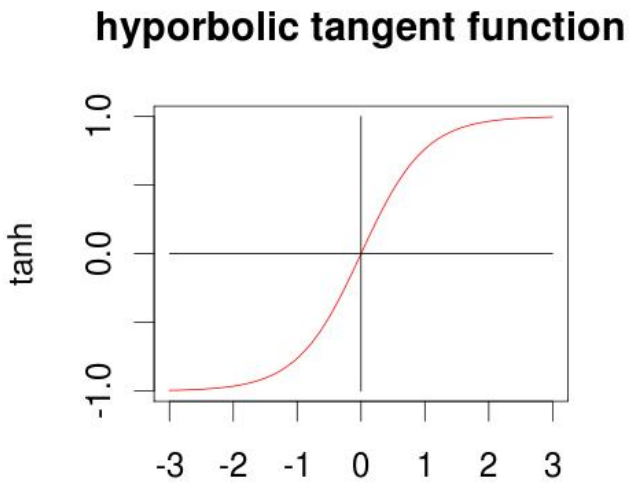
Traditional Pupillometry:
overall pupil size
linked to cognitive load.

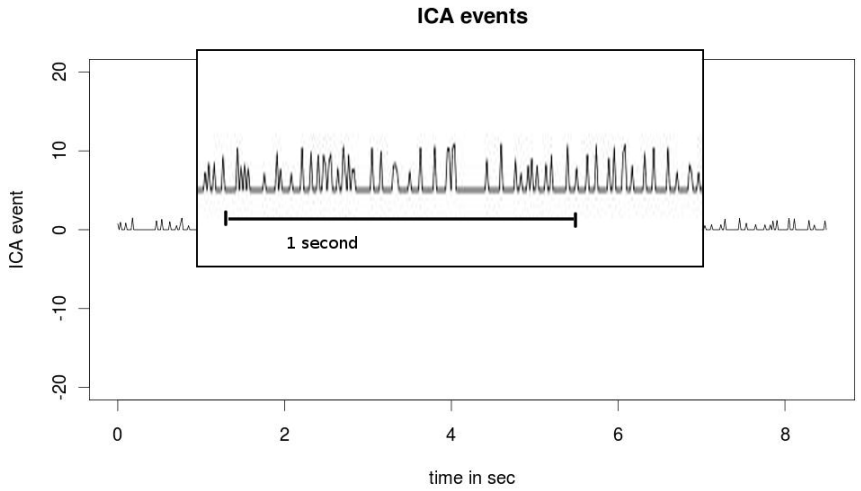
Index of Cognitive Activity:
frequency of rapid dilations
interpreted as sign of cognitive load.

How is the ICA calculated?

Calculation of ICA (Marshall, 2000)

- $\text{ICA} = \tanh(\frac{\#\text{ of rapid dilations per sec}}{30})$

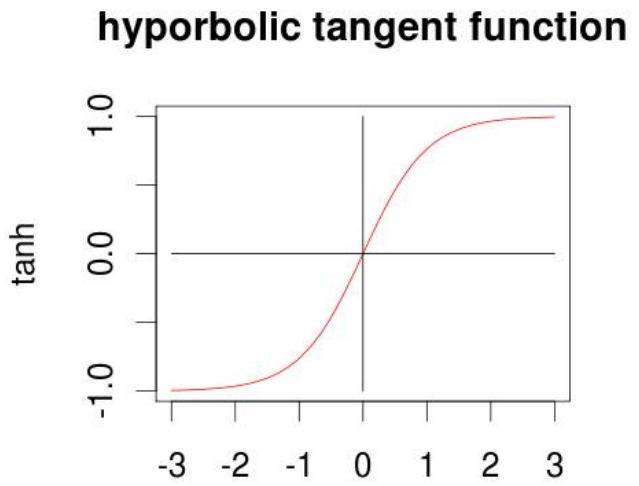


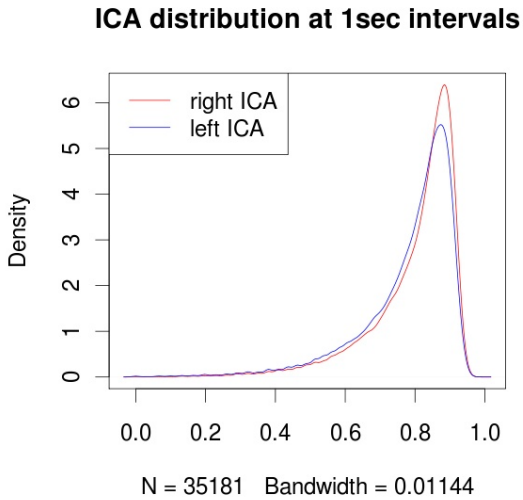


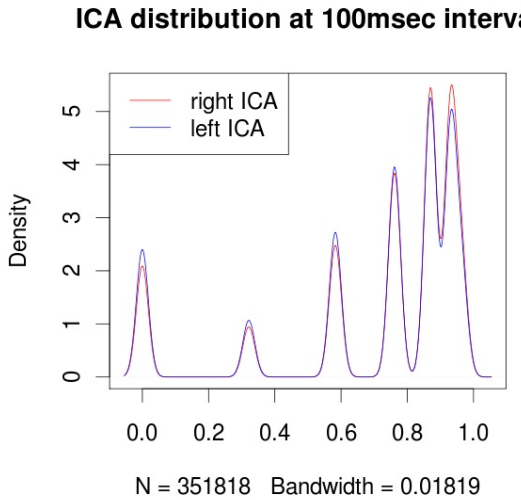
How is the ICA calculated?

Calculation of ICA (Marshall, 2000)

- $\text{ICA} = \tanh(\frac{\#\text{ of rapid dilations per sec}}{30})$



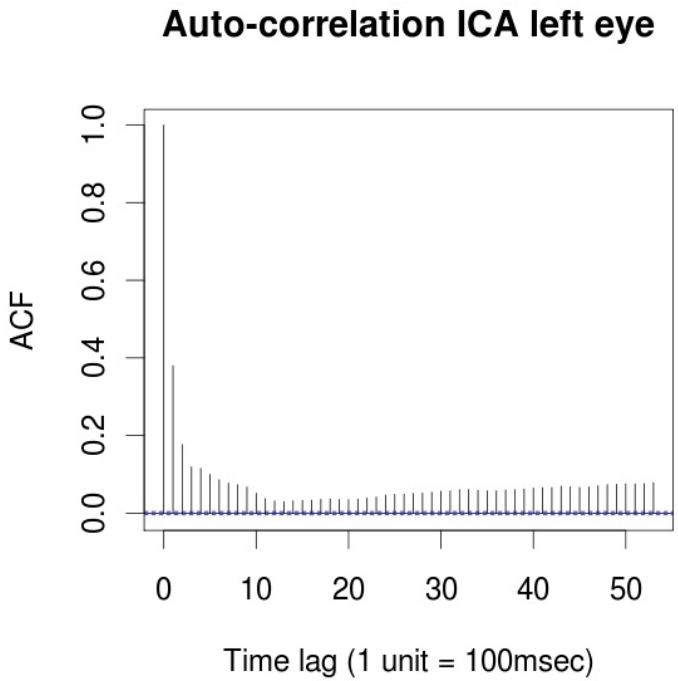


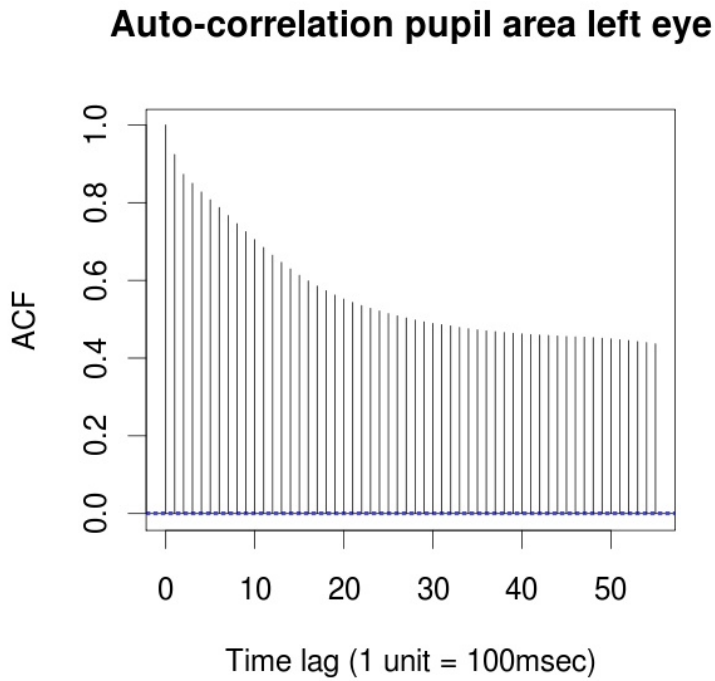


Properties of the ICA

Properties of ICA

- more robust wrt. changes in light / movement
- more dynamic





Overview of Experiments

Can the ICA tell us anything about language processing?

Self-paced reading with eye-tracking

Experiment 1:
subject vs. object relative clauses

Experiment 2:
semantic anomalies

Experiment 3:
gender mismatch

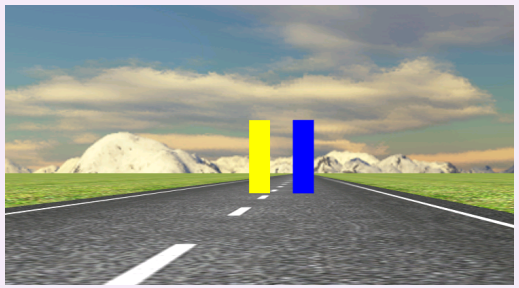
Visual world and ICA

Visual world

causal vs. concessive discourse connectors

Dual task: driving and listening

Simulated Driving



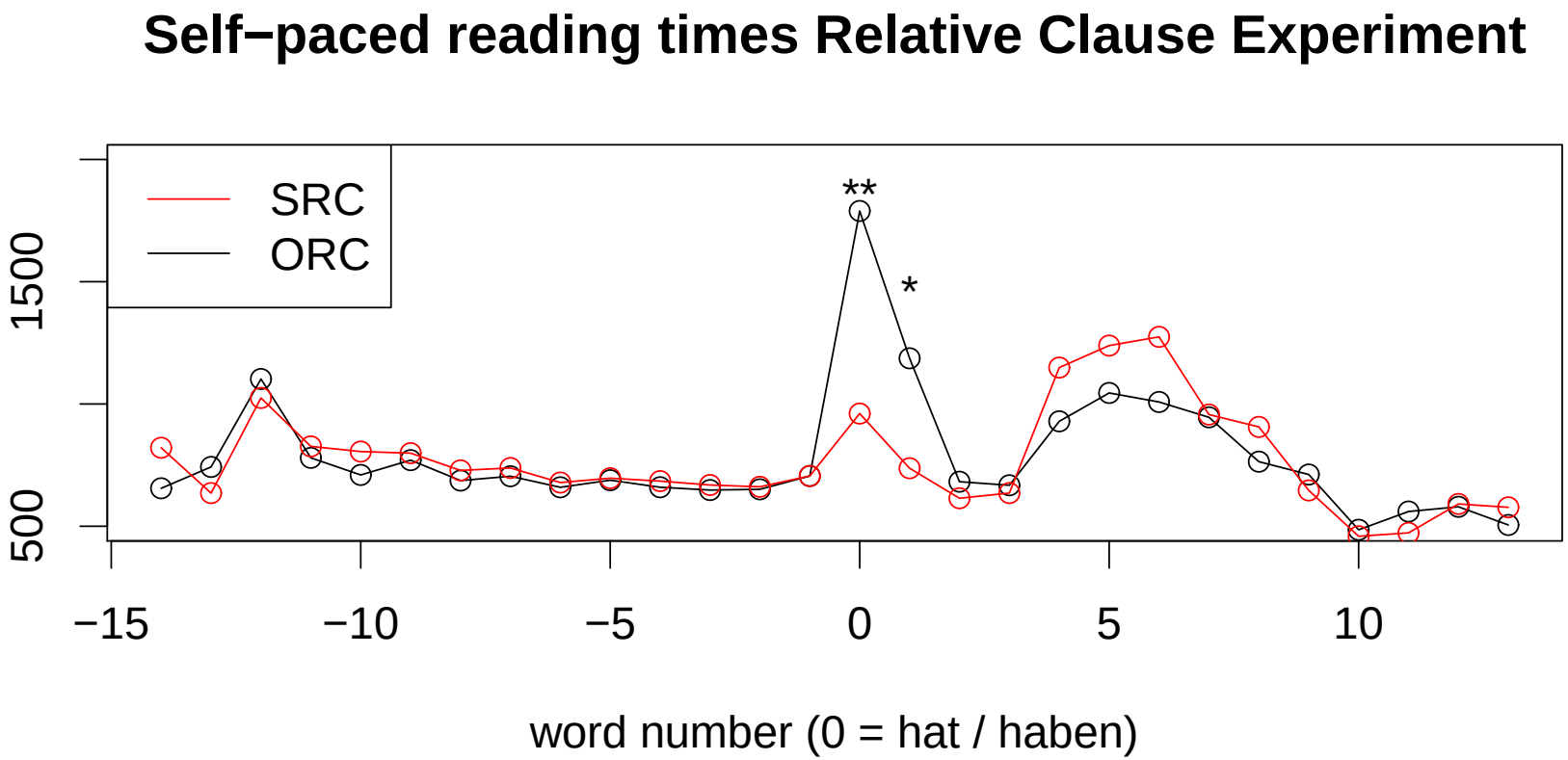
+

subject vs. object relative clauses

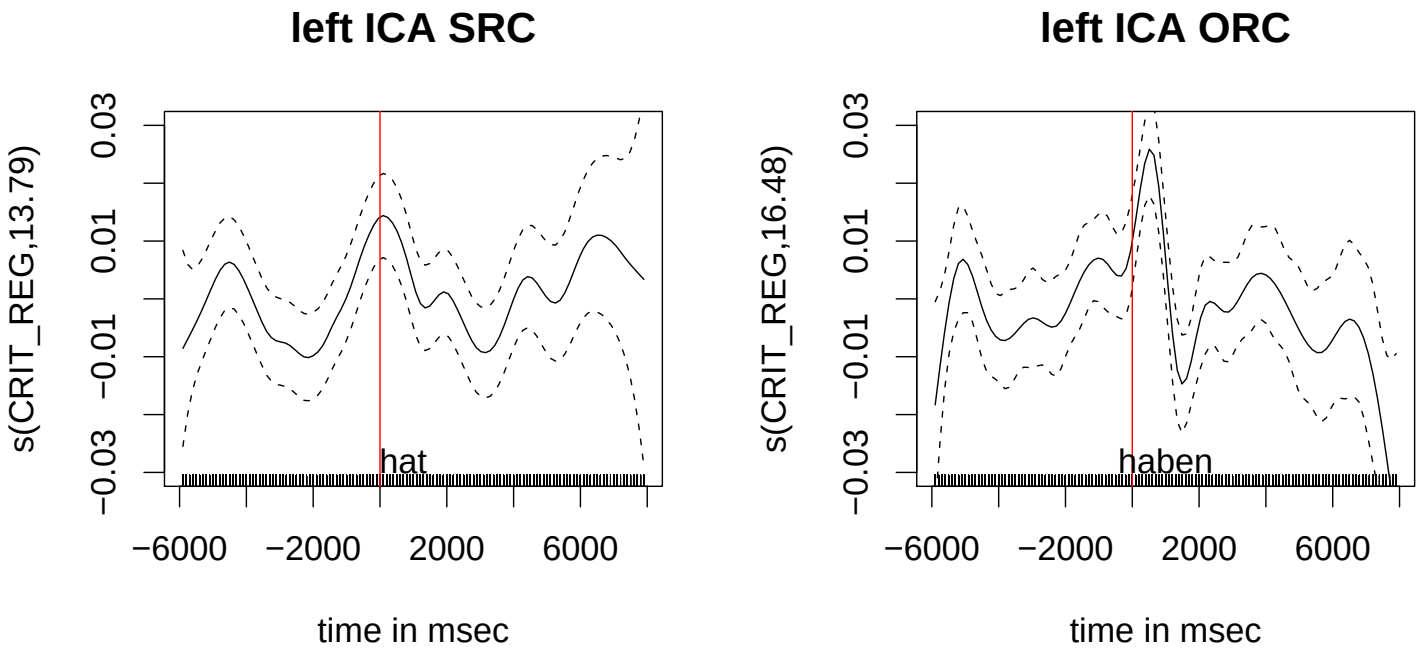
Locally Ambiguous Subject vs. Object Relative Clause

Die Nachbarin, [die$_{sg,n/a}$ einige$_{pl,n/a}$ der Mieter auf Schadensersatz verklagt **hat$_{sg}$ / haben$_{pl}$**$]_{RC}$*, traf sich gestern mit Angelika.*

"The neighbor, [whom some of the tenants sued for damages / who sued some of the tenants for damages]$_{rc}$, met Angelika yesterday."

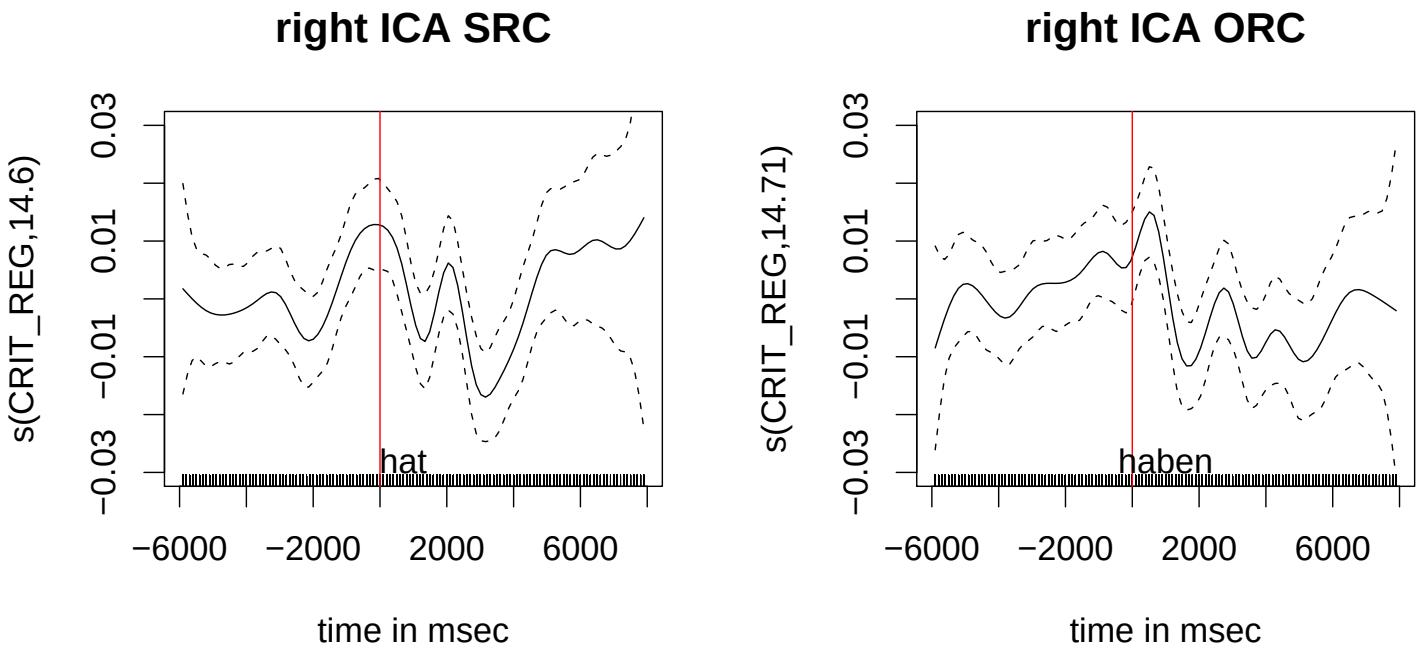


ICA results relative clauses – left eye



left ICA	β	sdev	t val	sign
(Intercept)	0.831	0.008	100.60	***
Subject RC	-0.013	0.006	-2.16	*

ICA results relative clauses – right eye



right ICA	β	sdev	t val	sign
(Intercept)	0.821	0.009	90.58	***
Subject RC	-0.008	0.005	-1.49	

Summary Relative Clauses

Difficulty effect on target region found for

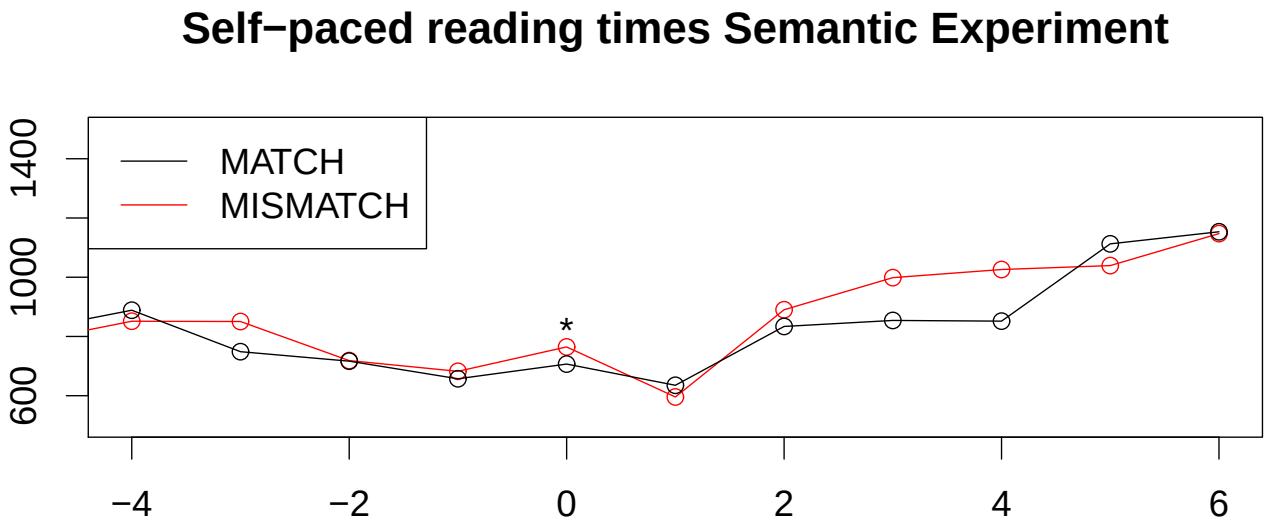
- self-paced reading
- ICA left eye (peaks 1s after stimulus)
- pupil dilation speed for both eyes (during first 2secs)

Semantic Anomalies

Max singt / arbeitet als **Rechtsanwalt** *bei einer großen Firma.*

"Max is singing / working as a **lawyer** for a large company."

Results:

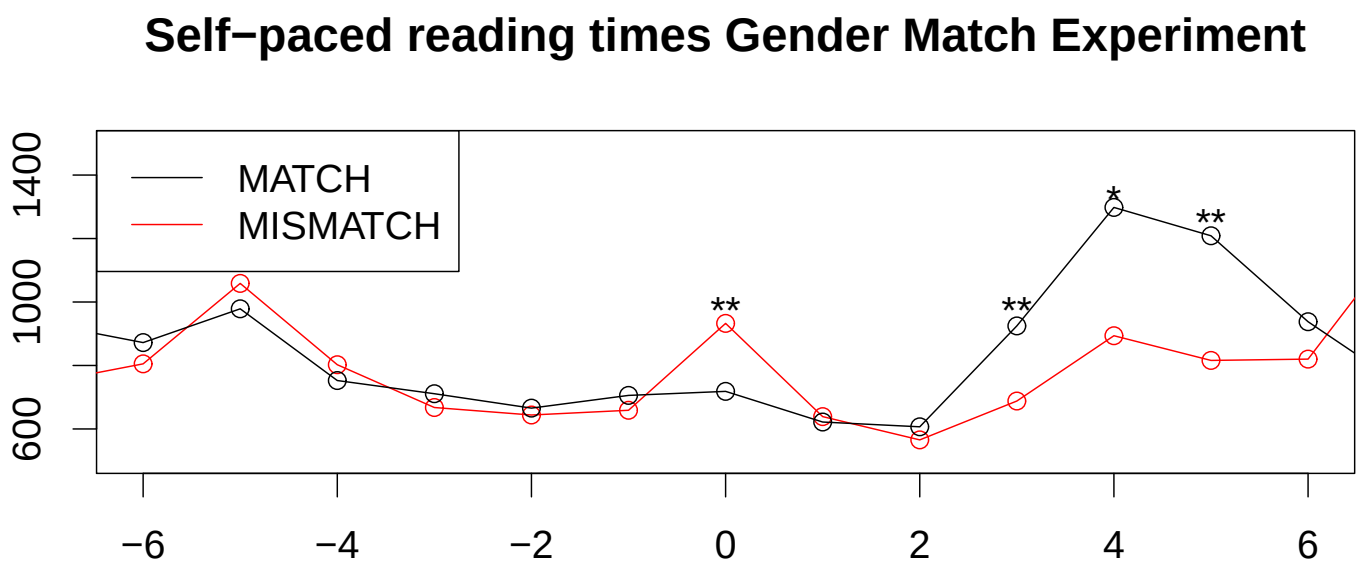


- ICA effect significant on both eyes, effect larger on left eye.
- Significant effect of faster pupil dilation both eyes

Gender Mismatch

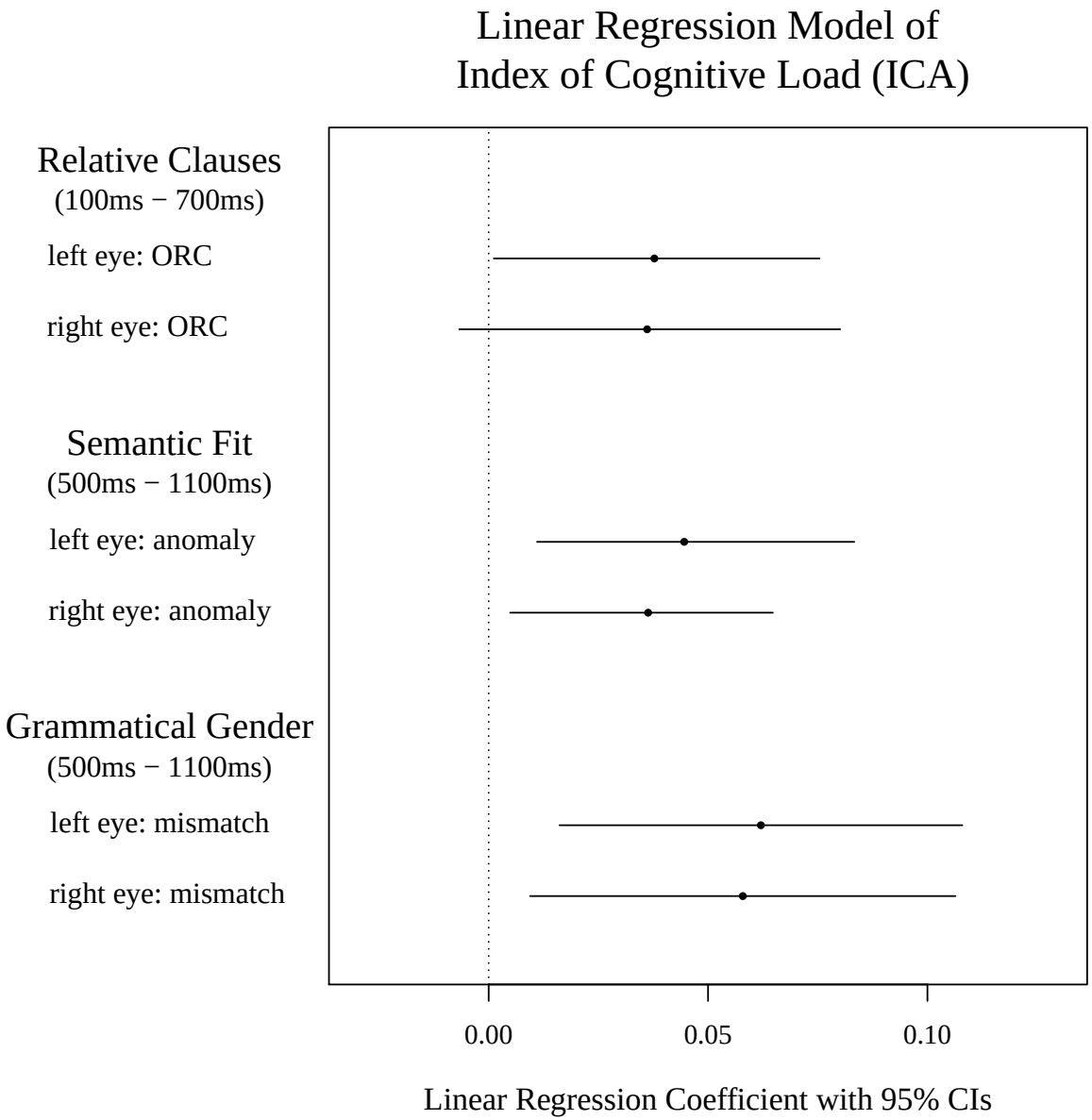
Simone hatte eine(n) schreckliche(n) **Traum** und stand auf.
"Simone had a$_{[masc/fem]}$ horrible$_{[masc/fem]}$ dream and got up."

Results:



- ICA effect significant on both eyes
- Significant effect of faster pupil dilation both eyes

Single Task Language Comprehension results



More subtle manipulation: discourse connectors

Mark denkt über einen kleinen Snack nach. Er hat gerade Lust, etwas Süßes zu essen. **Daher/Dennoch** holt er sich aus der Küche die/den appetitliche(n) Waffel/Kuchen/ Käse/Bretzel.

"Marc fancies a snack. He feels like having something sweet.
Therefore/However he gets from the kitchen the delicious waffle/cake/cheese/pretzel."

ICA und Visual World Experiment

ICA effects during connector region:

N=24	left ICA			right ICA		
	β	t val	p val	β	tval	pval
(Intercept)	6.218e-01	21.669	***	5.909e-01	19.837	***
concessive	4.476e-02	2.263	*	4.936e-02	2.142	*
X position	-3.860e-05	-2.747	**	-1.663e-07	-0.012	
Y position	-1.193e-04	-4.728	**	-1.442e-04	-5.715	***

"Marc fancies a snack. He feels like having something sweet.
Therefore / However he gets from the kitchen the delicious waffle."

ICA und Visual World Experiment

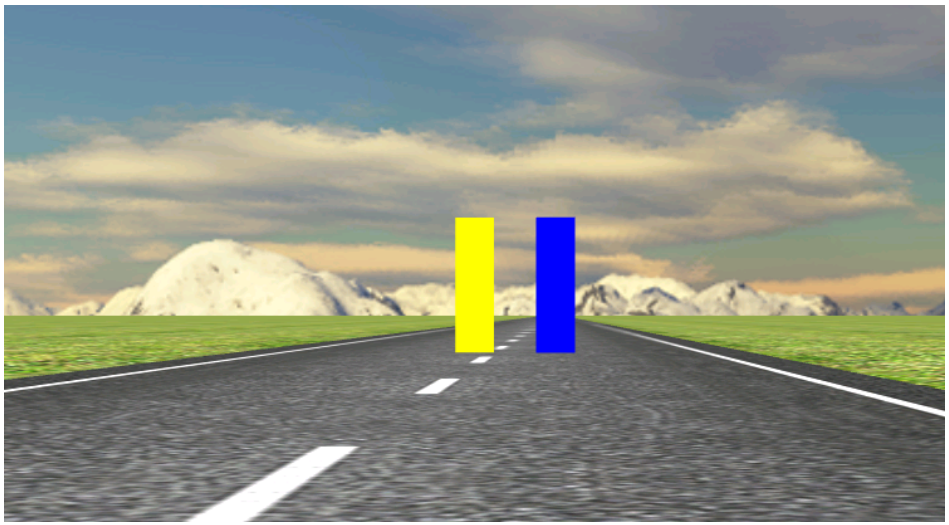
ICA effects during connector region:

- evidence for larger processing difficulty at concessive connector.
- consistent with P600 effect on same region in similar EEG expt
  $\rightarrow$ might reflect search for alternatives.
- allows one to assess both visual attention and processing difficulty

"Marc fancies a snack. He feels like having something sweet.
Therefore / However he gets from the kitchen the delicious waffle."

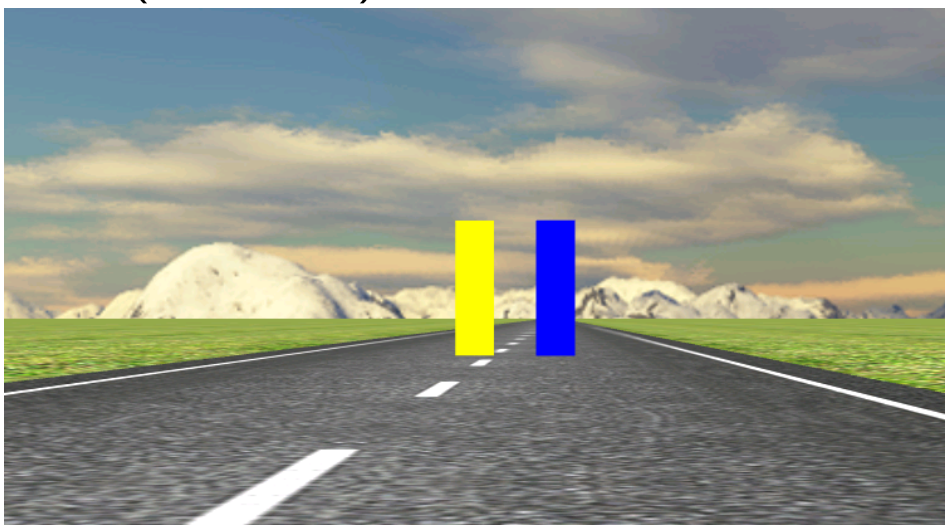
ICA and driving (single task)

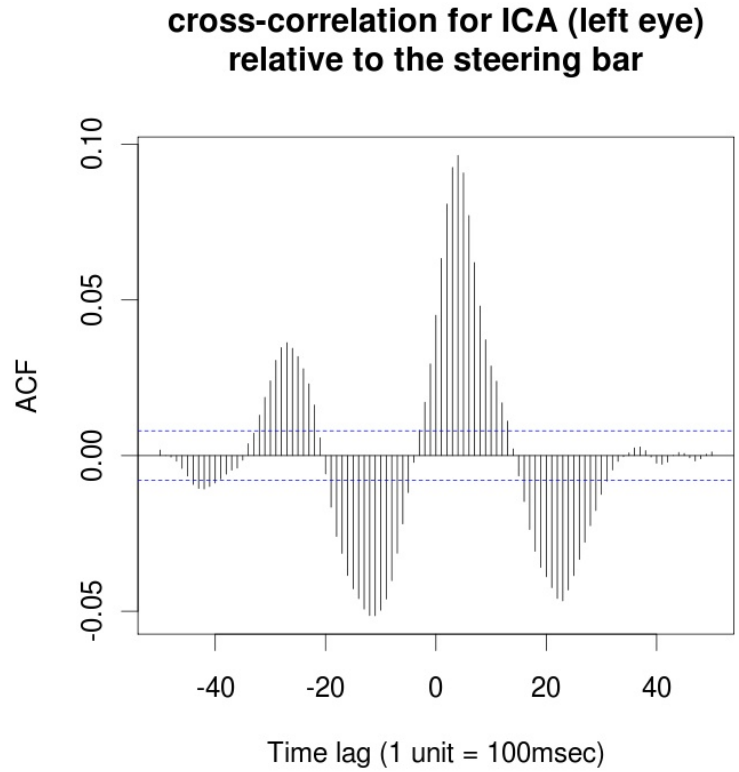
Continuous Tracking and Reaction task (ConTRe)



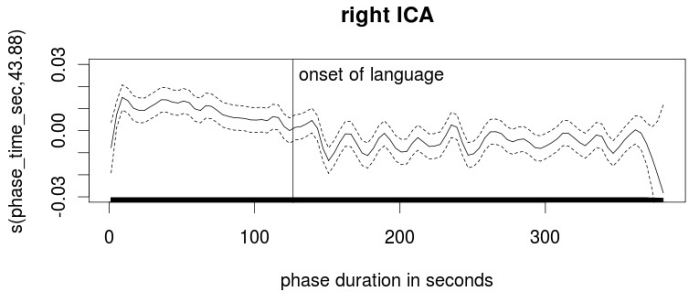
ICA and driving (single task)

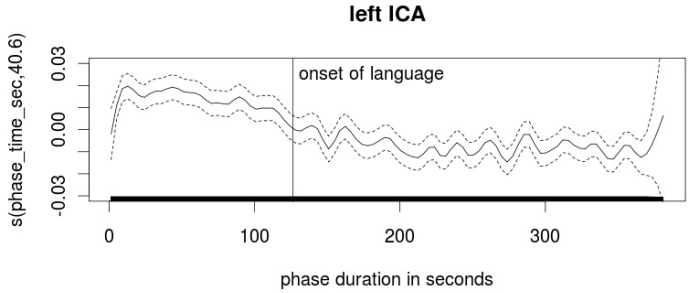
Continuous Tracking and Reaction task (ConTRe)

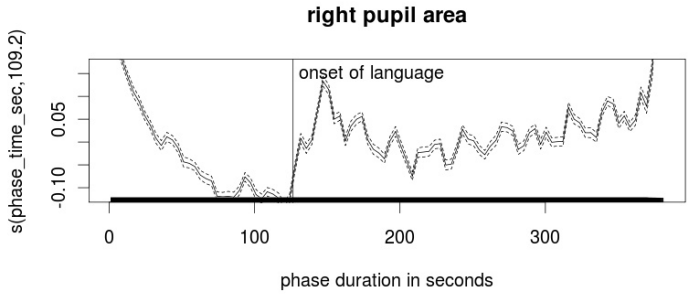


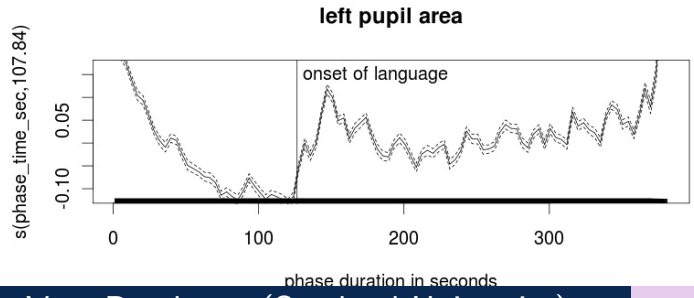


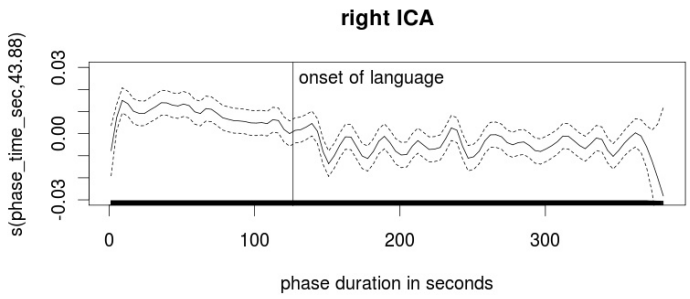
- More pupil twitch right after steering movement.
- No measurable effect on skin conductance or pupil size.

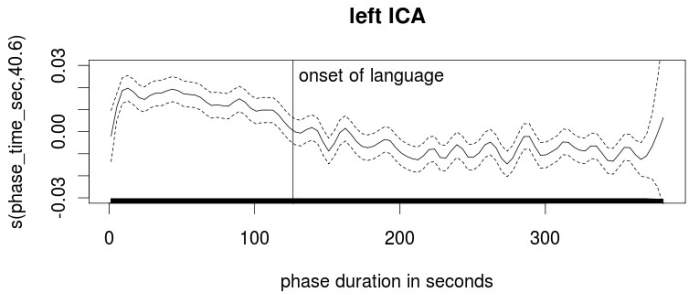

right ICA
s(phase_time_sec,43.88)
0.03
0.00
-0.03
onset of language
0
100
200
300
phase duration in seconds

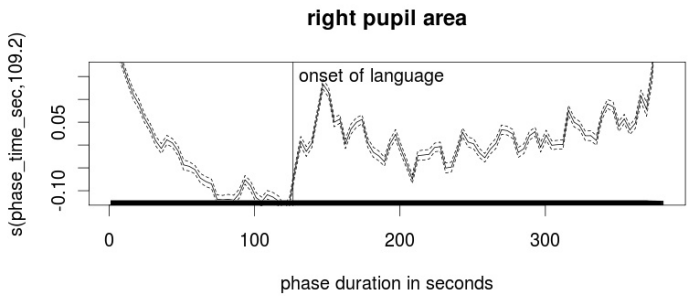

left ICA
s(phase_time_sec,40.6)
0.03
0.00
-0.03
onset of language
0
100
200
300
phase duration in seconds

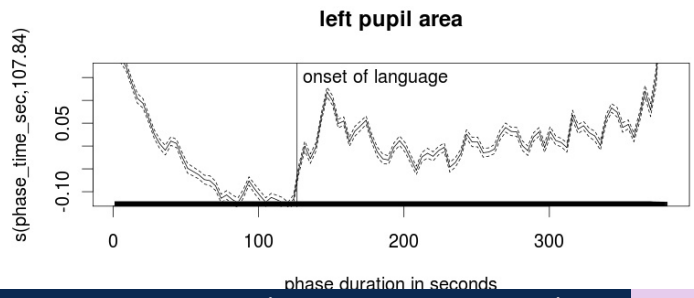

right pupil area
s(phase_time_sec,109.2)
0.05
-0.10
onset of language
0
100
200
300
phase duration in seconds


left pupil area
s(phase_time_sec,107.84)
0.05
-0.10
onset of language
0
100
200
300
phase duration in seconds









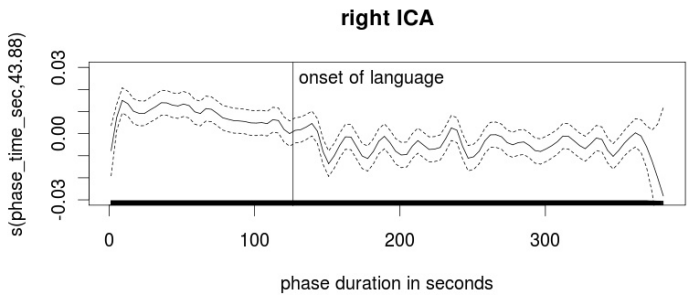
effect of language within dual task section

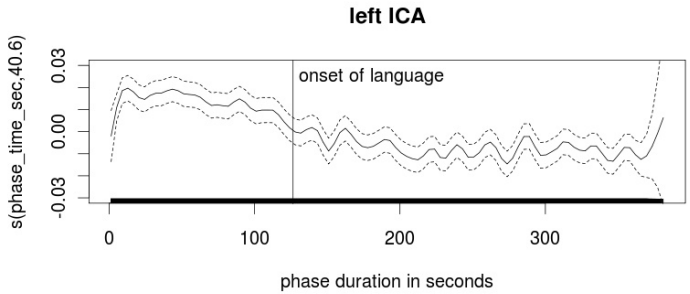
right ICA

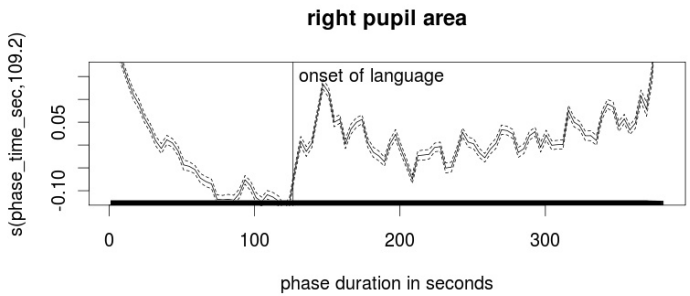
	coef	t value	signif
(Intercept)	0.730	50.49	***
sound file playing	0.033	8.99	***
easy driving	-0.012	-2.08	*

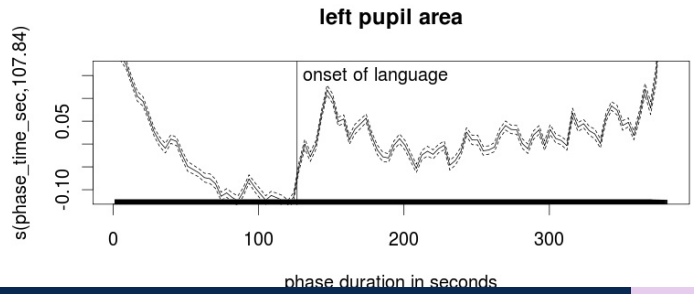
left ICA

	coef	t value	signif
(Intercept)	0.704	49.30	***
sound file playing	0.034	9.18	***
easy driving	-0.008	-1.01	









effect of language within dual task section

right ICA

	coef	t value	signif
(Intercept)	0.730	50.49	***
sound file playing	0.033	8.99	***
easy driving	-0.012	-2.08	*

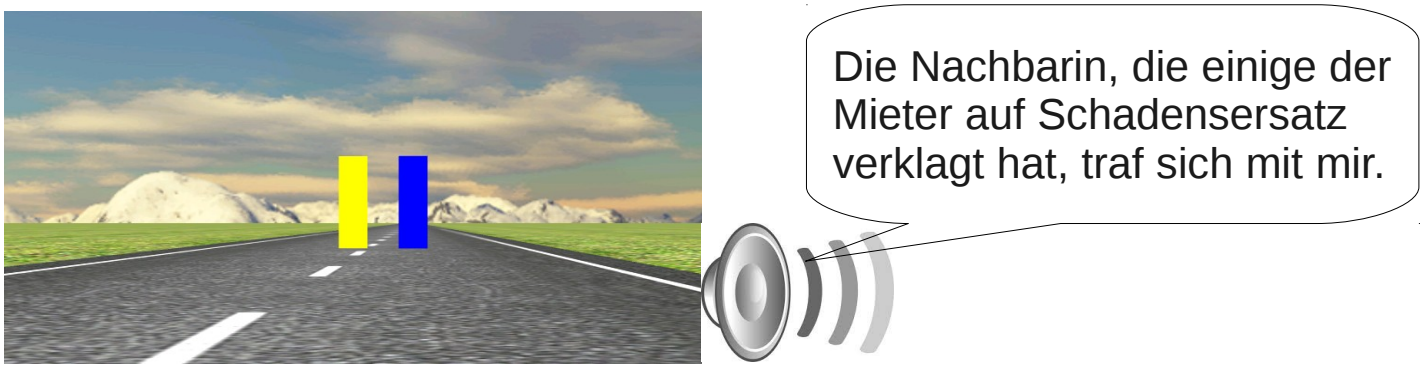
left ICA

	coef	t value	signif
(Intercept)	0.704	49.30	***
sound file playing	0.034	9.18	***
easy driving	-0.008	-1.01	

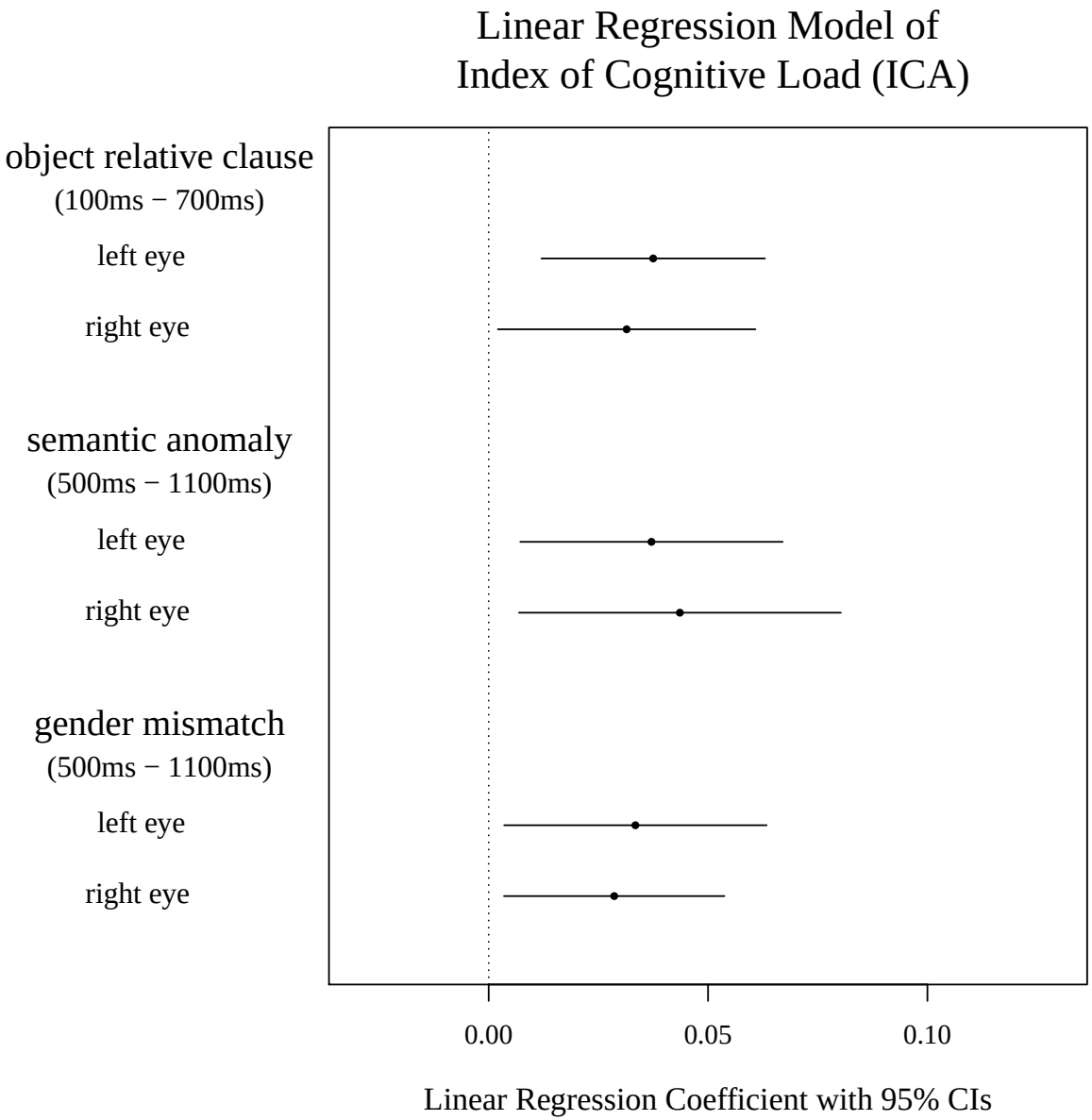
no effect of pupil size

But does it work in the dual task condition?

Dual Task: Driving and Language



Dual Task Language Comprehension results



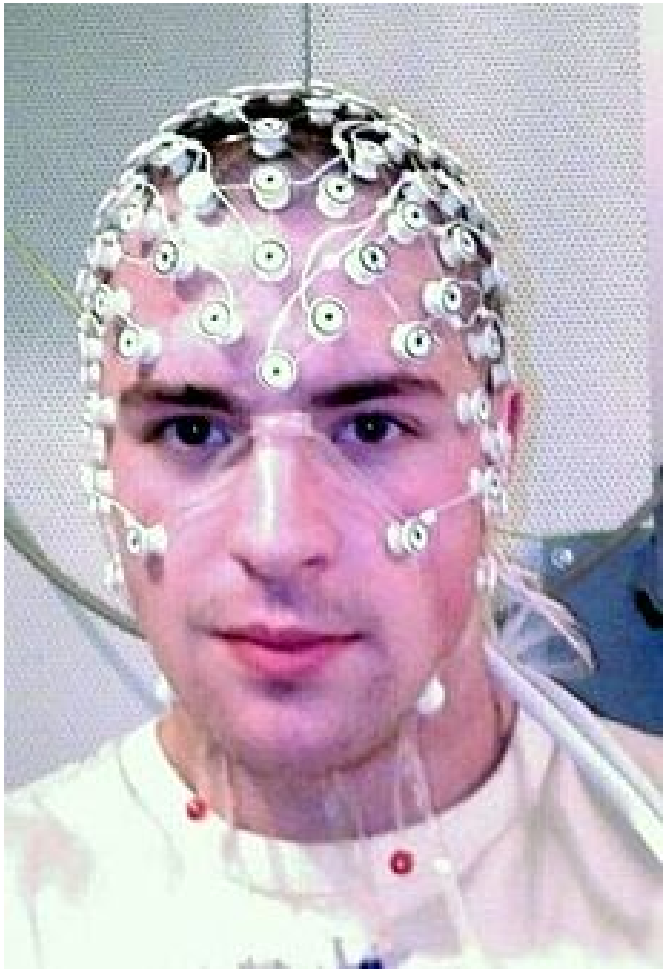
Conclusions ICA

Conclusions:

- sensitive to linguistic manipulation
- dynamic
- quite fast
- sensitive also to driving task
- both effects can be separated
- not so sensitive to lighting / movement

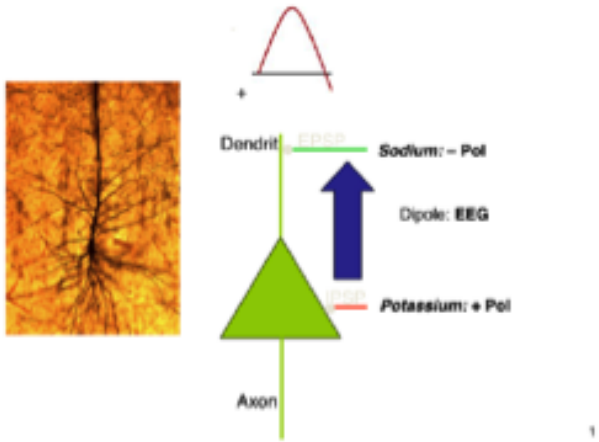
$\rightarrow$ suitable measure for our dual-tasking setting with driving and language.

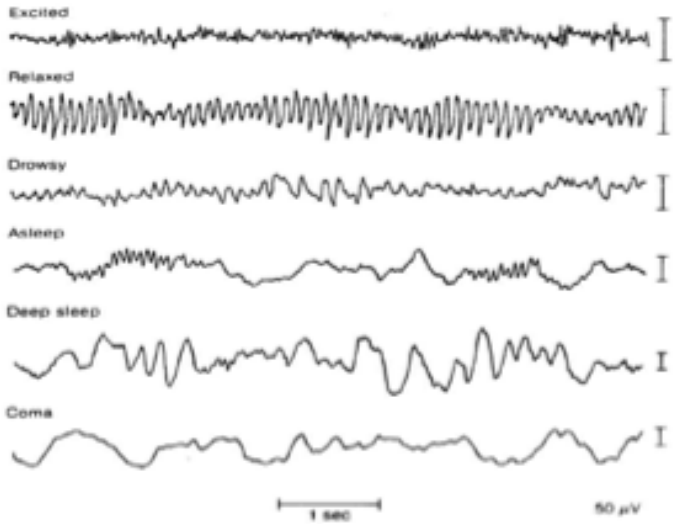
Table of Contents

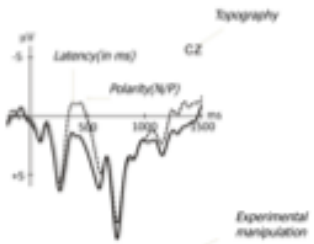


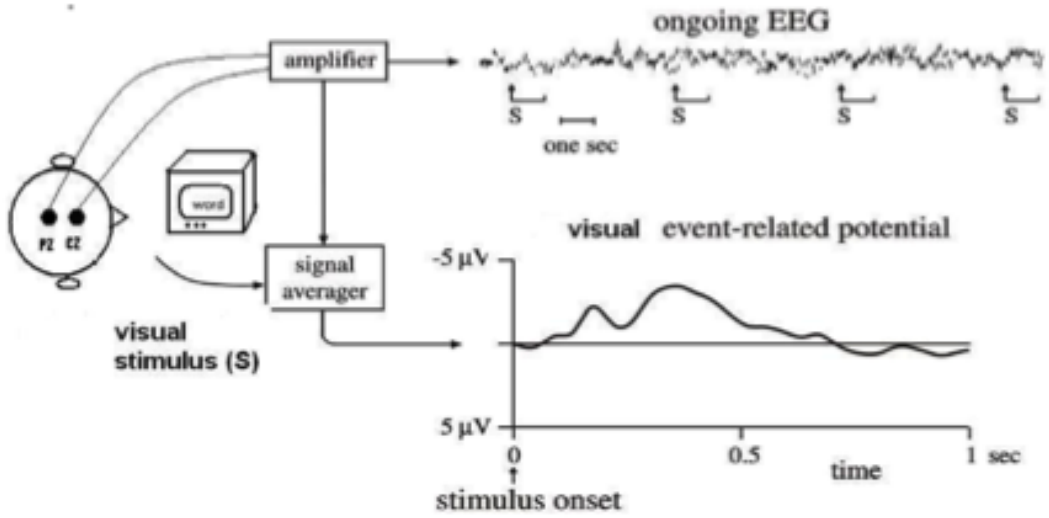
EEG (Elektroencephalography) measures electrical potential on the scalp.

EEG/ERP: Quick and dirty

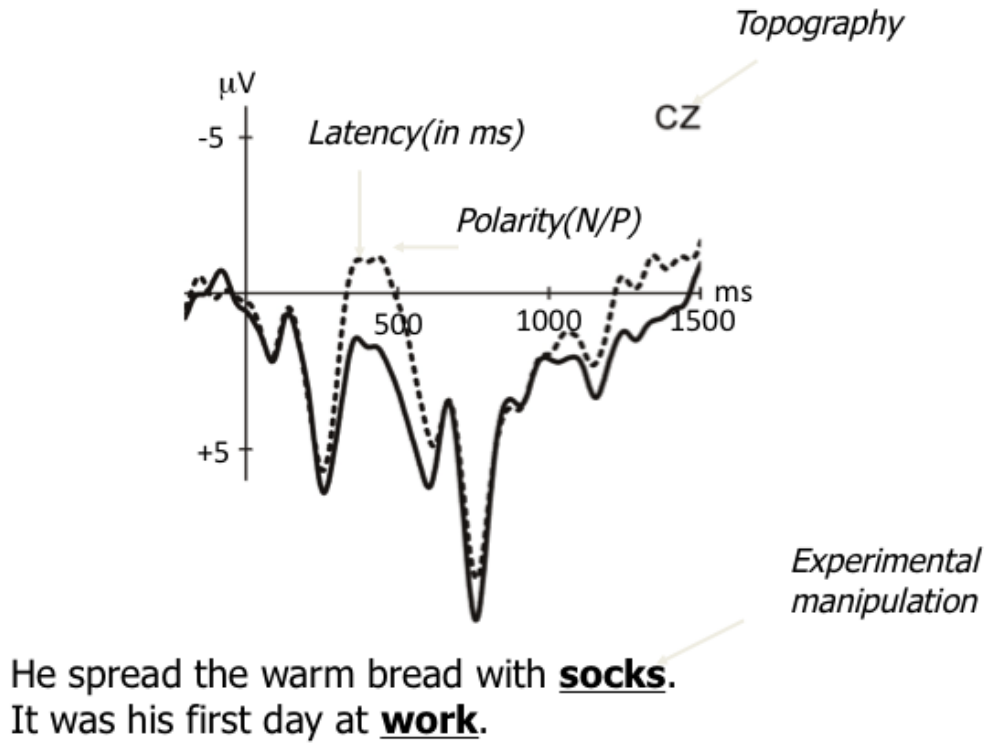








Event-related brain potentials (ERPs)



- Very high temporal (millisecond-by-millisecond) resolution
- ERP effects (so-called components) characterized by a set of
  - quantitative parameters (amplitude, latency)
  - qualitative parameters (polarity, topography, experimental sensitivity)

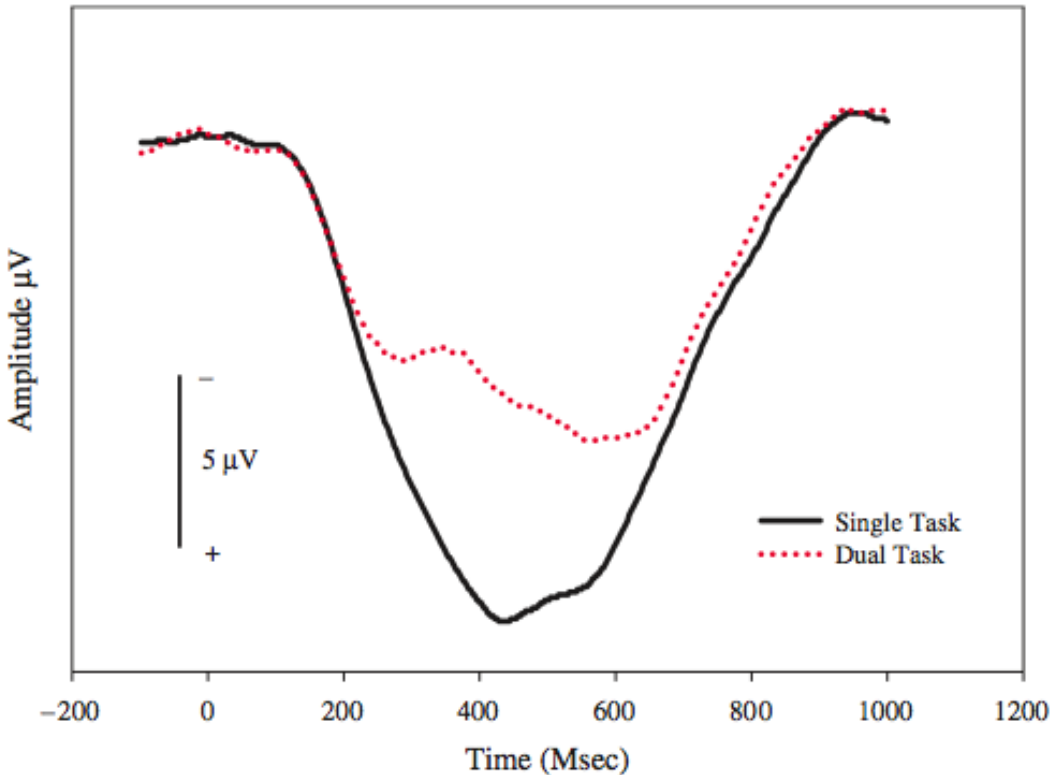
EEG and ERPs in dual tasking

- EEG frequency bands: observe increased activity in specific frequency bands which are related to memory storage / retrieval / ...
- ERPs: do subjects react to a specific stimulus? (e.g., are they surprised? do they notice something odd or dangerous about it?)

Challenges:

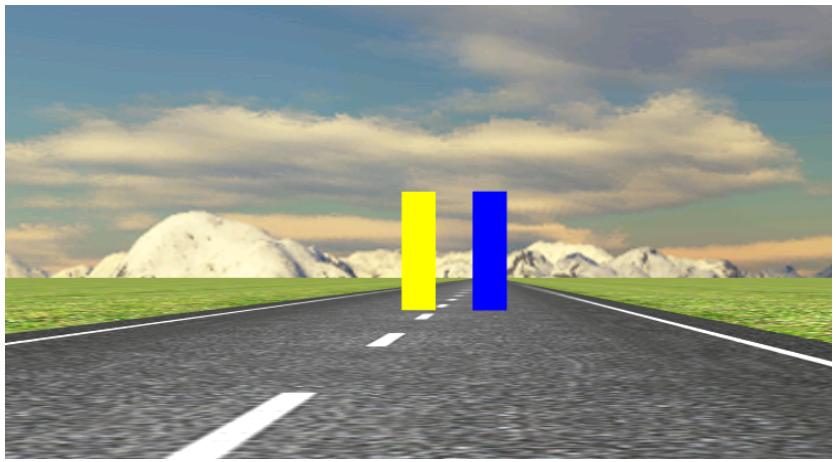
- movement artefacts
- electronic noise from driving simulator
- $\rightarrow$ bad quality of recordings / much data loss

Strayer and Drews, 2007:



- P300 is sensitive to the attention allocated to a task (Sirevaag et al, 1989; Wickens et al, 1983)
- memory performance is superior for objects eliciting larger-amplitude P300s during encoding (Fabiani et al. 1986; Otton & Donchin, 2000)
- P300 component sensitive to task difficulty: decreasing as task demands increase (Kramer et al, 1987; Sirevaag et al., 1993)

Conclusions – Methods



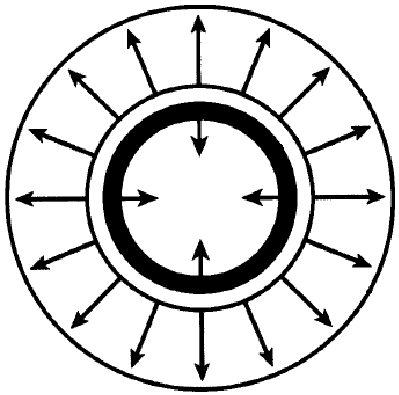
ConTRe driving task

- continuous task; well-established paradigm (tracking task) in psychology
- fine control over difficulty settings
- larger steering deviation in dual task condition
- found larger steering deviation in during linguistic critical region

Conclusions – Methods II

Index of Cognitive Activity (ICA)

- Frequency of rapid small pupil dilations
- More robust than overall pupil size
- Showed that it correlates with linguistic processing difficulty in seven experiments
- Only measure to reflect actions in steering task



from: Beatty & Lucero-Wagoner 2000

Other measures:

- steering wheel reversal
- skin conductance, heart rate
- eye-gaze, blink rate, pupil size
- EEG / ERPs

Conclusions

Summary

- Linguistic complexity has a measurable effect on driving performance.
- Linguistically induced work load can have an effect on driving safety.
- Dialog systems should adapt the way they speak to user and situation.
- tomorrow: closer look at how language comprehension and dual tasking ability differs between individuals; focus on young vs. old adults