Pupillometry in Linguistic Research

Psycholinguistics: The Eye-tracking Method and its Applications in Language Research
UCL Linguistics Autumn School, University College London, 2023

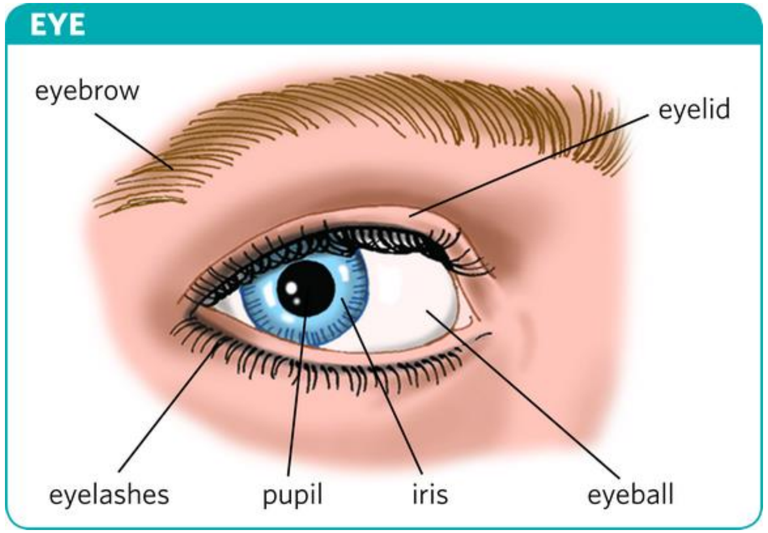
Kayla Keyue Chen, 9 November 2023

Content

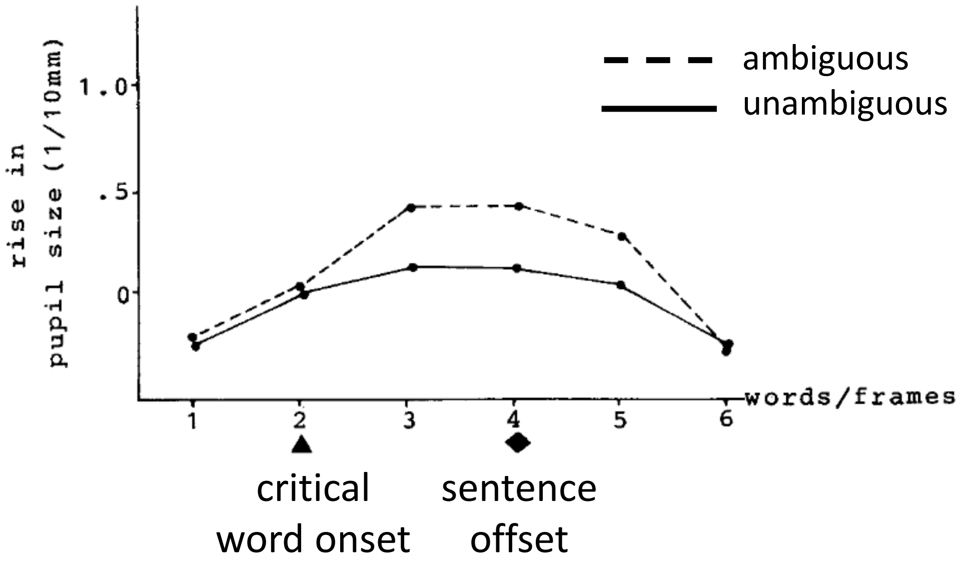
- What is pupillometry?
  - A typical pupillary response
  - Advantages and disadvantages of pupillometry
- What does pupil dilation index?
  - The pupil and the brain
- Pupillometry Research
  - Language processing
  - Speech perception
- Considerations on experimental design
  - Stimuli, procedure, environment, participants
- Data processing
  - Raw data, preprocessing, statistical analysis

What is pupillometry?

- Pupillometry measures the variation of the pupil diameter.
- Old way: "*Pictures were taken starting 0.5 sec before sentence beginning, throughout the sentence and until 1 sec after sentence end. …… frames were spaced at about* ***every 0.5 sec*** *throughout that period. …… The pictures were obtained on a negative film,* ***enlarged ×12.5*** *by a magnifying machine. This device also projected them onto a small table having a* ***movable scale****. Pupil diameter was measured in each frame.*" (Ben-Nun, 1986, p. 3)
- Today, most eye-trackers also record the pupil diameter with a temporal resolution of 60 to 1,000 Hz.



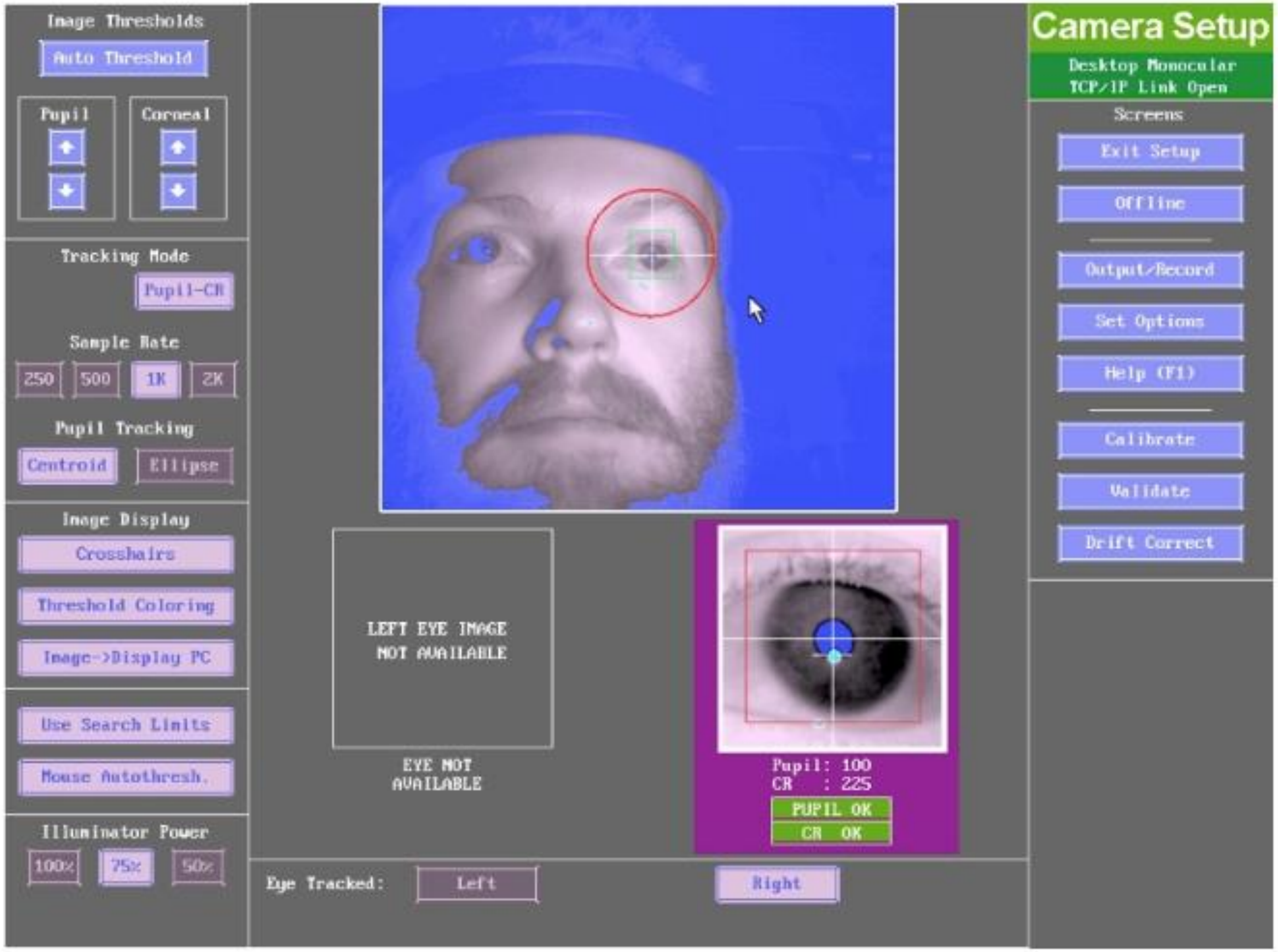
Picture from: https://www.ldoceonline.com/dictionary/pupil



What is pupillometry?

- EyeLink 1000 (Plus) desktop mount

Picture from: https://bss.au.dk/en/cognition-and-behavior-lab/for-researchers/labs-equipment/eye-tracking



Picture from: https://usermanual.wiki/Document/EyeLink20100020Plus20User20Manual201012.1370337621/view

What is pupillometry?

• A typical pupillary response

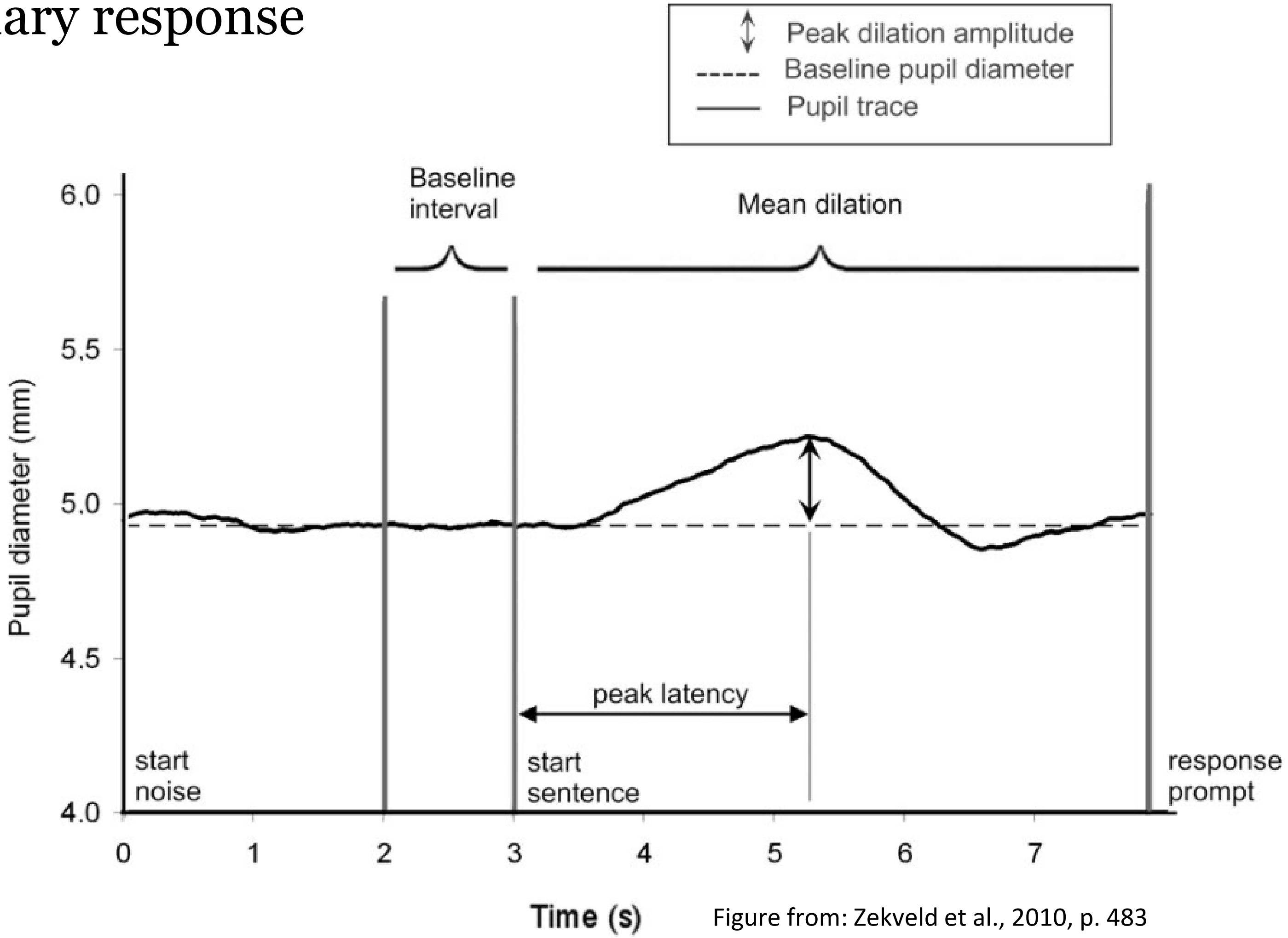


Figure from: Zekveld et al., 2010, p. 483

Advantages and disadvantages of pupillometry

- Advantages
  - A continuous measure of cognitive activity with high temporal resolution.
  - The cost is much lower compared to electro-physiological recording equipment, especially for departments that already possess eye-tracking equipment.
- Disadvantages
  - Pupillary responses can be slow
    - The pupil will start to dilate between roughly 0.5 to 1.3 s following the stimulus onset
    - The peak dilation occurs typically roughly 0.7 to 1 s following the end of the stimulus
  - Many factors can influence pupil size (see the next page)

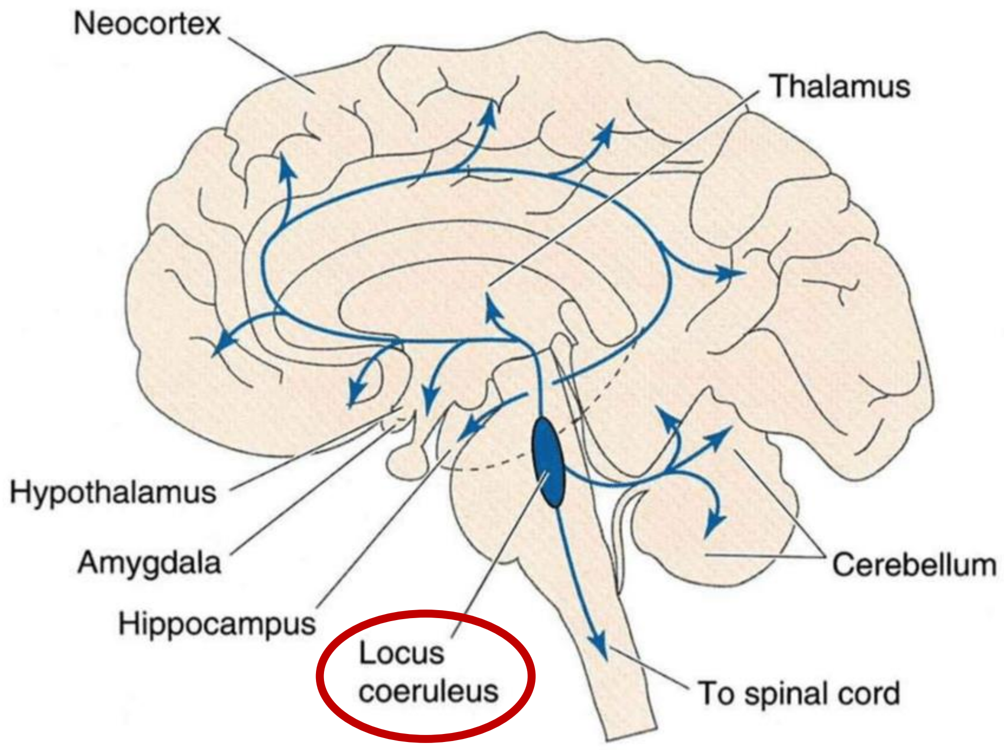
What does pupil dilation index?

- Pupil dilation as an index of effort
  - Task-evoked pupil dilation
    - Cognitive tasks (e.g., working memory and cognitive control tasks, math problems)
    - Speech perception (e.g., listening effort related to intelligibility)
    - Language processing (e.g., structural and lexical ambiguity)
- But also …
  - Illumination
    - When changing from light to dark environments, pupil diameter can increase by as much as 3-4 mm, or 120%.
    - Compared to that, cognitive processes typically evoke 0.1-0.5 mm change, depending on testing conditions and tasks.
  - Emotion: Positive and negative emotion can evoke larger pupil dilation.
  - Motivation: Fatigue and boredom lead to smaller pupil dilation.
  - Attention, motion, age, substance use, etc.
- Pupil size is not a monotonic direct index of effort but rather **a complicated mixture** that reflects the combined contributions of the autonomic nervous system (ANS)

The pupil and the brain

- The *autonomic nervous system* (ANS)
  - A component of the peripheral nervous system that regulates involuntary physiologic processes including *heart rate*, *blood pressure*, *respiration*, *digestion*, and *sexual arousal*.

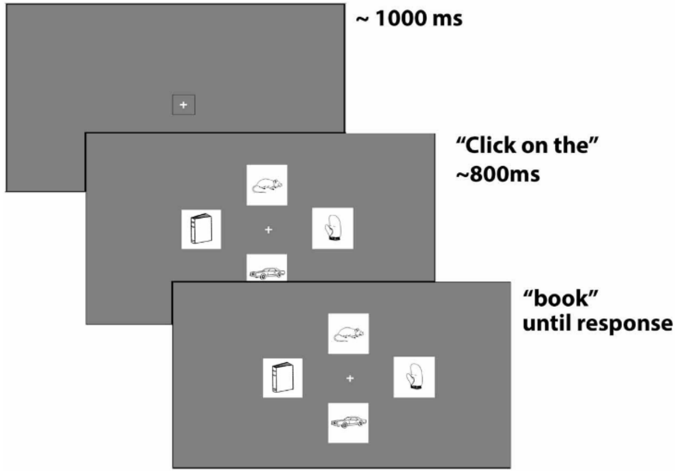
- Changes in pupil size are correlated to changes in activity in neurons of the locus coeruleus (LC)
  - A small nucleus in the pons of the brainstem.
  - The locus coeruleus (LC) contains norepinephrine (NE)-synthesizing neurons that send diffuse projections throughout the CNS. The LC-NE system has a major role in *arousal*, *attention*, and *stress response*.

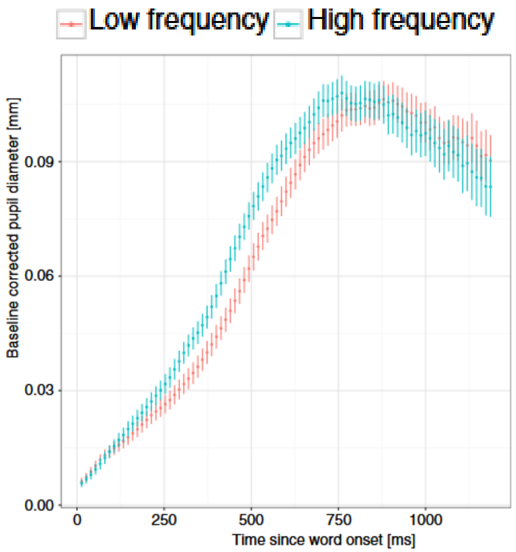


Picture from: https://www.researchgate.net/figure/The-Locus-Coeruleus-Norepinephrine-LC-NE-System_fig1_325625804

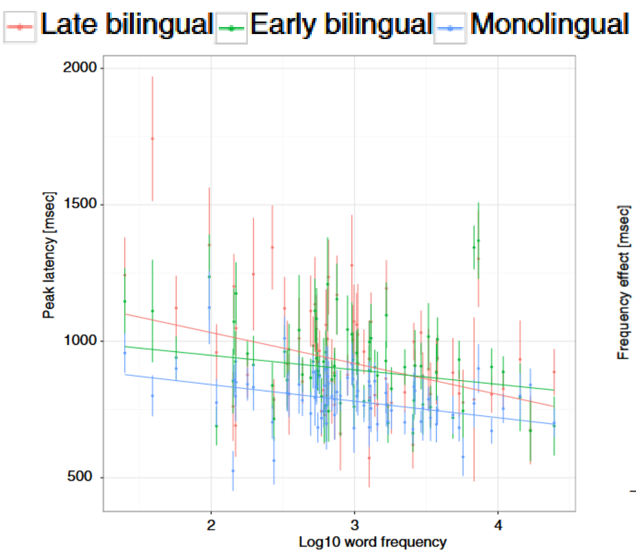
Linguistics research: Word-level processing

- Schmidtke (2014) used pupillometry to measure word retrieval effort associated with **frequency** and **neighbourhood density** in monolinguals and bilinguals.
  - Participants heard the instruction “click on the [target word]” and selected the mentioned picture.
  - Peak amplitude: the largest dilation in a trial
  - Peak latency: the time elapsed from word onset to the peak amplitude
  - Results about frequency:

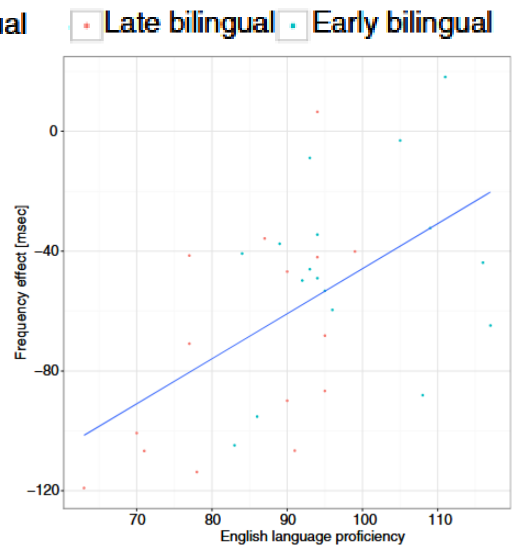




monolinguals



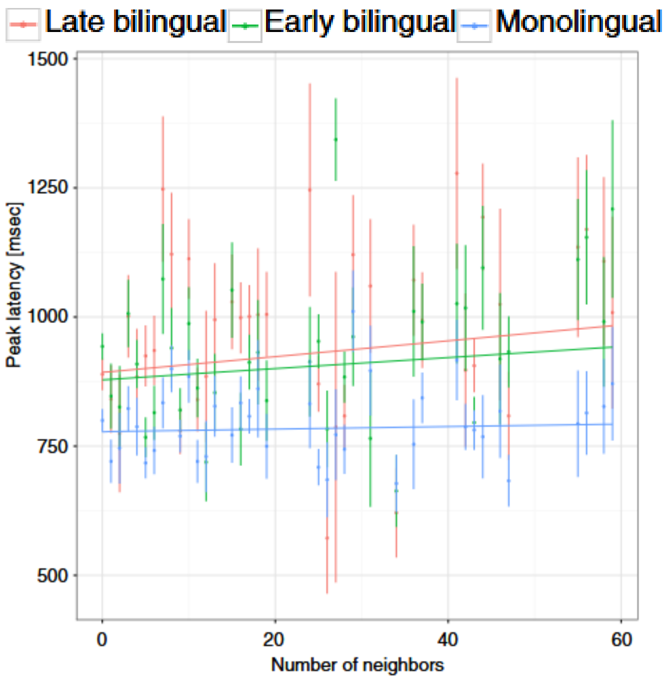
mono & bilinguals



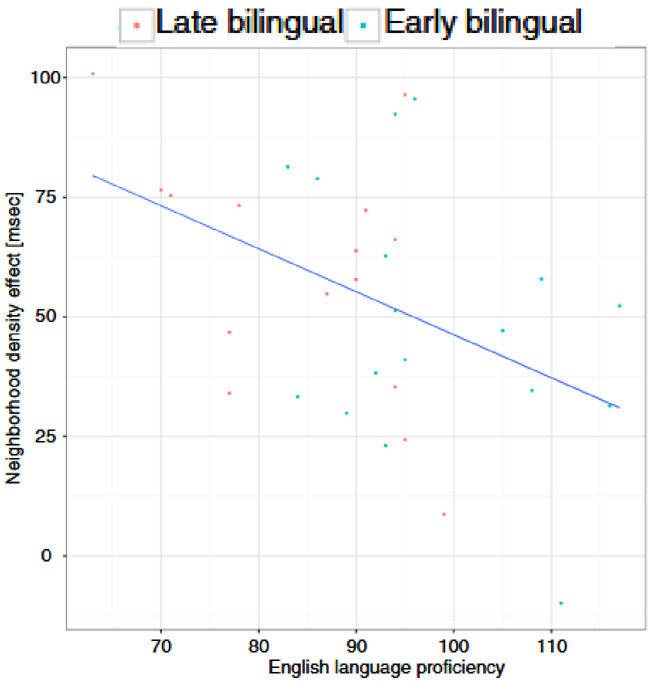
bilinguals

Linguistics research: Word-level processing

- Schmidtke (2014)
  - Results about neighbourhood density:



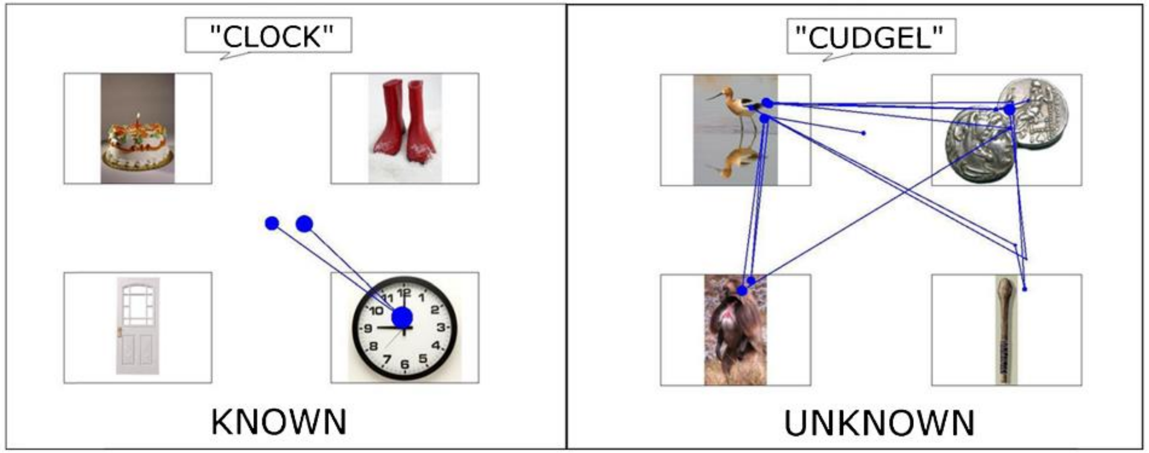
mono & bilinguals



bilinguals

Linguistics research: Word-level processing

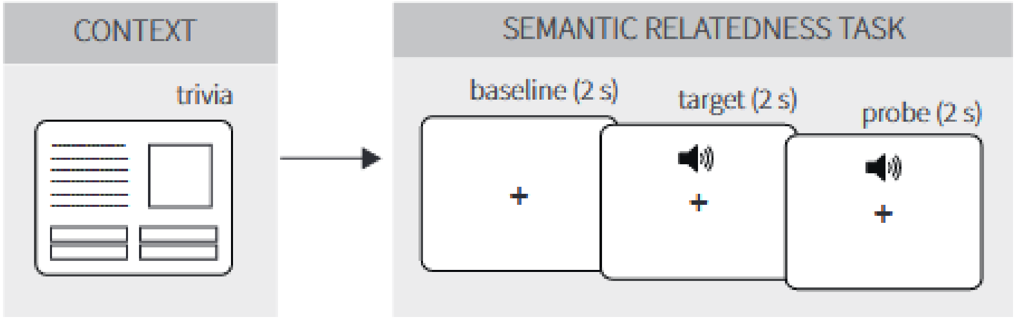
- Ledoux et al. (2016) studies **receptive vocabulary knowledge** using concurrent eye movement and pupillometry
  - Participants heard a word and selected the mentioned picture
  - Known words: *cat*, *airplane*, and *camera*
  - Unknown words: *cherimoya*, *agouti*, and *cainito*



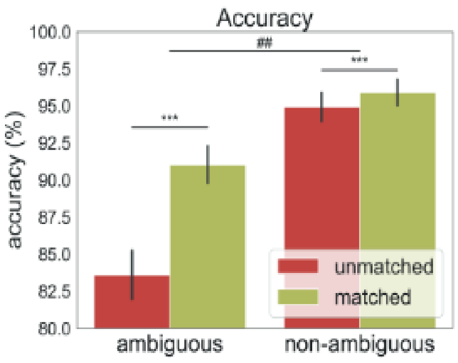
	Known		Unknown	
Dependent Variable	*M*	*SD*	*M*	*SD*
Pupillary Dilation				
Peak dilation (mm)	5.43	1.51	7.53	2.30
Mean change in pupil size (mm)	0.01	0.70	1.31	0.74
Max percent change in pupillary dilation (%)	15.78	3.82	22.10	5.45

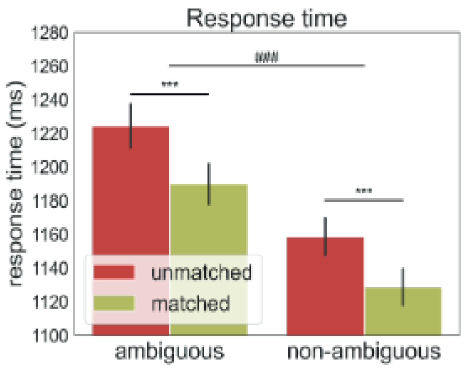
Linguistics research: Word-level processing

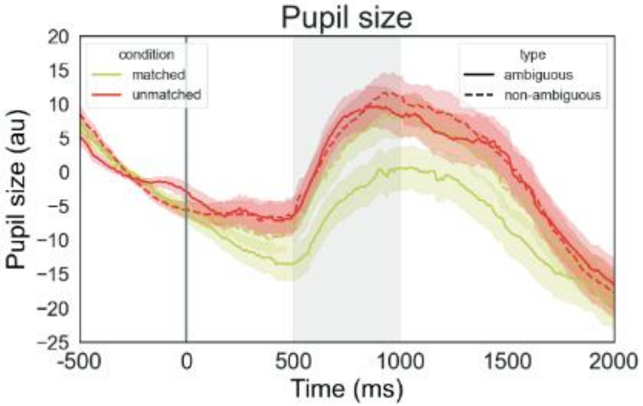
- Laurino et al. (2023) studied whether contextual information facilitate ambiguity resolution of a semantically **ambiguous word**.
  - Context: a trivia related to a topic (e.g., music)
    - Context matched target: e.g., note, guitar
    - Context unmatched target: e.g., bank, money
  - Ambiguous words
    - Ambiguous: note, bank
    - Unambiguous: guitar, money

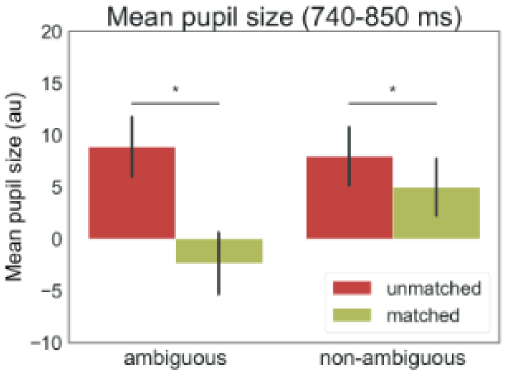


CONTEXT CONDITION	TARGET TYPE	E.g.: CONTEXT	TARGET	PROBE
matched	ambiguous	music	'note'	'song'
unmatched	ambiguous	music	'bank'	'finance'
matched	non-ambiguous	music	'guitar'	'string'
unmatched	non-ambiguous	music	'money'	'coin'



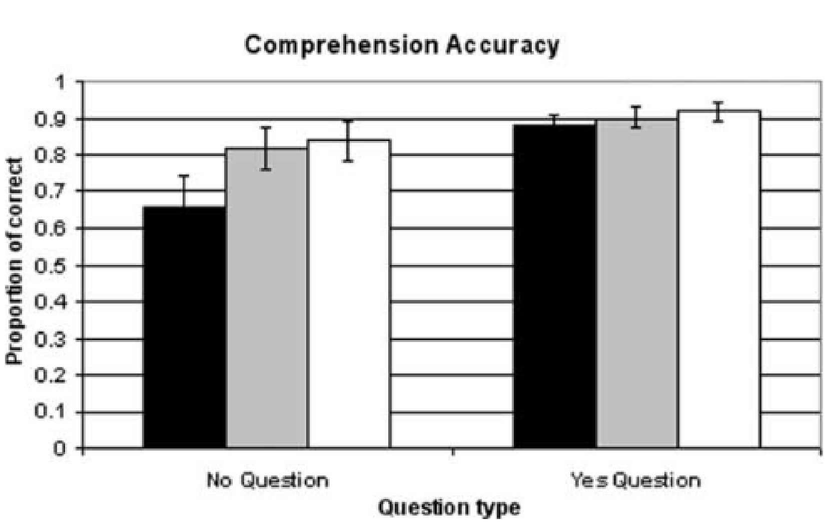


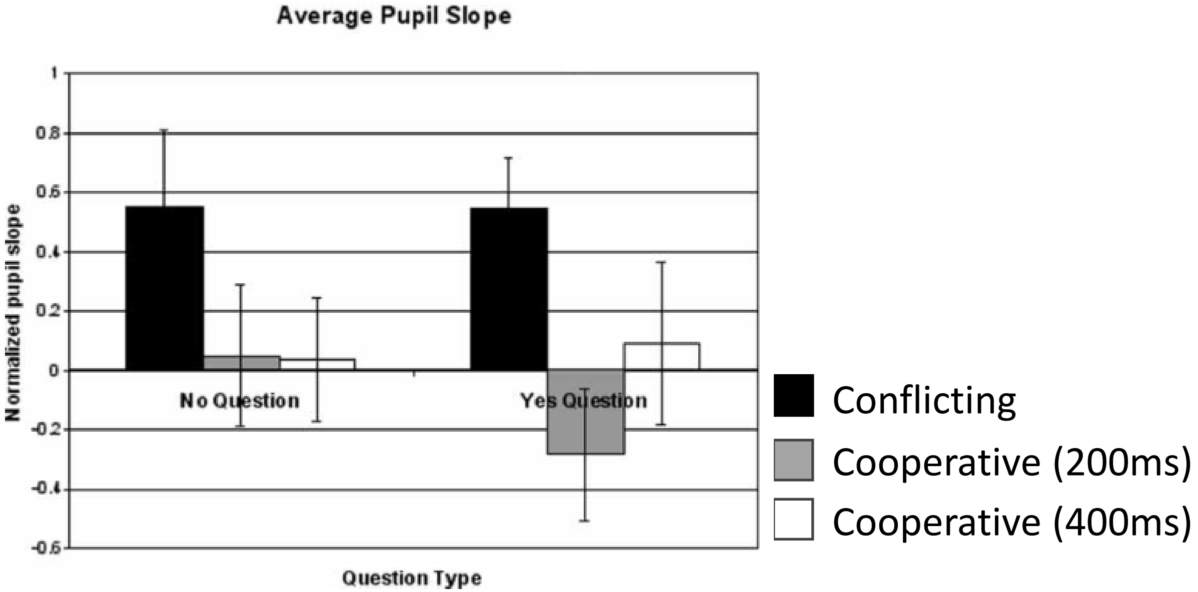




Linguistics research: Sentence-level processing

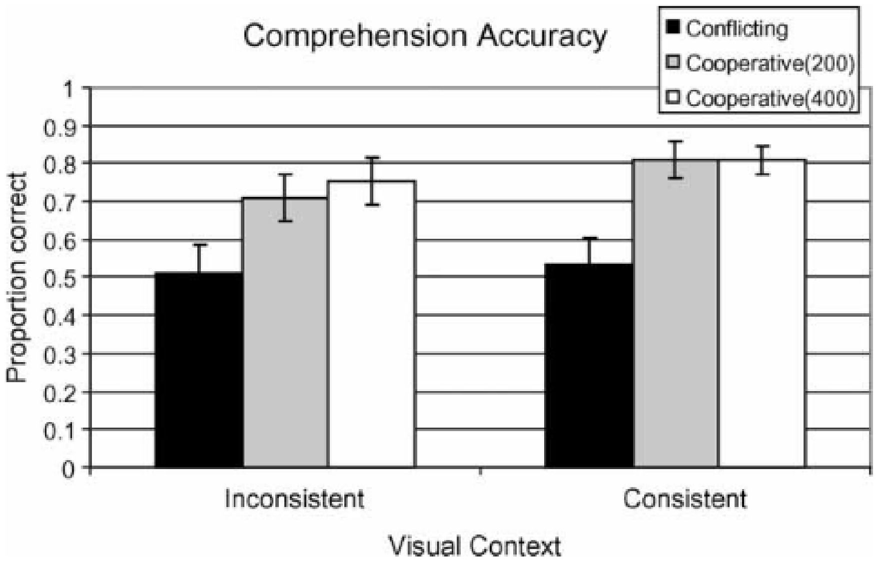
- Engelhardt et al. (2010) investigated processing effort when listening to sentences containing a **temporary syntactic ambiguity**. They tested how **prosody** and **visual context** interacted in parsing.
  - *While the woman cleaned (#) the dog that was big and brown stood in the yard.*
    - *Did the woman clean the dog? ("No" question)*
    - *Did the dog stand in the yard? ("Yes" question)*
  - **Experiment 1:** 2 question types × 3 prosody types
    - Cooperative prosody: a prosodic break between *cleaned* and *the dog* (the pause was 200 or 400 ms)
    - Conflicting prosody: as if *the dog* were the object of the verb *cleaned*

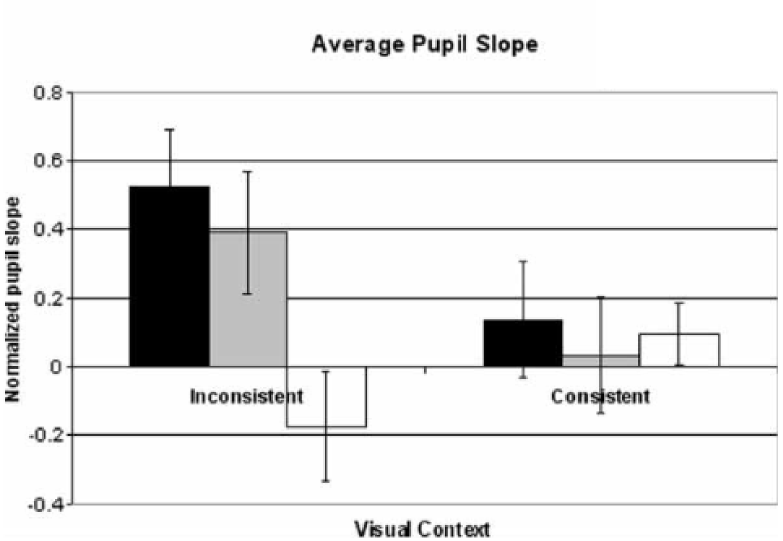




Linguistics research: Sentence-level processing

- Engelhardt et al. (2010)
  - *While the woman cleaned (#) the dog that was big and brown stood in the yard.*
  - **Experiment 2:** 2 visual context × 3 prosody types
    - Consistent visual context: a woman cleaning something that was not a dog (e.g., a window)
    - Inconsistent visual context: a picture of a woman cleaning a dog
    - For the filler trials, half of the pictures were related to the sentence, and half were not.
    - (in experiment 2, participants only saw “no” questions)



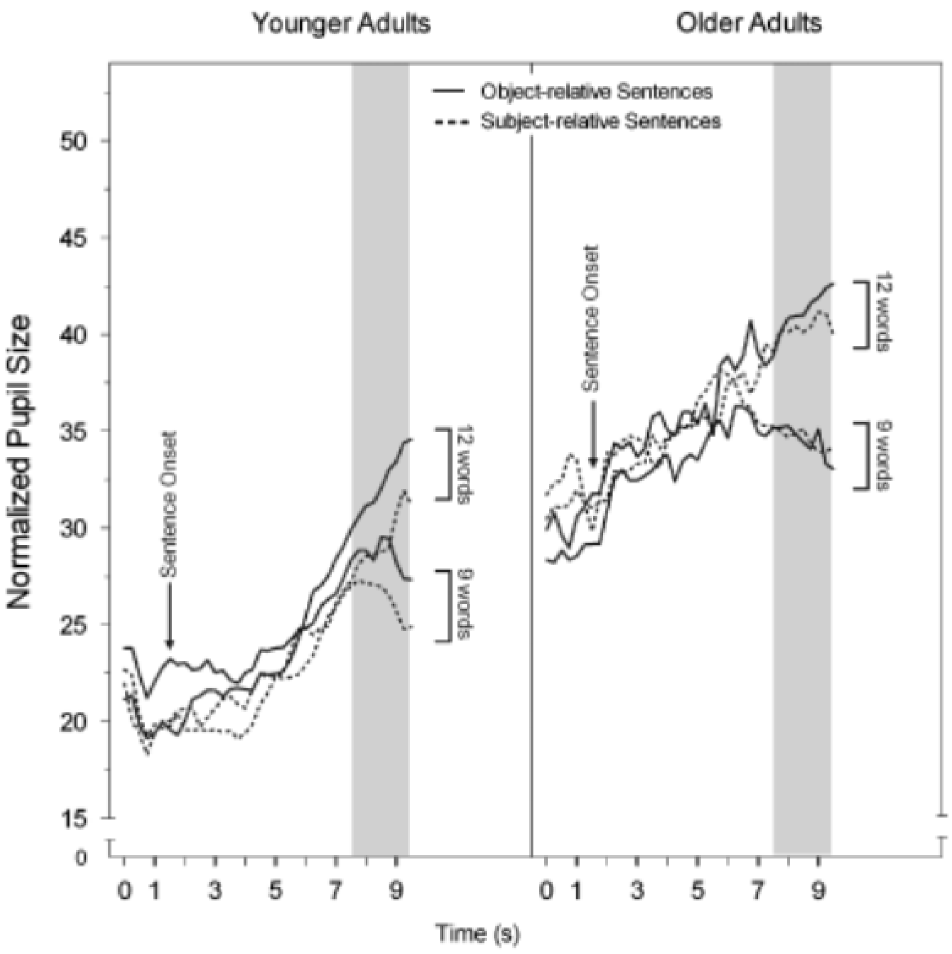


Conflicting
Cooperative (200ms)
Cooperative (400ms)

Linguistics research: Sentence-level processing

- Piquado et al. (2010) investigated how younger and older adults' pupillary responses can be affected by **syntactic complexity (subject- and object-relative clause)** and **sentence length**.
    - Participants heard and repeated sentences. There was 2-second retention interval between the end of each sentence and the signal to respond.

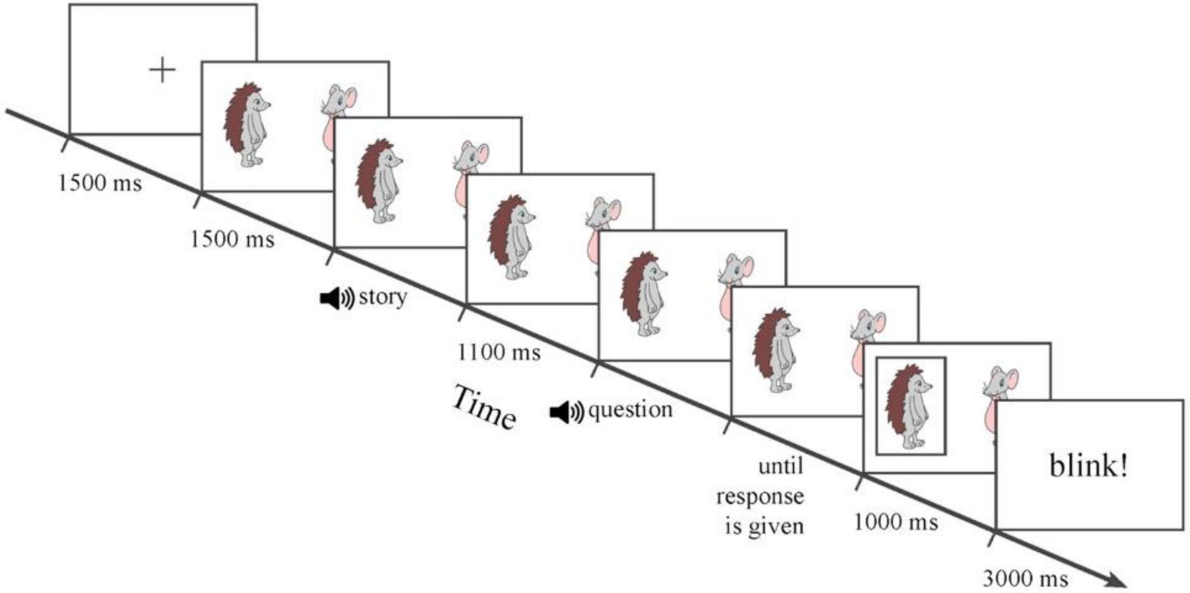
Sentence type	Without modifiers	With modifiers
Subject-relative	The gambler that signaled the dealer revealed the card.	The professional gambler that signaled the suspicious dealer revealed the perfect card.
Object-relative	The gambler that the dealer signaled revealed the card.	The professional gambler that the suspicious dealer signaled revealed the perfect card.

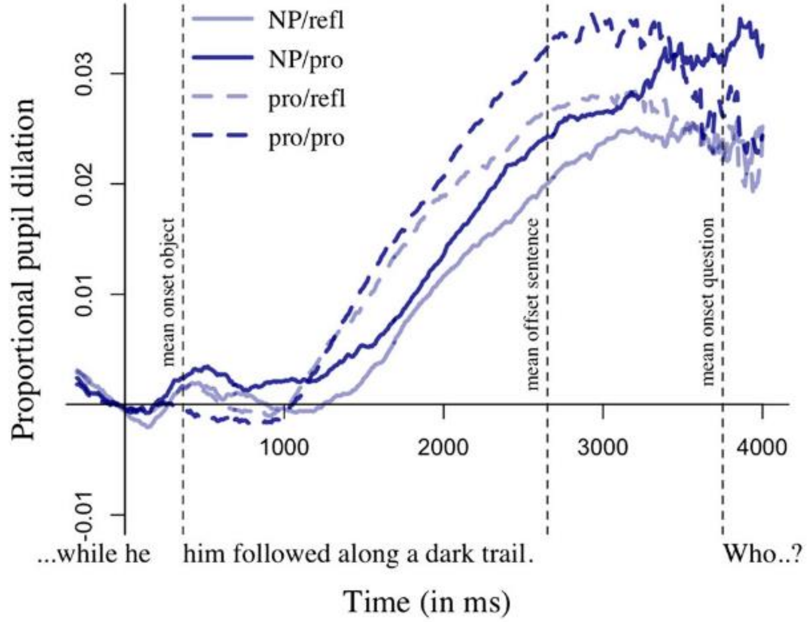


Linguistics research: Discourse-level processing

- Vogelzang et al. (2016) investigated **the resolution of ambiguous pronouns**.
  - (originally in Dutch) The hedgehog has built a tree-house. Last Tuesday, the hedgehog walked with the mouse through the forest to home, while ......

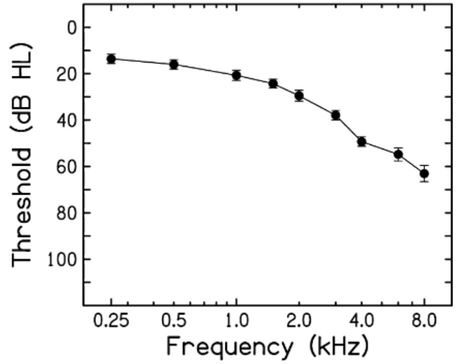
S	O	V		CONDITION
the hedgehog	himself	hurried	along a dark trial	(NP-reflexive)
the hedgehog	him	followed	along a dark trial	(NP-pronoun)
he	himself	hurried	along a dark trial	(pronoun-reflexive)
he	him	followed	along a dark trial	(pronoun-pronoun)

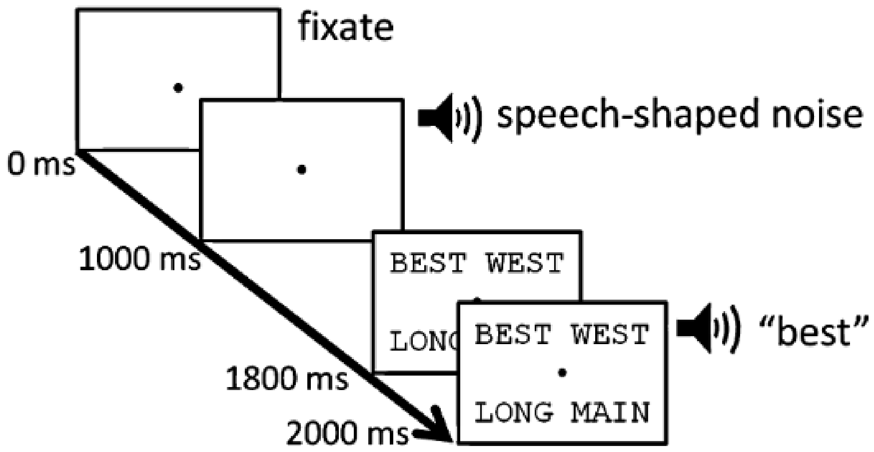


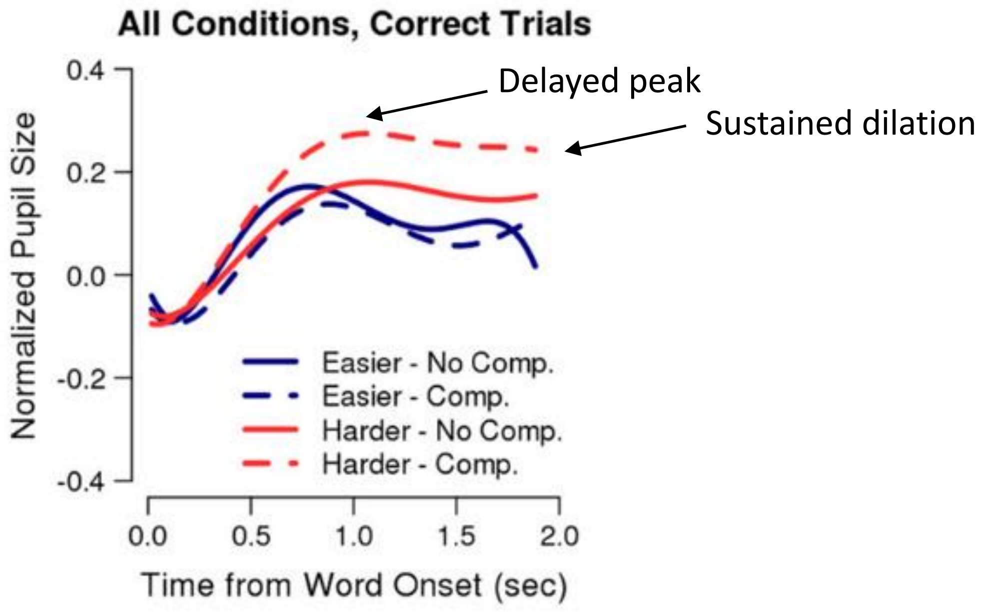


Speech perception: Word perception

- Kuchinsky et al. (2013) found that pupil size varies with **word listening difficulty** in older adults with hearing loss
  - Healthy older adults aged 61 to 88
  - Participants heard a word and selected the word they heard
    - Background noise: spectrally shaped speech but completely unintelligible (easier vs. harder)
    - Competition: whether the target word (e.g., *best*) was presented with a competitor (e.g., *west*)

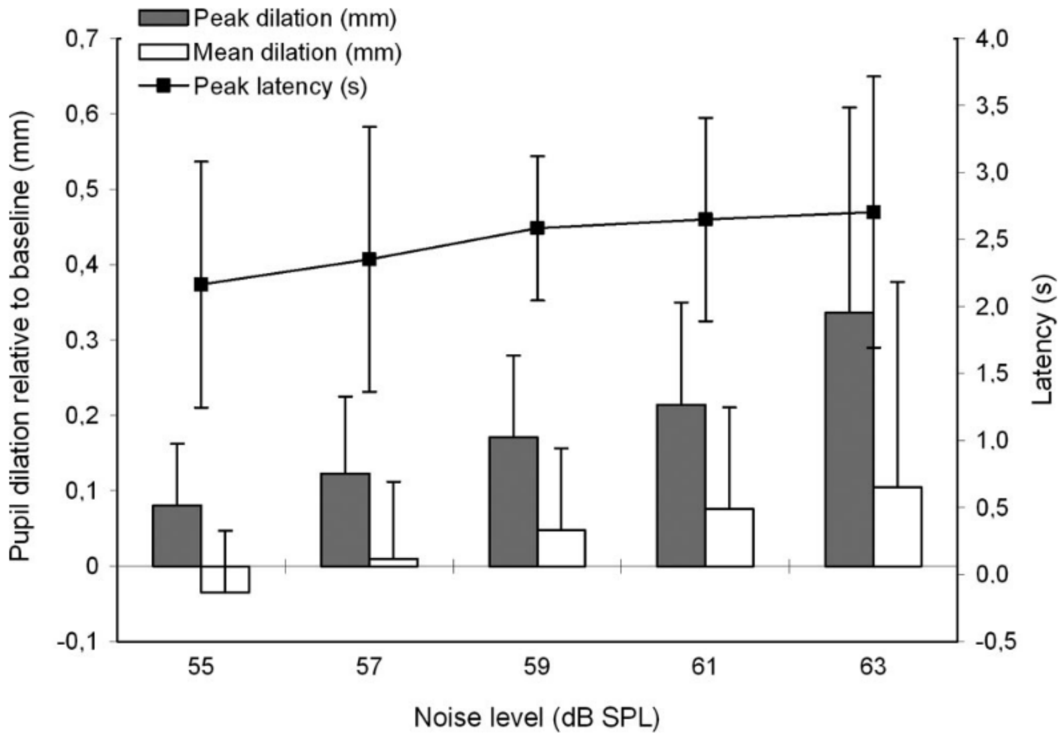






Speech perception: Sentence perception

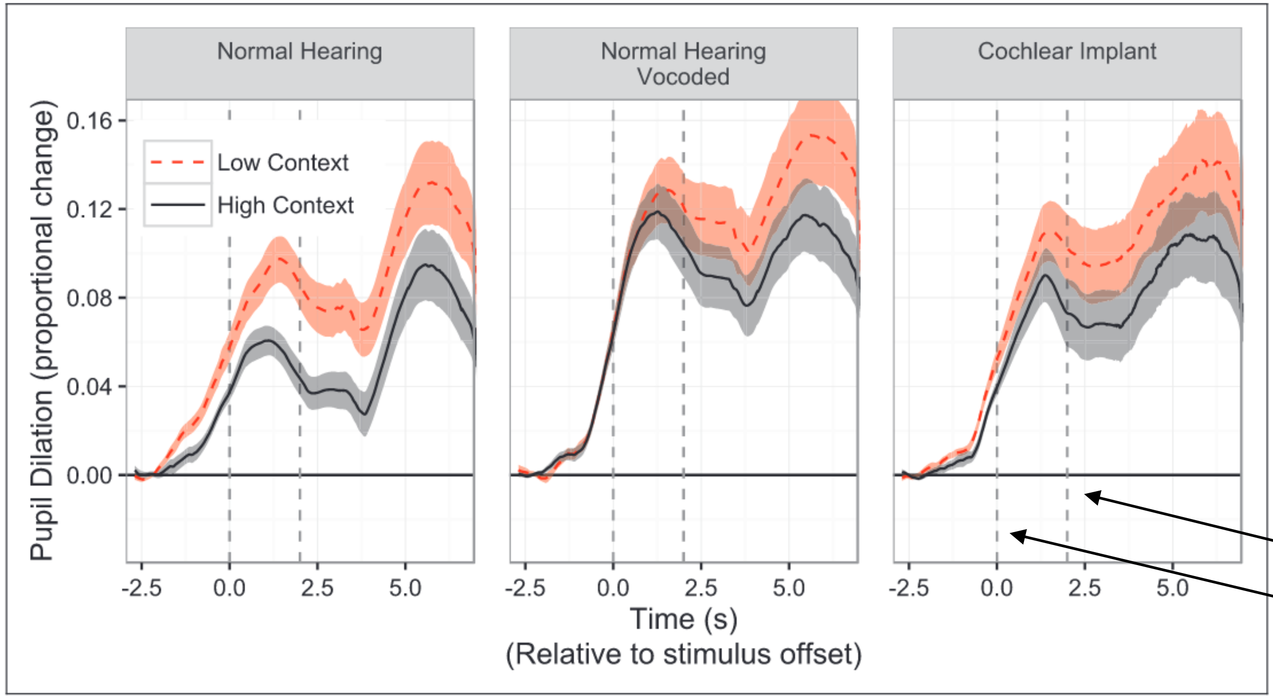
- Zekveld et al. (2010) evaluated the influence of **sentence intelligibility** on the pupil dilation response during listening.
  - Participants listened to sentences and repeated them
  - Results:
    - The peak dilation amplitude, peak latency, and mean pupil dilation systematically increase with decreasing speech intelligibility.
    - Irrespective of signal to noise ratio (SNR), the pupil response was higher for incorrectly repeated sentences than for correctly repeated sentences

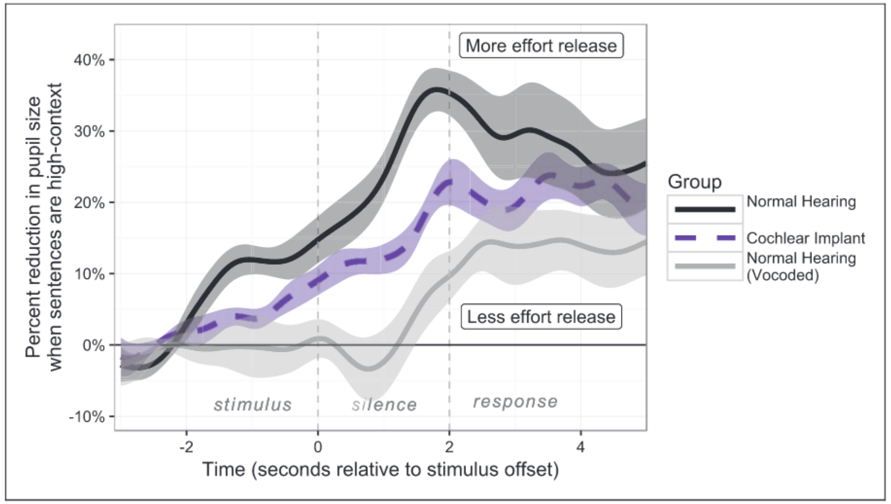


	Pupil Measure					
	Peak Dilation, mm		Peak Latency, sec		Mean Dilation, mm	
	M	SD	M	SD	M	SD
Correct	0.15	0.12	2.35	0.57	0.02	0.12
Incorrect	0.24	0.08	2.97	0.80	0.09	0.07

Speech perception: Sentence perception

- Winn (2016) explored whether cognitive load can be reduced due to **predictive processing**, and **how hearing impairment and difficulty** can affect the process.
  - Participants heard and repeated sentences
  - Normal hearing – normal speech, normal hearing – vocoded speech, cochlear implant – normal speech
  - Predictable (high context): *stir your coffee with a spoon*
  - Unpredictable (low context): *Jane thought about a spoon*





the timing of the response prompt

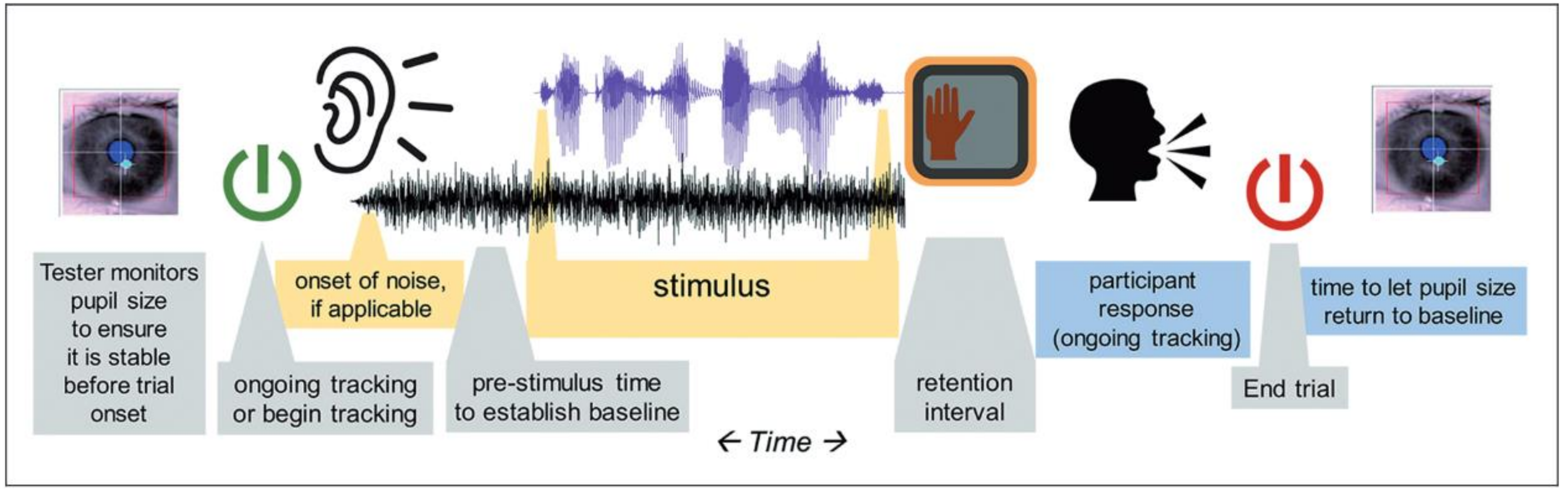
the offset of the sentence

Considerations on experimental design

- Stimuli
  - Stimuli content
    - Visual properties (if images) and auditory properties (if audio) should ideally be the same between conditions
    - Stimuli should not evoke strong emotions (if emotion is not the research question)
    - Try not to be too boring
  - Stimuli difficulty
    - Not so easy as to demand too little cognitive effort
    - Not so difficult as to make cognitive effort meaningless
  - Number of trials
    - At least 16 to 18 good recordings of pupil size for each condition
    - For sentence-perception tasks, 20 to 25 trials are normally a safe starting number
    - For more subtle differences between similar conditions, a larger number of trials is recommended

Considerations on experimental design

- Procedure
  - Trial event and timing
    - Retention interval (time after stimuli presentation and before any response)
      - The pupil will start to dilate between roughly 0.5 to 1.3 s following the stimulus onset
      - The peak dilation occurs typically roughly 0.7 to 1 s following the end of the stimulus
      - An interval of 2 to 3 s during which no other events happen is recommended
    - Pupil size should return to the baseline level before the next trial starts
      - An intertrial interval of at least 3 s is recommended, to reduce (but not eliminate) carryover effects in the pupil response from one trial to the next



Events in a basic pupillometry experiment

Figure from: Winn et al., 2018, p. 8

Considerations on experimental design

- Procedure
  - Eye position should ideally be constant between conditions
    - Eye movements affect pupil size
      - For most eye trackers, the angle from which the camera records the eye affects the recorded pupil size (but not the actual pupil size)
      - Eye movements are followed by pupil constriction (presumably due to change in luminance)
    - Can be corrected using some algorithms developed by researchers, e.g., Gagl et al. (2011)
  - Combining pupillometry with eye-tracking
    - Pupil size data can be contaminated by changes in luminance and gaze position
    - Stimuli should be controlled to have equal luminance and presented before acoustic stimulus

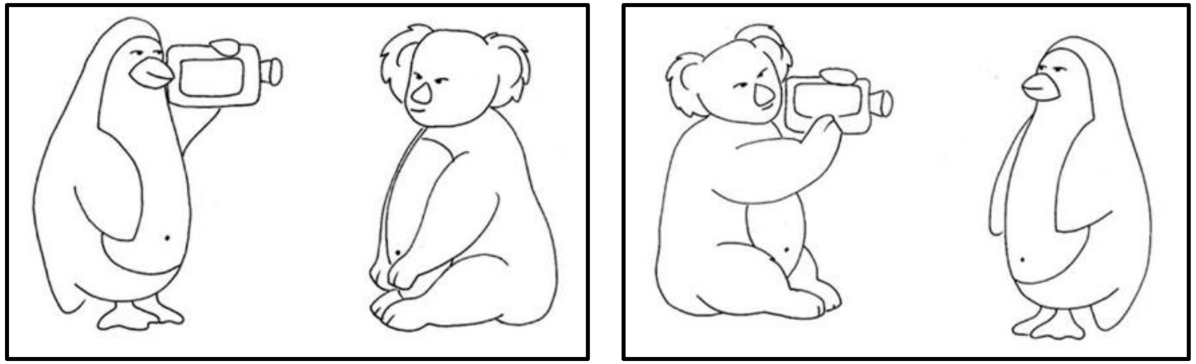


Figure from: Wendt et al., 2016, p. 4

Considerations on experimental design

- Environment
    - Visual field
        - Pupil will excessively constrict when it's too bright, and dilate when it's too dark
        - A moderate mid-range grey colour background on a computer monitor
    - Room lighting:
        - Soft lighting in the room is the best
        - Baseline pupil sizes between 3 and 7 mm are suitable
- Participants
    - Eye colour
        - Most eye trackers are robust to differences in iris colour, but there are occasional difficulties with very dark and very light irises
    - Age
        - Older listeners show generally weaker pupil dilation responses to light
        - A control task that measures pupil dynamic range is recommended when comparing younger and older adults
    - Hearing status
        - Smaller amounts of pupil dilation are routinely observed in listeners with hearing loss

Data processing

- Raw data
  - Sample rate: 30Hz or higher is sufficient.
    - You can acquire the data at a higher sampling rate and down-sample the data after preprocessing.
  - Monocular and binocular tracking: Both are okay.
    - You might have a chance to select the eye which provides better data quality if you use binocular tracking
- Preprocessing
  - De-blinking
    - Normal kind of blinks is transient and easy to detect
  - Interpolating
    - Connects ends of deleted data points
  - Filtering
    - Changes faster than 10 Hz are uncorrelated across the eyes
    - Low-pass filtering at 10 Hz, or n-point smoothing average filtering

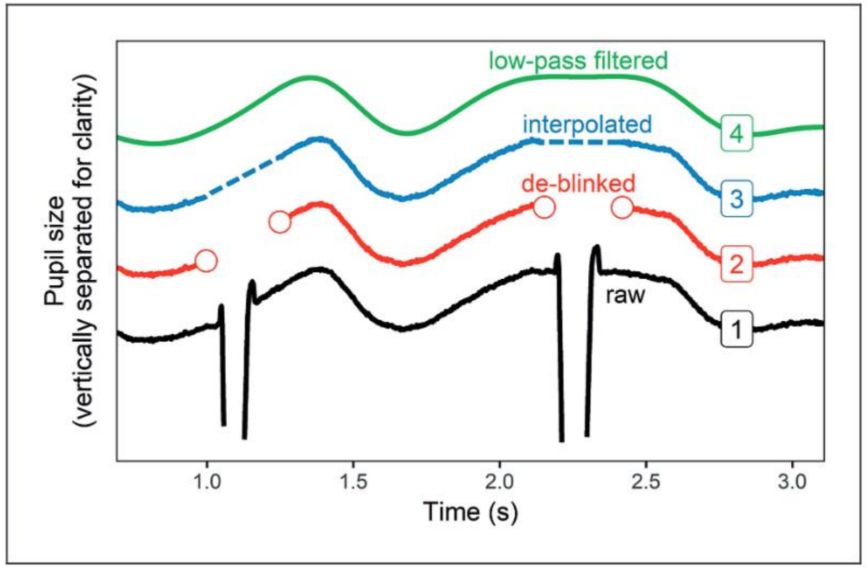


Figure from: Winn et al., 2018, p. 14

Data processing

- Baseline correction
  - Pupil size typically will vary across people, vary within people across time, and will gradually diminish over the course of a testing session
  - Common practice of baseline window is 1 s (100~2000 ms), though variation in the baseline duration should play no substantial role
  - Methods:
    - <u>Absolute subtraction</u>: each data point minus baseline average
    - <u>Proportional transformation</u>: each data point divided by baseline average, then minus 1

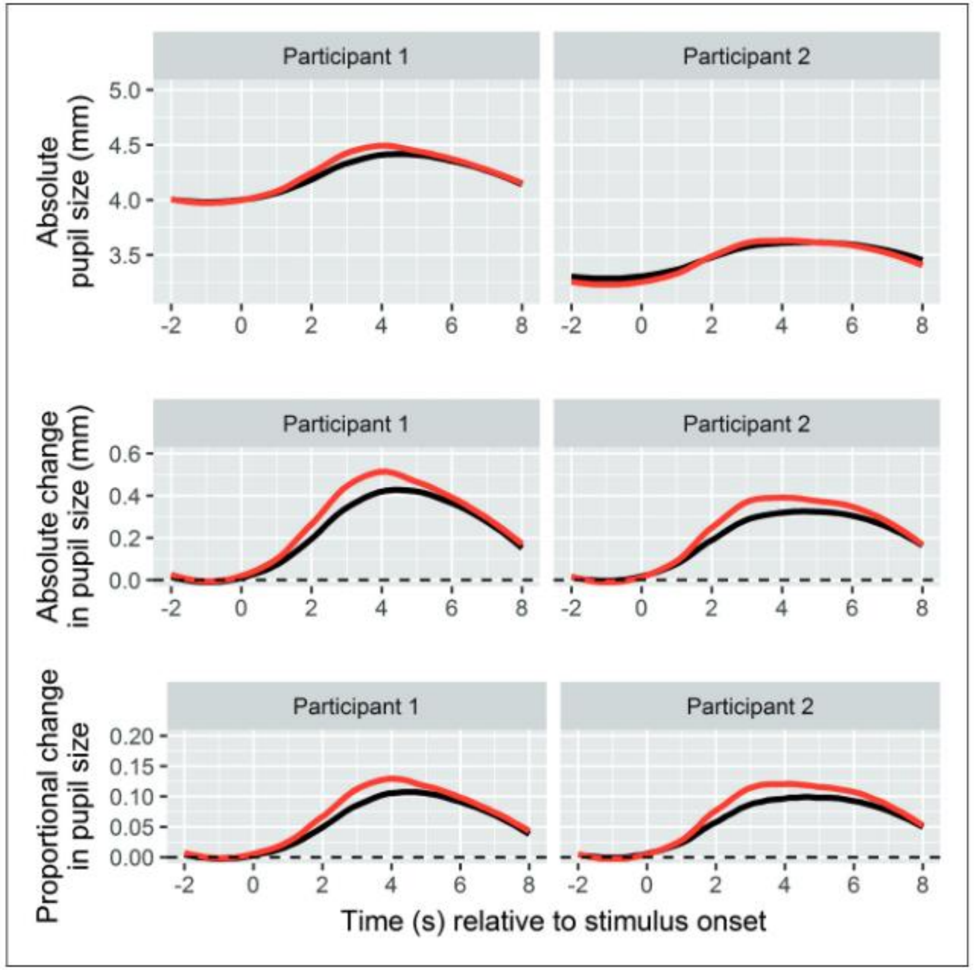


Figure from: Winn et al., 2018, p. 17

Data processing

- Normalisation
  - Older individuals tend to have pupils that are smaller in size, with a more restricted range of dilation, and which take longer to reach maximum dilation or constriction
  - Many different methods, e.g., change of pupil size as a proportion of the individual's dynamic range of pupil size (Piquado et al., 2010)

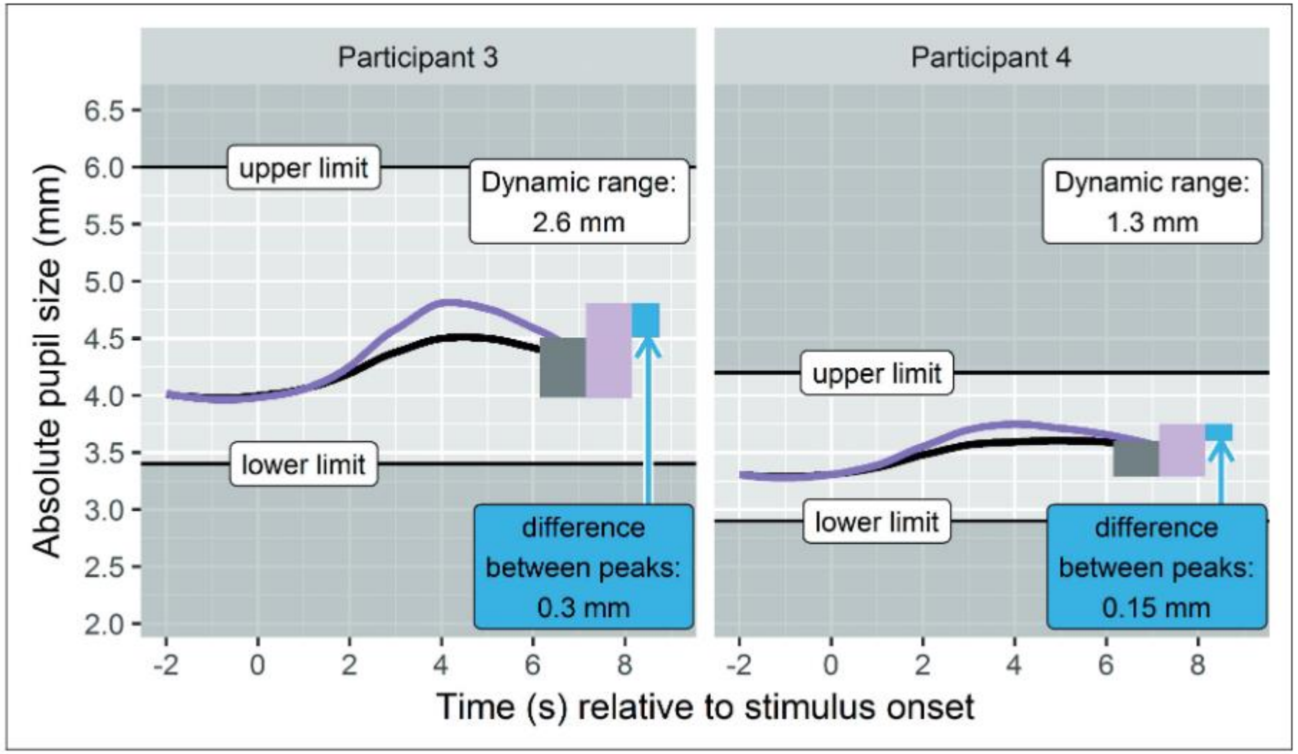
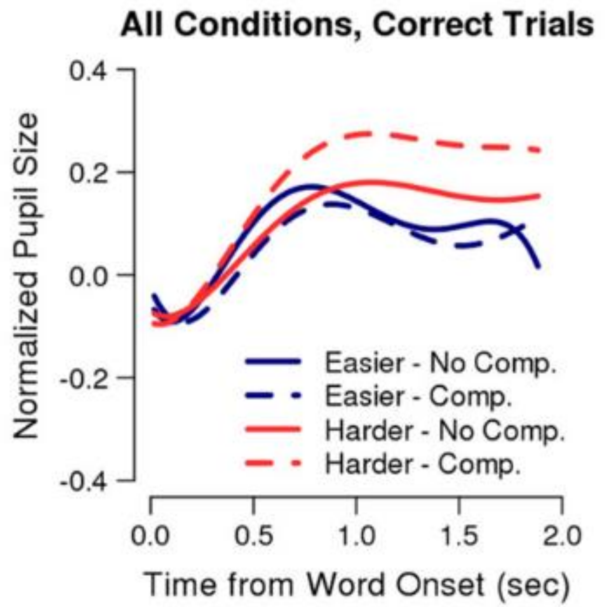


Figure from: Winn et al., 2018, p. 18

Data processing

- Statistics
  - Time window analysis
    - Peak pupil dilation, peak pupil latency, and mean pupil dilation in a fixed window of time around stimulus presentation
    - Using a predetermined time window
    - Cluster-based permutation testing (Mathôt & Vilotijević, 2023)
  - Time course analysis
    - Analyse the shape of the pupillary response over time
    - Method:
      - Growth-curve analysis (Mirman, 2014) https://www.danmirman.org/gca
      - Generalised additive models (van Rij et al., 2019)



Linear
Quadratic
Cubic
Quartic
Quintic

Figure from: Kuchinsky et al., 2013, p. 30

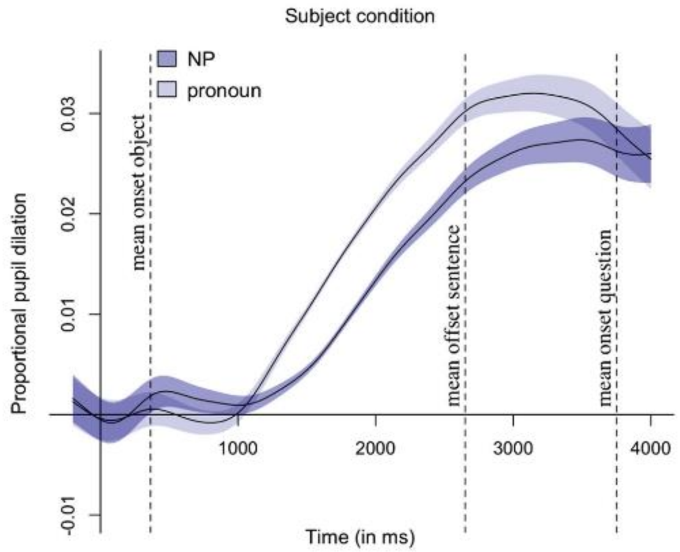


Figure from: Vogelzang et al., 2016, p. 882

Readings and resources

- Guideline papers
  - Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best Practices and Advice for Using Pupillometry to Measure Listening Effort: An Introduction for Those Who Want to Get Started. *Trends in Hearing*, *22*, 1–32. https://doi.org/10.1177/2331216518800869
  - Mathôt, S., & Vilotijević, A. (2023). Methods in cognitive pupillometry: Design, preprocessing, and statistical analysis. *Behavior Research Methods*, *55*(6), 3055–3077. https://doi.org/10.3758/s13428-022-01957-7
- Review papers
  - Schmidtke, J. (2018). Pupillometry in linguistic research: An introduction and review for second language researchers. *Studies in Second Language Acquisition*, *40*(3), 529–549. https://doi.org/10.1017/S0272263117000195
  - van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, *25*(6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y
- Tutorials
  - Drew J. McLaughlin: Building EyeLink-Compatible PsychoPy Experiments, Preparing Pupillometry Data with R https://sites.google.com/view/drewjmclaughlin/pupillometry
  - Geller, J., Winn, M. B., Mahr, T., & Mirman, D. (2020). GazeR: A Package for Processing Gaze Position and Pupil Size Data. *Behavior Research Methods*, 52(5), 2232–2255. https://doi.org/10.3758/s13428-020-01374-8
  - van Rij, J., Hendriks, P., van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the Time Course of Pupillometric Data. *Trends in Hearing*, *23*, 2331216519832483. https://doi.org/10.1177/2331216519832483

References

Ben-Nun, Y. (1986). The use of pupillometry in the study of on-line verbal processing: Evidence for depths of processing. *Brain and Language*, *28*(1), 1–11. https://doi.org/10.1016/0093-934X(86)90086-6

Engelhardt, P. E., Ferreira, F., & Patsenko, E. G. (2010). Pupillometry reveals processing load during spoken language comprehension. *Quarterly Journal of Experimental Psychology*, *63*(4), 639–645. https://doi.org/10.1080/17470210903469864

Gagl, B., Hawelka, S., & Hutzler, F. (2011). Systematic influence of gaze position on pupil size measurement: Analysis and correction. *Behavior Research Methods*, *43*(4), 1171–1181. https://doi.org/10.3758/s13428-011-0109-5

Geller, J., Winn, M. B., Mahr, T., & Mirman, D. (2020). GazeR: A Package for Processing Gaze Position and Pupil Size Data. *Behavior Research Methods*, *52*(5), 2232–2255. https://doi.org/10.3758/s13428-020-01374-8

Kuchinsky, S. E., Ahlstrom, J. B., Vaden JR., K. I., Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. A. (2013). Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology*, *50*(1), 23–34. https://doi.org/10.1111/j.1469-8986.2012.01477.x

Laurino, J., Traverso, A., Jurado, D., & Kaczer, L. (2023). *The role of context in semantic ambiguity processing: Insights from behavioural and pupillometry measures* [Poster presentation]. The 2023 Meeting of Argentine Society for Research in Neurosciences, San Luis, Argentina.

Ledoux, K., Coderre, E., Bosley, L., Buz, E., Gangopadhyay, I., & Gordon, B. (2016). The concurrent use of three implicit measures (eye movements, pupillometry, and event-related potentials) to assess receptive vocabulary knowledge in normal adults. *Behavior Research Methods*, *48*(1), 285–305. https://doi.org/10.3758/s13428-015-0571-6

Mathôt, S., & Vilotijević, A. (2023). Methods in cognitive pupillometry: Design, preprocessing, and statistical analysis. *Behavior Research Methods*, *55*(6), 3055–3077. https://doi.org/10.3758/s13428-022-01957-7

Mirman, D. (2017). *Growth Curve Analysis and Visualization Using R*. CRC Press.

Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology*, *47*(3), 560–569. https://doi.org/10.1111/j.1469-8986.2009.00947.x

References

Schmidtke, J. (2014). Second language experience modulates word retrieval effort in bilinguals: Evidence from pupillometry. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00137

Schmidtke, J. (2018). Pupillometry in linguistic research: An introduction and review for second language researchers. *Studies in Second Language Acquisition*, *40*(3), 529–549. https://doi.org/10.1017/S0272263117000195

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, *25*(6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y

van Rij, J., Hendriks, P., van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the Time Course of Pupillometric Data. *Trends in Hearing*, *23*, 2331216519832483. https://doi.org/10.1177/2331216519832483

Vogelzang, M., Hendriks, P., & van Rijn, H. (2016). Pupillary responses reflect ambiguity resolution in pronoun processing. *Language, Cognition and Neuroscience*, *31*(7), 876–885. https://doi.org/10.1080/23273798.2016.1155718

Wendt, D., Dau, T., & Hjortkjær, J. (2016). Impact of Background Noise and Sentence Complexity on Processing Demands during Sentence Comprehension. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00345

Winn, M. B. (2016). Rapid Release From Listening Effort Resulting From Semantic Context, and Effects of Spectral Degradation and Cochlear Implants. *Trends in Hearing*, *20*, 2331216516669723. https://doi.org/10.1177/2331216516669723

Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best Practices and Advice for Using Pupillometry to Measure Listening Effort: An Introduction for Those Who Want to Get Started. *Trends in Hearing*, *22*, 1–32. https://doi.org/10.1177/2331216518800869

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil Response as an Indication of Effortful Listening: The Influence of Sentence Intelligibility. *Ear and Hearing*, *31*(4), 480. https://doi.org/10.1097/AUD.0b013e3181d4f251