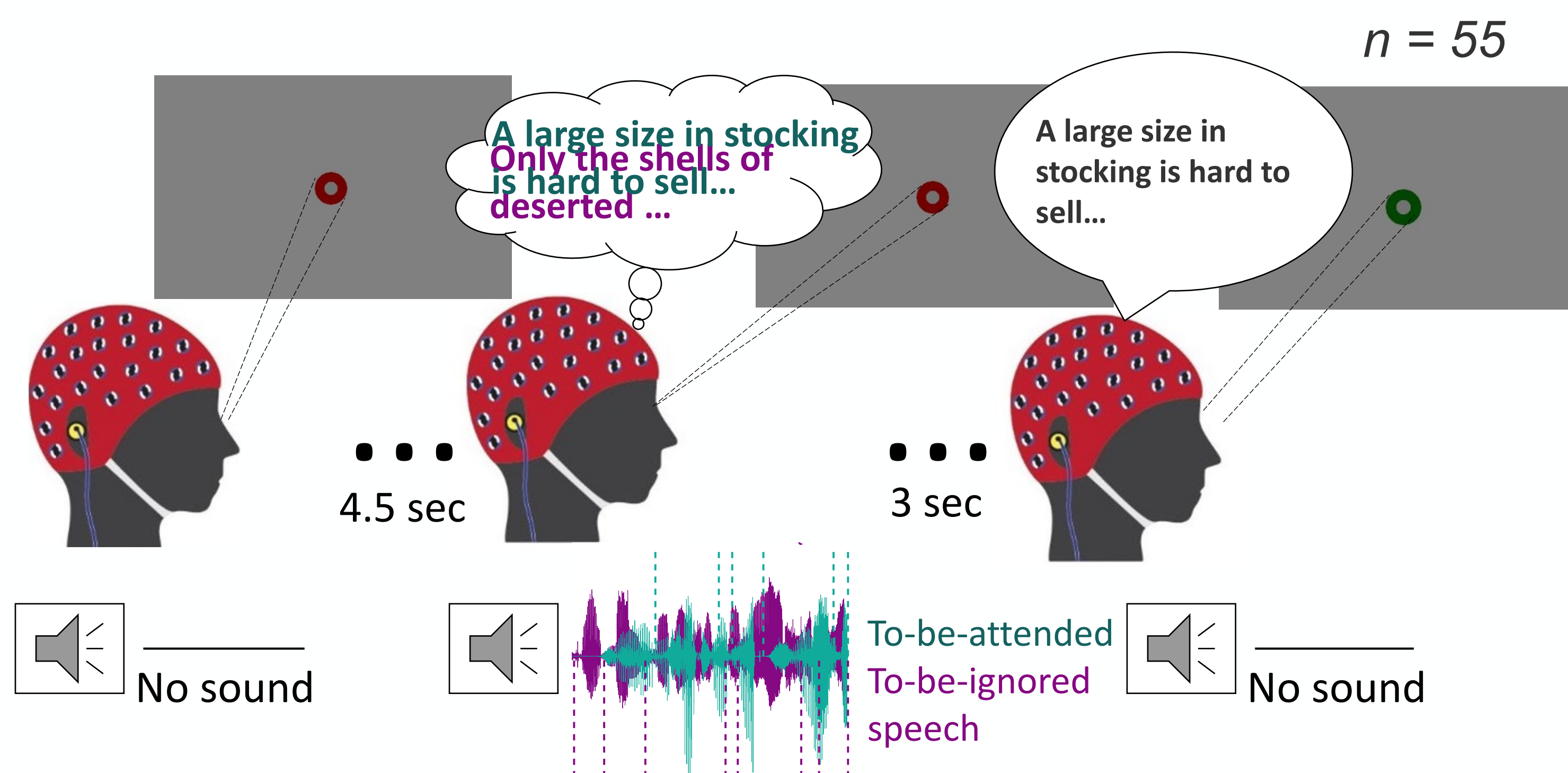


INTRODUCTION

- Pupillometry is commonly used to quantify listening effort (LE) in speech-in-noise perception tasks (Generally, more LE → Higher pupil dilation).
- However, task-evoked pupillometric responses (TEPR) fatigue over the course of the experimental session, introducing (to most researchers) undesirable variance into the data.
- **It is unclear if TEPR fatigue modulates effects of experimental manipulations** (e.g., does fatigue interact change of SNR affects TEPR). Such interactions would have important implications for the design choices in pupillometric paradigms (e.g., experiment duration).
- Here we examine this possible interaction effect in the context of a sentence recognition task with speech-on-speech masking and an SNR manipulation of LE.

METHOD

- **Paradigm:** Diotic speech-in-speech recognition at four SNR levels (5, 0, -5, and -10 dB SNR) with 50 trials/condition, performed over 10 blocks of 20 trials. SNR varied on a trial-by-trial basis.
- **Stimuli:** IEEE target sentences with simultaneous audiobook distractor.
- Distractor level was fixed at 65 dB SPL, and in each trial its onset preceded target audio by one second to facilitate streaming.



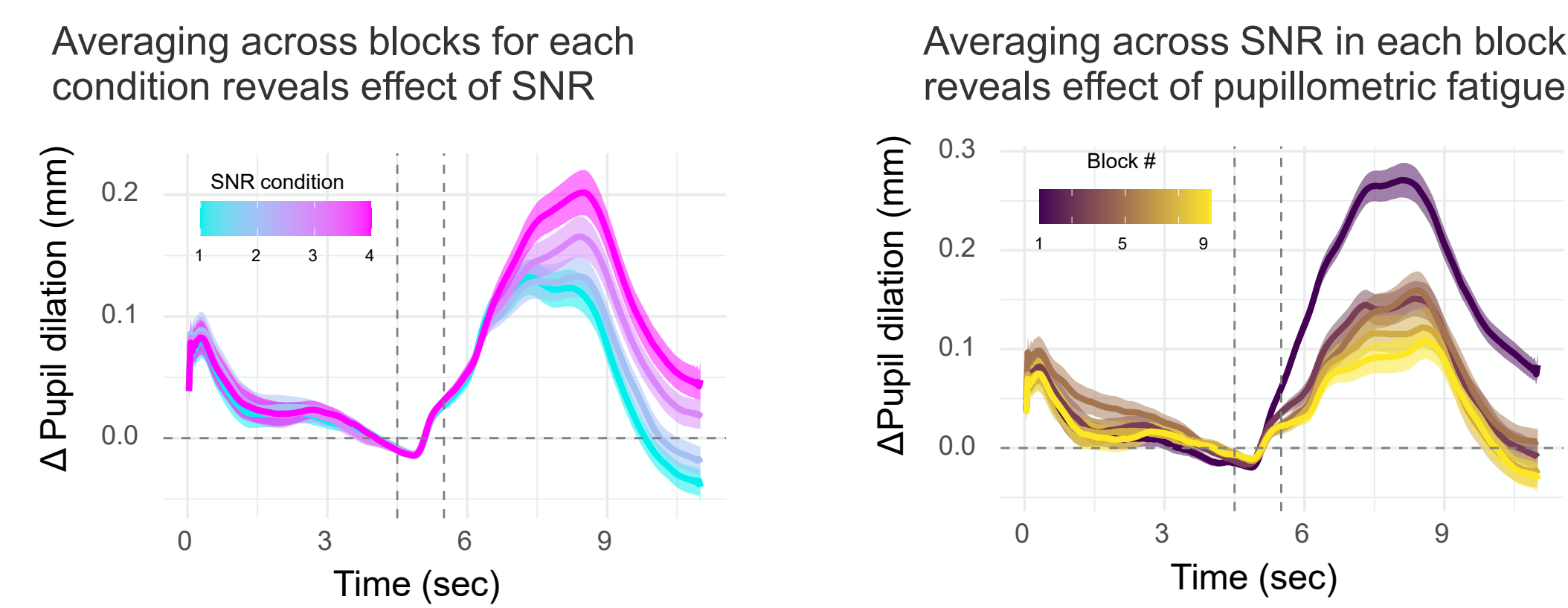
Behavioral Measures: Speech intelligibility and self-reported LE

Pupillometry Data: Tobii Pro Spectrum with binocular tracking at 300 Hz

Preprocessing: PUPIL Toolbox (Relaño-Iborra & Bækgaard, 2020) used to preprocess the pupil dilation and correct for eye blink-related distortions. Data was downsampled to 30 Hz for LME analyses, and 10 Hz for GAMM analyses

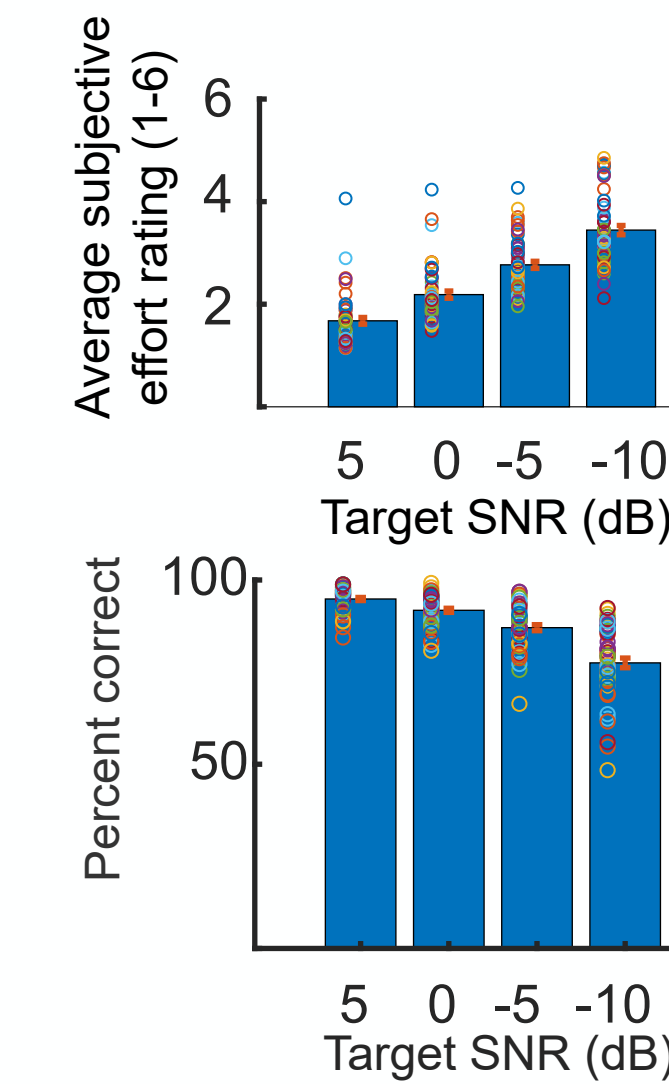
Each trial was baseline corrected at -1.5 to 0 sec relative to target speaker onset. Eye with higher average % of valid samples was used from each participant.

TEPR GRAND AVERAGE

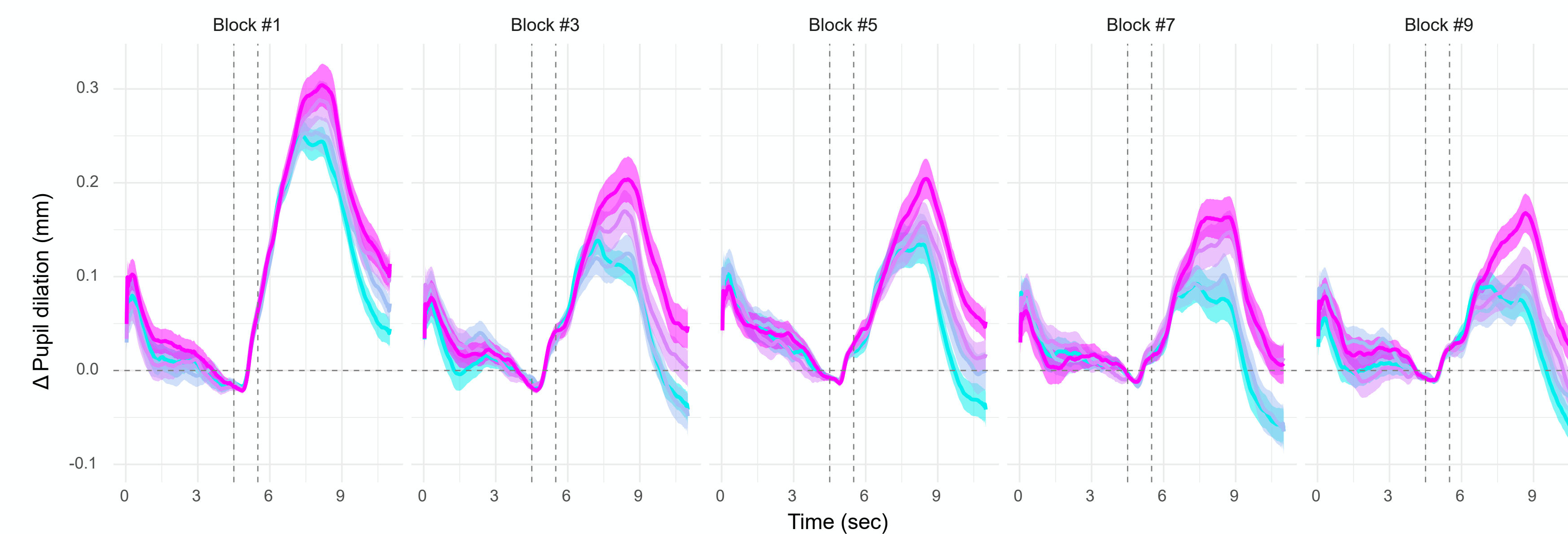


BEHAVIOR

Consistent with past work, behavioral data shows that more adverse SNR conditions were associated with higher subjective LE ratings, and moderate reduction in intelligibility.



While the experimentally manipulated variable, SNR, has a robust effect on pupil dilation, so does experimental block (i.e., time within the session). The fatigue effect appears rapid early on, and more gradual later. If pupillometric fatigue caused a reduction in the effect of SNR, this would lead to diminishing returns from longer session. Does it?



Visual inspection suggests that SNR effects are preserved across the entire duration of the experiment. If anything, the SNR effect may be weaker in the first block (maybe effect of task novelty?). Next goal: Statistically assess whether this is really the case.

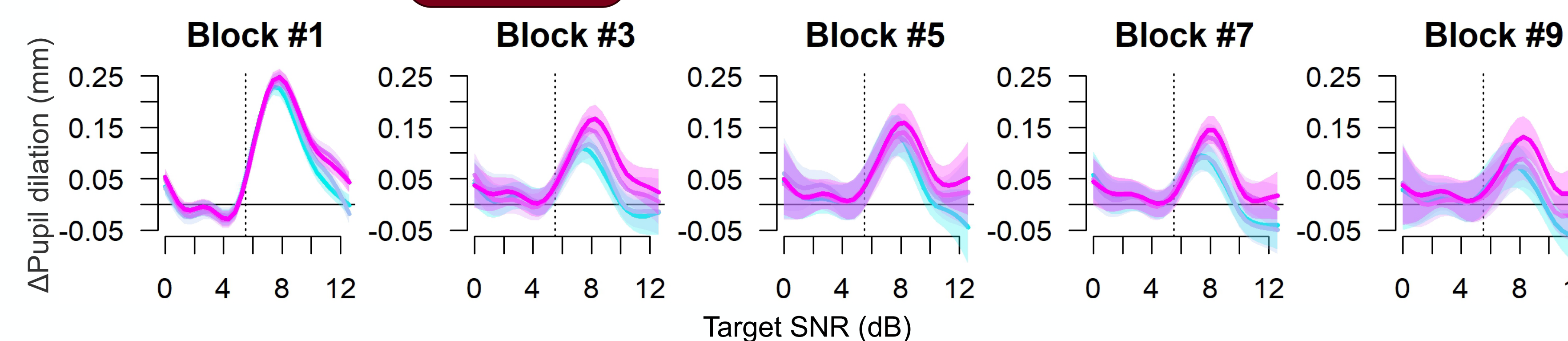
PRELIMINARY GAMM ANALYSIS (WORK-IN-PROGRESS)

Model: $pupil \sim s(\text{timeSeconds}) + s(\text{snr}, \text{timeSeconds}, \text{bs} = "sz") + s(\text{block}, \text{timeSeconds}, \text{bs} = "sz") + s(\text{block}, \text{snr}, \text{timeSeconds}, \text{bs} = "sz", \text{id} = 1) + s(\text{subject}, \text{timeSeconds}, \text{by} = \text{block}, \text{fs} = "fs")$

Why GAMMs?

- Pupil response is observed to be nonlinear over time
- Pupil size varies by subject
- Variability in SNR response exists over time

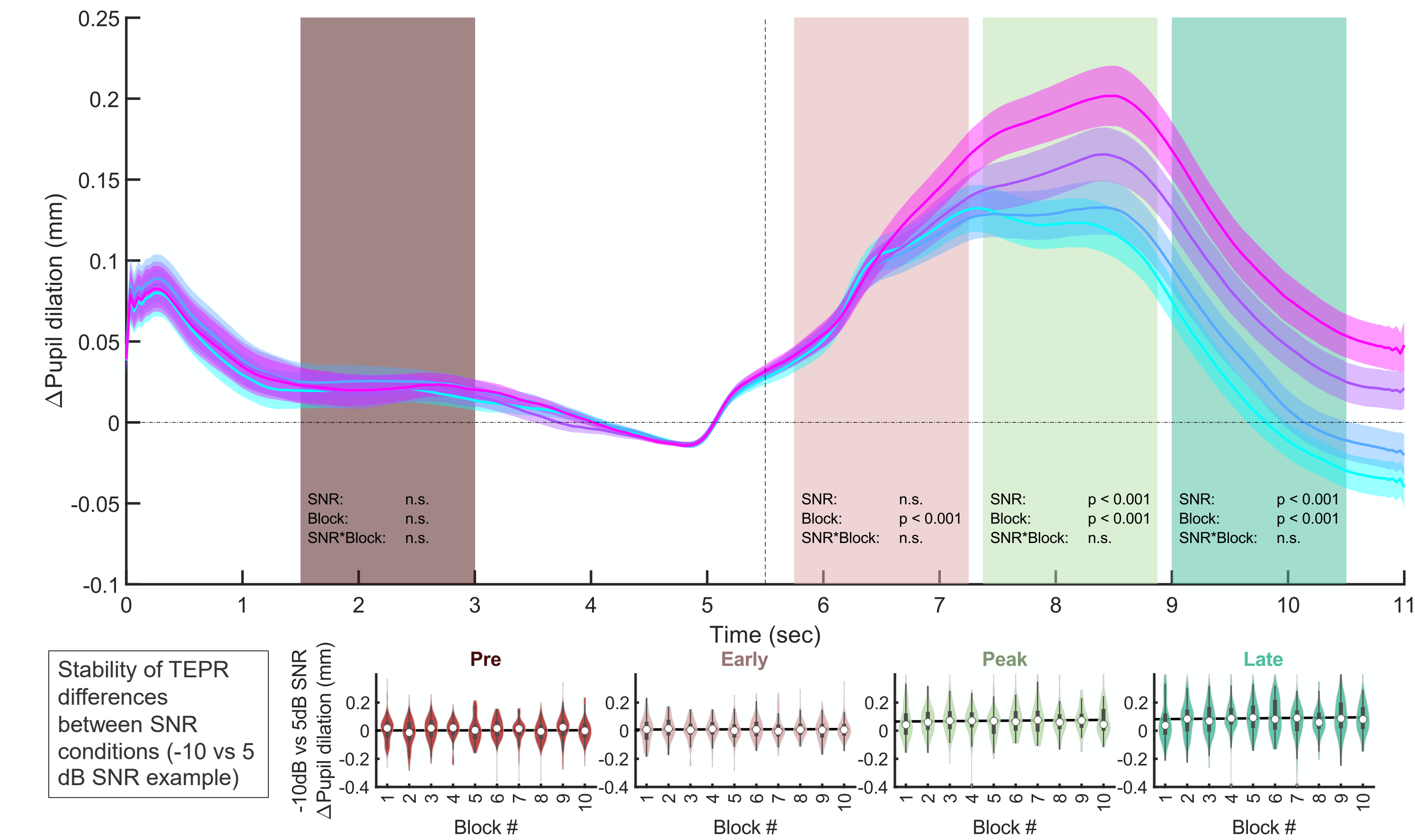
Can observe how much SNR response variability exists across block



Preliminary GAMM model captures general patterns of SNR and fatigue effects on TEPR. However, work is ongoing on finalizing model specification and assessing if differences between SNR conditions remain stable across blocks. Advice and suggestions from GAMM experts are welcome and appreciated!

LME ANALYSIS

- LME analysis used to test whether fatigue and SNR interact during 4 pre-selected periods of TEPR:
 - "Pre-onset", "Early rise", "Peak", "Late decrease"
- Model: $pupil \sim SNR * Block + (1 + SNR + Block | Subject)$



LME analysis of averaged TEPR time course segments showed strong effects of SNR and block number, but no interaction in any segment. Same outcome in offset-aligned TEPR analysis.

While encouraging (for long sessions), the coarse nature of this analysis may miss temporally more localized effects. A more suitable approach: GAMMs, which allow for modelling of non-linear patterns, while accounting for autocorrelation inherent in the data.

DISCUSSION

- TEPR is a commonly used measure of LE, but is prone to pupillometric fatigue that is difficult to mitigate through experimenter interventions (McLaughlin et al., 2023).
- Our data and preliminary analyses are encouraging in suggesting that at least in traditional sentence-based paradigms with SNR manipulations of LE, TEPR fatigue may be independent of effects of SNR (in ~2.5 hr sessions).
- However, recent work by Widmann et al. (2025) showed that the sensitivity of TEPR to experimental condition can diminish over time in a continuous speech paradigm with long (10+ min) speech narratives.
- Ongoing GAMM analyses will explore how the fine-scale time course of TEPR is affected by the time-within-experiment.
- Future studies should assess effects of TEPR fatigue in different paradigms and investigate the cognitive and physiological contributors to TEPR fatigue.

Baayen, R. H., Fasiolo, M., Wood, S., & Chuang, Y.-Y. (n.d.). *A note on the modeling of the effects of experimental time in psycholinguistic experiments.* *Journal of Experimental Psychology: Applied*, 29, 1-12. <https://doi.org/10.1037/xap0000425>

McLaughlin, D. J., Zink, M. E., Gaunt, L., Reilly, J., Sommers, M. S., Van Engen, K. J., & Peelle, J. E. (2023). Give me a break! Unavoidable fatigue effects in cognitive pupillometry. *Psychophysiology*, 60(7), e14256. <https://doi.org/10.1111/psyp.14256>

Widmann, A., Herrmann, B., & Scharf, F. (2025). Pupillometry is sensitive to speech masking during story listening: A commentary on the critical role of modeling temporal trends. *Journal of Neuroscience Methods*, 413, 110299. <https://doi.org/10.1016/j.jneumeth.2024.110299>

Winn, M. B. (2023). Time Scales and Moments of Listening Effort Revealed in Pupillometry. *Seminars in Hearing*, 44(02), 105-123. <https://doi.org/10.1055/s-0043-178741>

Relaño-Iborra, H., & Bækgaard, P. (2020). PUPILS pipeline: A flexible Matlab toolbox for eye-tracking and pupillometry data processing (No. arXiv:2011.05118). arXiv: <https://arxiv.org/abs/2011.05118>

Smith, M. L., & Winn, M. B. (2025). Repairing Misperceptions of Words Early in a Sentence is More Effortful Than Repairing Later Words, Especially for Listeners With Cochlear Implants. *Trends in Hearing*, 29, 23312165251320789. <https://doi.org/10.1177/23312165251320789>

Wood, S. N., Coude, Y., & Shaw, S. (2015). Generalized Additive Models for Large Data Sets. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 64(1), 139-155. <https://doi.org/10.1111/rssc.12068>