

# Listening effort under two types of auditory masking conditions, as measured by pupillometry

Ayesha Alam, Tyler Perrachione, Gerald Kidd, Jr., & Sarah Villard

Department of Speech, Language, and Hearing Sciences at Boston University



## BACKGROUND

### Cocktail Party Problem/Auditory Masking

- The task of attempting to selectively listen to and understand an auditory target in the presence of distracting sounds, or maskers, is known as “the cocktail party problem” (Cherry, 1953).
- Many experiments on the cocktail party problem have focused on trying to identify the target-to-masker ratio (TMR) at which a listener can understand a certain percentage of a target utterance under specific auditory masking conditions (whether noise or speech).
- However, even though TMRs provide important information about speech intelligibility, they do not provide meaningful information about the important topic of listening effort.

### What is Listening Effort and what is Pupillometry?

- The term “listening effort” has been defined as “the mental exertion required to attend to, and understand, an auditory message” (Miles et al., 2017). Two listeners can achieve the same results (e.g., same TMRs) on a speech intelligibility task, but exert different amounts of effort to do so.
- Measuring listening effort can provide us with a better understanding of the resources that a listener requires in order to be successful in a given listening situation, which in turn may affect the availability of resources for other ongoing cognitive-communicative tasks.
- One way to measure listening effort is through the use of pupillometry, or change (dilation) in pupil size over time.

### Energetic Masking (EM) vs. Informational Masking (IM)

- Listening effort may differ depending on what type of masking is present.
- Two types of auditory masking have been identified, energetic masking (EM) and informational masking (IM).
- EM is masking caused by spectrotemporal overlap between target and masker energy. It is based in the peripheral auditory system.
- IM is additional masking that cannot be accounted for by EM. It is often the result of confusion between target and masker and is considered to be related to central processing (Kidd & Colburn, 2017).

## PARTICIPANTS

- Participants were recruited from the Boston University community through online postings and word of mouth.
- All participants were native English speakers, did not have a diagnosis of ADHD or ADD, and did not have sustained any past head injury that resulted in a loss of consciousness. In addition, the participants were administered a hearing screening to ensure that their hearing is within normal limits for the study.

Subject	Gender	Age	Race
YNH1	M	19	Asian
YNH2	F	23	White
YNH3	F	19	White
YNH4	F	20	White
YNH5	F	22	White
YNH6	M	18	Asian and White
YNH7	F	21	Asian
YNH8	F	20	Asian
YNH9	M	20	Asian and White
YNH10	F	21	White
YNH11	F	21	Black
YNH12	M	21	Asian
YNH13	F	24	White
YNH14	F	23	Asian

## STIMULI

The auditory stimuli in this experiment consisted of single-word recordings spoken by eight different female talkers and concatenated into five-word sentences. Table 1 below shows all the words in the experimental corpus.

Subject	Verb	Number	Adjective	Object
Bob	bought	2	big	bags
Gene	found	3	cheap	cards
Jane	gave	4	green	gloves
Jill	held	5	hot	hats
Lynn	lost	6	new	pens
Mike	saw	8	old	shoes
Pat	sold	9	red	socks
<b>Sue</b>	took	10	small	toys

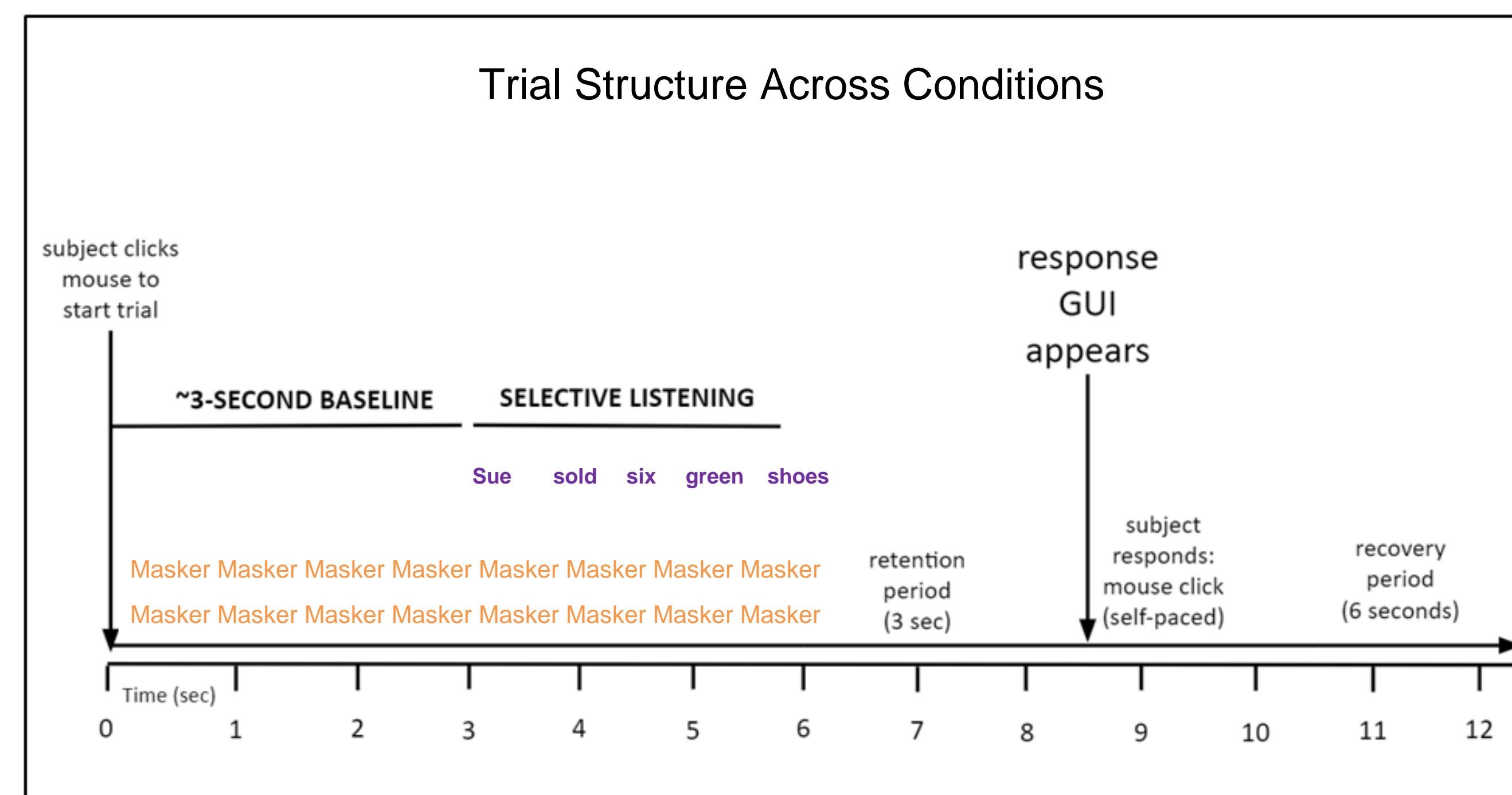
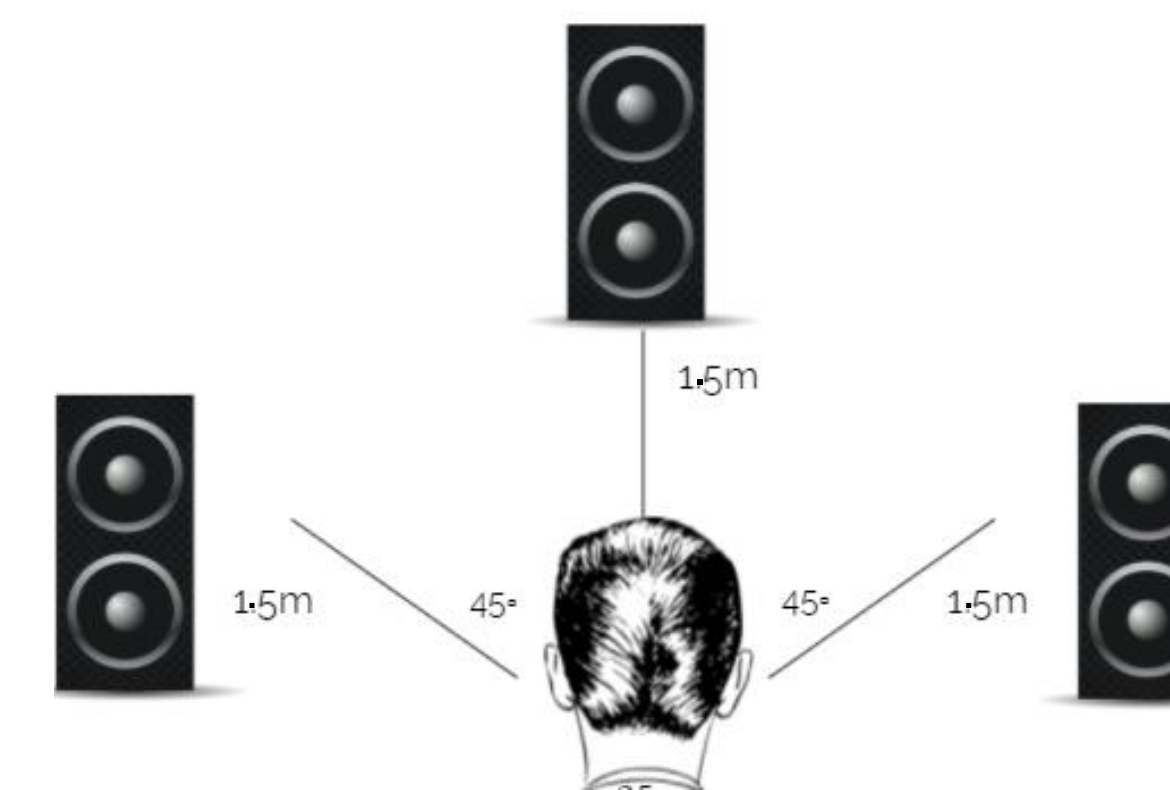
Table 1 (Note: “Sue” is bolded because it will always be the start of the target sentence.)

## CONDITIONS

Participants listened to target sentences drawn from the corpus above and starting with the word “Sue” (e.g., “Sue saw 8 red pens”), in three conditions, each involving a different type of masking:

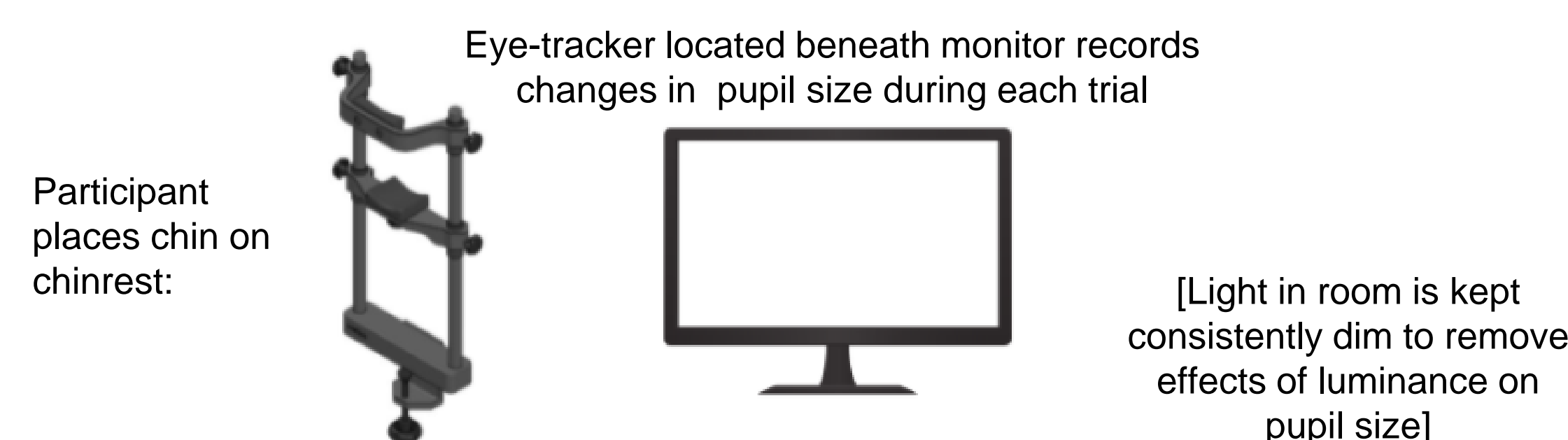
ISM	SSSMN	SNM
Intelligible Speech Maskers (ISM): maskers consist of 2 intelligible target sentences drawn from the same corpus as the target sentence, each one preceded by three names which comprise the baseline (e.g., “William, Peter, Kathy, Sue sold 3 green shoes”).	Speech Shaped, Speech Envelope Modulated Noise Maskers (SSSMN): maskers consist of speech-shaped noise that is modulated using the envelopes of words in the corpus.	Stationary Noise Maskers (SNM): maskers consist of the same speech-shaped noise played in the SSSMN condition, but unmodulated.
High-IM	High-EM	High-EM

Stimuli were presented through three loudspeakers, each placed approximately 1.5 meters from the listener’s head. A loudspeaker located straight in front of the listener (0 degrees azimuth) always presented the target sentence, and two loudspeakers located 45 degrees to the left and right of the listener (+/- 45 degrees azimuth) always presented the two (simultaneous) maskers.



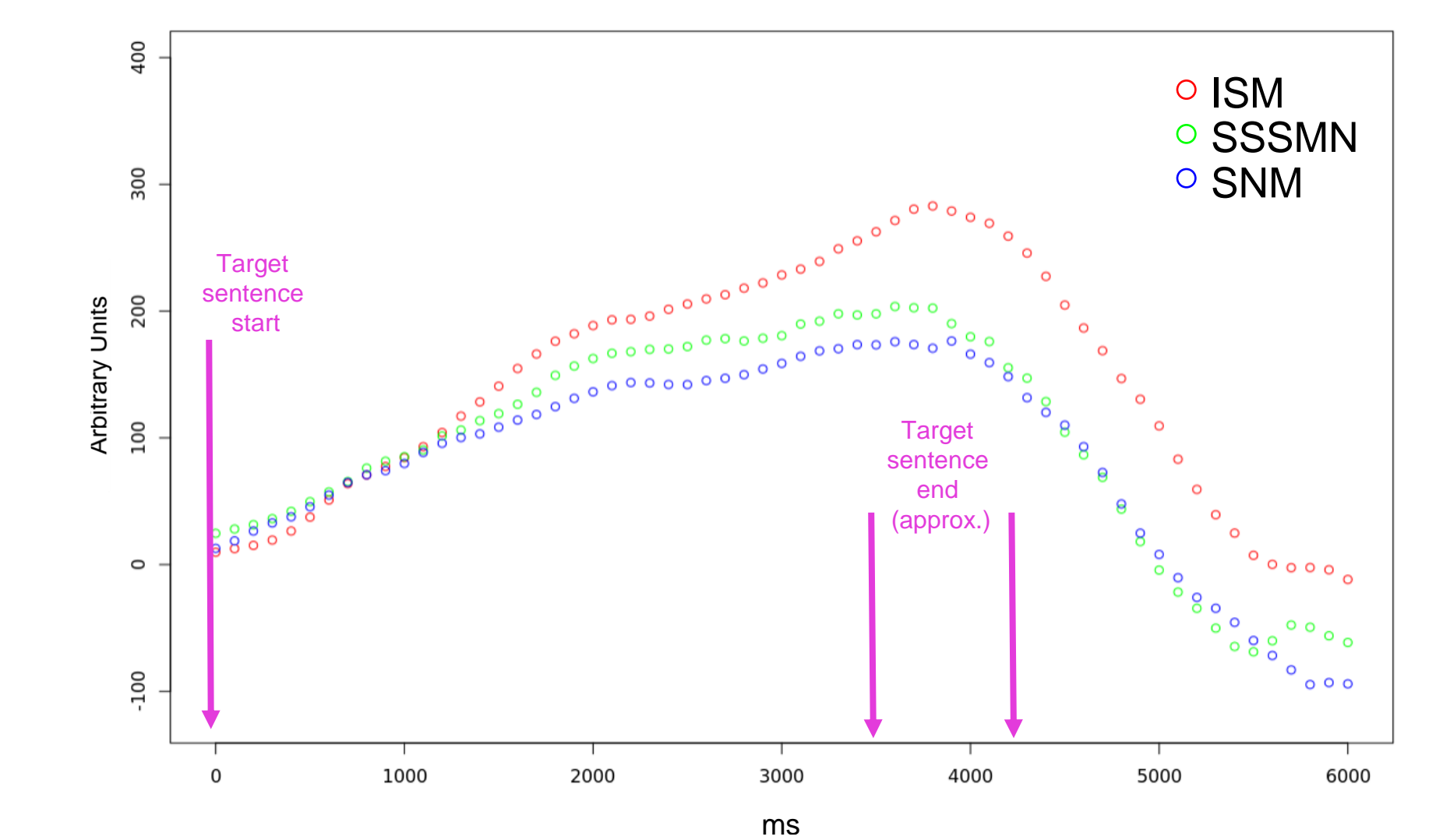
## PROCEDURES

- Participants completed three 1-up, 1-down adaptive tracks in each condition, resulting in estimates of the TMR representing their individual 50% correct point on the psychometric function for each condition.
- Participants completed 24 trials in each condition at their individually-determined 50% correct TMR, with changes in pupil size recorded.

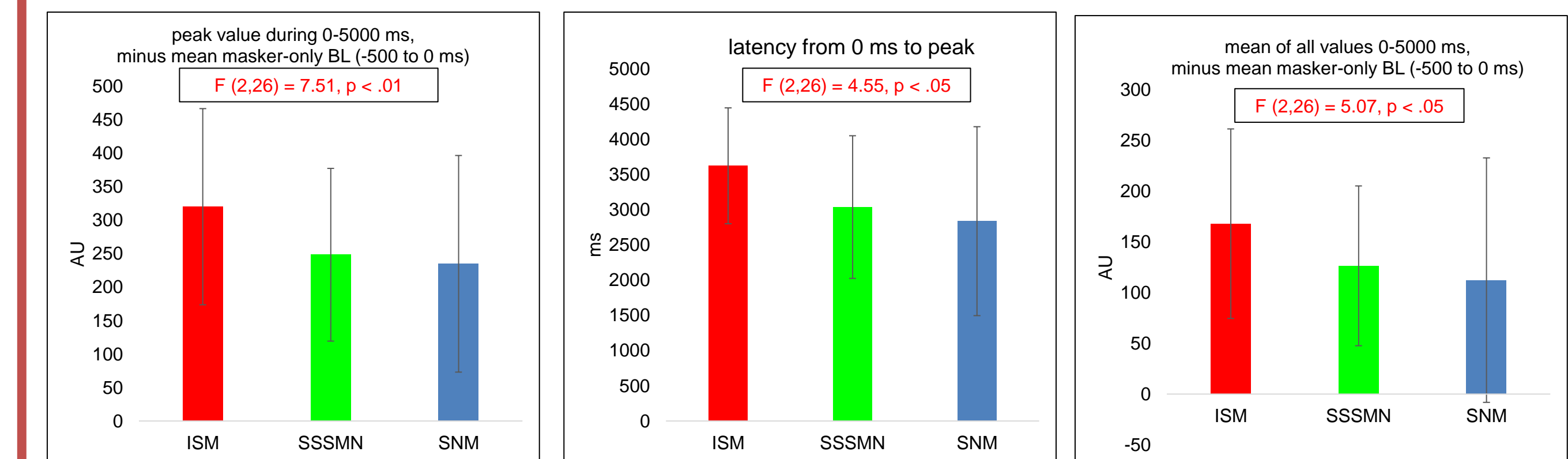


## RESULTS

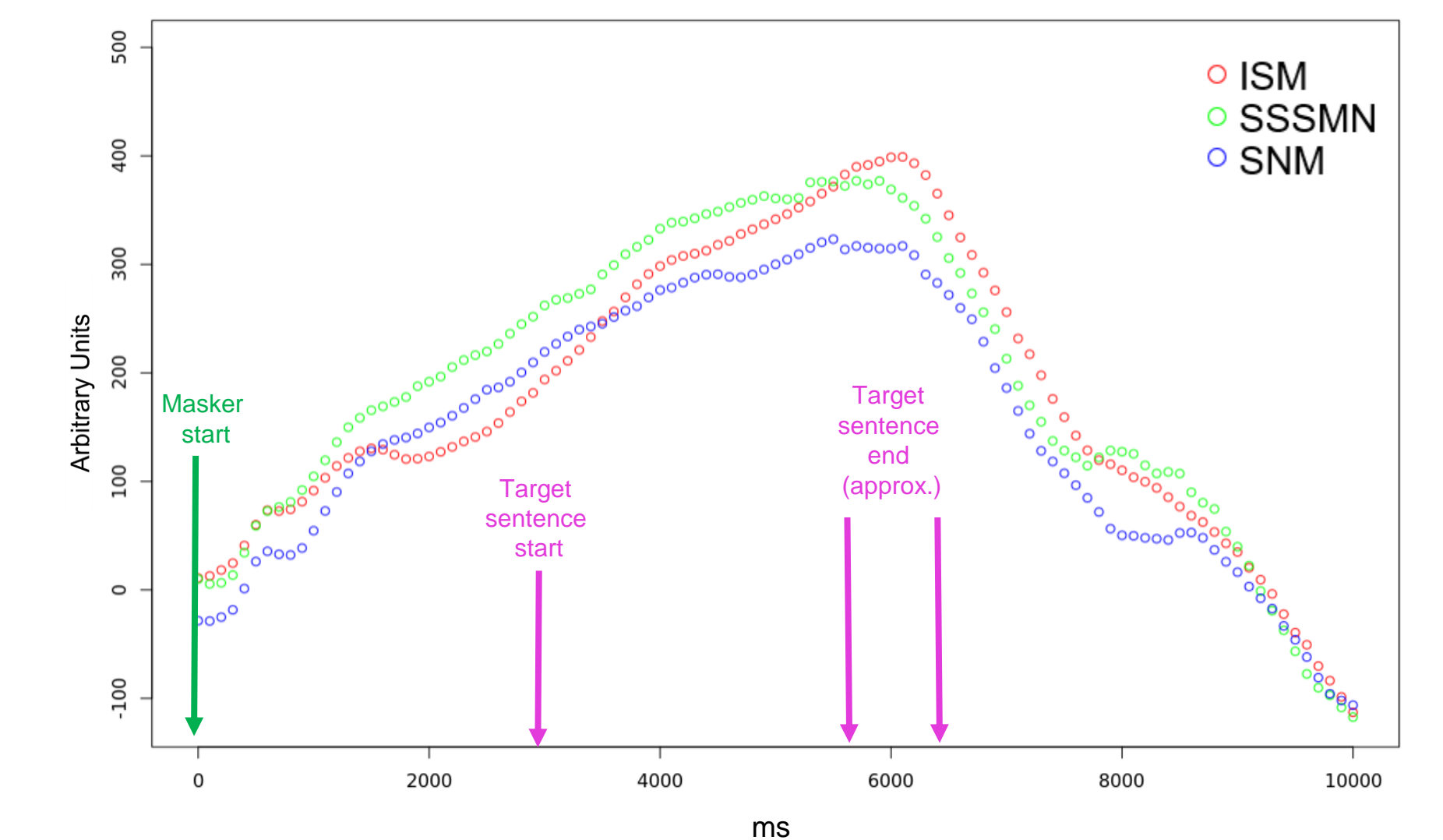
Pupil traces over time for each condition, relative to 500 ms masker-only baseline (baseline calculated separately for each condition)



3 repeated-measures ANOVAs were conducted to compare peak value, latency to peak, and mean value (relative to masker-only baseline) between the 3 conditions:



Pupil traces over time for each condition, relative to 500 ms SILENT baseline prior to masker start (baseline calculated as a grand average across conditions)



## DISCUSSION

- Results suggest that selectively listening to a target sentence in a high-IM condition (ISM) elicited greater changes in pupil size, relative to baseline, than in either of the high-EM conditions (SSSMN, SNM). Significant differences in pupil peak, latency to peak, and mean pupil size were observed.
- However, some of these differences may have been a result of the fact that during the masker-only baseline, pupil sizes were somewhat smaller in the ISM condition than in the other two conditions. Further research is needed to better understand the reasons behind this.
- These results help establish a foundation for future studies of effort in individuals with hearing loss and/or cognitive-linguistic deficits (e.g., aphasia).

## REFERENCES

- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, 25(5), 975-979.
- Kidd, G., & Colburn, H. S. (2017). Informational masking in speech recognition. In J. C. Middlebrooks, J. Z. Simon, A. N. Popper, & R. R. Fay (Eds.), *The auditory system at the cocktail party* (pp. 75-109). New York: Springer.
- Miles, K., McMahon, C., Boisvert, I., Ibrahim, R., de Lissa, P., Graham, P., & Lyxell, B. (2017). Objective assessment of listening effort: Coregistration of pupillometry and EEG. *Trends in Hearing*, 21, 1-13.

## ACKNOWLEDGEMENTS

The authors would like to thank the members of the Psychoacoustics Laboratory and Communication Neuroscience Laboratory at Boston University, as well as our funding source, NIH/NIDCD K99DC018829.