



Poster ID
T155

Effects of Contextual Cues on Speech Recognition and Spatial Release from Masking in Children with Typical Hearing using Vocoded Speech



*Nimesha Didulani Dantanarayana, Shelly P. Godar, Sara M. Misurelli, Ruth Y. Litovsky
University of Wisconsin-Madison, Waisman Center, Madison, Wisconsin

*Presenting Author e-mail : ndantanaraya@wisc.edu



Association for Research in Otolaryngology - 2024
Anaheim - California

Introduction

- Children with cochlear implants (CIs) show significant variability in outcome measures including speech understanding in quiet and in noise. Numerous factors can contribute to such variability, including;
 - Auditory experience prior to the onset of deafness and implantation
 - Downstream effects of deafness including;
 - Neurocognitive abilities
 - Neural health
 - Integrity of the auditory pathway
- Much of the research to date on speech understanding in children with CIs has utilized standardized test lists or sentence materials consisting of highly predictable content.
- However, when navigating realistic listening situations, children are likely to be presented with information that is complex and unpredictable. To fully understand speech understanding outcomes in children with CIs, the content of sentences should vary in level of predictability, or context.
- To investigate the extent to which children with bilateral CIs benefit from spatial separation of target speech from background noise, spatial release from masking (SRM) was investigated, comparing performance on sentences with high- and low-predictability¹.
- Prior to testing children with CIs, SRM was examined in a previous study for children with TH using sentences vary in the level of predictability, or context (Figure 1)².

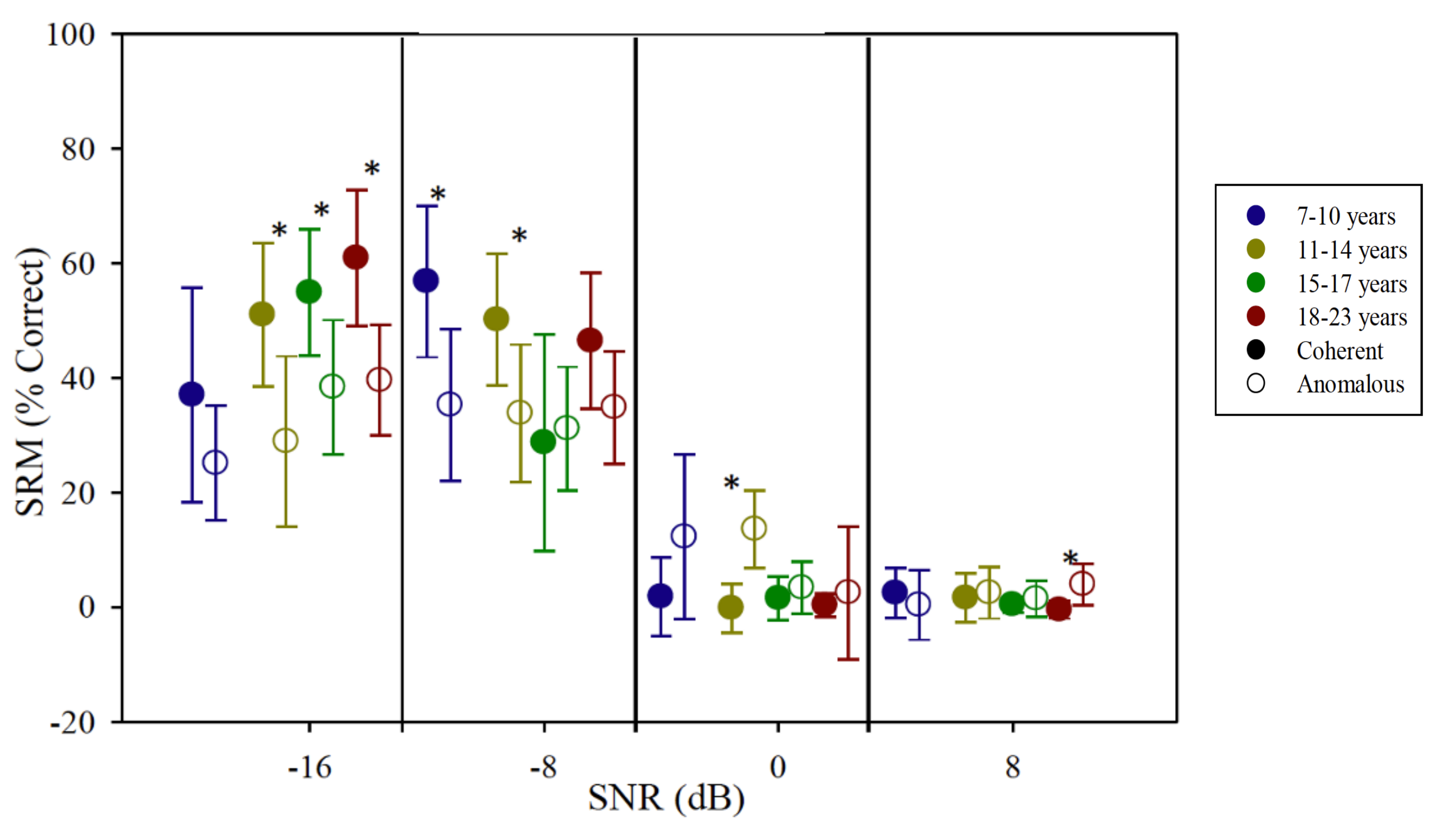


Figure 1: Data from Misurelli's PhD dissertation are shown here (2015). Mean (±SD) SRM is plotted at each SNR, for different age groups for high-predictable (coherent) and low-predictable (anomalous) sentences with Speech Interferer. Significant differences within each group between the coherent SRM vs. anomalous SRM ($p < 0.0125$) are highlighted (*). The sentence materials are provided with permission by Ingrid Johnsrude.

- In the current study we investigated effect sizes in typically-hearing (TH) children, who listened to speech that was spectrally degraded via vocoding, to estimate aspects of CI processing.
- Because TH children are not accustomed to listening to vocoded speech, they were first tested during an exposure period followed by post exposure testing³.
- Further, poor bilateral hearing in children with CIs is expected to hinder efficient and flexible cognitive processing.
- But TH children can use their cognitive skills such as attention, cognitive flexibility of using information and ability to monitor and suppress information which aid them to selectively attend and understand the target speech in noisy listening environments. Therefore, cognitive aspects of executive functioning were also measured.

Study Objectives:

- To examine speech recognition and spatial release from masking (SRM) in TH children for unprocessed and vocoded speech.
- To investigate these effects with speech materials that have low- vs. high-context (coherent & anomalous).

Methods

Participant Characteristics and Test Environment

- Six (6) TH native English speakers aged 9 to 15 years
- Passed audiometric hearing screening at 20 dB HL, bilaterally, from 250Hz – 8000Hz
- Speech recognition task was conducted in a standard, carpeted sound-proof booth
- All stimuli was presented through loudspeakers ($0^\circ / +90^\circ$ azimuth)

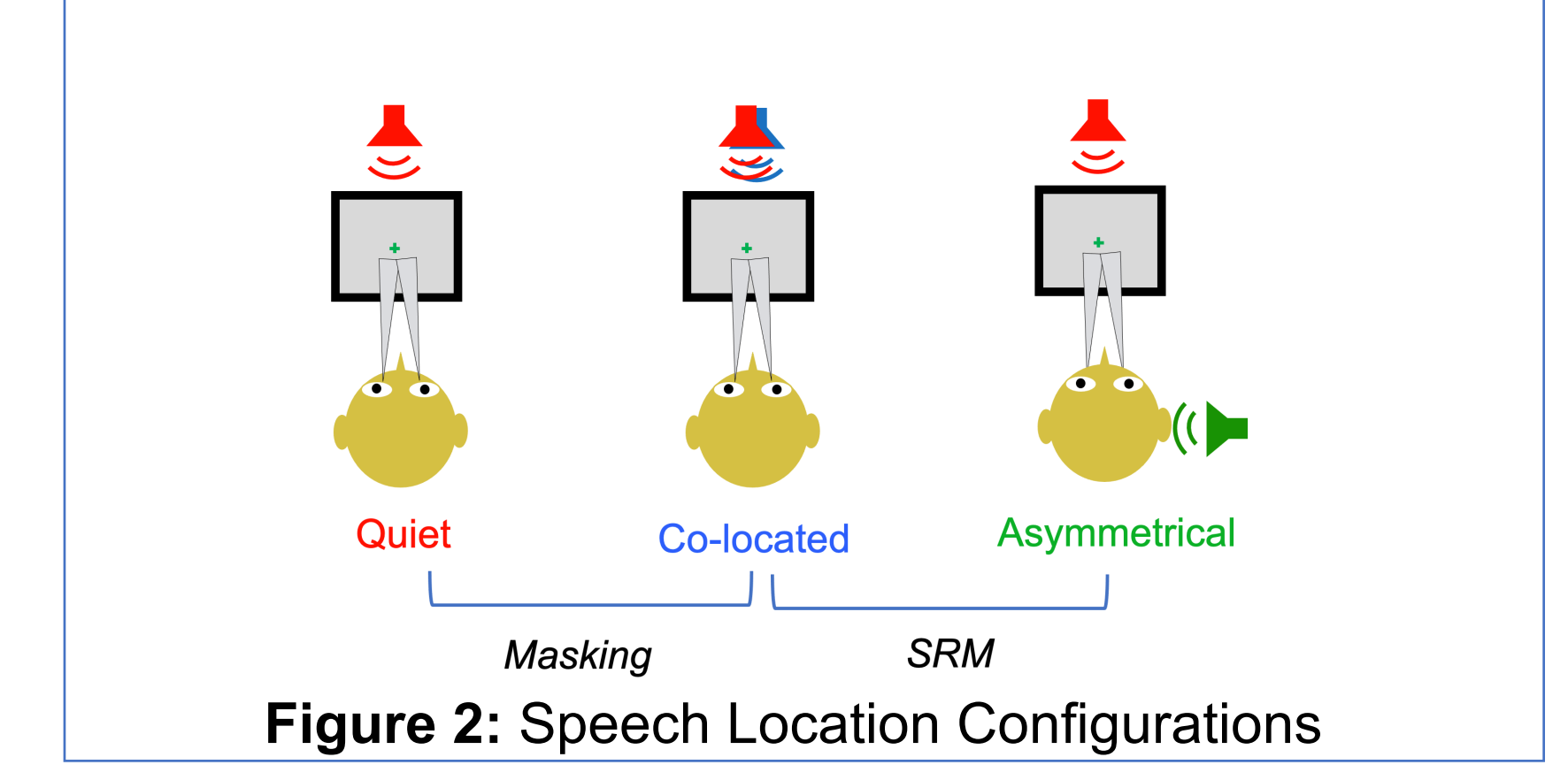


Figure 2: Speech Location Configurations

Design

- Participants attended two, 2.5-hour sessions
 - Session 1: vocoder exposure
 - Session 2: post exposure testing
- Participants were instructed to listen to the target sentences in Quiet and in Noise, with a speech masker, and to repeat back the target sentences. An experimenter transcribed responses in real-time.
- Target sentences were presented at 0° azimuth from the front speaker and the speech masker was either co-located with the target or spatially separated at $+90^\circ$ azimuth.
- Percent correct (accuracy) was determined from the number of correct words in each transcription.
- Visual feedback was provided during the vocoder exposure session. No feedback was provided during post exposure testing.
- This study has used an 8-channel noise vocoder with the Praat Software to vocode the sentence materials.

Vocoder Exposure (Phase I)

- Auditory stimuli**
- Harvard IEEE sentences
 - Coherent, phonetically balanced, five key-words
 - All sentences vocoded
 - Half the sentences presented in quiet at 60 dB SPL, half presented with background speech maskers at +8 dB SNR (Target 60 dB SPL and Masker at 52 dB SPL)

- Familiarization**
- 10 coherent vocoded sentences in Quiet and 10 coherent vocoded sentences with speech maskers at +8 dB SNR

- Coherent conditions**
- Quiet, vocoded
 - +8 dB SNR, co-located
 - Randomized
 - 2 hours of vocoder exposure

Post Exposure Testing (Phase II)

- Auditory stimuli**
- Coherent (high-context)/ Anomalous (low-context) sentences
 - Coherent Sentence e.g. "The whole sky was full of birds"
 - Anomalous Sentence e.g. "The success moved to hope the milk"
 - Phonologically, lexically, and syntactically balanced, 6 to 13 words
 - Ten sentences per condition
 - Half the sentences are coherent, and half are anomalous

- Familiarization**
- 4 coherent and 4 anomalous sentences in Quiet in unprocessed and vocoded conditions

- Coherent & Anomalous conditions**
- Quiet (Target: 60 dB SPL)
- Vocoded
 - Unprocessed
- +8 dB SNR (Target: 60 dB SPL; Masker: 52 dB SPL)
- Vocoded
 - Unprocessed
- +8 dB SNR Asymmetric Right
- Vocoded
 - Unprocessed

Figure 3: Study Design

Results

Phase I: Vocoder Exposure with IEEE sentences

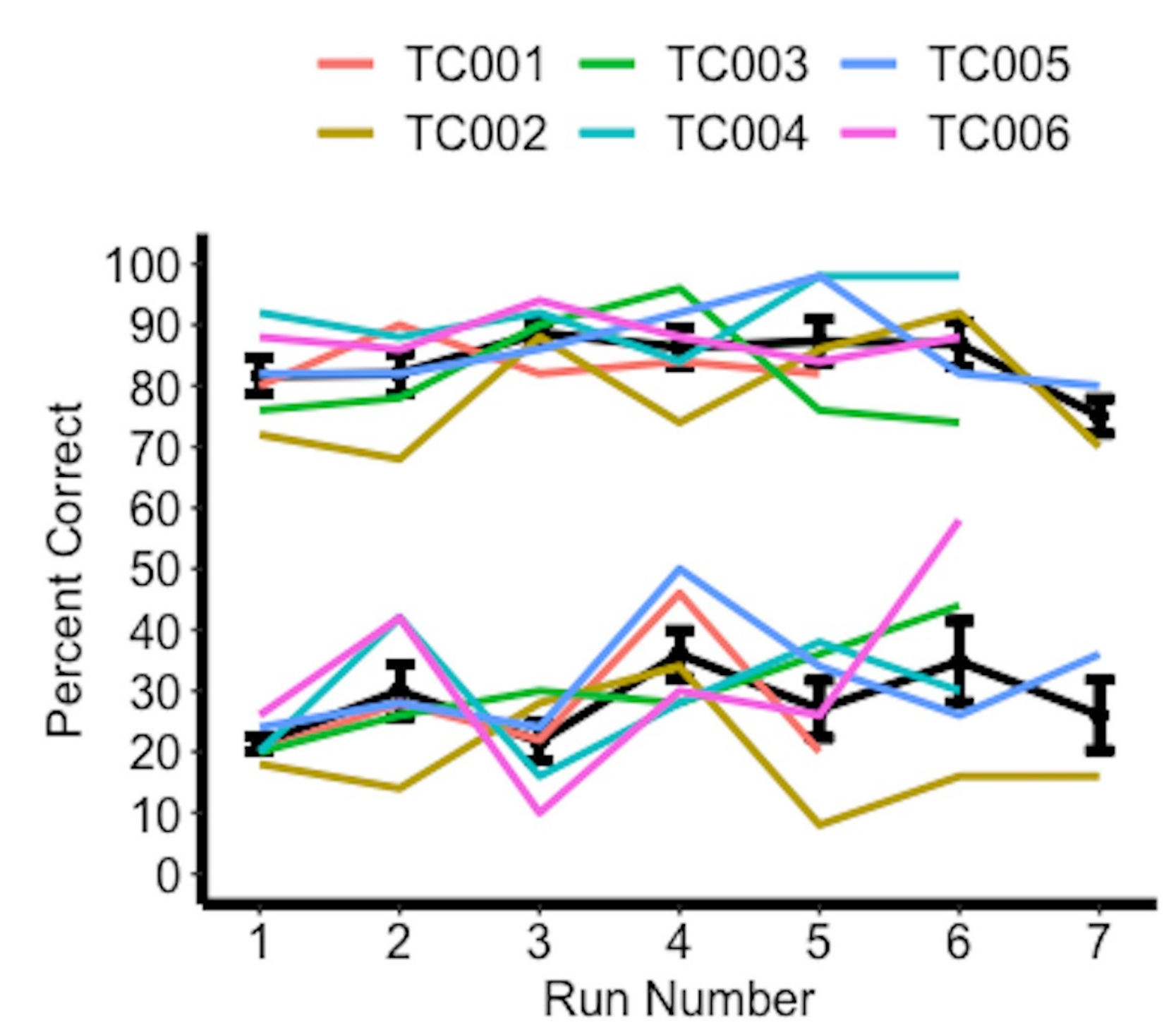


Figure 4: Percent Correct Words for the Vocoded IEEE Sentences in Quiet and +8dB SNR Co-located Conditions

- Speech intelligibility remained constant during the two hours of exposure with vocoded IEEE sentences. This indicates that there may not be a learning effect for these stimuli, which differs from a prior study (Goupell et al., 2020) in which children showed a gradual improvement.

Quiet

- Performance varied among the six participants tested and between runs, but overall performance in the +8dB SNR remained substantially lower than in Quiet.

Co-located +8dB SNR

- With the large degree of masking observed, the next phase in the study will be to test performance with the vocoded and unprocessed sentences that are either coherent or anomalous.

Results (cont.)

Phase II: Testing with Coherent and Anomalous Sentences

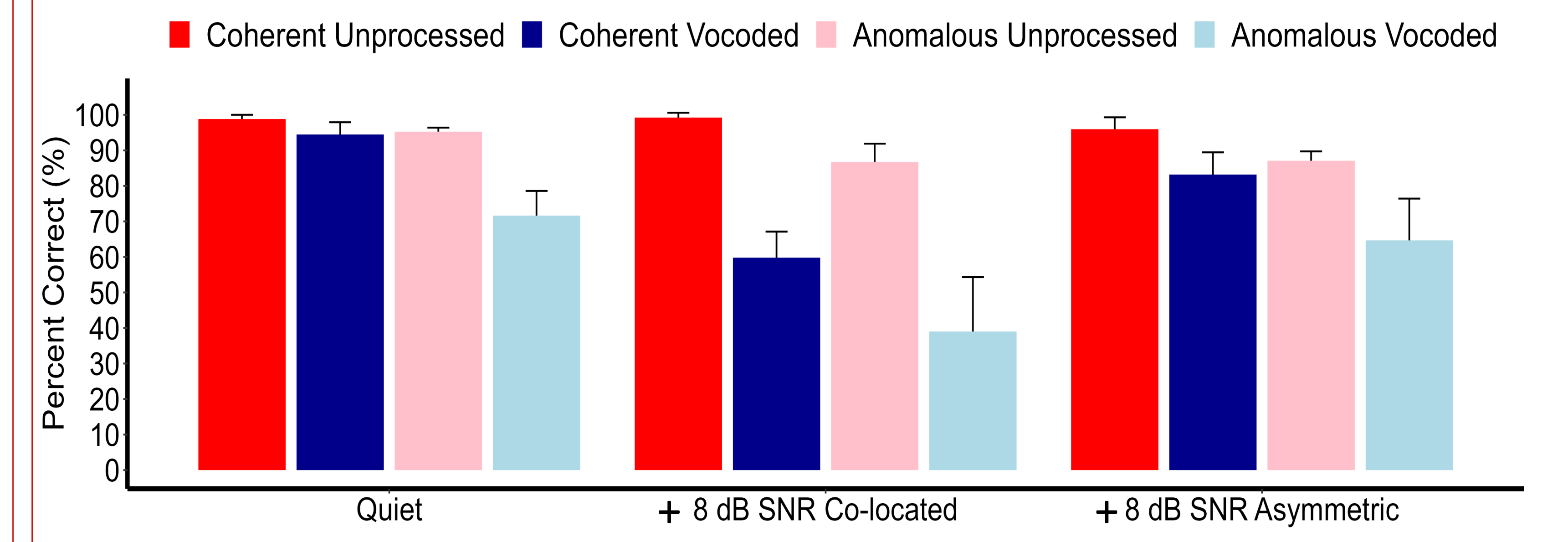


Figure 5: Percent Correct for Vocoded and Unprocessed Coherent and Anomalous Sentences

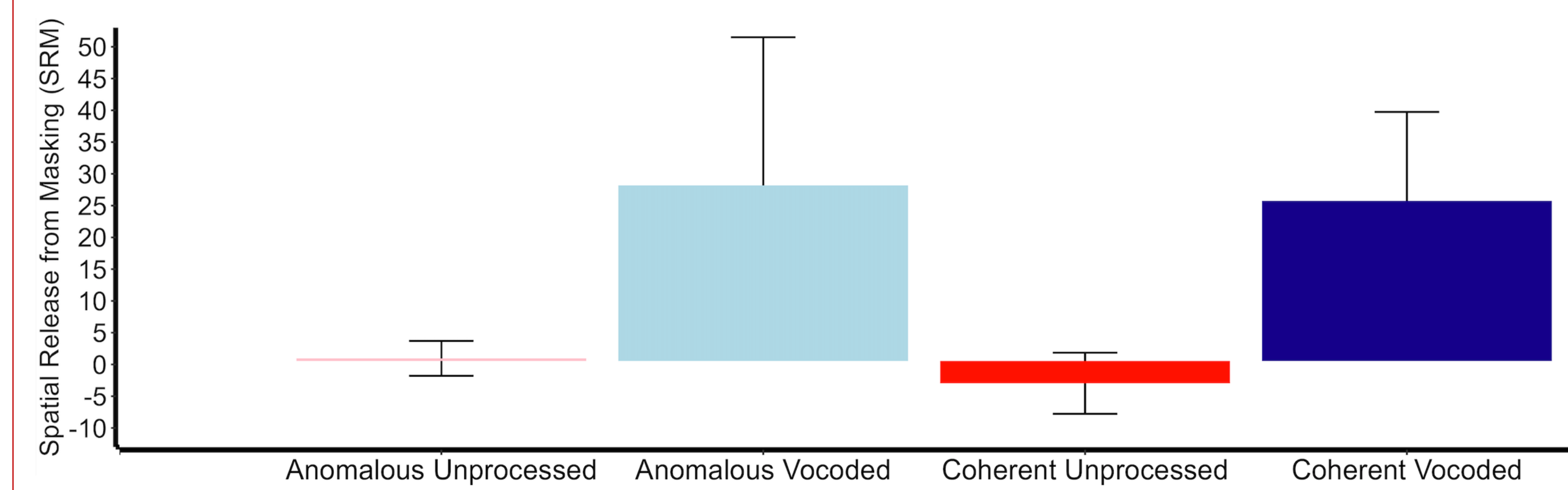


Figure 6: Spatial Release Mean and Standard Error for data from Figure 4

- Speech was most intelligible in Quiet, irrespective of the predictability of the sentences, and least intelligible in the +8 dB SNR co-located condition, specifically when the stimuli were vocoded.
- In the Vocoded conditions, Anomalous sentences were less intelligible than Coherent sentences.
- SRM was observed in both the Coherent and Anomalous Vocoded conditions. Preliminary data suggests that spatial separation can equally improve speech intelligibility when stimuli is vocoded, regardless if the target sentences are Coherent or Anomalous.

Discussion

- Understanding of the vocoded speech was increased with added exposure.
- In the exposure phase of the study there was no observable effect of improvement over a 2-hour exposure to vocoded speech.
- As expected, performance was better in Quiet than in +8 dB SNR unprocessed compared with vocoded speech.
- Semantically predictable sentences (Coherent) were better understood than semantically unpredictable sentences (Anomalous) in vocoded conditions.
- SRM was observed in both coherent and anomalous sentences in vocoded conditions. Lack of SRM in Unprocessed conditions is likely due, at least in part, to ease of the task.
- These are preliminary data from 6 participants and data collection is ongoing. Larger samples of children are needed to more fully investigate the effects of contextual cues on SRM with spectrally degraded speech.
- This work will also investigate the contribution of executive functioning skills on speech recognition.
- Future work will also focus on the effects of contextual cues on speech understanding and SRM in children with cochlear implants.

References

- Davis, M. H., Ford, M. A., Kherif, F., and Johnsrude, I. S. (2011). "Does semantic context benefit speech understanding through "top-down" processes? Evidence from time-resolved sparse fMRI." *J. Cogn. Neurosci.* 23, 3914-3932.
- Misurelli, S.M., (2015). "The effects of contextual cues, signal-to-noise ratio, and spatial separation on speech understanding for children and adults with normal hearing" (PhD Dissertation, University of Wisconsin).
- Goupell, M.J., Draves, G.T., & Litovsky, R.Y. (2020). Recognition of vocoded words and sentences in quiet and multi-talker babble with children and adults. *PLoS ONE*, 15(12).
- Davis, M. H., Johnsrude, I.S., Hervais-Adelman, A., Taylor, K., & McGestigan, C., (2005) Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, 134(2), 222-241.
- O'Neil, E.R., Parke, M.N., Krefl, H.A., & Oxenham, A.J. (2021). "Role of Semantic Context and Talker Variability in Speech Perception of Cochlear-Implant users and Normal Hearing Listeners". *J. Acoust.Soc.Am.* 149, 1224-1239.

Acknowledgements

NIH-NIDCD to R.Y. Litovsky (R01DC019511, R01DC020355) and in part by a core grant to Waisman Center from the National Institute of Child Health and Human Development (P50HD105353). Special thanks to BHSL lab members - Malcolm Worrell, Stephanie Sellner, Molly Osborn & Anshu K and Participants.

