

The effect of speaking style adaptations on speech perception in noise by native and non-native listeners

Michelle DuBois, Irene Smith, Kirsten Meemann & Rajka Smiljanic
Department of Linguistics, The University of Texas at Austin

Background

- Speech recognition in noise is difficult for both **native (Na)** and **non-native (NN)** listeners
- Native listeners outperform non-natives at speech recognition presented in speech shaped noise (SSN) [1,2,3,4]
- Two kinds of noise:
 - Speech-shaped noise (SSN)**, a white noise signal shaped to match the long-term spectrum of speech and
 - 2 talker babble (2T)**, competing speech
- SSN compromises audibility of the speech signal; 2T additionally presents linguistic/lexical competition with the target speech [5,6]
- Speakers can improve listener's ability to perceive their speech in adverse conditions by using speaking style adaptations [7,8,9,10]
- 4 speaking styles**
 - Conversation (CO)** - normal conversational speech
 - Clear (CL)** - speech adapted to address someone with low English proficiency [7,8,9]
 - Quiet Speech (Q)** - speech produced in a quiet environment
 - Noise Adapted (N)** - speech adapted in response to background noise [10]

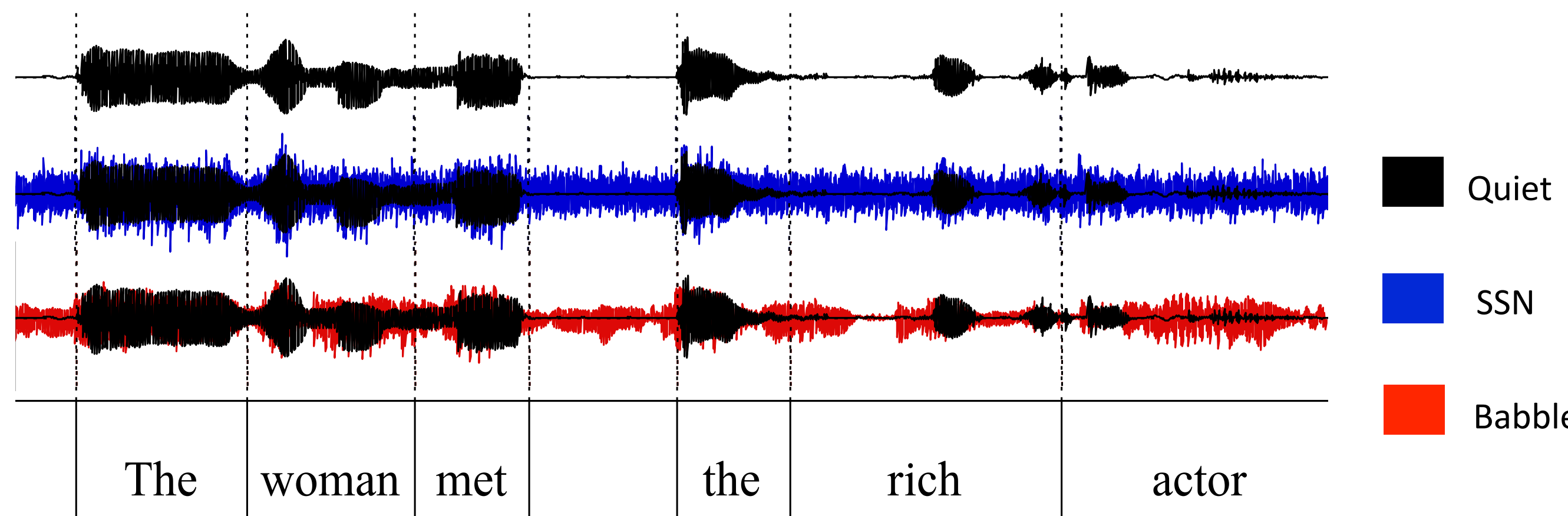
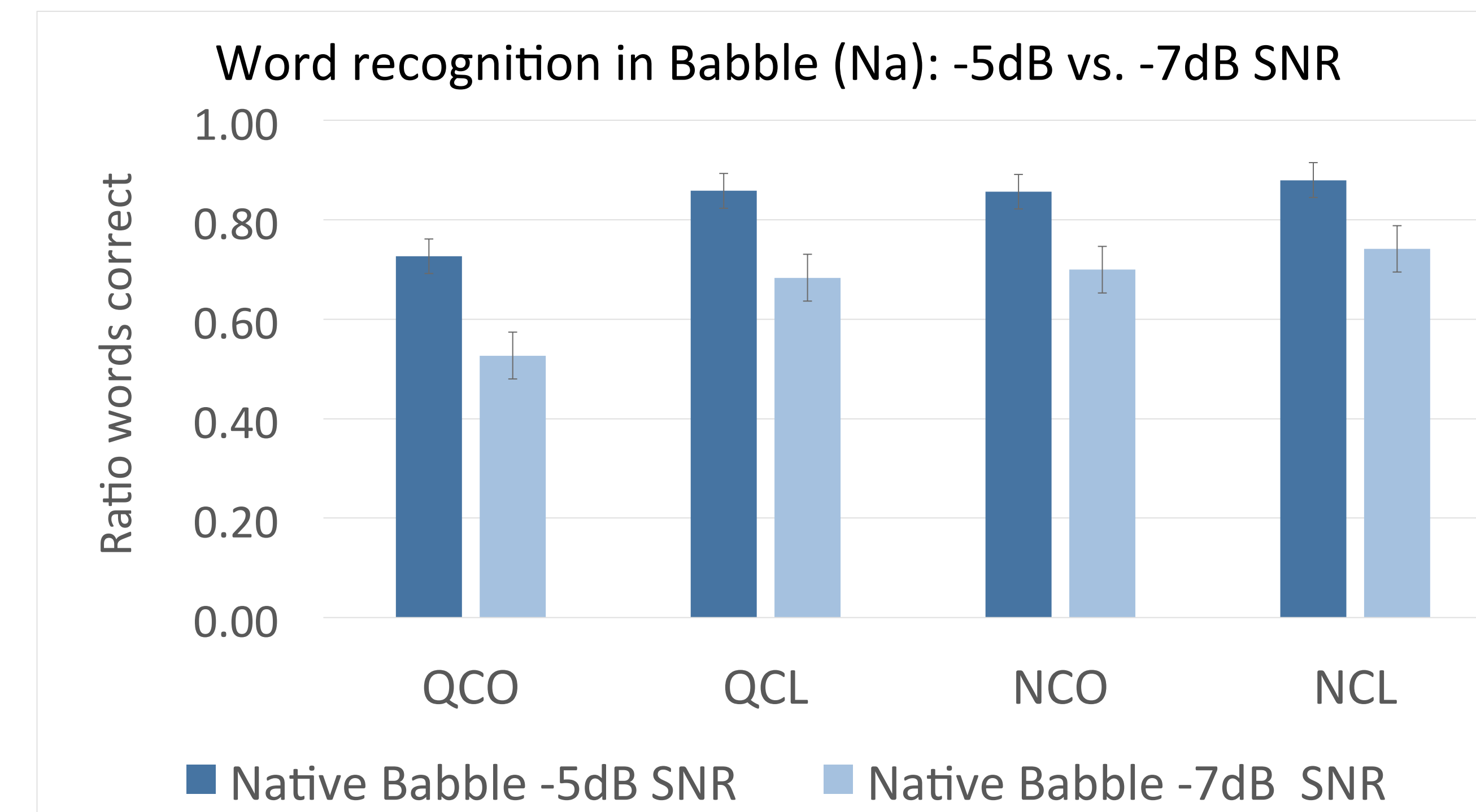
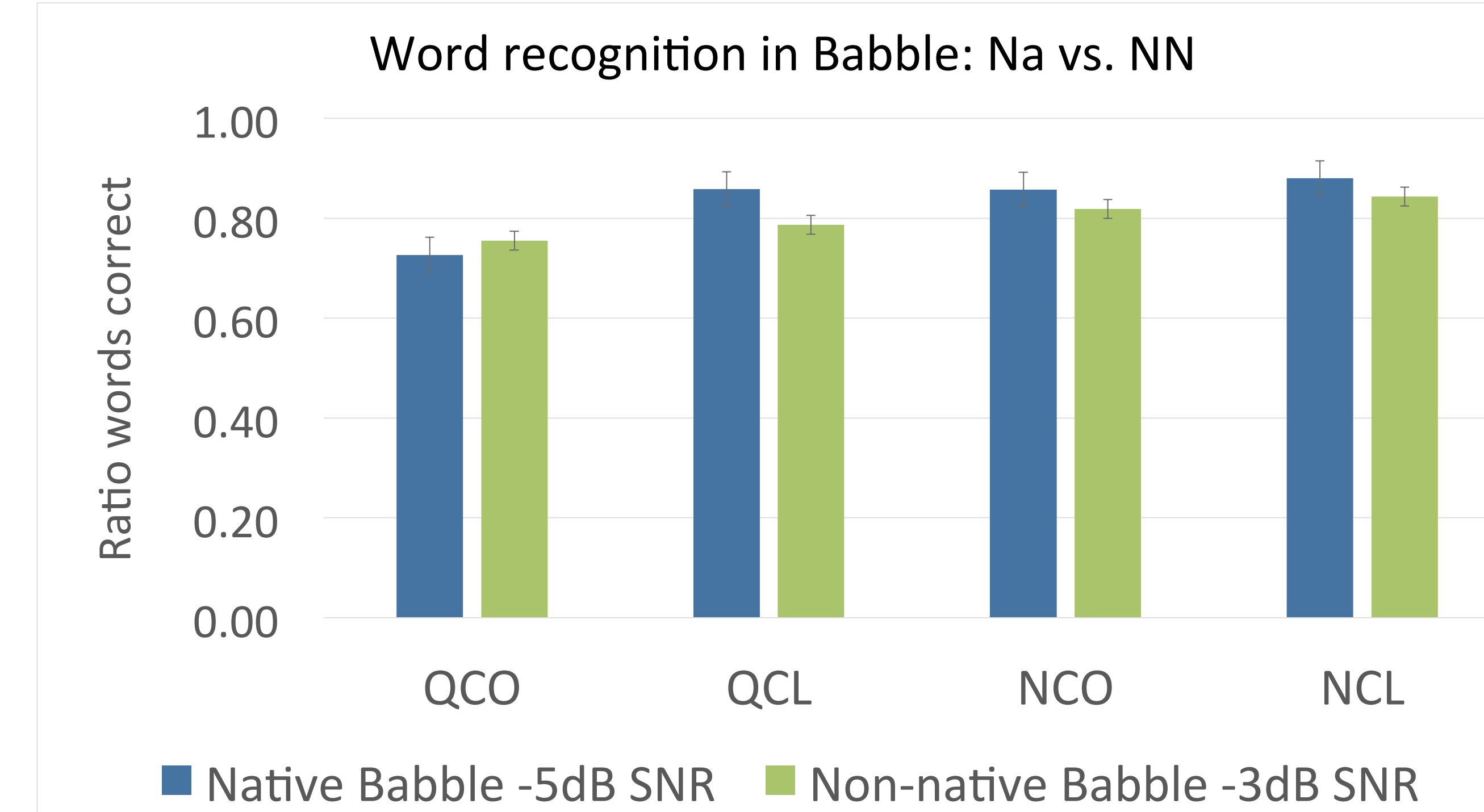
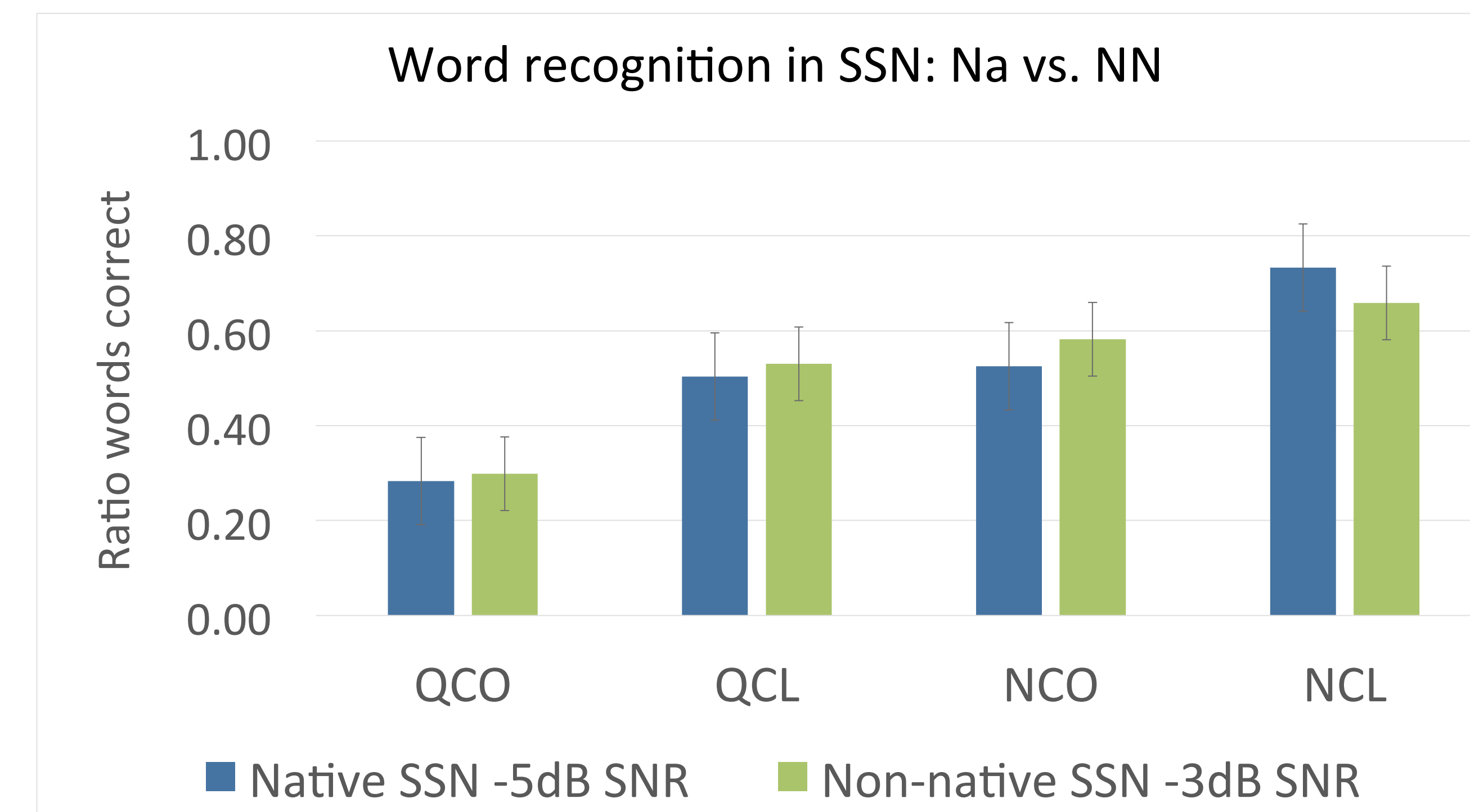
Research Questions

- Can word recognition in noise be improved through N and CL modifications equally for N and NN listeners?
- How do energetic masking (EM) and informational masking (IM) interact with the speaking adaptations to determine the intelligibility for the two listener groups?

Methods and Materials

- Talkers**
 - 10 young adults
 - Native English speakers
- Stimuli**
 - 60 meaningful sentences (Fallon et al. 2002)
 - Produced in either CO or CL
 - Elicited in Quiet first and then with SSN interference
 - Two listening conditions:
 - SSN condition:** -5dB signal-to-noise ratio (SNR) for Na listeners and -3dB SNR for NN listeners
 - Babble condition:** 20 Na -5 dB SNR, 20 native -7 dB SNR, 20 NN -3 dB SNR
- Listeners**
 - 100 speakers of English: 60 Na and 40 NN
 - NN listeners –UT students that started learning English between ages 6-13, L1 varied
- Procedure**
 - Participants heard 40 sentences total: 4 sentences from each of 10 speakers; 4 speaking styles per speaker (QCO, QCL, NCO, NCL)
 - Sentences were presented randomly, and no one sentence was heard twice by the same listener
 - Participants asked to type what they heard

Legend: QCO = Quiet Conversational, QCL = Quiet Clear, NCO = Noise Conversational, NCL = Noise Clear, Na = Native, NN = Non-native



Waveforms for the target sentences (top), sentence mixed with SSN (middle) and sentence mixed with 2T babble (bottom)

Results

- Word recognition was significantly enhanced for:
 - speech produced in noise (N) than in quiet (Q)
 - clear speech (CL) than conversational speech (CO)
- Intelligibility in **SSN**:
 - Both Na and NN listeners benefit from CL and NAS modifications
 - Na and NN perform at the same level for speech produced in quiet
 - Na performed worse than NN in CO and better than NN in CL for speech produced in noise
- 2T babble**:
 - Overall word recognition is better in 2T than in SSN for both Na and NN
 - Na and NN performed at the same level in QCO
 - Na benefited more from all speaking style adaptations compared to NN
 - Na listeners showed larger intelligibility enhancement from CL and noise-adapted speech in more difficult listening condition (-7 dB SNR)

Discussion

- Real-world communication occurs under adverse conditions (e.g., noise), so understanding what helps and hinders successful communication can help us improve otherwise difficult communicative situations, especially for vulnerable groups such as nonnative listeners
- Na listeners are better able to use acoustic-phonetic enhancements in CL and noise-adapted speech than NN
 - Speaking style adaptations are geared toward Na listeners?
- All listeners are better able to recognize words in 2T babble than SSN
 - This may be due to the moments of quiet that naturally occur in babble and which allow the listener to latch onto the target signal
 - IM is less disruptive to speech processing than EM (at least 2T babble)
- In babble, CL and N speech provided the same benefit to both Na and NN
 - This could be because listeners performed at the ceiling level and no further improvement could be observed
 - Further intelligibility benefit could be observed when Na listeners performed the task at a more difficult SNR

Further investigation

- Decrease 2T SNR to -5 dB for NN listeners to remove ceiling effect
- Introduce 6-talker babble condition which, in addition to IM, provides greater EM

References

- Mayo, L.H., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *J. Speech Lang. Hear. Res.*, 40, 686-693.
- Bradlow, A.R., & Alexander, J.A. (2007). Semantic and phonetic enhancements for speech-in-noise by native and non-native listeners. *The Journal of the Acoustical Society of America*, 121 (4), 2339-2349.
- Cutler, A., Garcia Lecumberri, M.L., & Barker, J. (2008). Consonant identification in noise by native and non-native listeners: effects of local context. *The Journal of the Acoustical Society of America*, 124 (2), 1264-1268.
- Bradlow, A. R., and Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America*, 112 (1), 272-284.
- Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal of the Acoustical Society of America*, 84, 917-928.
- Nábělek, A. K., Czyzewski, Z., & Crowley, H. (1994). Cues for perception of the diphthong /ai/ in either noise or reverberation. Part I. Duration of the transition. *The Journal of the Acoustical Society of America*, 95, 2681-2693.
- Van Engen, K.J., Chandrasekaran, B., & Smiljanic, R. (2012). Effects of speech clarity on recognition memory for spoken sentences. *PLoS One*, 7, e43753.
- Smiljanic, R. (2012). Can older adults enhance the intelligibility of their speech? *The Journal of the Acoustical Society of America*, 133(2), EL 129-EL135.
- Gilbert, R.C., Chandrasekaran, B., & Smiljanic, R. (2014). Recognition memory in noise for speech of varying intelligibility. *The Journal of the Acoustical Society of America*, 135, 389-399.
- Pichora-Fuller, M.K., Goy, H., & van Lieshout, P. (2010). Effect on speech intelligibility of changes in speech production influenced by instructions and communication environments. *Seminars in Hear.* 31, 77-94.