The Motor System’s [Modest] Contribution to Speech Perception

University of California, Irvine

Ryan C. Stokes; Jonathan H. Venezia; Gregory Hickok
University of California, Irvine

Introduction

Motor theory of speech perception

The motor theory of speech perception claims that the listener maps variable acoustic speech inputs to invariant gestural (i.e. motor) commands to recover the identity of incoming speech sounds.[1,2]

- The strong version of the motor theory claims that the objects of speech perception are the vocal tract gestures, which then engage the motor system for speech perception.[1] However, damage to speech motor systems tends to leave speech perception intact:
  - Congenital anarthria, cerebral palsy, bilateral inferior frontal lesions, and left hemisphere anesthesia, destroy or prevent the normal functioning of the speech motor system without compromising speech perception[3, 4, 5, 6]
- A weak version, as revised by Liberman, proposes the *intended* gestures as the objects of speech perception[2]
  - Unclear what the representational nature of an “intended gesture” is
- Another weak version claims the motor system assists during speech perception, but is not essential, suggesting that the motor system plays an ancillary role in perception, being engaged primarily in noisy or degraded conditions[7, 8]

Transcranial magnetic stimulation (TMS) studies have demonstrated that disrupting the motor cortex affects phonetic discrimination during noisy listening conditions.[9]

- Effect size is small and disappears in quiet listening conditions[10]
- A constant signal to noise ratio (SNR) was used, which does not capture changes outside of the value used

Psychometric functions

Estimating psychometric functions allows for a more complete view of performance differences across conditions, and avoids several issues with testing at a single SNR.

- When slopes differ, a measurement of a lower SNR (X_1) will have the opposite measurement of an SNR at a higher SNR (X_2)
- When thresholds differ, the difference in accuracy tells us nothing about the threshold shift (X_4 vs X_5)

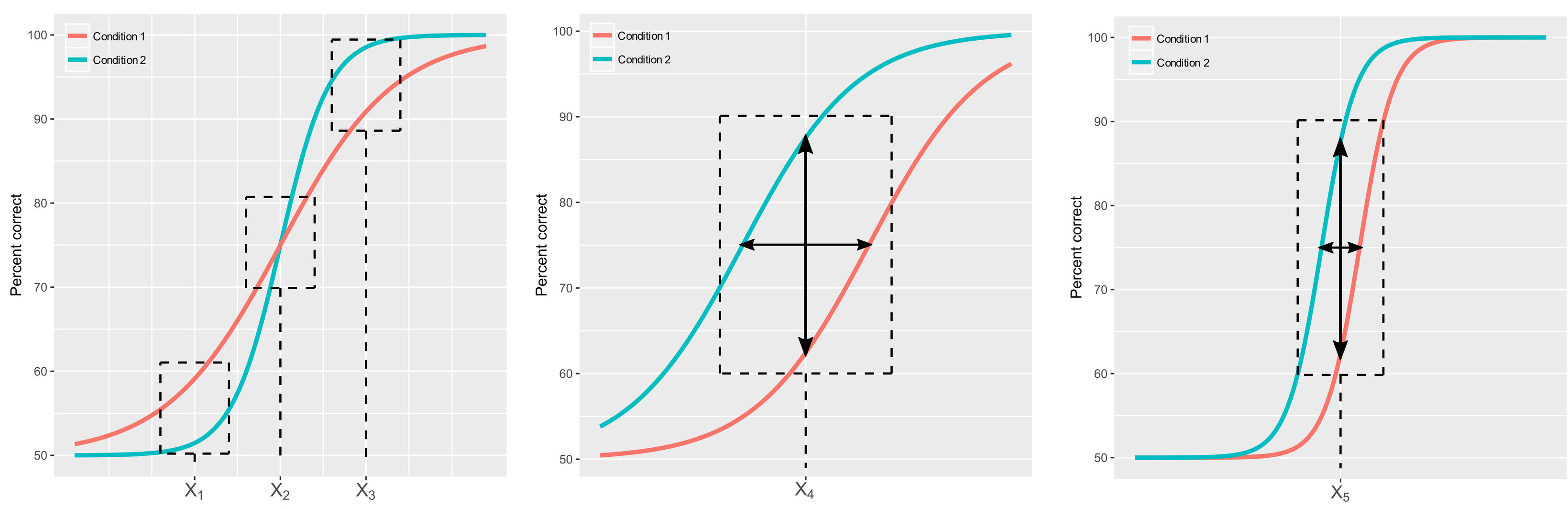

Figure 1. Example experiments highlighting the misleading interpretation of a single SNR.

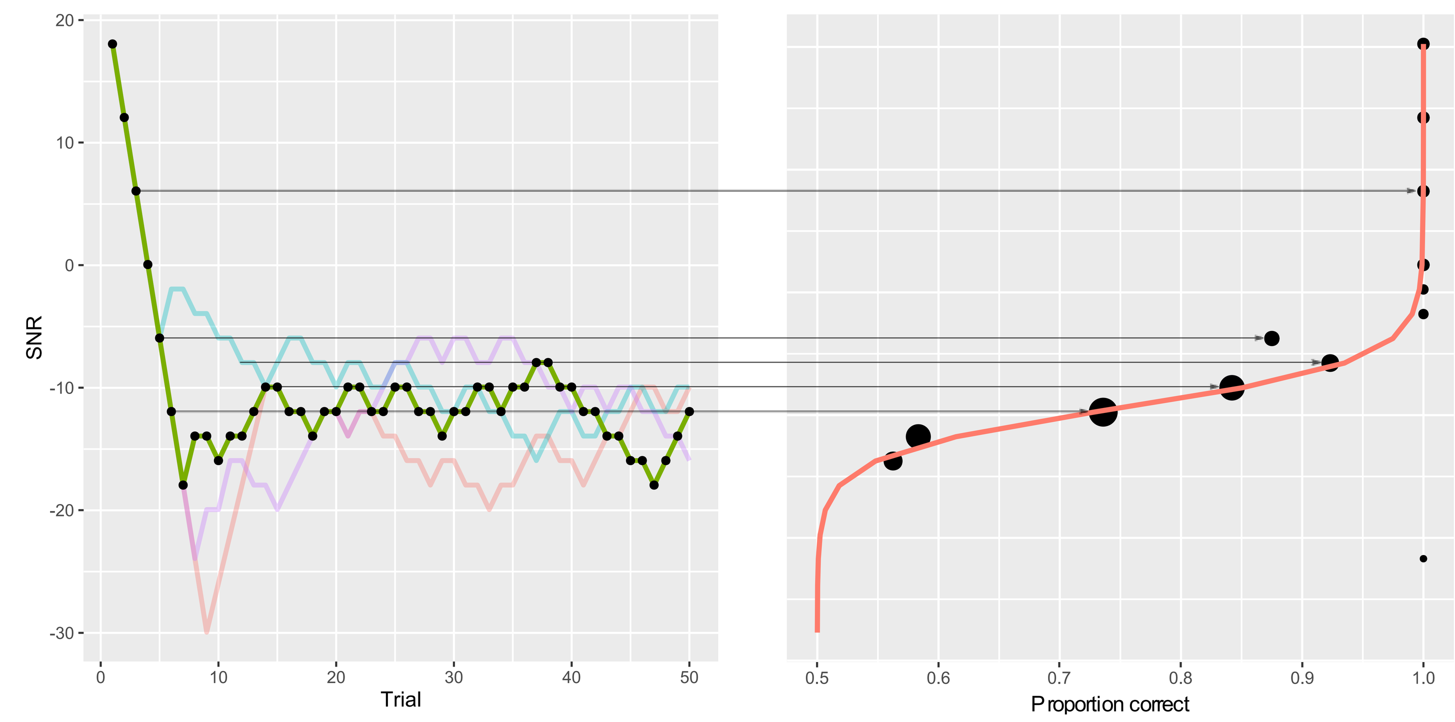

Figure 2. Example block with four staircases (left) is used to estimate a psychometric function (right)

Methods and Materials

Participants

17 participants recruited from the University of California at Irvine student population through the Human Subjects Lab Pool. Four participants were excluded from the study due to poor psychometric fits, or staircases not converging. Data from the remaining 13 participants were used for analysis (age 20.7 ± 2.5 years; 3 males).

Stimuli

Four minimal phonemic pairs consisting of monosyllabic, consonant-vowel English words were selected:

- place of articulation distinction (bilabial vs. alveolar: buy/dye and pie/tie)
- manner of articulation distinction (voiced vs. unvoiced: buy/pie and dye/tie)

Speech stimuli at a variable level were embedded in constant noise to generate various SNR conditions.

Procedure

Experiment consisted of eight blocks, each block comprised one of the four minimal pairs either with passive listening or articulatory suppression.

- Articulatory suppression: Participants were instructed to repeat the word *the* subvocally every second for the duration of the block
- Passive listening: Participants instructed to listen to the stimuli without subvocal rehearsal

An adaptive staircase procedure was used to sample from informative parts of the psychometric function.

Procedure for fitting psychometric functions

For each condition and subject a psychometric function was estimated by using the data from up to four adaptive staircase runs. This method allows for larger amounts of data to be sampled around the most informative parts of the psychometric, allowing for greater accuracy in its estimation. More data at each SNR will contribute more weight to the respective SNR of the psychometric fit. (Figure 2.)

Results

The horizontal difference, or difference in SNR, indicates the size of the articulatory suppression effect in dB at each performance level. The largest horizontal difference between articulatory suppression and passive listening was 1.39 dB for the Buy/Dye pair and 4.62 for the Pie/Tie pair.

The vertical difference, or difference in performance, indicates how much better a participant would perform in the passive listening condition compared to the articulatory suppression condition at a fixed SNR. The maximum performance difference between the articulatory suppression condition and the passive listening condition was 6.17% at SNR -7.36 for the Buy/Dye pair and 8.39% at SNR -4.29 for the Pie/Tie pair.

Table 1. Threshold and slope of each condition

Pair	Condition	Threshold	Slope
Buy/Dye	Articulatory Suppression	-7.12 (3.48)	0.49 (0.39)
	Passive Listening	-8.11 (3.27)	0.51 (0.40)
Pie/Tie	Articulatory Suppression	-4.91 (4.06)	0.38 (0.19)
	Passive Listening	-6.37 (2.57)	0.47 (0.28)

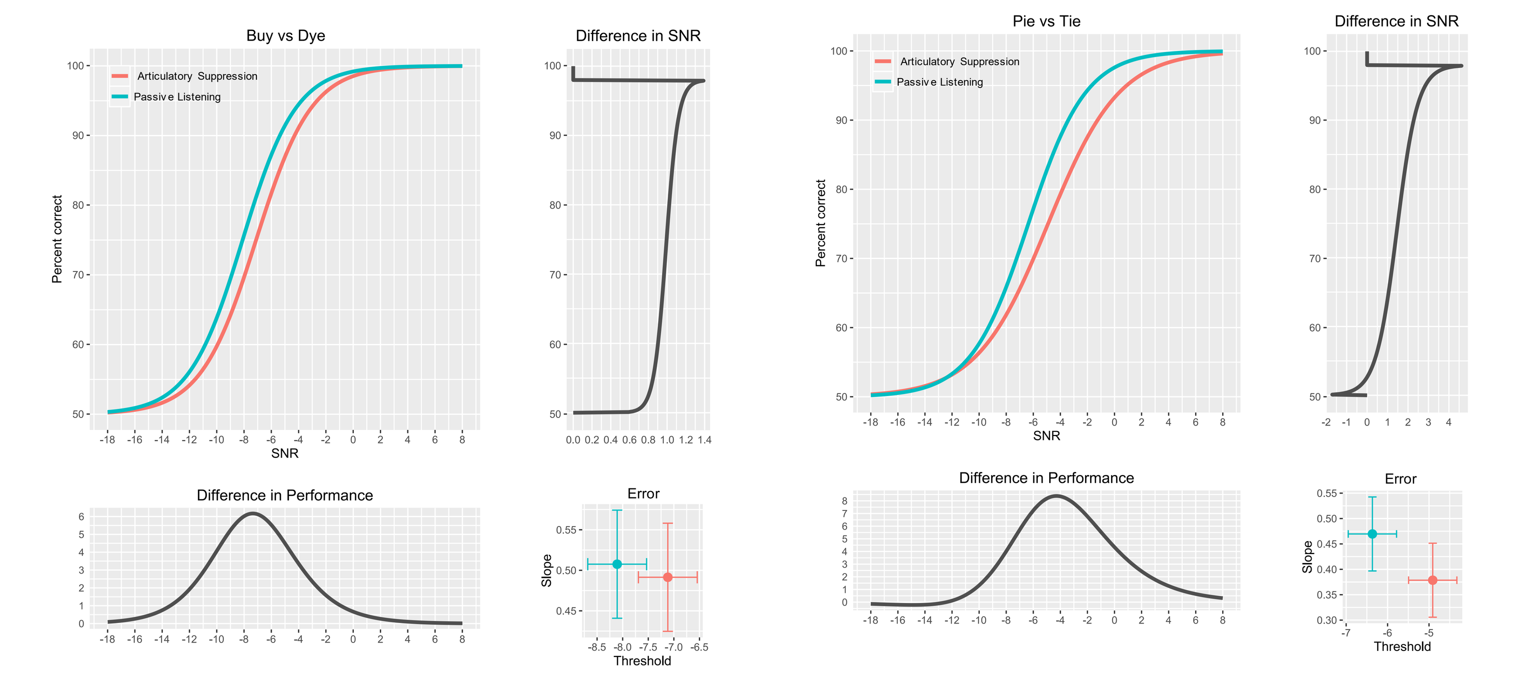

Figure 3. Difference in performance and SNR between articulatory suppression and passive listening

Conclusions

Speech in noise presented at one SNR condition has limited explanatory power. Performance differences can vary, sometimes significantly, as the SNR is changed. If we look at the SNR levels that provide the maximum performance difference presently, the results would look strikingly similar to that of Meister et al., who reported an effect of TMS-based speech motor suppression of 8.3%.[9] The maximum performance difference we found was 8.39% for the Pie/Tie pair, and 6.17% for the Buy/Die pair. Additionally, the steepness of the psychometric functions, an indication of the rapid change in perceptual acuity as the signal degrades, means that what appears a noteworthy difference in performance can be explained by only a small (1-2 dB) change in SNR. A pre-registered follow up will assess replicability of this effect and determine how much of the effect is driven by a secondary task generally, as opposed to a specific motor-speech effect.

Contact

Ryan Stokes
University of California, Irvine
Email: stokesr@uci.edu

References

1. Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychological review, 74(6), 431.
2. Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. Cognition, 21(1), 1–36.
3. Craighero, L., Metta, G., Sandini, G., & Fadiga, L. (2007). The mirror-neurons system: data and models. Progress in brain research, 164, 39–59.
4. Bishop, D., Brown, B. B., & Robson, J. (1990). The relationship between phoneme discrimination, speech production, and language comprehension in cerebral-palsied individuals. Journal of Speech, Language, and Hearing Research, 33(2), 210–219.
5. Lenneberg, E. H. (1962). Understanding language without ability to speak: a case report.
6. Hickok, G. (2010). The role of mirror neurons in speech perception and action word semantics. Language and Cognitive Processes, 25(6), 749–776.
7. Wilson, S. M. (2009). Speech perception when the motor system is compromised. Trends in cognitive sciences, 13(8), 329.
8. Wu, Z.-M., Chen, M.-L., Wu, X.-H., & Li, L. (2014). Interaction between auditory and motor systems in speech perception. Neuroscience bulletin, 30(3), 490–496.
9. Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. Current Biology, 17(19), 1692–1696.
10. D’Ausilio, A., Maffongelli, L., Bartoli, E., Campanella, M., Ferrari, E., Berry, J., & Fadiga, L. (2014). Listening to speech recruits specific tongue motor synergies as revealed by transcranial magnetic stimulation and tissue-doppler ultrasound imaging. Phil. Trans. R. Soc. B, 369(1644), 20130418.