Speaking clearly improves speech segmentation in optimal listening conditions

Zhe-chen Guo, Rajka Smiljanic

Department of Linguistics, The University of Texas at Austin

1. Background

- To understand speech, listeners need to segment fast, continuous, highly variable speech into the **component words**.
- Speech segmentation is even more complex in everyday life where listeners recognize speech under perceptual load such as **in noise** or in their **second language (L2)** [1, 2].
- Hyper-articulated, listener-oriented **clear speech** (CS) is more intelligible than **conversational speech** (CO) in quiet and in noise for native and non-native listeners [3, 4, 5, 6].
- **Goal:** Examine whether CS facilitates speech segmentation for native and non-native listeners in quiet and in noise.

2. Research Questions

1. Does **CS** facilitate speech segmentation **in quiet** and **in noise**?

Hypothesis 1: CS > CO.
Segments in CS are associated with longer duration, enhanced phonetic contrasts, and decreased coarticulation [6, 7]. CS word recognition is increased in noise [6, 8].

Hypothesis 2: CO > CS.
Listeners perceive coarticulated segments as belonging to a unit [9] and coarticulation is increased in CO [7]. Perceptual load is increased in noise [2].

2. Do **non-native listeners** benefit from CS in speech segmentation?

Hypothesis: NA > NN.
NN listeners benefit less from CS in word recognition in noise [8]. L2 listening is more effortful [1].

3. Methods

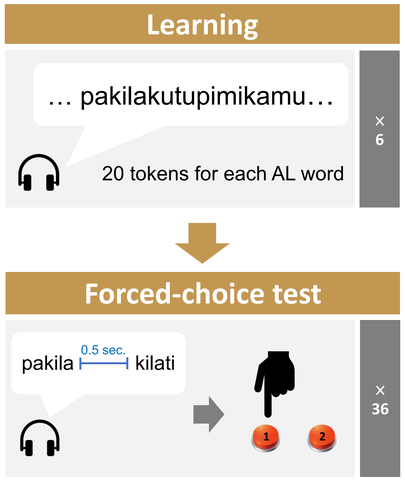
Design and Stimuli:

- **Artificial language (AL) learning** experiment [10]
- Six words of the AL:
  /pakila/, /timani/, /kutupi/, /mikamu/, /nuluta/, /lipuna/
- Six partwords spanned the boundary of two AL words:
  e.g., /kilati/ (= /pakila/ + /timani/)
- All tokens were read conversationally and clearly by a female speaker of American English in "*The word I said yesterday was* ___."
- All tokens were equalized for intensity and F0 was flattened.
- **Acoustic analyses** confirmed that CS vowels were longer and more hyper-articulated than CO vowels. V to V coarticulation within words was smaller in CS than in CO.

Listeners:

- 94 English native (56 F, mean age: 21.0): English only from birth
- 64 English early bilingual (41 F, mean age: 20.5): English and another language from birth
- 42 Non-native (29 F, mean age: 22.4): English after 6

Procedure:



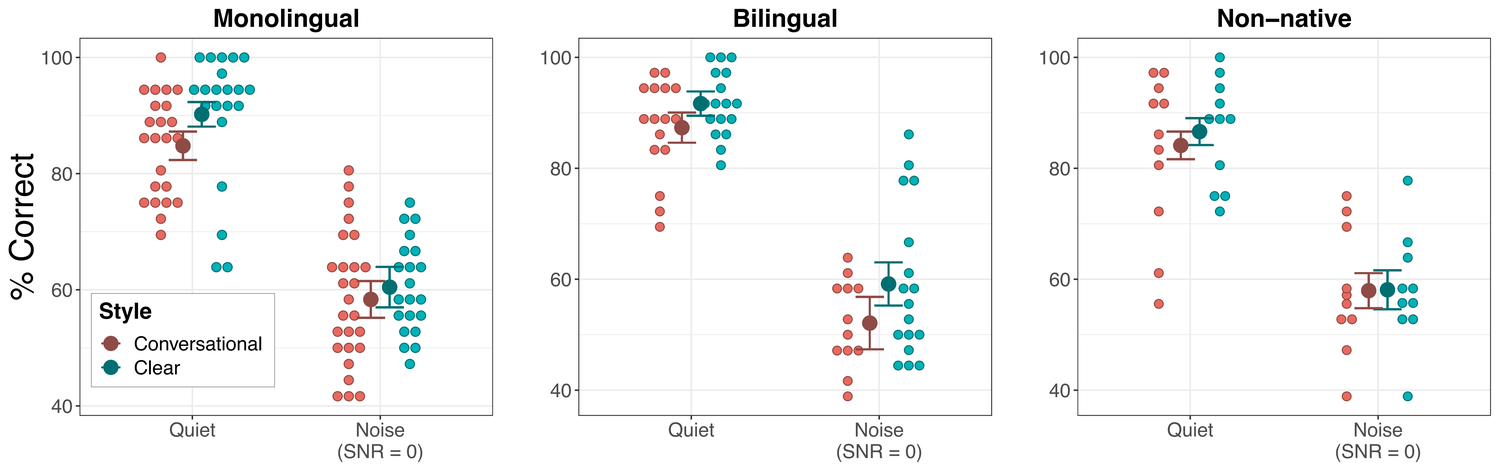
- Each subject heard either CS or CO speech streams either in quiet or in speech-shaped noise (SSN) with a 0 dB signal-to-noise ratio (SNR).
- Each subject heard test stimuli in quiet in the same speaking style as in the learning phase. For each word/partword pairs, they had to indicate which one was the word of the AL.

Analysis:

- Linear mixed-effects logistic regression analyses were performed on test responses separately for each group.
- Formula: Response ~ Style ($CO_{[ref]}$ vs. CS) * Listening condition ($quiet_{[ref]}$ vs. noise) + (1|Subject) + (1| ALWord)

4. Word Segmentation Accuracy

Fig. 1 Mean accuracy rates (%) with 95% confidence intervals for clear and conversational speaking styles in quiet and in noise for the three listener groups. Lighter dots indicate individual subjects' accuracy mean rates. Subjects with an accuracy rate of 2 SDs above or below the mean were excluded as outliers.



- Style (p = .008): CS > CO
- Style × Listening condition (p = .078)

- Style (p = .055): CS = CO
- Style × Listening condition (p = .620)

- Style (p = .453): CS = CO
- Style × Listening condition (p = .590)

- For all groups, Quiet > Noise (all ps < .001)

5. Discussion

For **native** listeners:
- CS improves segmentation.
- CS benefit is reduced in noise, suggesting that they might fail to use acoustic-phonetic enhancements under perceptual load.
  1. F0 and intensity cues are absent; however, the relative weight of prosodic information increases in noise [11].
  2. Lexical information is crucial for perceptual learning of unfamiliar speech in noise [12].
  3. 0 dB SNR may be too hard.

For **non-native** listeners:
- No evidence for CS benefit in quiet and in noise.
  1. Some CS enhancements are native-listener oriented [8].
  2. Listening effort is increased in L2; fewer cognitive resources are available for speech segmentation and storing information in memory [1].

For **bilingual** listeners:
- CS benefit seems to be reduced.
  1. Even early bilinguals experience more difficulty with speech in noise [13, 14].

References

➤[1] Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*, *52*(11–12), 864–886. ➤[2] Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*(7–8), 953–978. ➤[3] Behrman, A., Ferguson, S. H., Akhund, A., & Moeyaert, M. (2019). The effect of clear speech on temporal metrics of rhythm in Spanish-accented speakers of English. *Language and Speech*, *62*(1), 5–29. ➤[4] Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *J Acoust Soc Am*, *121*(4), 2339–2349. ➤[5] Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *J Speech, Language, and Hearing Research*, *50*(5), 1241–1255. ➤[6] Smiljanić, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *J Acoust Soc Am*, *118*(3), 1677–1688. ➤[7] Scarborough, R., & Zellou, G. (2013). Clarity in communication: "Clear" speech authenticity and lexical neighborhood density effects in speech production and perception. *J Acoust Soc Am*, *134*(5), 3793–3807. ➤[8] Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *J Acoust Soc Am*, *112*(1), 272–284. ➤[9] Fernandes, T., Ventura, P., & Kolinsky, R. (2007). Statistical information and coarticulation as cues to word boundaries: A matter of signal quality. *Perception & Psychophysics*, *69*(6), 856–864. ➤[10] Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *J Memory and Language*, *35*(4), 606–621 ➤[11] Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *J Experimental Psychology: General*, *134*(4), 477–500. ➤[12] Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *J Experimental Psychology: General*, *134*(2), 222–241. ➤[13] van Wijngaarden, S. J., Steeneken, H. J. M., and Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *J Acoust Soc Am*, *111*(4), 1906–1916. ➤[14] Werner, L. A. (2007). Issues in auditory development. *J Communication Disorders*, *40*(4), 275–283.

Acknowledgements

Thanks to all RAs at UTSoundLab for assistance with data collection.