Coarticulation across communicative contexts:
An acoustic analysis of the LUCID corpus using spectral and temporal measures

Zhe-chen Guo, Rajka Smiljanic
Department of Linguistics, The University of Texas at Austin

1. Background

- Speakers continuously vary their speech output in response to the communicative context, reflecting a dynamic balance between **hypospeech** and **hyperspeech** (H&H theory [1]; Adaptive Speaker Framework [2]).
- **Coarticulation**, or the overlap between articulatory gestures, is a low-cost motor behavior [1, 3].
- Listener-oriented, intelligibility-enhancing **clear speeches** are expected to show coarticulatory resistance.
  - Evidence from read-speech [4, 5, 6] and spontaneous speech to a real listener [7] is mixed.

2. Research Questions

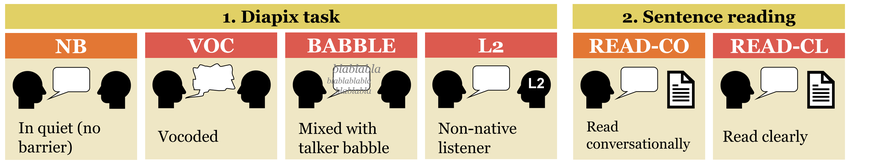
How do speaking style adjustments impact the degree of coarticulation when the speaker is:

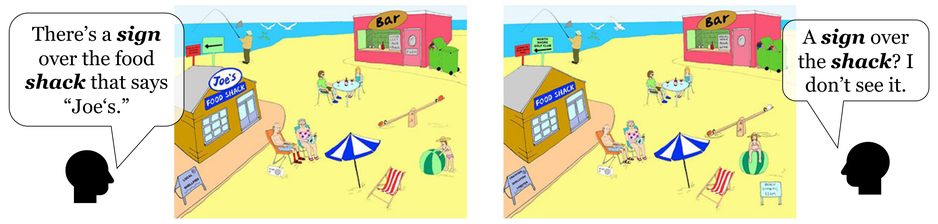
1. **producing spontaneous speech in response to a real communication barrier** (e.g., speech to the communication partner is masked by background noise)?
2. **reading aloud to an imagined listener with perceptual difficulty** (e.g., the listener is hearing-impaired)?

3. Methods

Speech data

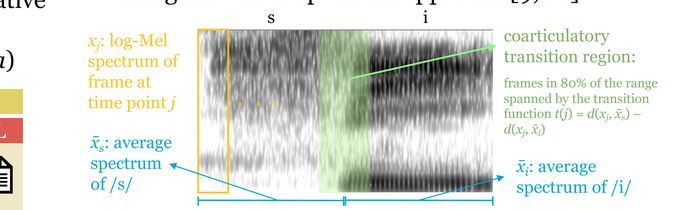
- 40 Southern British speakers
- Tasks: 1) an interactive Diapix task completed in four different communicative conditions; 2) sentence reading in two speaking styles (LUCID corpus [8])
- Elicited productions of 36 monosyllabic CV(C) keywords (e.g., *sign, shack, pea*)





Coarticulation measures and analysis

- Spectral and temporal measures of coarticulation using the whole-spectrum approach [9, 10]

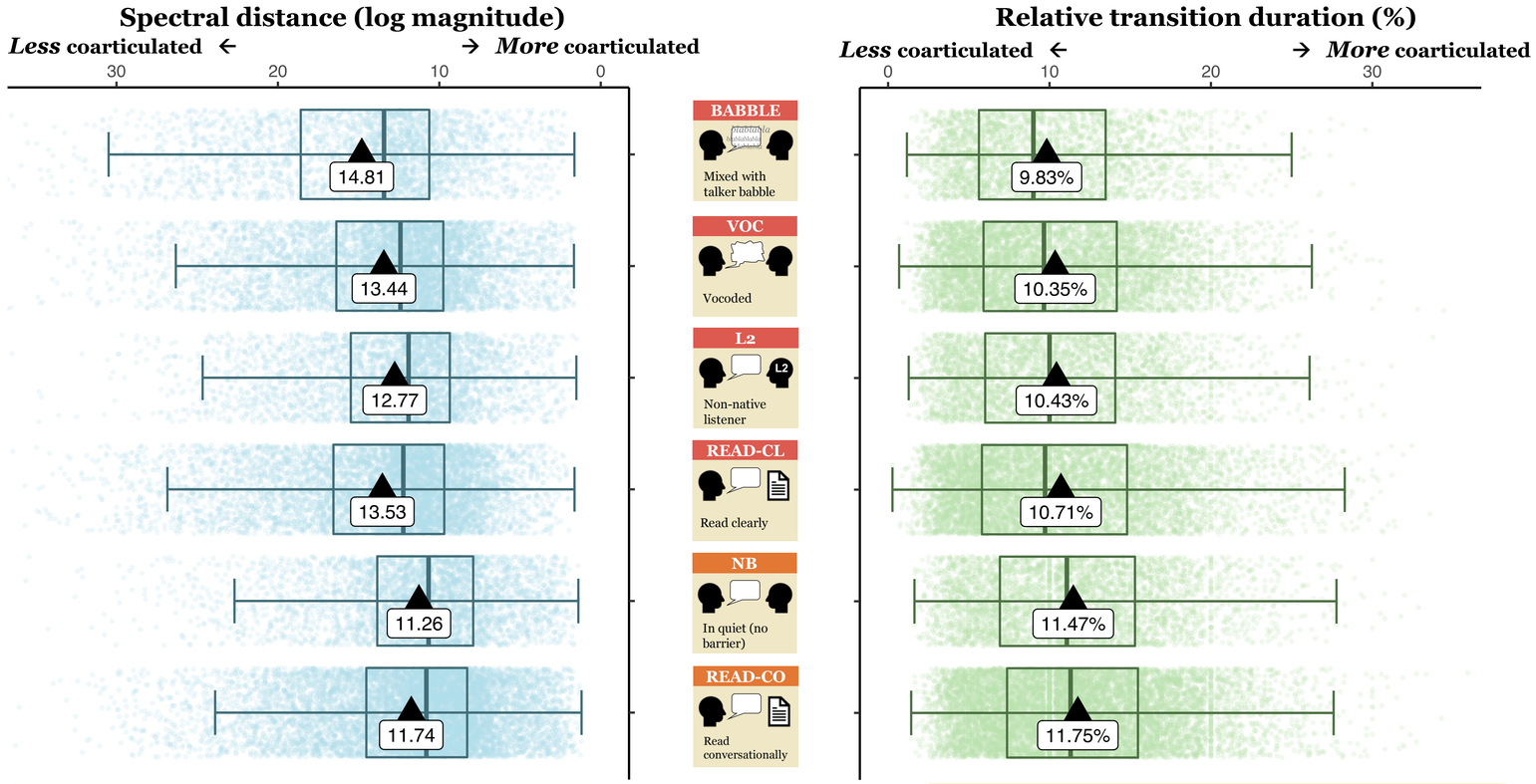


- *Spectral distance*: Euclidean distance d between average spectral shapes ($d(\bar{x}_s, \bar{x}_i)$)
- *Relative transition duration*: Proportion of the duration of coarticulatory transition
- Computed for every diphone in each keyword token (32,478 measurements total)
- Bayesian hierarchical modeling [11]:

  measure ~ **condition** + keyword repetition + word frequency + (1 + condition | speaker) + (1 + condition | diphone)

4. Results

Fig. 1 Spectral distance (left) and relative transition duration (right) mean values (triangle) and individual data points (circle) by communication condition. Significance assessed with the 95% highest density interval criterion.



- BABBLE > VOC > L2 > NB, READ-CO
- READ-CL > NB, READ-CO

(>: sig. greater spectral distance / greater coarticulatory resistance)

Keyword repetition was n.s. for both measures.

- BABBLE, READ-CL < VOC, L2 < NB, READ-CO

(<: sig. shorter coarticulatory transition / greater coarticulatory resistance)

- Shorter transition in more frequent words

5. Discussion

- Speech produced in response to communicative barriers, whether they are real or not, shows increased coarticulatory resistance relative to speech in the absence of such barriers.
- Overall, read clear speech and speech by talkers whose voices were masked by babble are the least coarticulated.
  - Speech in these conditions is also the most hyperarticulated [12].
- Talkers adjust coarticulatory patterns dynamically in response to the specific communication challenges (e.g., BABBLE is less coarticulated than VOC).
- Consistent with the view of coarticulation as a low-cost motor behavior [1].
- Spectral vs. temporal measures:
  - Spectral distance is more sensitive to differences among different communicative barriers (BABBLE, VOC, L2, NB).
  - Relative transition duration, but not spectral distance, distinguishes READ-CL from VOC and L2.
  - Shorter coarticulatory transition in more frequent words may reflect greater articulatory precision for words produced more often [13].
- Future work: Do less coarticulated clear speeches improve listeners' word segmentation?

References

➤**[1]** Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403–439). Springer Netherlands. ➤**[2]** Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *J Memory and Language*, *89*, 68–86. ➤**[3]** Farnetani, E., & Recasens, D. (2010). Coarticulation and connected speech processes. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences: Second Edition* (pp. 316–352). Blackwell Publishing. ➤**[4]** Moon, S. J., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *J Acoust Soc Am*, *96*(1), 40–55. ➤**[5]** Bradlow, A. R. (2002). Confluent talker- and listener-oriented forces in clear speech production. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 241–273). Mouton de Gruyter. ➤**[6]** Matthies, M., Perrier, P., Perkell, J. S., & Zandipour, M. (2001). Variation in anticipatory coarticulation with changes in clarity and rate. *J Speech, Language, and Hearing Research*, *44*(2), 340–353. ➤**[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]** Baker, R., & Hazan, V. (2010). LUCID: a corpus of spontaneous and read clear speech in British English. In *Proc of the DiSS-LPSS Joint Workshop 2010* (pp. 3–6). ➤**[9]** Cychosz, M., Edwards, J. R., Munson, B., & Johnson, K. (2019). Spectral and temporal measures of coarticulation in child speech. *J Acoust Soc Am*, *146*(6), EL516–EL522. ➤**[10]** Gerosa, M., Lee, S., Giuliani, D., & Narayanan, S. (2006). Analyzing children's speech: An acoustic study of consonants and consonant-vowel transition. *Proc 2006 ICASSP* (pp. 393–396). ➤**[11]** Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *J Statistical Software*, *80*(1), 1–28. ➤**[12]** Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *J Acoust Soc Am*, *130*(4), 2139–2152. ➤**[13]** Tomaschek, F., Wieling, M., Arnold, D., & Baayen, R. H. (2013). Word frequency, vowel length and vowel quality in speech production: An EMA study of the importance of experience. In *Proc ISCA 14th International Conference* (pp. 1302–1306).