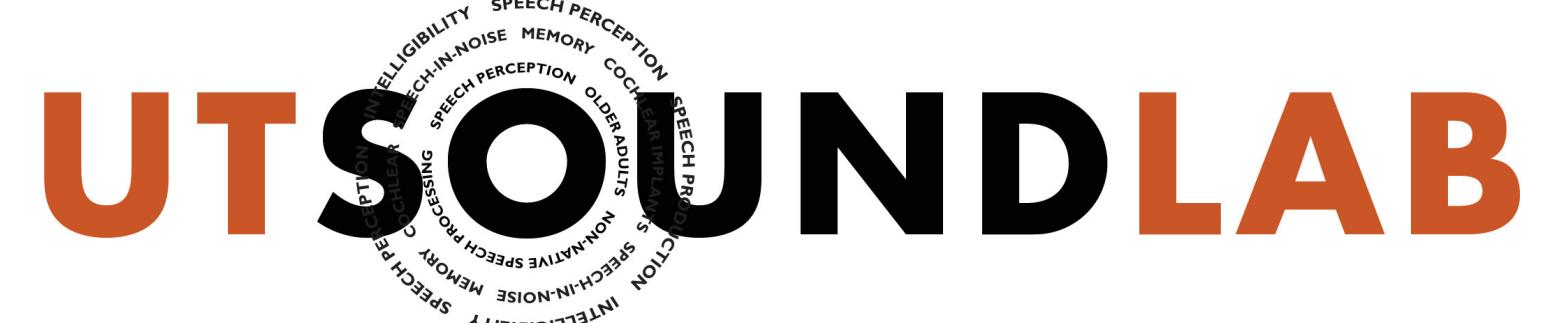


# Coarticulation is reduced in clear speech produced with protective face masks



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



## 1. Background

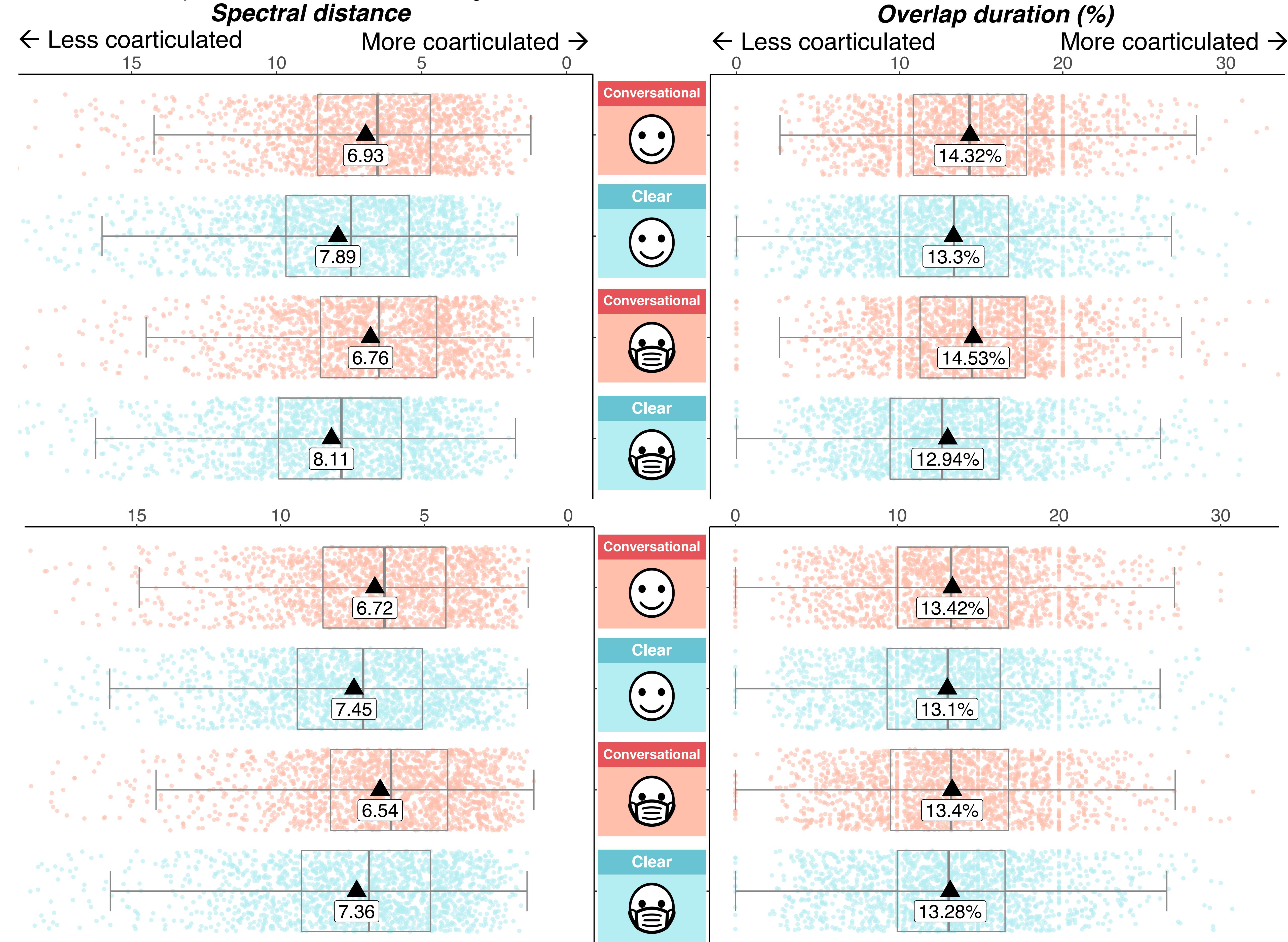
- Protective face masks degrade speech signal and impair speech intelligibility [1, 2].
- Listener-oriented hyperarticulated clear speech** [3] improves intelligibility and memory in noise even when produced with face masks [4]–[7].
  - The clear speech intelligibility benefit is larger for a native compared to a non-native speaker [6].
- Coarticulation**, or overlap between articulatory gestures, is reduced in hyperarticulated clear speech.
  - Evidence from CV coarticulation in clear speech produced without face masks [8, 9].

## 2. Research Questions

- Is coarticulatory resistance different for clear speech produced with a mask than without a mask?
- Does coarticulatory resistance in hyperarticulated clear speech differ for native and non-native speakers?

## 4. Results

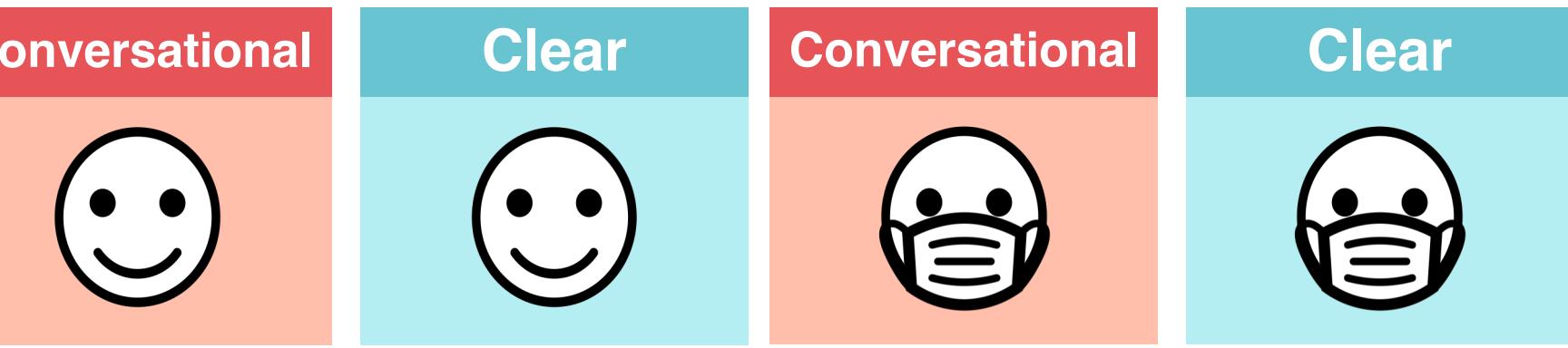
**Fig 1.** Results for the two whole-spectrum measures of coarticulation: 1) Spectral distances (left) and 2) overlap durations (right) for the native (top row) and non-native (bottom row) speakers. Each row shows results for the two mask conditions (mask-on, mask-off) and two speaking styles (clear, conversational). The dots represent individual spectral distance or overlap duration values and the triangles indicate the mean value for each condition.



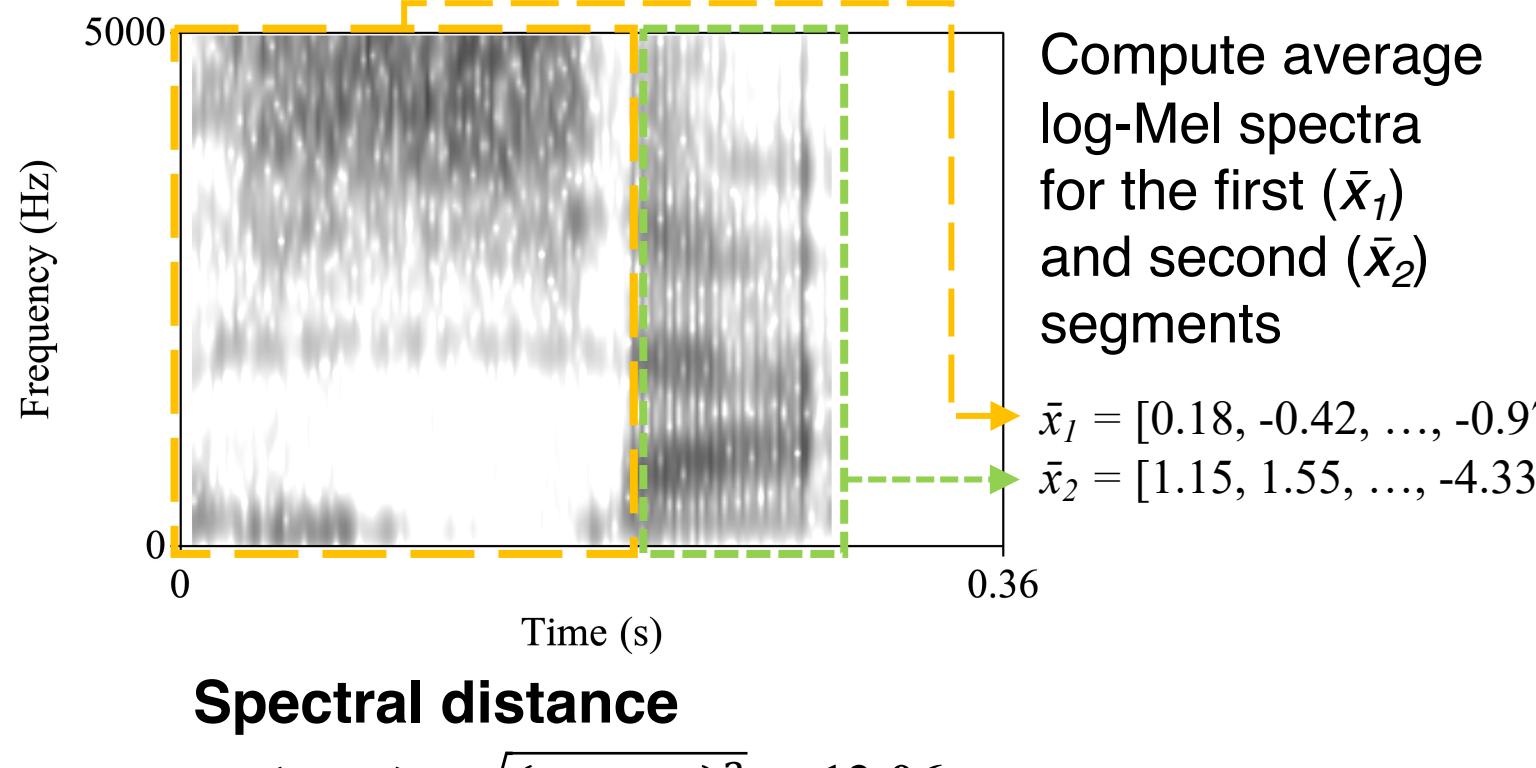
## 3. Methods

### Speech corpus

- Audio recordings from Smiljanic et al (2021) Toucan Audio-Visual Corpus [6].
- Forty-five sentences from a textbook essay about toucans [10].
  - E.g., There are approximately forty Toucan species indigenous to tropical America.*
- Read by one native and one non-native American English speaker in clear and conversational speaking styles with and without a surgical mask:



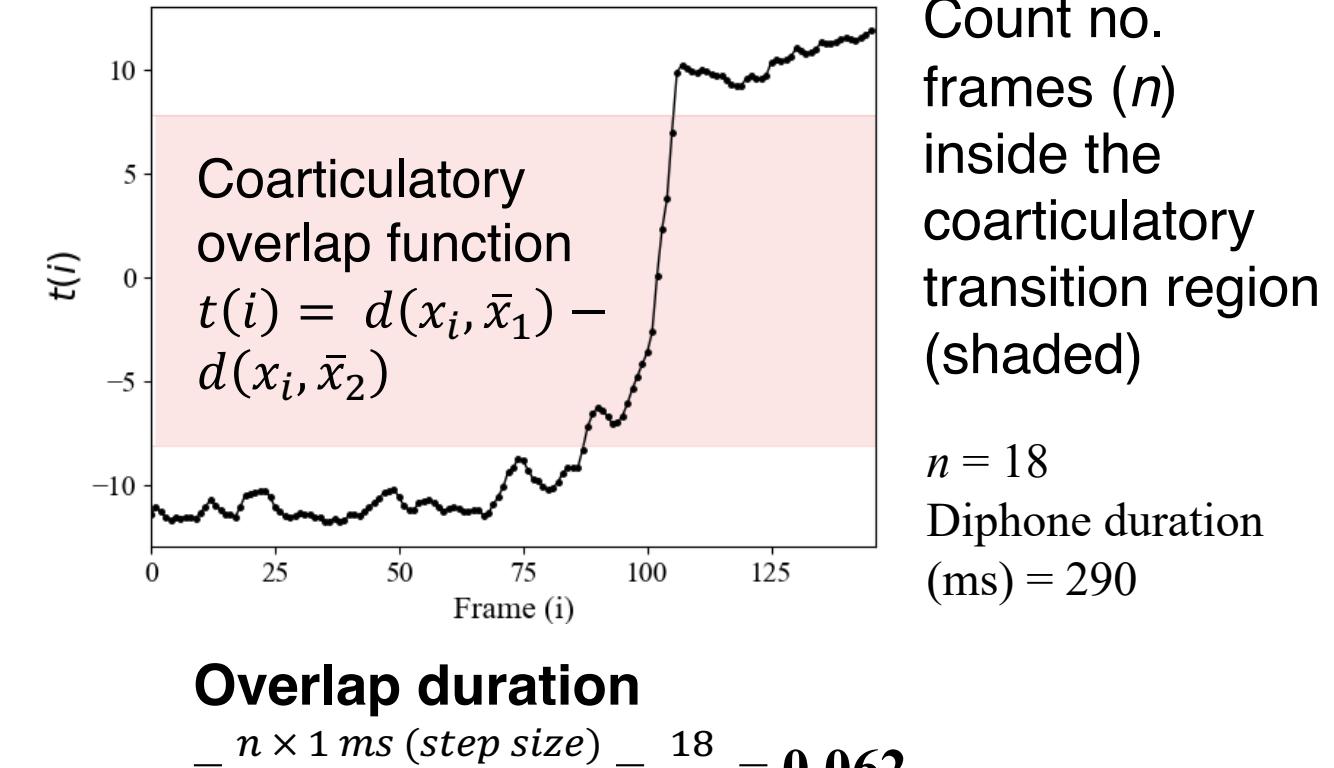
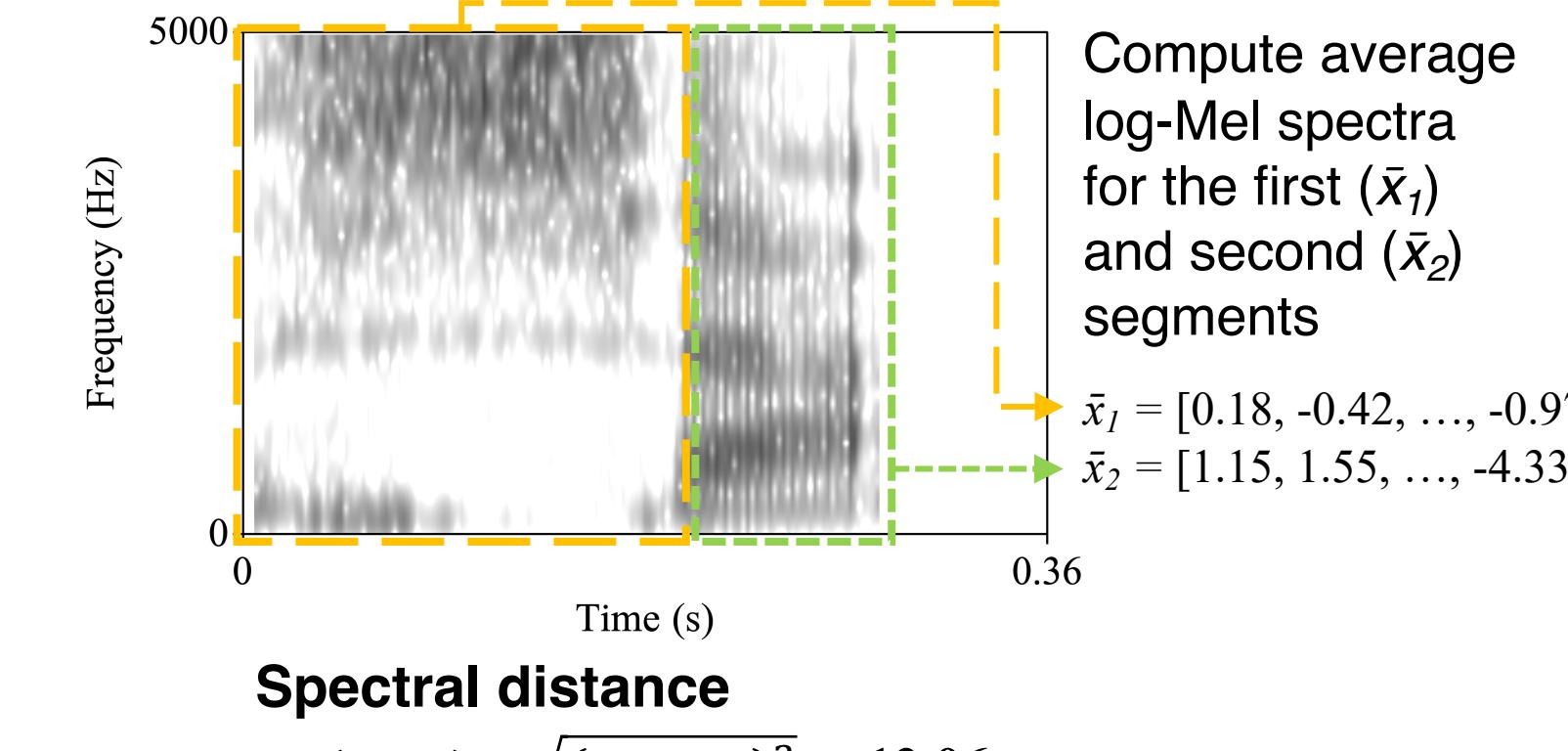
### Coarticulation analysis

- Spectral and temporal measures of coarticulation for word-internal diphones ( $N = 13,594$ ) using the whole-spectrum method [11, 12].
 

Compute average log-Mel spectra for the first ( $\bar{x}_1$ ) and second ( $\bar{x}_2$ ) segments

$\bar{x}_1 = [0.18, -0.42, \dots, -0.97]$

$\bar{x}_2 = [1.15, 1.55, \dots, -4.33]$
- Larger spectral distance or smaller overlap duration = less coarticulated
- Analyzed separately for each speaker with Bayesian hierarchical modeling [13]: measure  $\sim \text{Style}^*\text{Mask} + (\text{Style}^*\text{Mask} | \text{diphone})$



## 5. Discussion

- Compared to conversational speech, hyperarticulated clear speech showed coarticulatory resistance [8, 9].
- Speaking with a mask did not automatically induce coarticulatory resistance: conversational speech with a mask did not differ from conversational speech without a mask.
  - Even though masks attenuate high-frequencies [1, 2].
  - But they do not affect measures such as segment duration, vowel formants, etc. [5].
- Coarticulatory resistance was adaptively reinforced in clear speech produced with a face mask.
  - Only by the speaker with more extensive experience with the target language.
  - Consistent with the larger perception-in-noise benefit for masked clear speech for the native compared to the non-native speaker [6].
- The results are consistent with H&H theory [3]: speech produced in response to a communicative barrier shows coarticulatory resistance relative to speech in the absence of such barriers.
- Speakers vary coarticulatory patterns in clear speech in a graded fashion depending on the specific communicative challenges (e.g., when speaking with a face mask or imagining perceptual difficulty on the part of the listener).

### Native speaker

- Clear speech was less coarticulated than conversational speech.
- Coarticulatory resistance was greater for clear speech produced with a mask than for speech produced without a mask.
- No sig. main effect of Mask
- Same patterns for both spectral measures

### Non-native speaker

- Clear speech was less coarticulated than conversational speech.
- No sig. main effect of Mask
- No. sig. Style x Mask interaction
- Same patterns for both spectral measures

## References

- [1] Toscano, J. C., & Toscano, C. M. (2021). Effects of face masks on speech recognition in multi-talker babble noise. *PloS one*, 16(2), e0246842.
- [2] Corey, R. M., Jones, U., & Singer, A. C. (2020). Acoustic effects of medical, cloth, and transparent face masks on speech signals. *The Journal of the Acoustical Society of America*, 148(4), 2371-2375.
- [3] Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In *Speech production and speech modelling* (pp. 403-439). Springer, Dordrecht.
- [4] Cohn, M., Pyscha, A., & Zellou, G. (2021). Intelligibility of face-masked speech depends on speaking style: Comparing casual, clear, and emotional speech. *Cognition*, 210, 104570.
- [5] Kim, Y., & Thompson, A. (2022). An acoustic-phonetic approach to effects of face masks on speech intelligibility. *Journal of Speech, Language, and Hearing Research*, 1-11.
- [6] Smiljanic, R., Keercock, S., Meemann, K., & Ransom, S. M. (2021). Face masks and speaking style affect audio-visual word recognition and memory of native and non-native speech. *The Journal of the Acoustical Society of America*, 149(6), 4013-4023.
- [7] Yi, H., Pingsterhaus, A., & Song, W. (2021). Effects of wearing face masks while using different speaking styles in noise on speech intelligibility during the COVID-19 pandemic. *Frontiers in Psychology*, 12, 682677.
- [8] Guo, Z. C., & Smiljanic, R. (2021). Speakers coarticulate less when facing real and imagined communicative difficulties: An analysis of read and spontaneous speech from the LUCID Corpus. In *Proceedings of Interspeech* (pp. 4009-4013).
- [9] Moon, S. J., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *The Journal of the Acoustical Society of America*, 96(1), 40-55.
- [10] Chan, J. C., McDermott, K. B., & Roediger III, H. L. (2006). Retrieval-induced facilitation: initially nonrecited material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, 135(4), 553.
- [11] Cyphosis, M., Edwards, J. R., Munson, B., & Johnson, K. (2019). Spectral and temporal measures of coarticulation in child speech. *The Journal of the Acoustical Society of America*, 146(6), EL516-EL522.
- [12] Gerosa, M., Lee, S., Giuliani, D., & Narayanan, S. (2006). Analyzing children's speech: An acoustic study of consonants and consonant-vowel transition. *2006 IEEE International Conference on Acoustics, Speech, and Signal Processing Proceedings*, 393-396.
- [13] Bürnér, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1-28.