Methods for assessing convergence and entrainment in acoustic measures of prosody

Jennifer Cole

Marissa Goldrich

University of Illinois

Entrainment: synchrony & convergence

Edlund, Heldner & Hirschberg , 2009, *Interspeech*

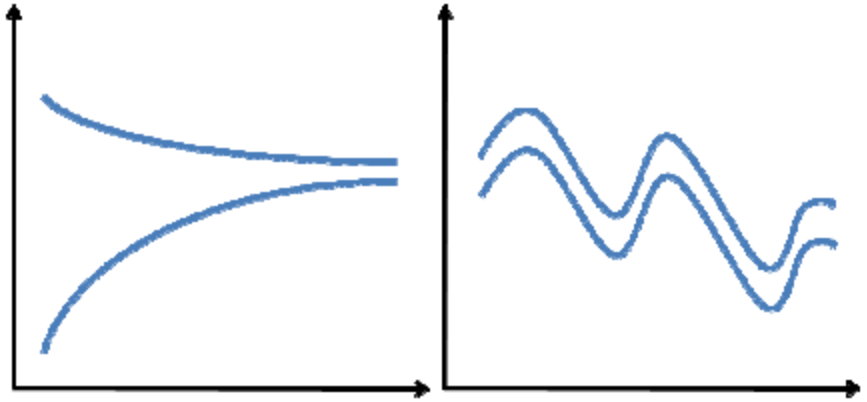


Figure 1: *Schematic illustrations of convergence (left pane) and synchrony (right pane) as they are used in this paper.*

- Between conversation partners
- Between Prosody & Gesture (w/in speaker; across speakers)
- Task effects (cooperative, competitive, conversation)
- Global and local trends

Prosodic entrainment at the phonological level

Requires prosodic annotation of prominences (or pitch accents) and boundaries

- Expert manual transcription (e.g., ToBI)
- Non-expert manual transcription –*marking only "P" and "B", no tonal features* (e.g., Rapid Prosody Transcription)
  - Crowd-sourcing, web-based
- automatic prosody detection (e.g., AuToBI)
- Hybrid method
  - small dataset transcribed by Expert and non-experts, comparison with AuToBI to determine viability of using AuToBI with new (untrained) languages.
  - Currently being tested for Hindi at Illinois by Preethi Jyothi.

Phonological measures of prosodic entrainment

- Frequency & rate of prominence and boundary expression
  - Rate analysis modeled as Poisson process?
- Pitch accent and boundary melody
  - Comparison of ToBI labels
  - Overall frequency of pitch accents, edge tones
  - f_0 modeling in region of pitch accent, boundary
- Alignment of prosodic features with syntactic landmarks
- Prosodic marking of speech preceding backchannels, discourse markers

Acoustic phonetic measures of prosodic entrainment

- f_0 (logarithmic scale)
  - mean, max, s.d., slope over measurement domain
  - contour measured at the word level with polynomial model (3rd order)
- intensity (mean, max, s.d.)
- speech rate (syll/sec)
- Pause/gap duration
- Frequency & rate of pause, gap, backchannels
  - Poisson process?

More challenging:

- Voice quality

Other vocal measures of entrainment

Frequency, duration, and rate of

- Laughter
- Backchannels
- Affirming expressions (*okay, uh-huh*)

Preliminary steps

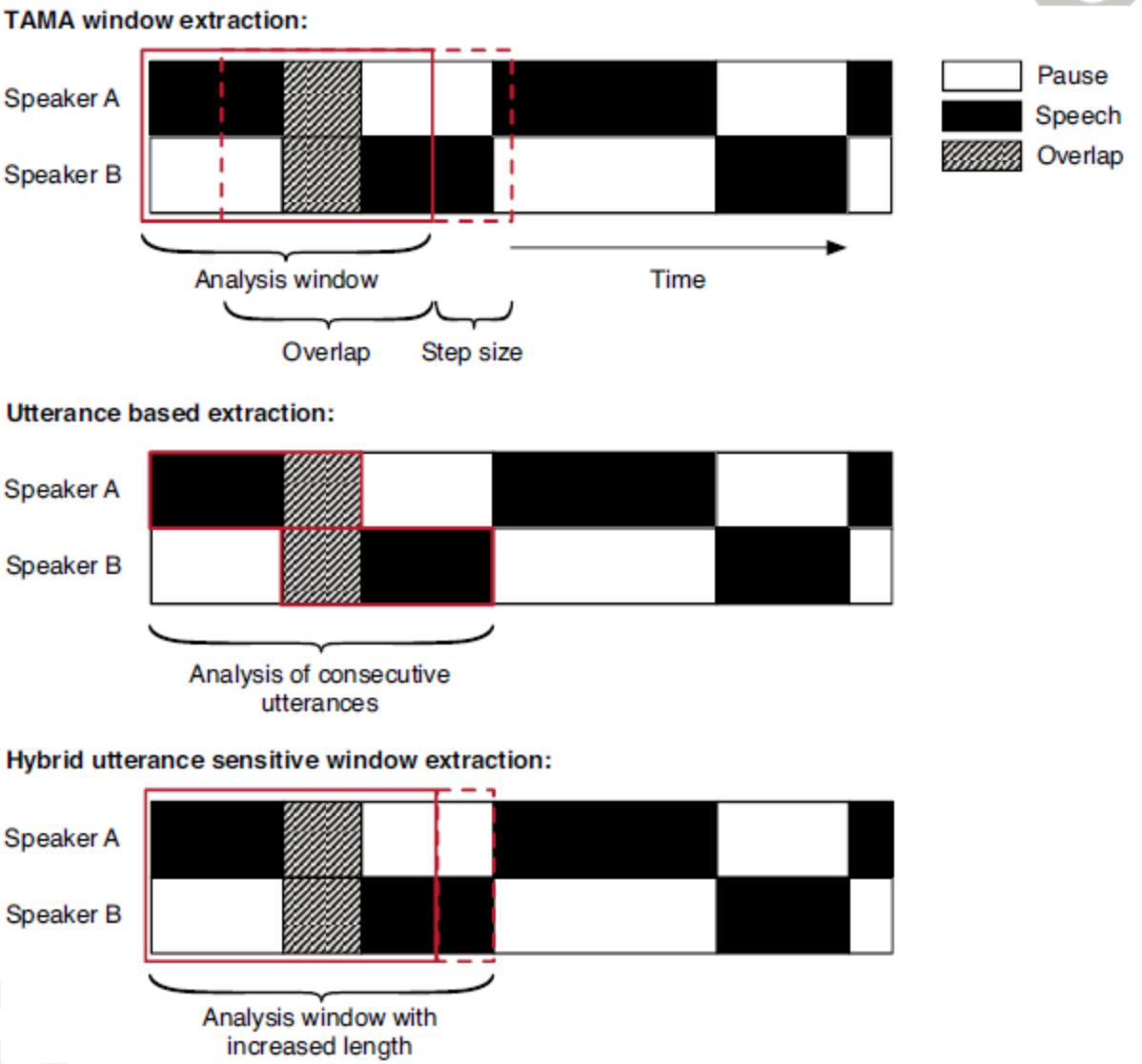
- Orthographic transcriptions converted to phone transcription; automate if digital dictionary is available. Syllabify phone transcription (use scripts to automate)
- Automatic detection and labeling of speech and non-speech intervals
- Set pause threshold; label non-speech intervals above threshold as pause.
- f_0
  - remove or correct halving and doubling errors
  - median filtering to smooth f_0 contour at the edges of sonorant and non-sonorant segments.
- Intensity
  - Measure in voiced intervals (auto detect)

Extract acoustic measures in small and large domains. Some example domain:

- Overlapping, fixed-length windows (e.g., 100 ms; 10 prosodic events) *--see ex. next slide*
- Word
- Inter-pause unit
- Turn
- Task or conversation session

Center the f_0 and intensity measurements for within-speaker normalization, using mean over all turns for that speaker

de Looze, et al. (in press). *Speech Communication*



Quantifying entrainment

Compute similarity measures to compare signals from two speakers

Single variable analysis (e.g., one acoustic measure)

- correlation and regression (eg., de Looze et al, *in press*)
- difference measures
- Compare similarity metrics for real vs. fake conversation partners.
- If the mean *sim*(Real) > mean *sim*(Fake), that's taken as evidence of entrainment

Look for dynamic trends within domain

Compare degree of entrainment across domains

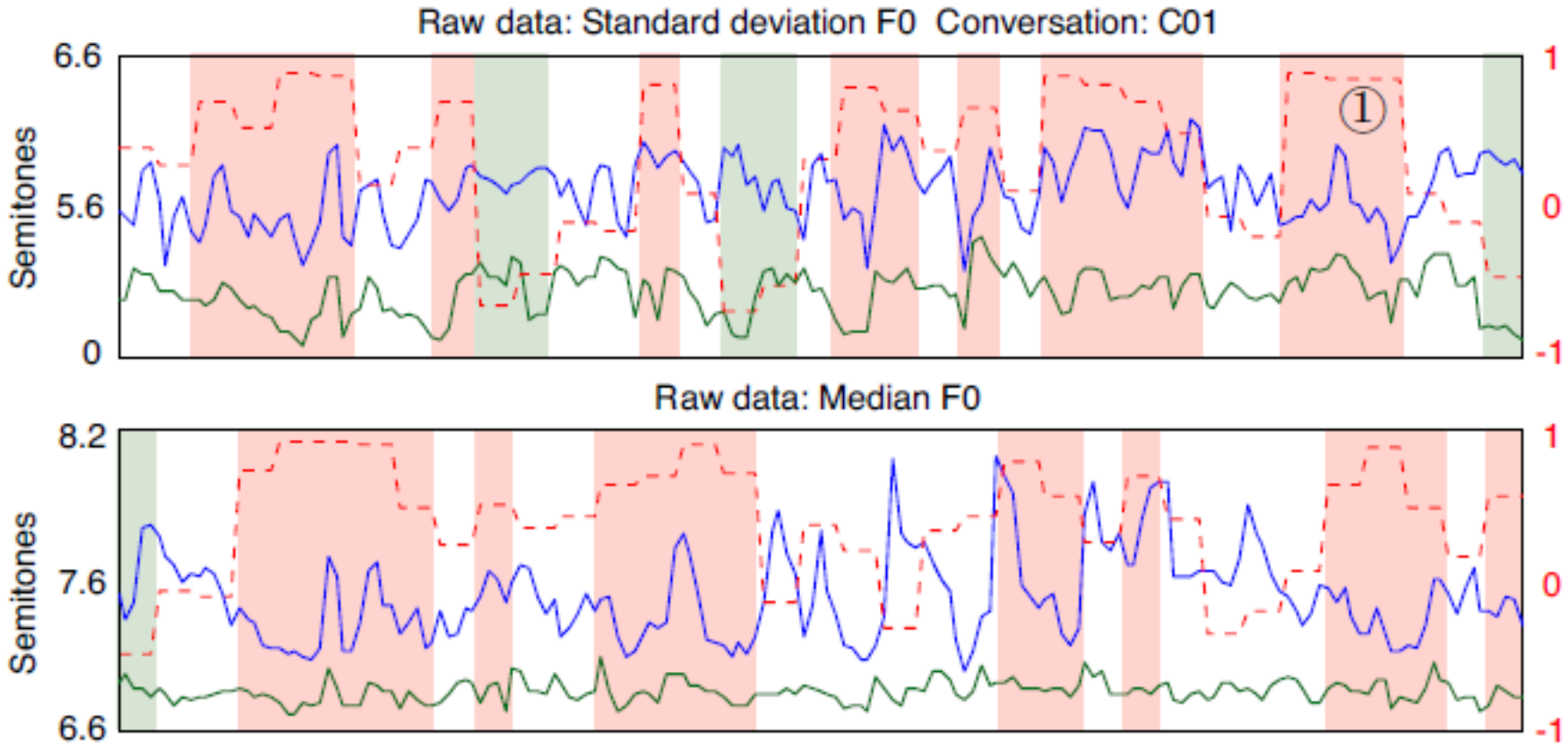


Figure 3: Raw values of Pair 1's first conversation of each extraction window for each speaker (Speaker 1 in blue; Speaker 2 in green) for the respective feature (left ordinate axis) along with Pearson's ρ values (dashed red line; right ordinate axis). Phases of synchrony are highlighted in red phases of asynchrony in green.

Part of Figure 3 from de Looze et al. (in press). *Speech Communication*

Quantifying entrainment

Multivariate analysis

Using Principal Component Analysis over set of acoustic variables (Lee et al., 2013, *Computer Speech and Language*).

- Similarity/dissimilarity metrics calculated over PCA components (details suppressed here)
- Compare similarity metrics for two conversation partners with similarity metrics for "fake" pairs (as for single-variable analysis)

Quantifying entrainment

Perceptual analysis

- Listeners rate similarity between conversation partners. Compare ratings before and after interaction (or early vs. late in interaction).
- Use linear mixed effects regression with acoustic measures as predictor variables for perceptual ratings.

Interesting comparisons:

- Familiar vs. unfamiliar conversation partners
- Same vs. prosodically distinct dialect

What is the relationship between entrainment measures and

  - Task outcome?
  - Qualitative ratings of the interaction (by participants, or by uninvolved third party)?

- Black, M., Katsamanis, A., Baucom, B., Lee, C.-C., Lammert, A., Christensen, A., … Narayanan, S. S. (2013). Toward automating a human behavioral coding system for married couples' interactions using speech acoustic features. *Speech Communication*, *55*.
- De Looze, C., Scherer, S., Vaughan, B., & Campbell, N. (2013). Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction. *Speech Communication*.
- Edlund, J., Heldner, M., & Hirschberg, J. (2009). Pause and gap length in face-to-face interaction.
- Lee, C.-C., Black, M., Katsamanis, A., Lammert, A., Baucom, B., Christensen, A., … Narayanan, S. S. (2010). Quantification of Prosodic Entrainment in Affective Spontaneous Spoken Interactions of Married Couples. *Proceedings of Interspeech 2010*.
- Lee, C.-C., Katsamanis, A., Black, M., Baucom, B., Christensen, A., Georgiou, P., & Narayanan, S. S. (2013). Computing vocal entrainment: A signal-derived PCA-based quantification scheme with application to affect analysis in married couple interactions. *Computer Speech and Language*.
- Levitan, R., Gravano, A., & Hirschberg, J. (2011). Entrainment in Speech Preceding Backchannels.
- Levitan, R., & Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions.
- McGarva, A., & Warner, R. (2003). Attraction and Social Coordination: Mutual Entrainment of Vocal Activity Rhythms. *Journal of Psycholinguistic Research*, *32*(3).
- Nenkova, A., Gravano, A., & Hirschberg, J. (2008). High Frequency Word Entrainment in Spoken Dialogue. *Proceedings of ACL-08*.
- Pardo, J. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, *119*(4).
- Pardo, J., Jordan, K., Mallari, R., Scanlon, C., & Lewandowski, E. (2013). Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *Journal of Memory and Language*, *69*, 183–195.