# Prosodic Entrainment in Dialogue: Language and Social Impact

Julia Hirschberg
Columbia University

# Real World Example

Waiting for a long time for the elevator on our floor. Two undergrads talking. One says she is a senior (let's call her Senior) and talks about what she has done this summer and her activities here on campus. Relatively high pitch, a light pleasant voice, normal speed. Other one is younger (let's call her Junior) and is giving opinions on the other person's activities. She has vocal fry/ creaky voice and low pitch and fast speech, which seems to be the norm of the day. As the conversation goes on, Senior gets faster and faster. The last part of it that I heard before they exited the elevator, Senior was now in Junior's pitch range and starting to show vocal fry. (Maxine Eskenazi, 9/13/17)

# Entrainment/Alignment/Adaptation: *The Chameleon Effect*

- Speech Accommodation Theory: *"In conversation, people tend to adapt their communicative behavior to that of their conversational partner." (Giles et al '87)*
  - ***Chameleon Effect:*** Non-conscious ***mimicry*** of the postures, mannerisms, facial expressions, and other behaviors of one's interaction partners (Chartrand & Bargh 1999).
  - ***Perception-behavior link***: the underlying mechanism for the Chameleon Effect --- " Unintentional, non-conscious ***effects*** of social perception on social behavior" (Chartrand, Maddux, & Lakin, 2005)

# Entrainment in Multiple Dimensions

- Lexical and syntactic (Brennan ‘00, Reitter et al ‘07)
- Acoustic/Prosodic (Matarazzo et al ‘68, Jaffe & Feldstein ‘70, Natale ‘77, Cappella & Planalp ‘81, Street ‘84, Sherlom & La Riviere ‘87, Guitar & Marchinkoski ‘01)
- Phonological/Phonetic (Pardo ‘06)
- Socio-cultural (Azuma ‘97, Roth ‘05)
- Jokes and laughter (Bales ‘50, Raganath et al ‘11)
- Facial expression and gesture (Mauer & Tindall ‘83, Hale & Burgoon ‘84, Chartrand & Bargh ‘99)
- Posture (Condon & Ogston ‘67)
- Brain oscillation and speech amplitude (Bosker & Kösem ‘17)

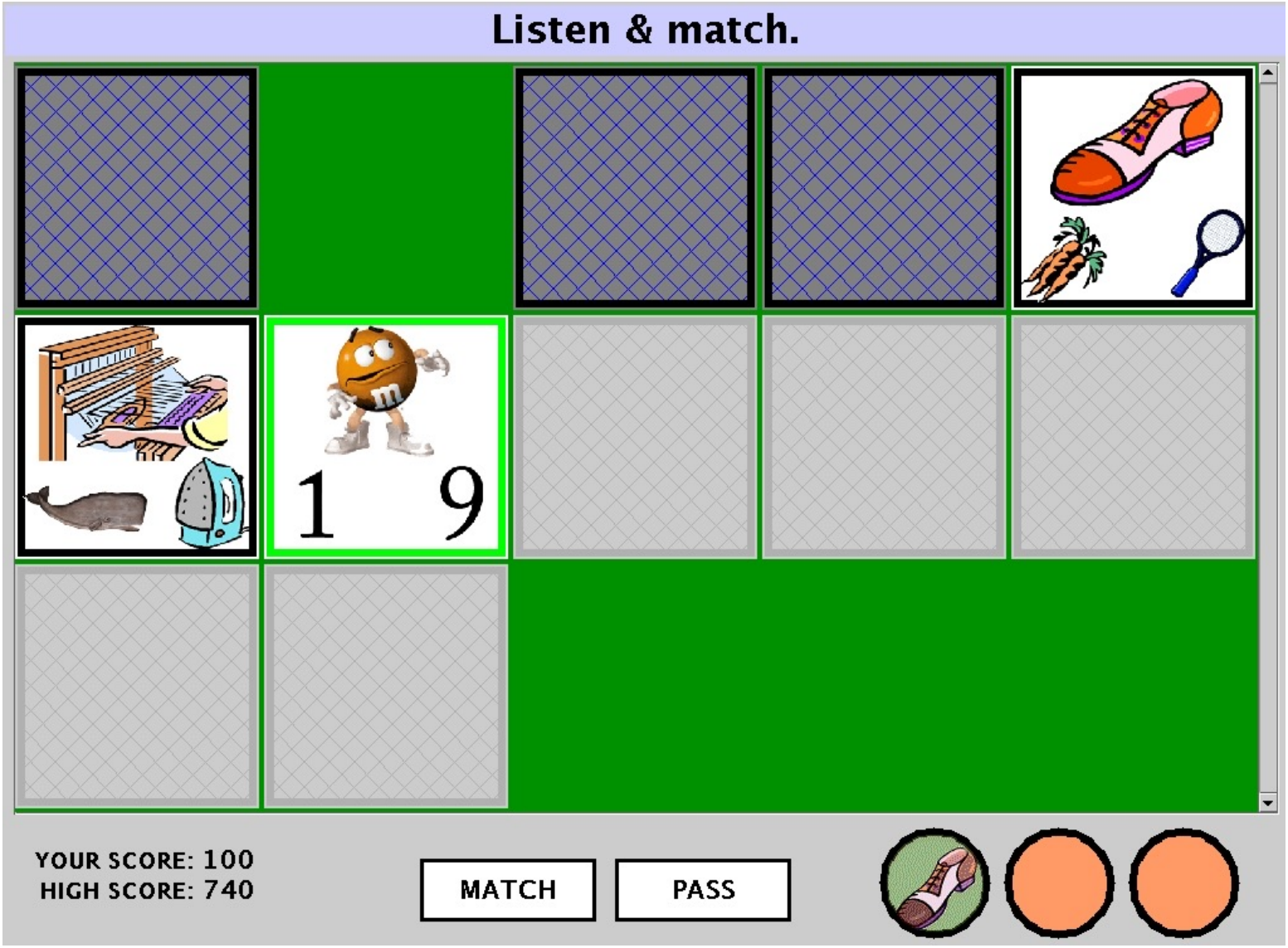
# Effects of Entrainment on Social Perception

- Subjects who entrain
  - Perceived as more socially ***attractive*** (Putnam & Street '84, Bourhis et al '75)
  - Perceived as more ***competent*** (Street '84)
  - Conversation perceived as more ***intimate*** (Buller & Aune '88)
- Entrainment leads subjects to ***like*** their conversational partners (and their computers) more and to perceive interactions as ***more successful*** (Nass et al '95, Chartrand & Bargh '99)
- Long-term syntactic entrainment is a good predictor of actual ***task success*** in Map Task (Reitter et al '07)

# The Columbia Games Corpus (Gravano '09)

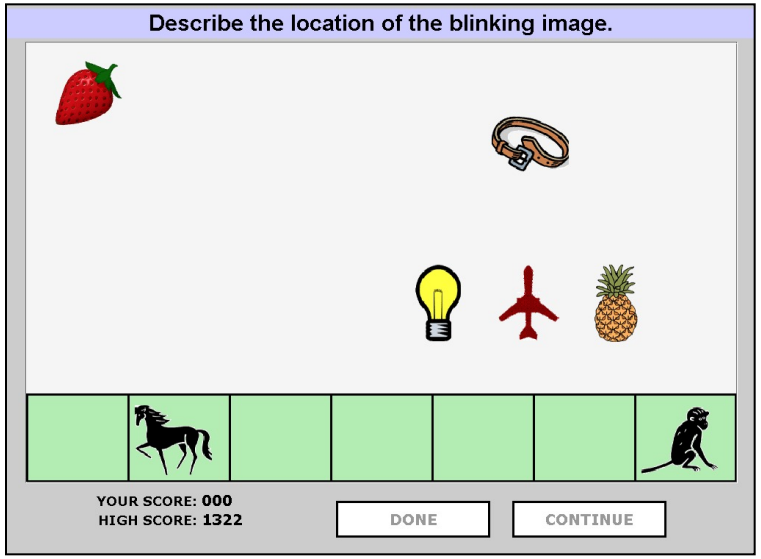
- Initial goal: study prosody of given/new items
  - 12 spontaneous task-oriented dyadic conversations (9h 8m speech)
  - 2 subjects play series of computer games, no eye contact (45m 39s mean session time)
  - 2 sessions per subject, *w/different partners*
  - Multiple games and types
- Recorded on separate channels in soundproof booth, digitized and down-sampled to 16k
- Features extracted with Praat

# The Cards Game

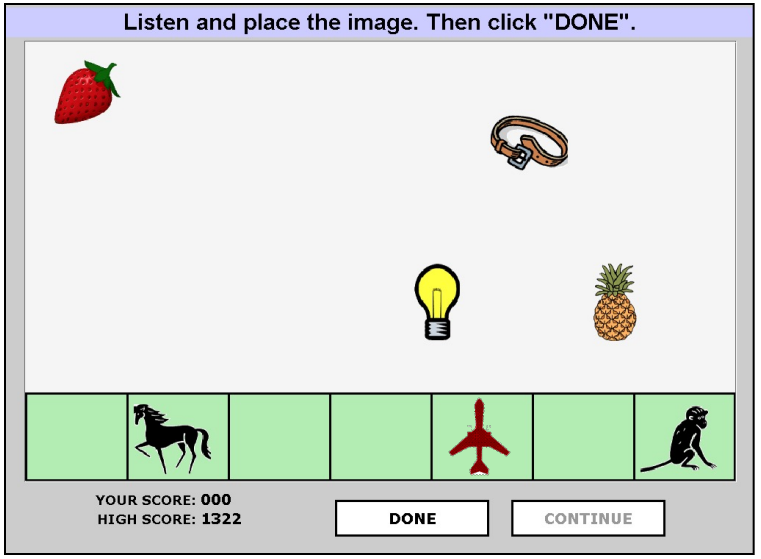


# The Objects Game

Describer:




Follower:



# Units of Analysis

- ***Inter-pausal unit*** (IPU): Pause-free segment of speech (50ms or more) from a single speaker
- speech <silence> speech <silence> speech
- ***Turn***: Sequence of speech from one speaker without intervening speech from the other speaker.
- ***Session***: Complete interaction between two subjects on one task

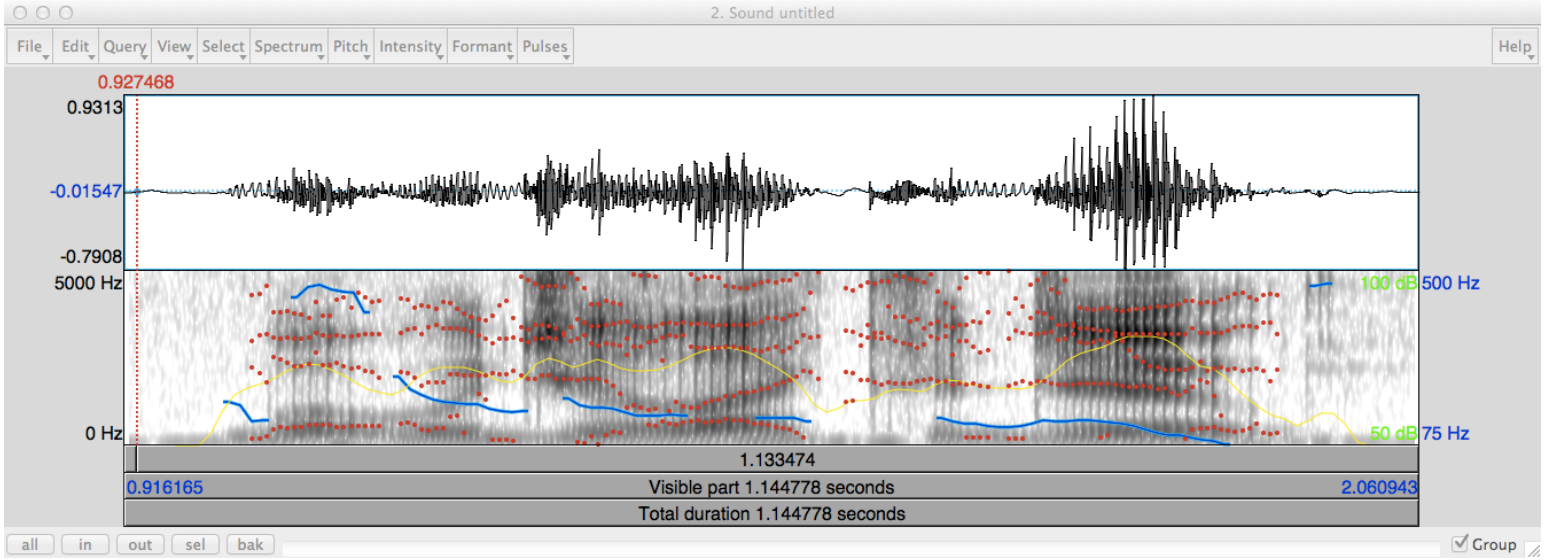
# Units of Analysis

- ***Inter-pausal unit*** (IPU): Pause-free segment of speech (50ms or more) from a single speaker

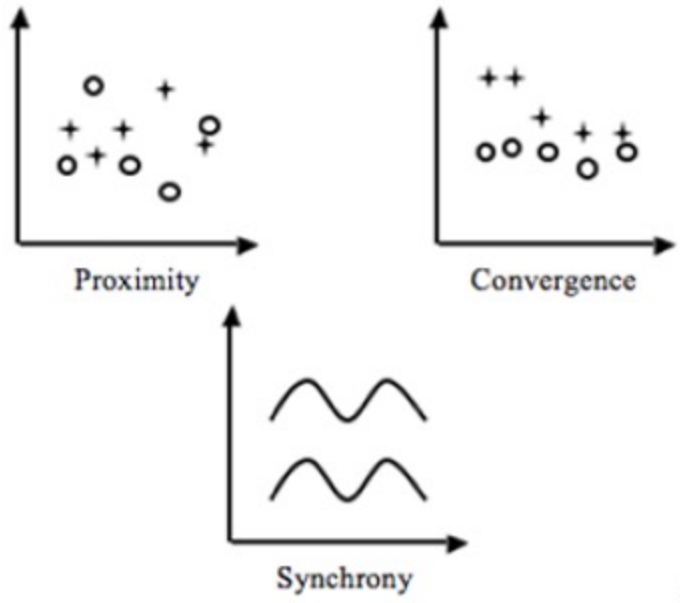

- ***Turn***: Sequence of speech from one speaker without intervening speech from the other speaker.
- ***Session***: Complete interaction between two subjects on one task

# Low Level Prosodic Features

- Intensity mean
- Intensity max
- Intensity min
- F0 mean
- F0 max
- F0 min
- speaking rate



# Forms of Entrainment (Levitan & Hirschberg '11)




***Proximity*** ---- significant similarity of partner features

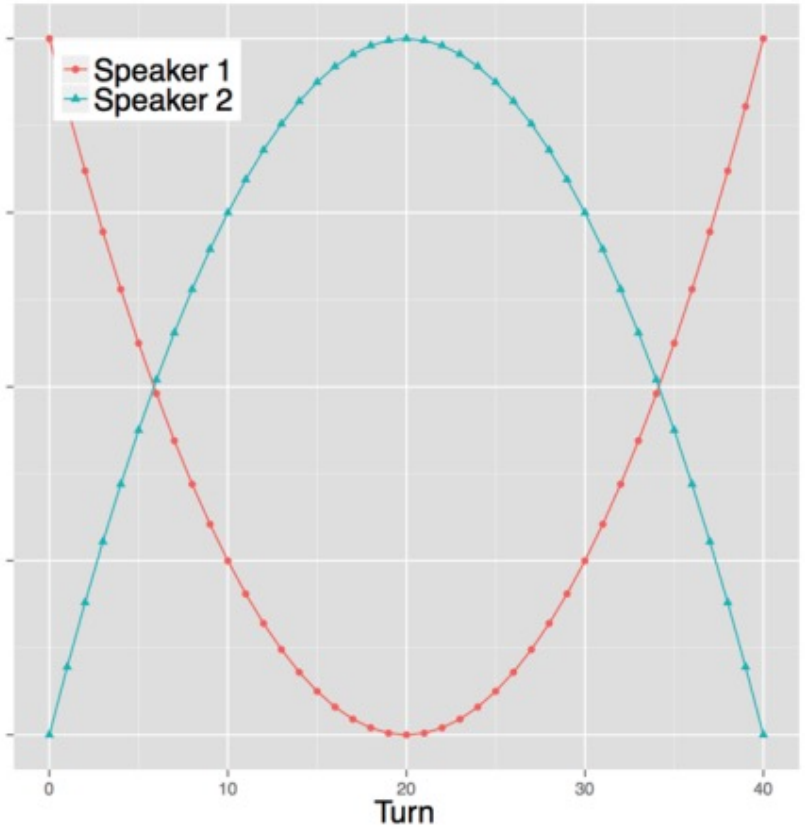
***Convergence*** ----significant increase in similarity of partner features over time

***Synchrony*** ---- correlated relative change in partner features

Correcting for Type 1 error (false pos)

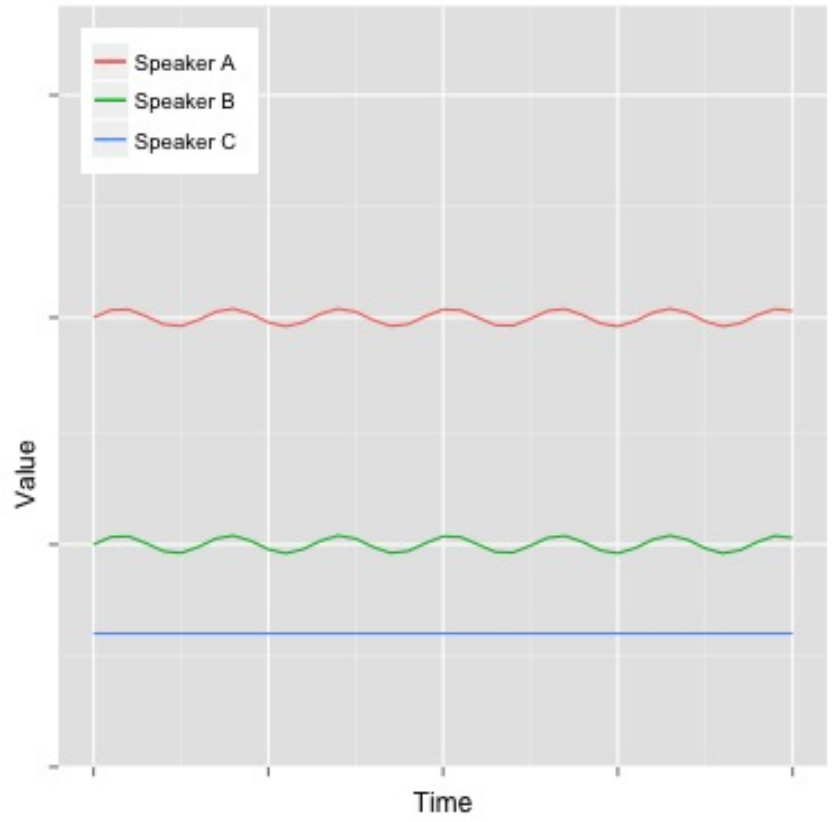
# Similarity/Proximity

- Global or local?
- Exact or relative?
- Convergent or constant?



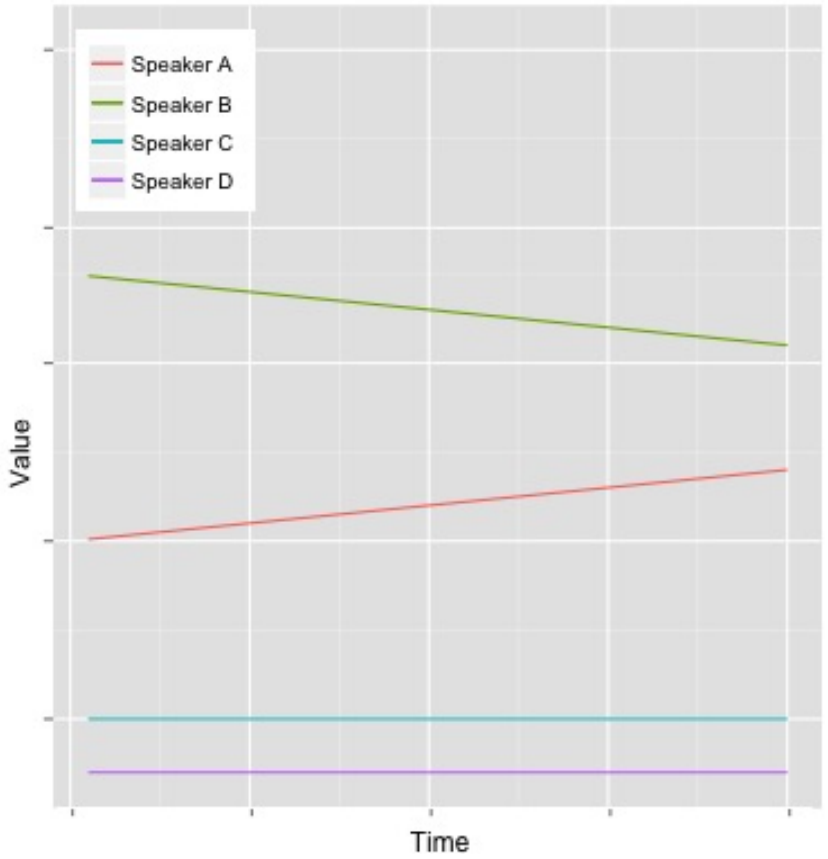
# Synchrony

- Global or local?
- Exact or relative?
- Convergent or constant?



# Convergence

- Global or local?
- Exact or relative?
- Convergent or divergent or constant?



# Defining Global Pairwise Entrainment

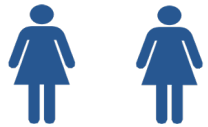
- Similarity

$$ENT_p = -|\text{speaker1}_f - \text{speaker2}_f|$$

*Partner* similarity

$$ENT_x = -\frac{\sum_i |\text{speaker1}_f - \text{X}_{i,f}|}{|\text{X}|}$$

*Non-partner* similarity

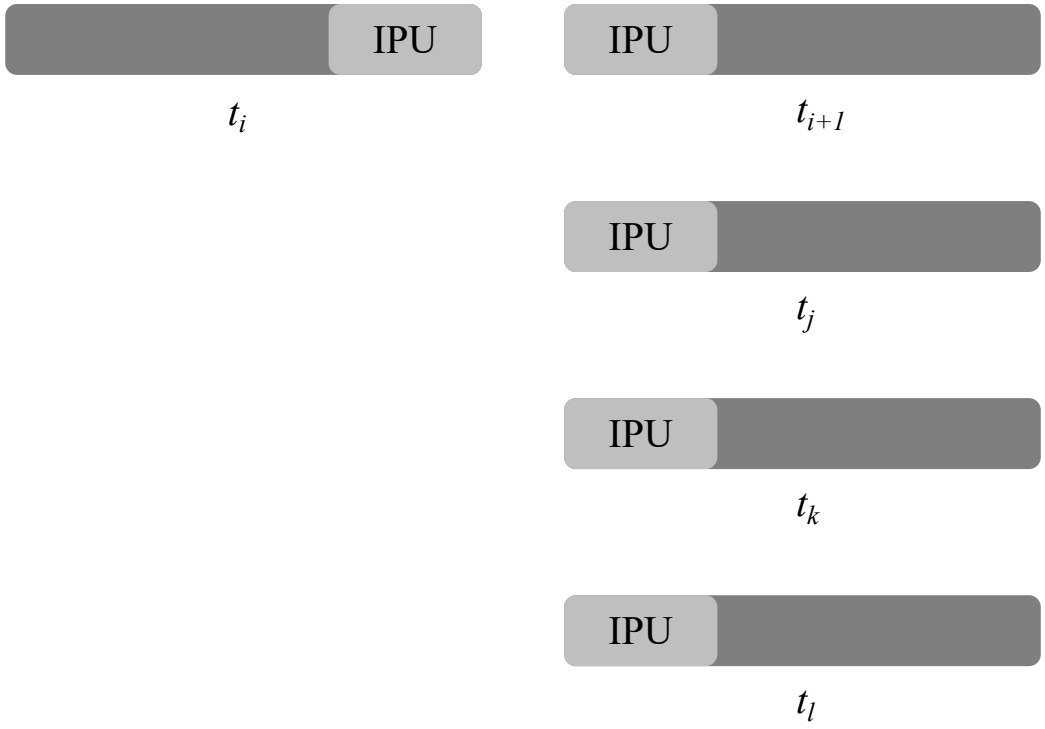


$$ENT_{self}(s, f) = -|s_f - s'_f|$$

*Self* similarity

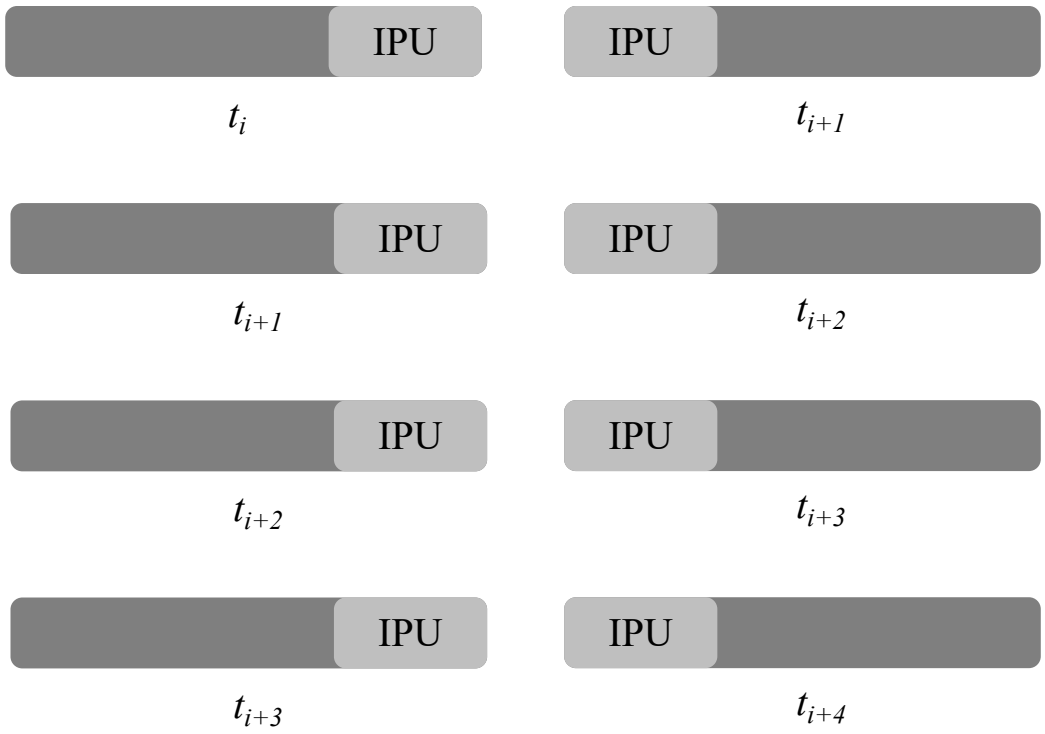
- Synchrony: positive correlation between partners
- Convergence: negative correlation between partners

# Local Entrainment: Proximity




Comparing to 10 random initial IPUs

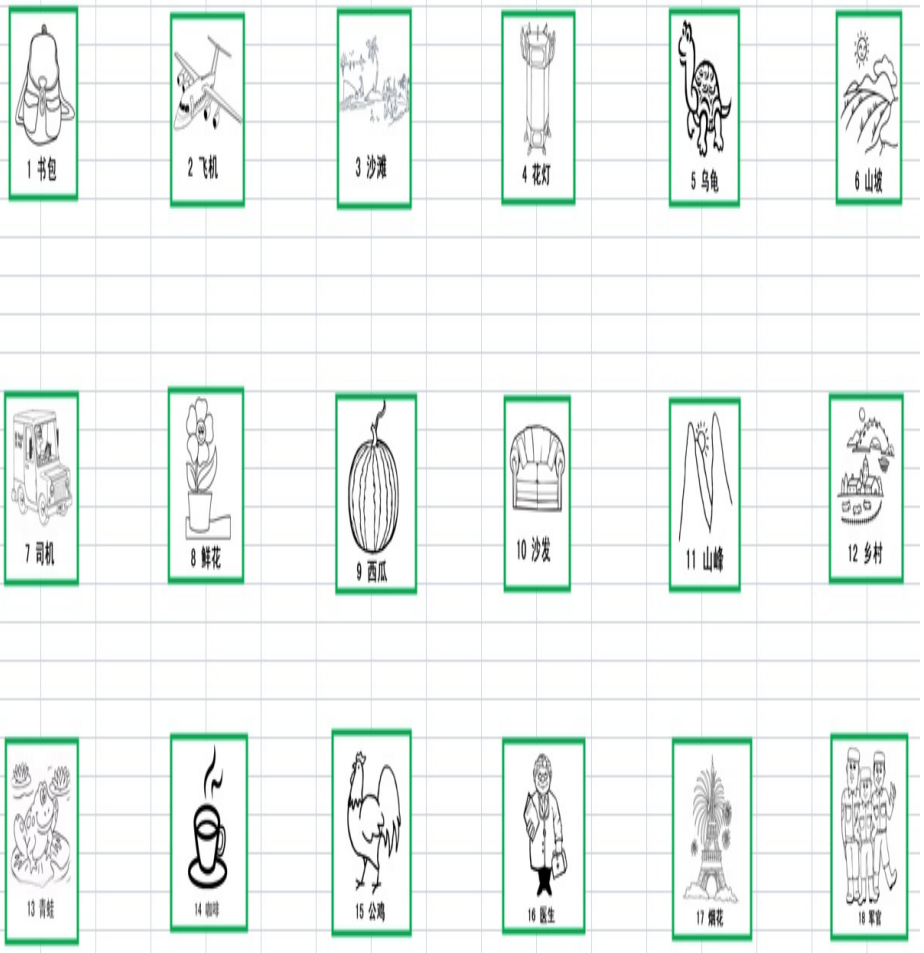
# Local Entrainment: Synchrony, Convergence



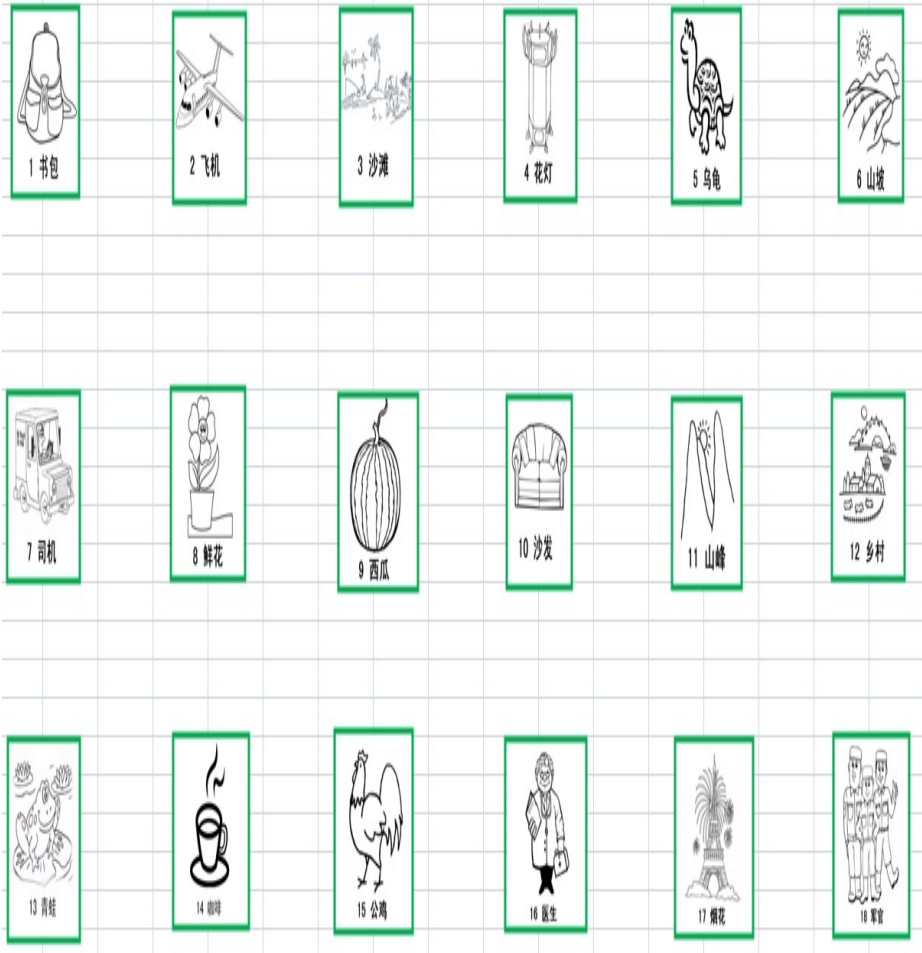
# Tongji Games Corpus (Xia et al ’14)

- Collaboration with Zihua (Shirley) Xia
  - Inspired by our work, she had recorded 115 spontaneous task-oriented sessions
    - 70 pairs of speakers (40 female, 30 male)
    - 12 hours of recorded dialogue
    - University students with a National Mandarin Test Certificate level 2, grade A or above
- Elicited using two games: Picture Ordering (role imbalance), Picture Classifying (cooperative)
- We compared entrainment in both corpora

# Picture Ordering Game

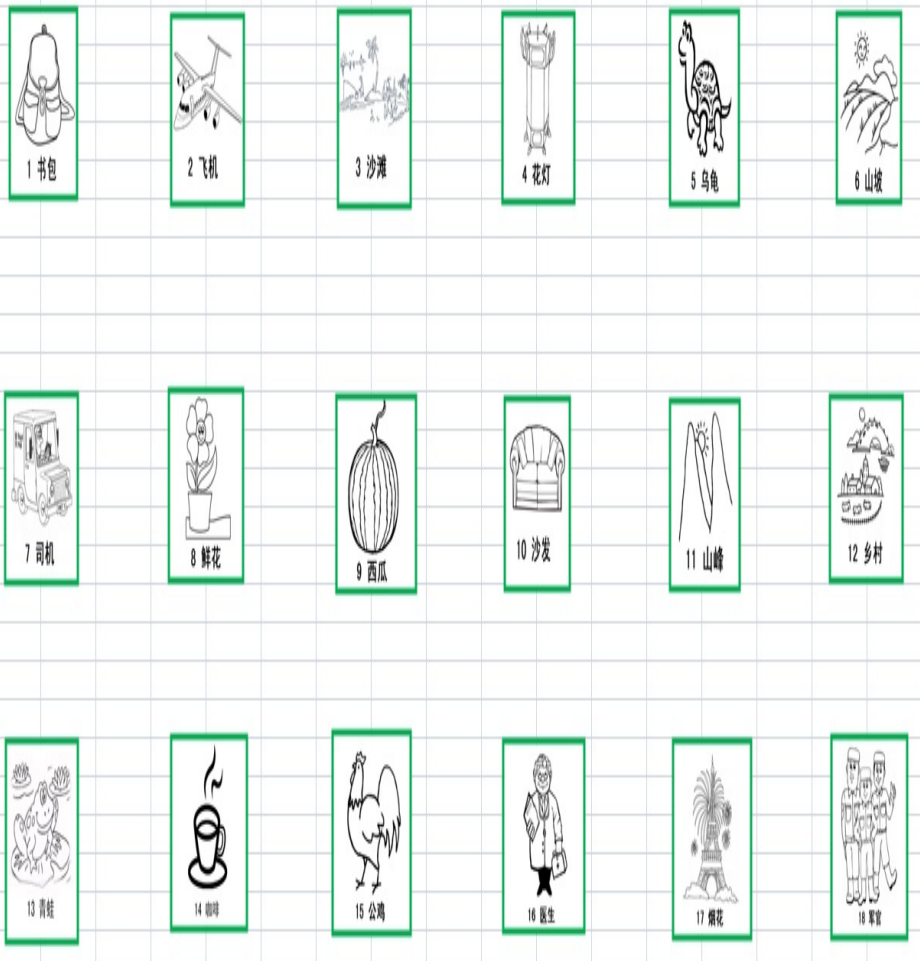



A

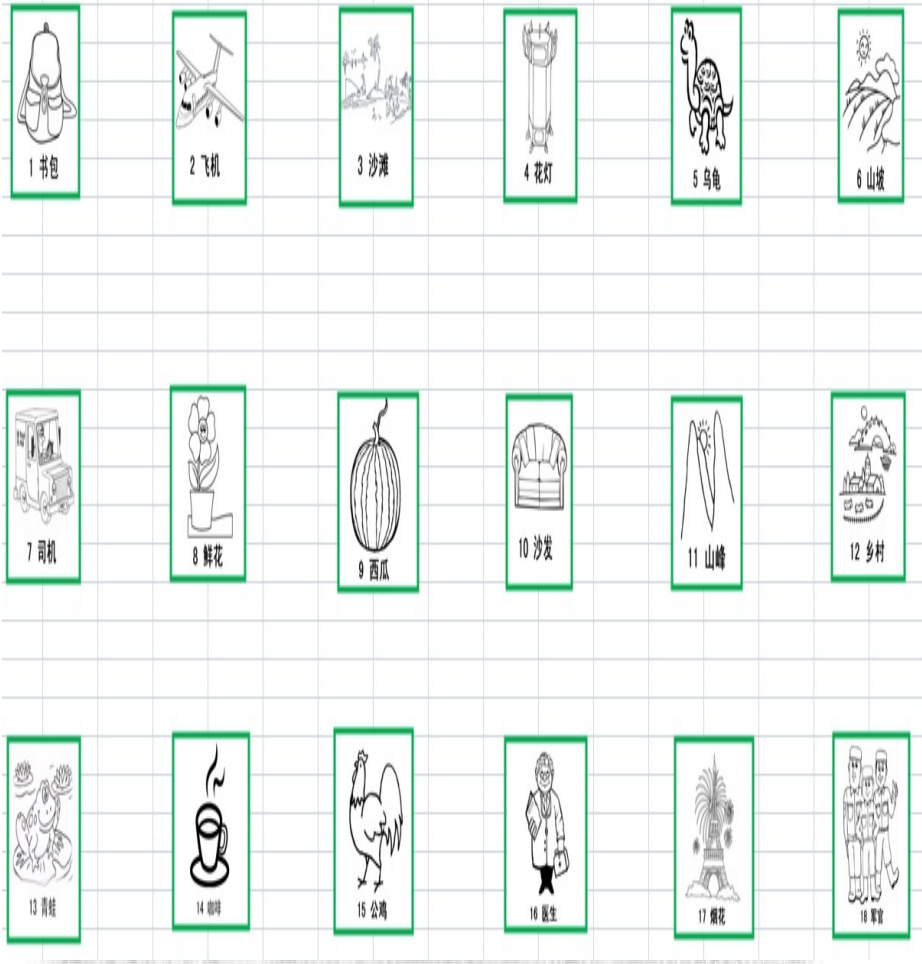



B

# Picture Classifying Game




A




B

# Comparing American with Mandarin Partners (Levitan '14, Xia et al '14)

| | *Global similarity* | | *Local similarity* | | *Synchrony* | | *Global convergence* | | *Local convergence* | |
|---|---|---|---|---|---|---|---|---|---|---|
| ***Feature*** | **SAE** | **MC** | **SAE** | **MC** | **SAE** | **MC** | **SAE** | **MC** | **SAE** | **MC** |
| Intensity mean | ✓✓ | ✓ | ✓ | ✓ | (✓) | ✓ | | | – | |
| Intensity max | ✓(✓) | ✓ | ✓ | ✓ | (✓) | ✓ | | | – | |
| Pitch mean | | | | | – | ✓ | | | – | – |
| Pitch max | | ✓ | | | – | ✓ | ✓ | | – | – |
| Jitter | | – | | – | – | – | | – | | – |
| Shimmer | | – | | – | – | – | | – | | – |
| NHR | | – | ✓ | – | – | – | ✓ | – | | – |
| Speaking rate | ✓ | ✓ | | ✓ | | | ✓ | | | – |

One check: significant difference from non-partner similarity
Two checks: also significant difference from self similarity

# Comparing American with Mandarin Partners (Levitan '14, Xia et al '14)

| | *Global similarity* | | *Local similarity* | | *Synchrony* | | *Global convergence* | | *Local convergence* | |
|---|---|---|---|---|---|---|---|---|---|---|
| ***Feature*** | **SAE** | **MC** | **SAE** | **MC** | **SAE** | **MC** | **SAE** | **MC** | **SAE** | **MC** |
| Intensity mean | ✓✓ | ✓ | ✓ | ✓ | (✓) | ✓ | | | – | |
| Intensity max | ✓(✓) | ✓ | ✓ | ✓ | (✓) | ✓ | | | – | |
| Pitch mean | | | | | – | ✓ | | | – | – |
| Pitch max | | ✓ | | | – | ✓ | ✓ | | – | – |
| Jitter | | – | | – | – | – | | – | | – |
| Shimmer | | – | | – | – | – | | – | | – |
| NHR | | – | ✓ | – | – | – | ✓ | – | | – |
| Speaking rate | ✓ | ✓ | | ✓ | | | ✓ | | | – |

One check: significant difference from non-partner similarity
Two checks: also significant difference from self similarity

# Entrainment in Different Languages and Situations

- Standard American English (SAE) vs. Mandarin Chinese (MC) showed surprising similarities on multiple metrics
- Similarity
  - Global : similar intensity, rate
  - Local: similar intensity
- Synchrony
  - Stronger synchrony for MC (intensity, pitch)
- Convergence
  - Global for SAE only (pitch, NHR, rate)
  - Stronger local convergence for MC (pitch)

# Social Dimensions of Entrainment

- Recall that subjects who entrain are
  - Perceived as more socially ***attractive*** (Putnam & Street '84, Bourhis et al '75)
  - Perceived as more ***competent*** (Street '84)
  - Speech perceived as more ***intimate*** (Buller & Aune '88)
- Entrainment leads subjects to ***like*** their conversational partners (and their computers) more and to perceive interactions as ***more successful*** (Nass et al '95, Chartrand & Bargh '99)
- Long-term syntactic entrainment a good predictor of actual ***task success*** in Map Task (Reitter et al ' 07)

# Annotation of Social Variables

- Amazon Mechanical Turk workers labeled 168 Columbia Games Corpus object games (5 turkers per game)
- Answered following questions about partners
  - Does s/he believe s/he is better than his/her partner?
  - Making it difficult for his/her partner to speak?
  - Seem engaged in the game?
  - Seem to dislike his/her partner?
  - Is s/he bored with the game?
  - Directing the conversation?
  - Frustrated with his/her partner?

  - *Encouraging his/her partner?
  - *Trying to dominate the conversation?
  - Making him/herself clear?
  - Planning what s/he is going to say?
  - Polite?
  - *Trying to be liked
- Questions about the conversation
  - *Does it flow naturally or is it awkward?
  - Are the participants having trouble understanding each other?
  - Which person do you like more?
  - Who would you rather have as a partner?

# Hypotheses from Literature

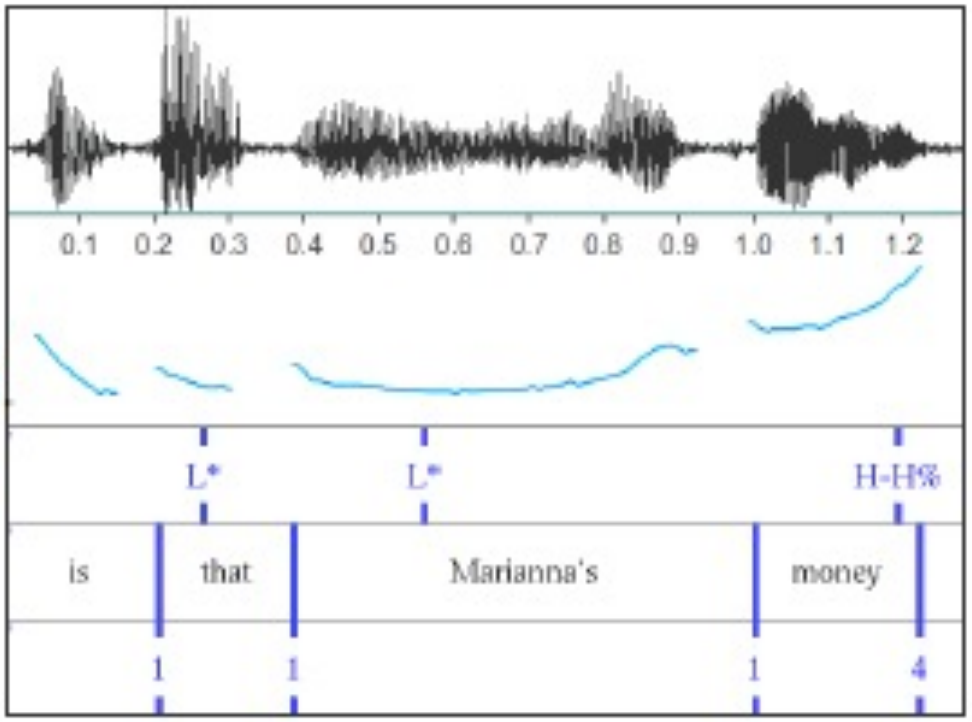
- Communication Accommodation Theory
  - Giving encouragement: positively correlated with entrainment
  - Conversational awkwardness: negatively
- Similarity-Attraction Theory
  - Trying to be liked: positively correlated?
- Dependency Over-Accommodation occurs when an interlocutor appears to be trying to dominate or control the conversation: excessive entrainment (West & Turner, 2009).

# Findings vs. Hypotheses

- From Communication Accommodation Theory
  - **Perceiving to be giving encouragement: positively correlated** with entrainment
  - **Perceived conversational awkwardness:** (in fact, a weak **positive** correlation)
- Based on Similarity-Attraction Theory
  - **Perceived as trying to be liked: positively correlated** (but raters **dis-preferred** these speakers)
  - **No correlation** between **perceived dominance** and entrainment: no Dependency-Over-Accommodation?

# Entrainment in Higher Level Prosodic Features (Gravano et al '14)

- **ToBI annotation** of Columbia Games Corpus
  - Three expert labelers using the ToBI conventions:
  - Tonal tier: targets in the F0 contour
    - Pitch accents: H*, L*, L+H*, L*+H, H*+L, downstep
    - Phrase accents: H-, L-, !H-
    - Boundary tones: H%, L%.
  - Orthographic tier: time-aligned words
  - Break index tier: degrees of juncture (0-4)
  - Misc tier: laughs, disfluencies, etc.


0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
L*
L*
H-H%
is
that
Marianna's
money
1
1
1
4

# Entrainment on Pitch Contours and Social Variables (Gravano et al ’14, ‘15)

- Measures of contour similarity between speakers: when I use contour X are you more likely to do the same?
  - **Perplexity of language models** of sequences trained on Speaker A and used to model prosodic sequences of Speaker B: *low perplexity indicates greater similarity*
  - **Levenshtein distance** of similar intonational phrase contours used by Speaker A and Speaker B: *low values show similar contours are uttered closer together*
  - **Kullback-Leibler divergence** between contours of Speaker A and Speaker B: *low values show that one is a subset of the other*
- How ***similar*** are Speaker A’s contours to Speaker B’s?

# Experiments

- Built 24D vectors with the value of each of the 3 metrics for each member of each speaker pair
- Built similar vector for each social variable v (e.g., bored-with-game) where Aj, Bj are the two speakers from the same session j
- Ran Pearson's correlation tests between entrainment vectors and social variable vectors

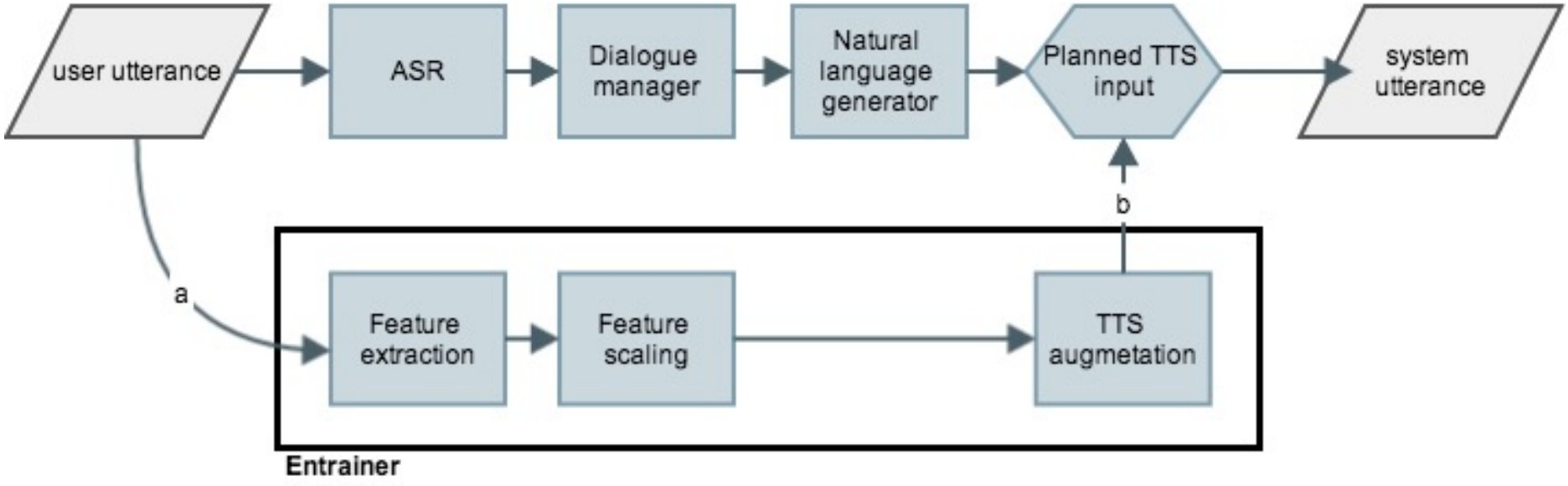
# Correlations for Different Prosodic Similarity Metrics with Social Variables

| Social Variable | Perplexity | Levenshtein Dist | KL Divergence |
|---|---|---|---|
| Making-self-clear | pos | pos | |
| Giving encouragement | | pos | pos |
| Engaged-in-game | neg | pos | pos |
| Contributes-to-successful-task-completion | pos | pos | pos |
| Trying-to-be-liked | | | pos |
| Planning-what-to-say | pos | pos | |
| Dislikes-partner | | neg | |
| Making-it-difficult-for-partner-to-speak | | pos | pos |
| Bored-with-game | neg | neg | neg |

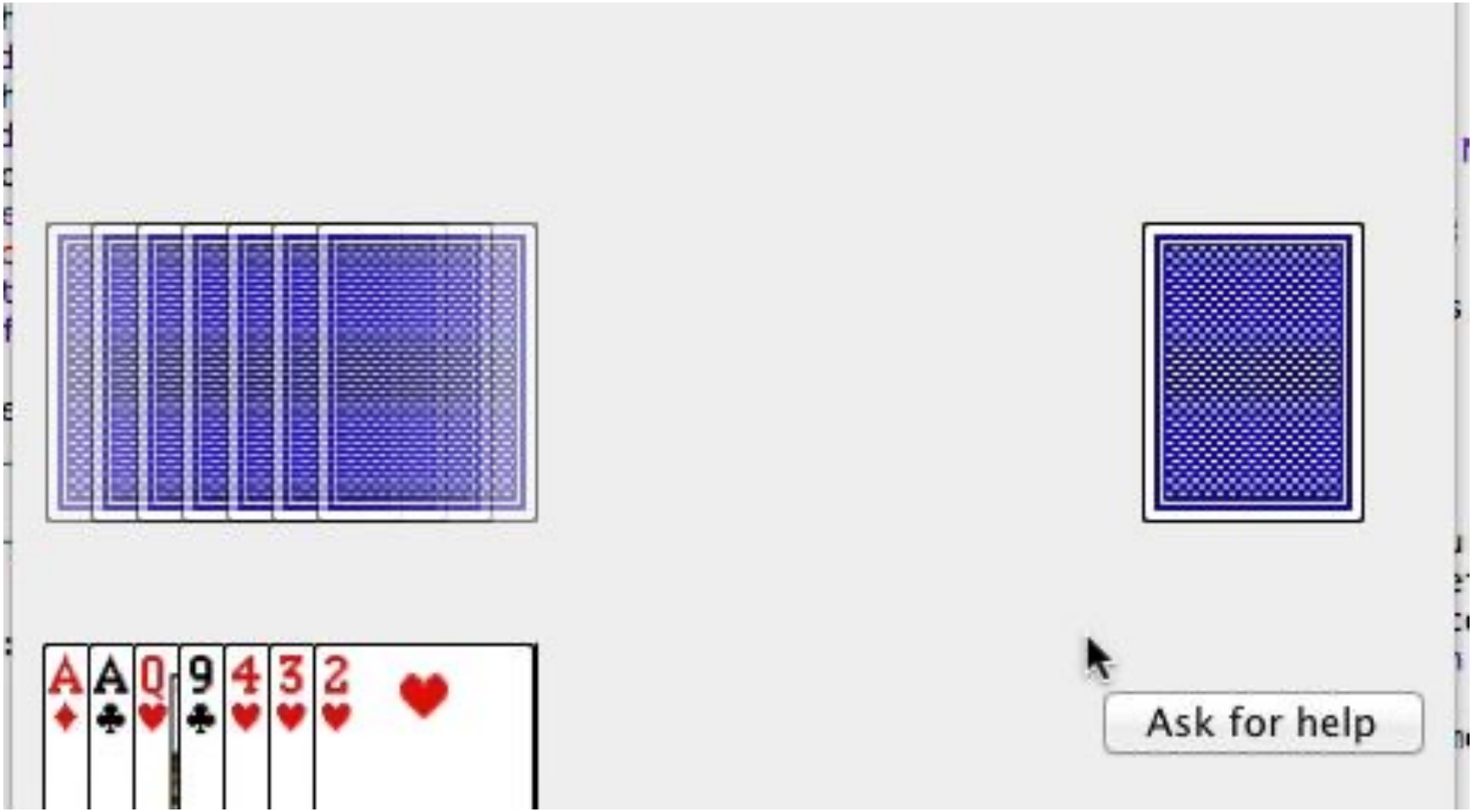
# Conclusions

- 3 novel metrics of entrainment on intonational contours annotated within the ToBI framework.
- Findings:  correlations of prosodic entrainment with perceived levels of
  - speaker engagement
  - positive partner-oriented features of social behavior (giving encouragement, making self clear, etc.)
- Future work: Automate computation of our measures using automatic prosodic labeling tools (e.g., AuToBI).

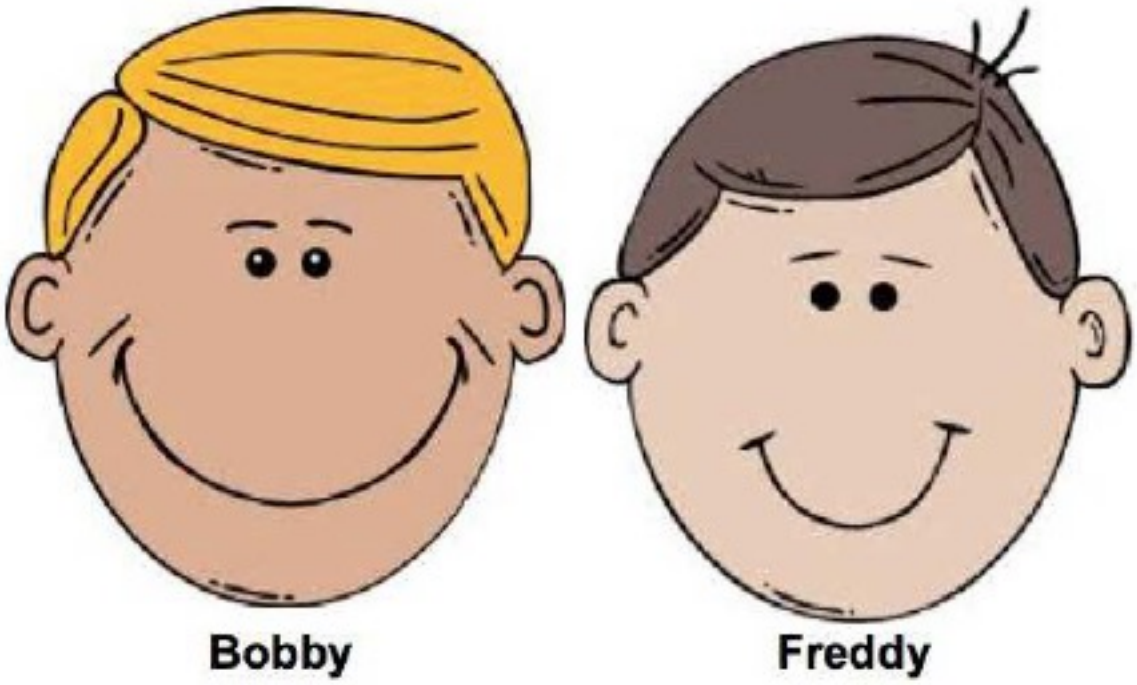
# Entraining on Rate and Intensity to Users in Spoken Dialogue Systems (Levitan et al '16)

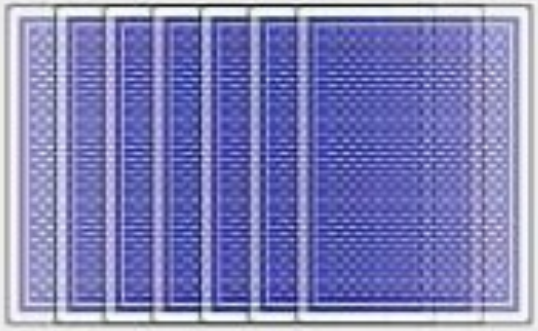


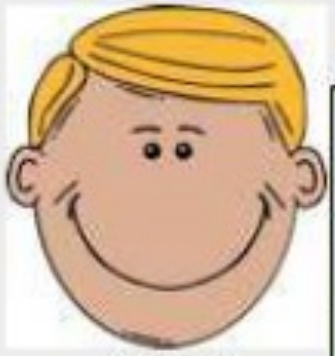
# Go Fish: Do Users Prefer an Entraining System?



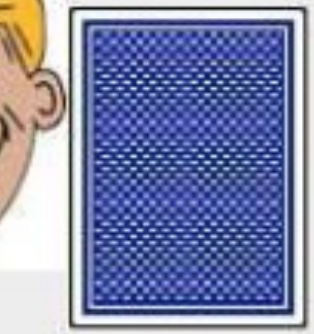
# Go Fish Helpers







Bobby



Ask for help

# Method

- 19 participants:
  - 9 female, 10 male
  - Ages 20—35
- Each session: ~45 user turns (entraining + control)
  - ~9 minutes
  - Acoustic-prosodic features extracted by Praat
  - Advice logged

# User Preferences for Entraining Helpers

- Trust
  - "Who gave better advice?" N.S.
  - Implicit trust (whose advice followed?) Entraining
- Liking
  - "Which advisor did you like better?" Entraining
- Voice
  - "Whose voice did you like better?" Entraining
  - "Strange" Non-Entraining
  - "Annoying" Non-Entraining

# Entrainment in Deceptive Speech

- CxC Corpus
  - 340 native speakers of English and Chinese, balanced by gender and native language
  - Taking turns as interviewer/interviewee with interviewer trying to detect deception
- Entrainment in lexical and acoustic/prosodic features
  - Some evidence that speakers become more similar to partner than to their own norming data in high frequency words, pitch, voice quality and intensity but..
  - *More pairs exhibited divergence over time than convergence*… not too surprising given task…
  - Interviewers guessed lying better when there was lexical entrainment between them and interviewees

# More Current and Future Research

- **Entrainment and trust:**
  - *GoFish*, *NavGame (*Harry Potter like adventure game), *GuessWho* (aka *TwentyQuestions*) games developed and tested for Slovak and Spanish
- **Entrainment in *code-switching:***
  - *Miami Bangor Corpus (Sp/Eng)* shows significant evidence of entrainment in CSW
- **Research on individual differences:**
  - Differences in gender, native language (e.g. Spanish, Slovak), culture, and personality may explain entrainment differences

# Thanks!


- This work has been supported by
  - NSF IIS-0803148
  - Humanities and Social Sciences Foundation of China Education Ministry (11YJA740113)
  - Air Force Office of Scientific Research, Air Force Material Command, USAF under Award No. FA9550-15-1-0055
  - UBACYT 20020120200025BA

# Publications


- Individual differences in Acoustic-Prosodic Entrainment in Spoken Dialogue, Andreas Weise, Sarah Ita Levitan, Julia Hirschberg, Rivka Levitan, Speech Communication , 115:78-87, December 2019.
- The Role of Cognate Words, POS Tags, and Entrainment in Code-Switching, Victor Soto, Nishmar Cestero and Julia Hirschberg, Interspeech, September 2018, Hyderabad, India.
- .Acoustic-Prosodic and Lexical Entrainment in Deceptive Dialogue, Sarah Ita Levitan, Jessica Xiang and Julia Hirschberg, Speech Prosody,Poznan, Poland, 2018.
- Prosodic entrainment and trust in human-computer interaction, Stefan Benus, Marian Trnka, Eduard Kuric, Luk Martk, Agustín Gravano, Julia Hirschberg, and Rivka Levitan, Speech Prosody, Poznan, 2018.
- Implementing acoustic-prosodic entrainment in a conversational avatar. R. Levitan, S. Benus, R. H. Galvez, A. Gravano, F. Savoretti, M. Trnka, A. Weise and J. Hirschberg. Interspeech 2016.
- Entrainment and Turn-Taking in Human-Human Dialogue. R. Levitan, S. Benus, A. Gravano and J. Hirschberg. AAAI Spring Symposium on Turn-Taking and Coordination in Human-Machine Interaction 2015.
- Acoustic-Prosodic Entrainment in Slovak, Spanish, English and Chinese: A Cross-Linguistic Comparison. R. Levitan, S. Benus, A. Gravano and J. Hirschberg. SIGDIAL 2015.
- Backward mimicry and forward influence in prosodic contour choice in Standard American English, A. Gravano, S. Benus, R. Levitan, J. Hirschberg. Interspeech 2015.

- Prosodic entrainment in Mandarin and English: a cross-linguistic comparison. Z. Xia, R. Levitan, J. Hirschberg. Speech Prosody 2014. (Best Student Paper Award)
- Three ToBI-based measures of prosodic entrainment and their correlations with speaker engagement. A. Gravano, S. Benus, R. Levitan, J. Hirschberg, SLT 2014.
- Acoustic-Prosodic Entrainment in Human-Human and Human-Computer Dialogue, R. Levitan, PhD Thesis, Columbia University, 2014.
- Entrainment in Slovak Collaborative Dialogues, S. Benus, R. Levitan, J. Hirschberg, A. Gravano, S. Darjaa. 5th IEEE International Conference on Cognitive Infocommunications, 2014.
- Entrainment, Dominance and Alliance in Supreme Court Hearings, S. Benus, A. Gravano, R. Levitan, S. I. Levitan, L. Willson, J. Hirschberg. Journal of Knowledge-Based Systems 2014.
- Acoustic-prosodic entrainment and social behavior. R. Levitan, A. Gravano, L. Willson, S. Benus, J. Hirschberg, A. Nenkova. NAACL 2012.
- Entrainment in spontaneous speech: The case of filled pauses in Supreme Court hearings. S. Benus, R. Levitan, J. Hirschberg. CogInfoCom 2012.
- Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. R. Levitan and J. Hirschberg. Interspeech 2011.
- Entrainment in speech preceding backchannels. R. Levitan, A. Gravano and J. Hirschberg. ACL 2011.
- Pitch similarity in the vicinity of backchannels. M. Heldner, J. Edlund, and J. Hirschberg. Interspeech 2010.
- High frequency word entrainment in spoken dialogue, A. Nenkova, A. Gravano and J. Hirschberg. ACL/HTL 2008.

# Other Research Interests

- Research on
  - Turn-taking signals in spoken dialogue
  - Emotion detection from speech
  - Deceptive vs. truthful, trusted vs. mistrusted speech
  - Production and perception of charismatic speech
  - Text-to-speech synthesis in Low Resource Languages and prosody prediction for TTS
  - Detecting hate speech and radicalization in videos

# Thanks!


- This work has been supported by
  - NSF IIS-0803148
  - Humanities and Social Sciences Foundation of China Education Ministry (11YJA740113)
  - UBACYT 20020120200025BA