
Thursday, July 6 | 7:00pm | Jacobs Science Building, Room 121

Fillmore Professor
Prosodic Entrainment Across Cultures



Julia Hirschberg
Columbia University



Reception to follow at Boone Center

Prosodic Entrainment Across Cultures

LSA Summer Institute

July 6, 2017

Julia Hirschberg, Columbia University

Collaborators

- *Rivka Levitan*, Brooklyn College
- *Zhihua Xia*, Tongji University
- *Agustín Gravano*, University of Buenos Aires
- *Štefan Beňuš*, Constantine the Philosopher University
- *Sarah Ita Levitan*, Columbia University
- *Nishmar Cestero*, Columbia University

The Chameleon Effect

- ***Entrainment/Alignment/Adaptation***

- *“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)

Why is Entrainment Important?

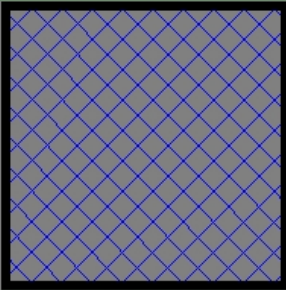

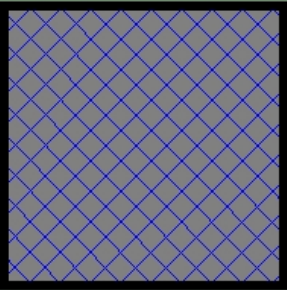
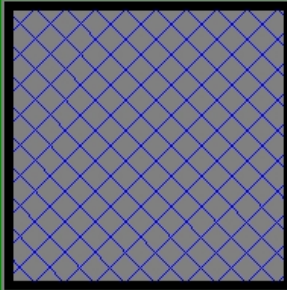



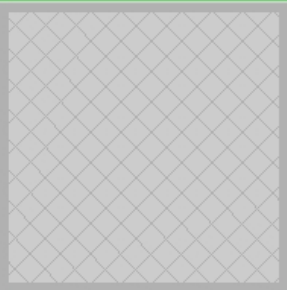
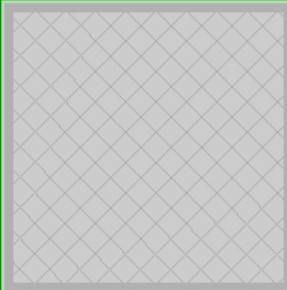
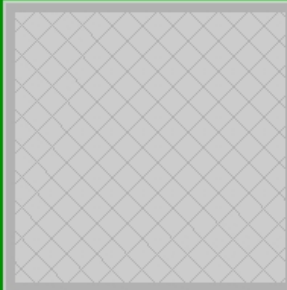
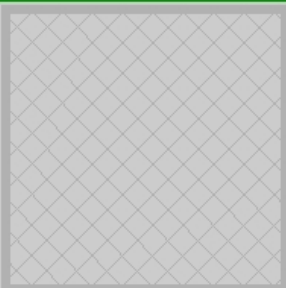
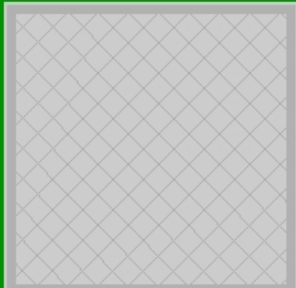

- Subjects who entrain
 - Perceived as **more socially attractive** (Putnam & Street '84, Bourhis et al '75)
 - Perceived as **more competent** (Street '84)
 - Conversation perceived as **more intimate** (Buller & Aune '88)
- Entrainment leads subjects to **like their conversational partners (and their 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)

- 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

Listen & match.

YOUR SCORE: 100
HIGH SCORE: 740

MATCH

PASS



The Objects Game



Describer:

Describe the location of the blinking image.

YOUR SCORE: 000
HIGH SCORE: 1322

DONE CONTINUE

Follower:

Listen and place the image. Then click "DONE".

YOUR SCORE: 000
HIGH SCORE: 1322

DONE CONTINUE

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

IPU

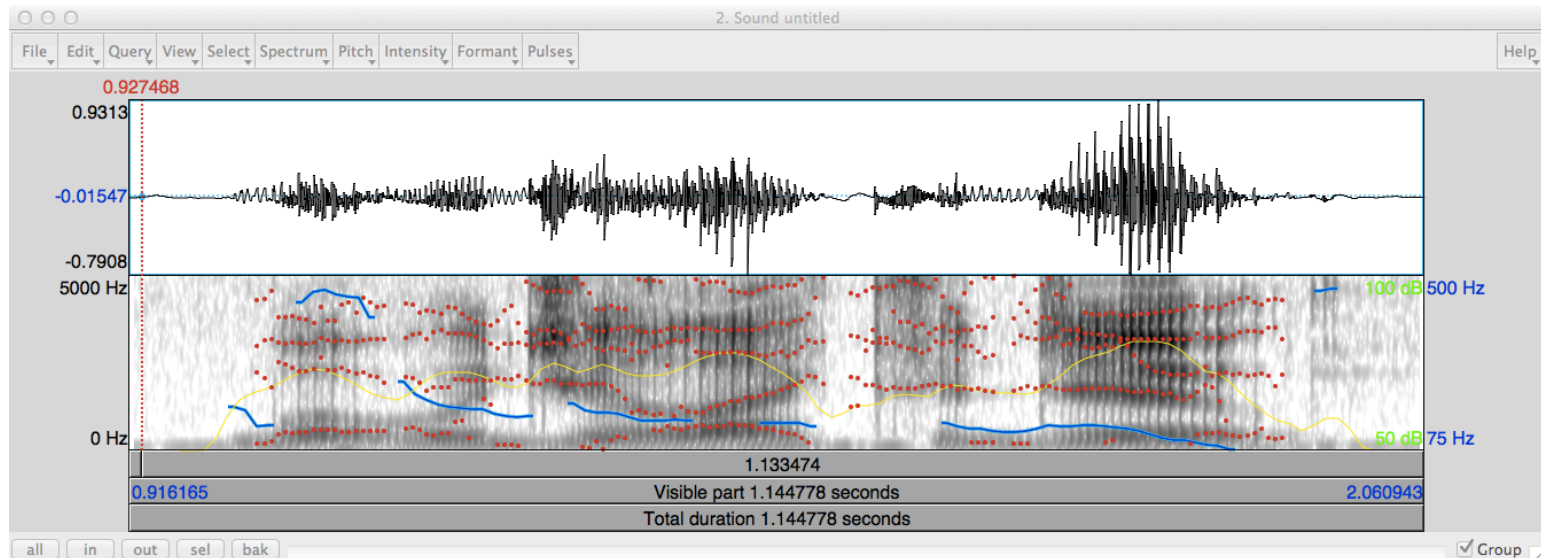
IPU

IPU

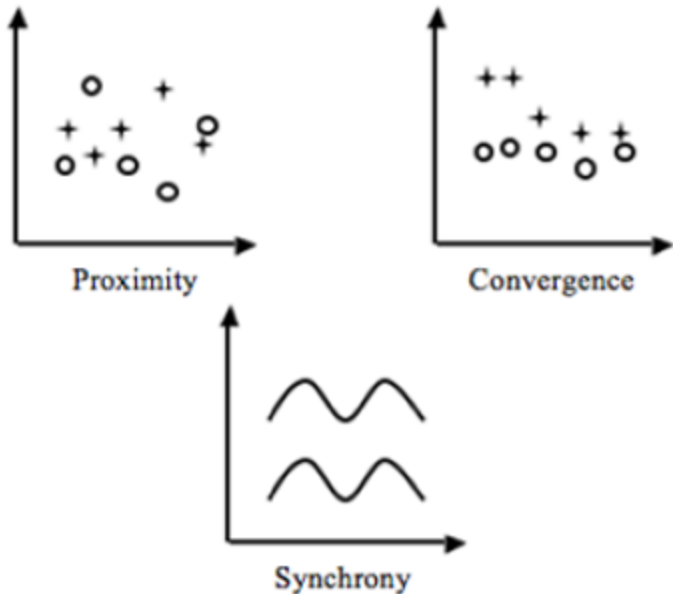
- ***Turn***: Sequence of speech from one speaker without intervening speech from the other speaker.
- ***Session***: Complete interaction between two subjects on one task

Low Level Prosodic Features

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



Forms of Entrainment (Levitan & Hirschberg '11)



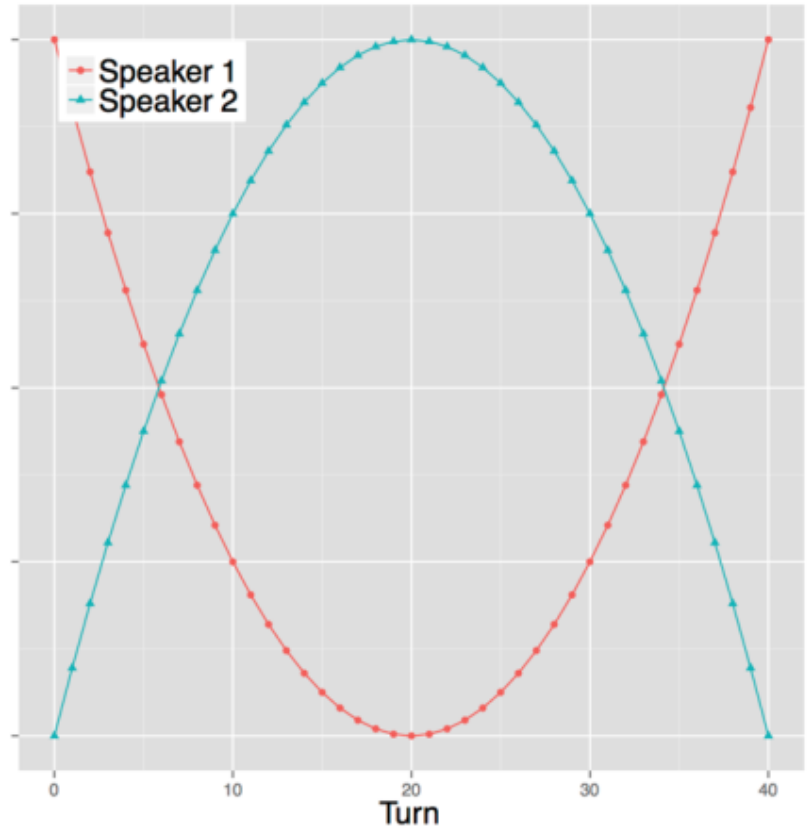
Proximity ---- significant similarity of partner features

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

Synchrony ---- correlated relative change in partner features

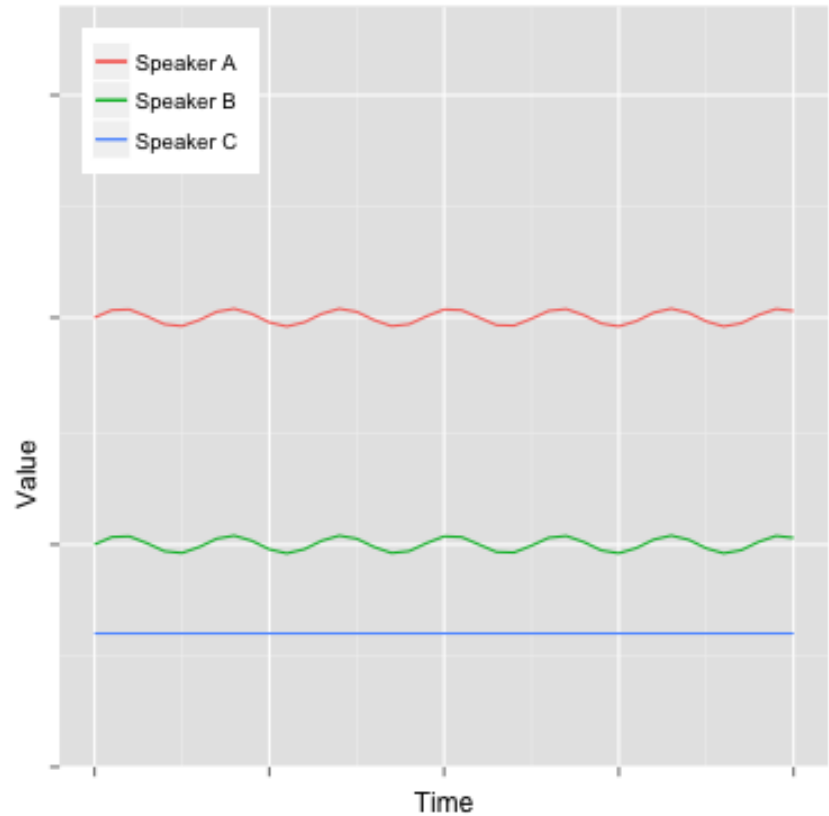
Similarity/Proximity

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



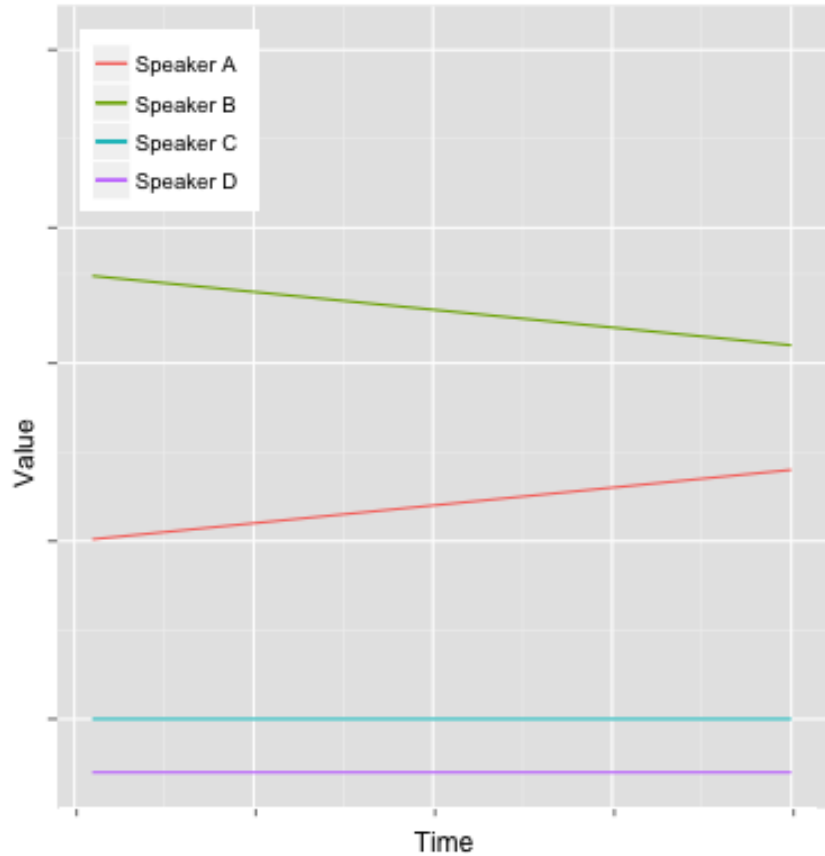
Synchrony

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



Convergence

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



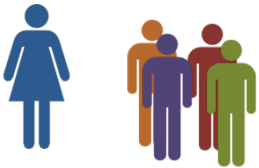
Defining Global Pairwise Entrainment

- Similarity



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

Partner similarity



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

Non-partner similarity

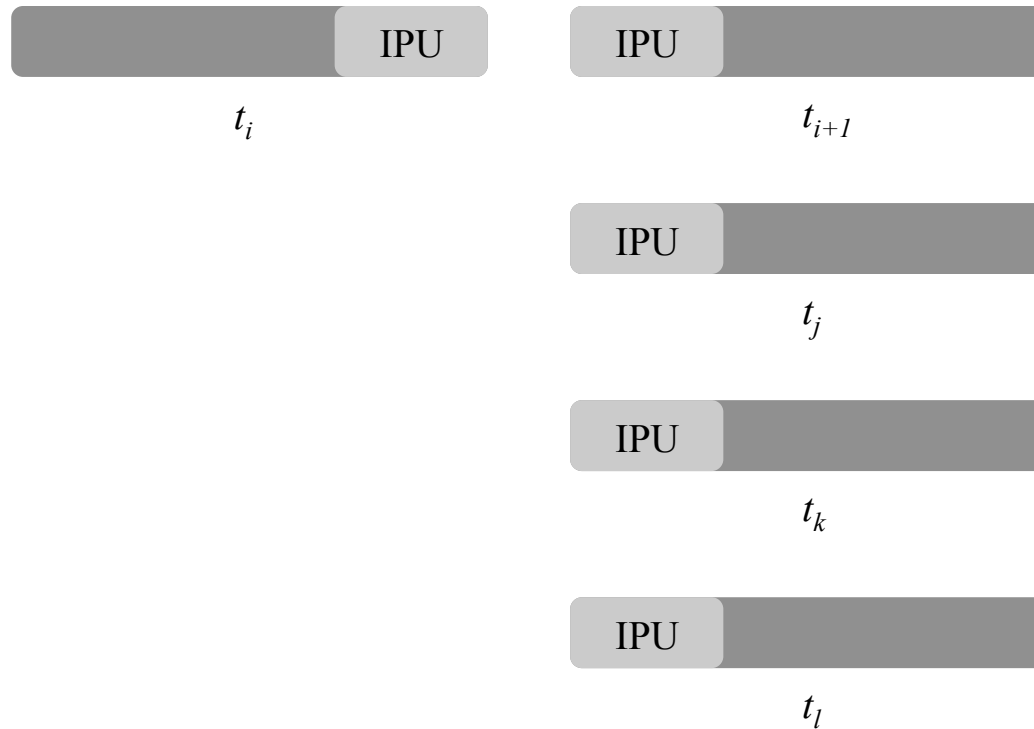


$$ENT_{self}(s, f) = -|s_f - s'_f|$$

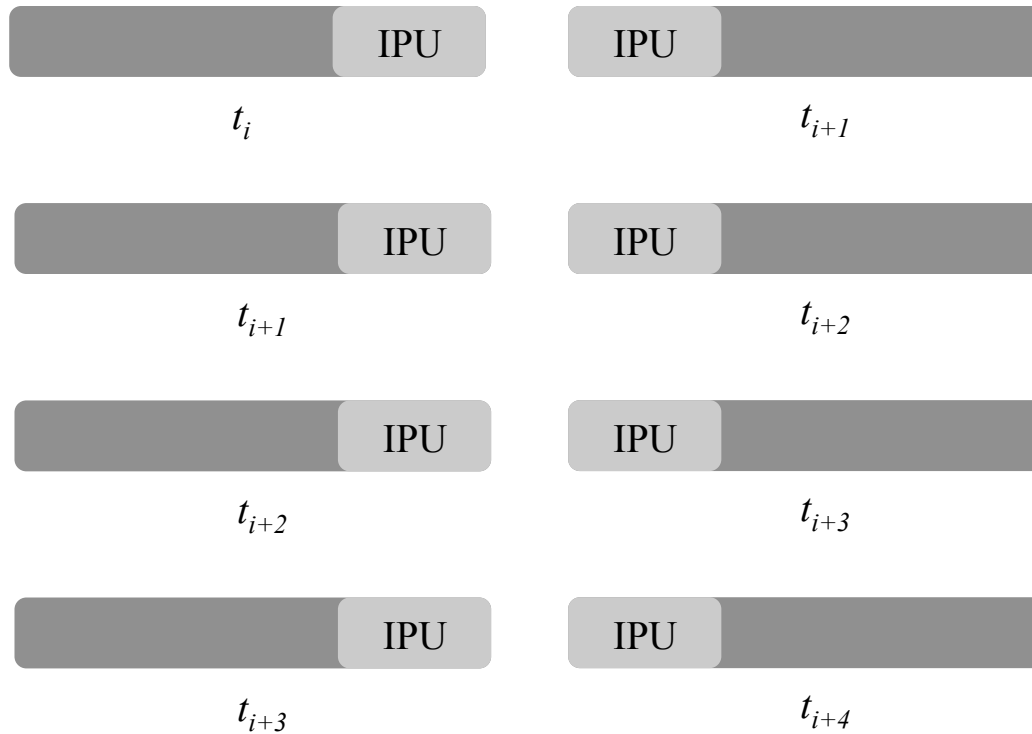
Self similarity

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

Local Entrainment: Proximity



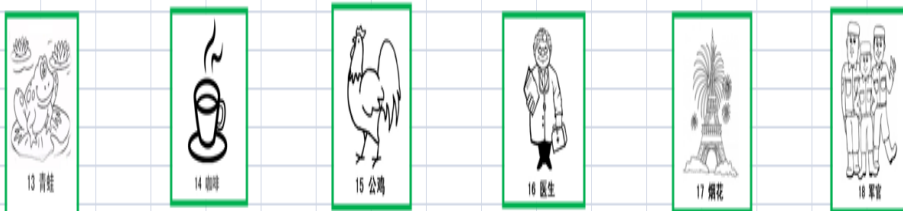
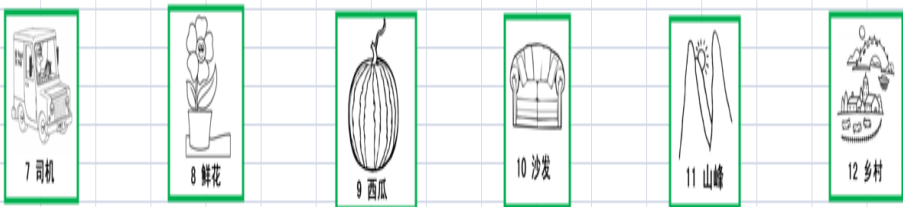
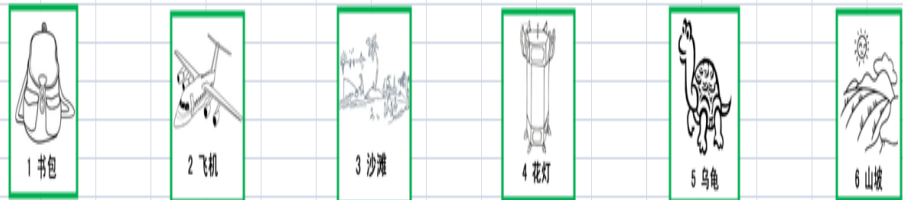
Local Entrainment: Synchrony, Convergence



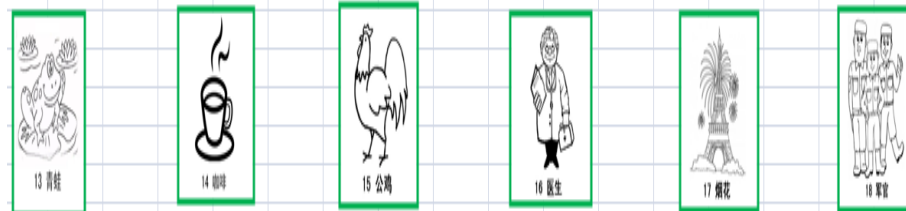
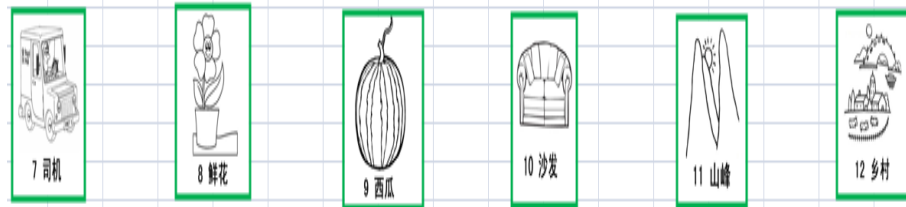
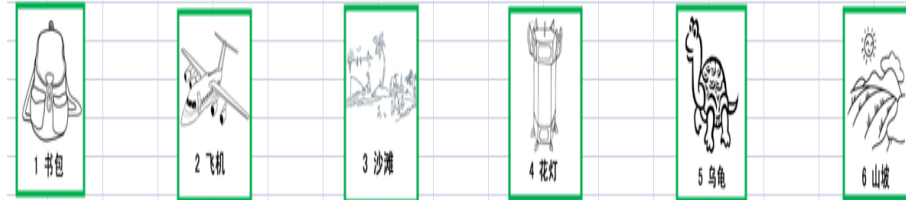
Tongji Games Corpus (Xia)

- A Chinese linguistics PhD student contacts us and asks to visit our lab
 - She has 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)

Picture Ordering Game

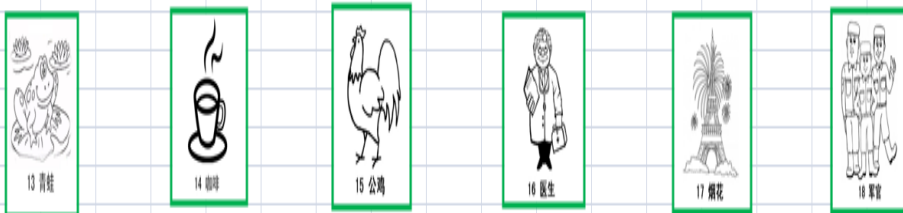
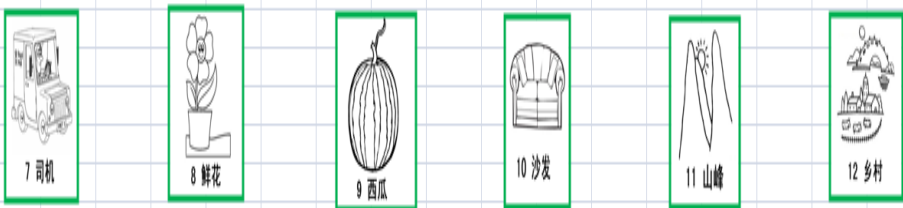
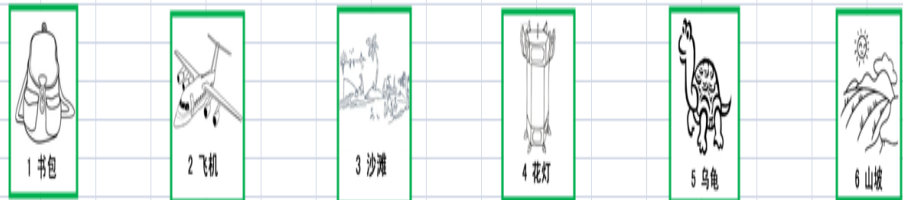


A

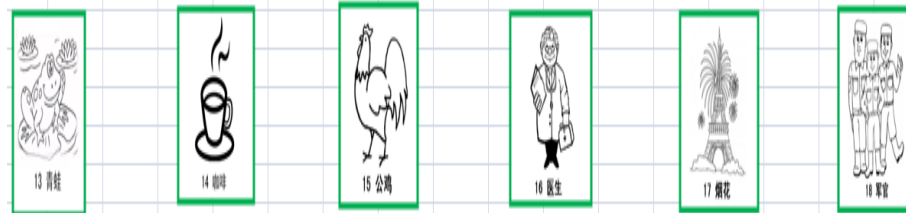
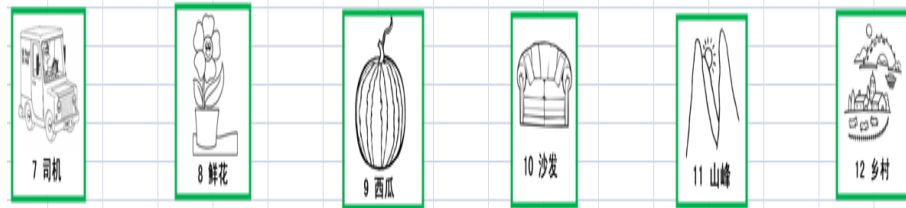
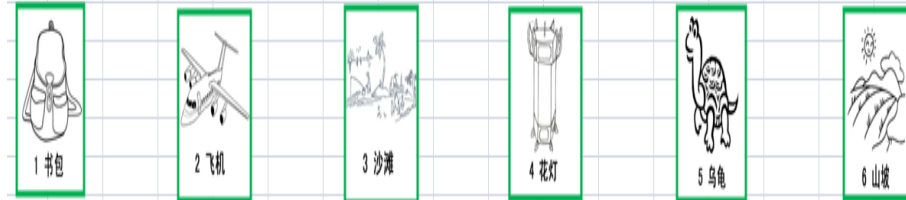


B

Picture Classifying Game



A



B

Comparing Entrainment in Different Languages and Cultures

- Employed similar speech analyses and metrics to identify entrainment in the Chinese corpus
 - Standard American English (SAE) vs. Mandarin Chinese (MC)
 - Surprising similarities on multiple metrics

Results (Levitan '14, Xia et al '14)

<i>Feature</i>	<i>Global similarity</i>		<i>Local similarity</i>		<i>Synchrony</i>		<i>Global convergence</i>		<i>Local convergence</i>	
	SAE	MC	SAE	MC	SAE	MC	SAE	MC	SAE	MC
Intensity mean	✓✓	✓	✓	✓	0.35	0.63			0.08	
Intensity max	✓(✓)	✓	✓	✓	0.33	0.55			0.08	
Pitch mean					0.07	0.66			0.08	0.22
Pitch max		✓			0.04	0.61	✓		0.10	0.24
Jitter		—		—	0.15	—		—		—
Shimmer		—		—	0.07	—		—	0.04	—
NHR		—	✓	—	0.12	—	✓	—	0.03	—
Speaking rate	✓	✓		✓			✓			0.13

Comparing Mandarin and Standard American English

- 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)

Theories of Entrainment and Gender

- Dominance
 - Male Dominance Hypothesis
 - Communication Accommodation Theory
- Perception
 - Perception-Behavior Link
 - Female perceptual sensitivity

Predictions and Previous Work

– Predictions

- In MF conversations, females should entrain more.
- There should be more entrainment in FF conversations than MM conversations.

– Previous work

- Bilous and Krauss (1988)
- Namy et al. (2002)
- Pardo (2006)

Partner vs. Non-Partner Differences

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

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

X = set of speakers of the same gender and role as speaker1's partner who are never paired with speaker1.

Entrainment and Gender

<i>Feature</i>	<i>FF</i>		<i>MM</i>		<i>FM</i>	
	<i>MC</i>	<i>SAE</i>	<i>MC</i>	<i>SAE</i>	<i>MC</i>	<i>SAE</i>
Intensity mean	✓	✓		✓	✓	✓
Intensity max	✓	✓		✓	✓	✓
Intensity min		--		--		--
F0 mean					✓	✓
F0 max					✓	✓
F0 min		--		--		--
Speaking rate	✓	✓	✓	✓	✓	✓

Entrainment and Gender

<i>Feature</i>	<i>FF</i>		<i>MM</i>		<i>FM</i>	
	<i>MC</i>	<i>SAE</i>	<i>MC</i>	<i>SAE</i>	<i>MC</i>	<i>SAE</i>
Intensity mean	✓	✓		✓	✓	✓
Intensity max	✓	✓		✓	✓	✓
Intensity min		--		--		--
F0 mean					✓	✓
F0 max					✓	✓
F0 min		--		--		--
Speaking rate	✓	✓	✓	✓	✓	✓

Entrainment and Gender

<i>Feature</i>	<i>FF</i>		<i>MM</i>		<i>FM</i>	
	<i>MC</i>	<i>SAE</i>	<i>MC</i>	<i>SAE</i>	<i>MC</i>	<i>SAE</i>
Intensity mean	✓	✓		✓	✓	✓
Intensity max	✓	✓		✓	✓	✓
Intensity min		--		--		--
F0 mean					✓	✓
F0 max					✓	✓
F0 min		--		--		--
Speaking rate	✓	✓	✓	✓	✓	✓

Entrainment and Gender

<i>Feature</i>	<i>FF</i>		<i>MM</i>		<i>FM</i>	
	<i>MC</i>	<i>SAE</i>	<i>MC</i>	<i>SAE</i>	<i>MC</i>	<i>SAE</i>
Intensity mean	✓	✓		✓	✓	✓
Intensity max	✓	✓		✓	✓	✓
Intensity min		--		--		--
F0 mean					✓	✓
F0 max					✓	✓
F0 min		--		--		--
Speaking rate	✓	✓	✓	✓	✓	✓

Conclusions

- $MM \ll FF \ll MF$
- Similar patterns for Mandarin Chinese and SAE

Social Dimensions of Entrainment (Levitan et al '12)

- 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
- 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

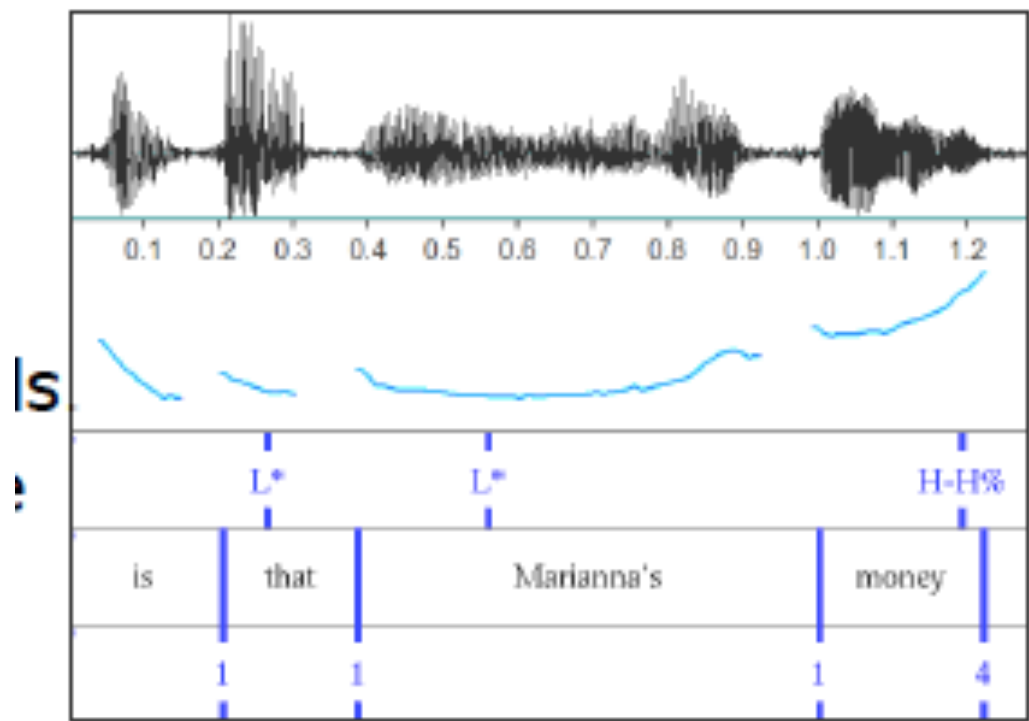
- **Communication Accommodation Theory**
 - **Giving encouragement** positively correlated with entrainment
 - Conversational **awkwardness** negatively
- **Similarity-Attraction Theory**
 - **Trying to be liked** should be positively correlated
- **Dependency Over-Accommodation** occurs when an interlocutor appears dependent on a speaker, giving the impression that the **speaker is controlling** the conversation (West & Turner, 2009).

Findings

- Based on Communication Accommodation Theory
 - **Giving encouragement** positively correlated with entrainment
 - **Conversational awkwardness** (weak positive correlation)
- Based on Similarity-Attraction Theory
 - **Trying to be liked** positively correlated
 - **No correlation** between **perceived dominance** and entrainment

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.



Entrainment on Pitch Contours and Social Variables (Gravano et al '14, '15)

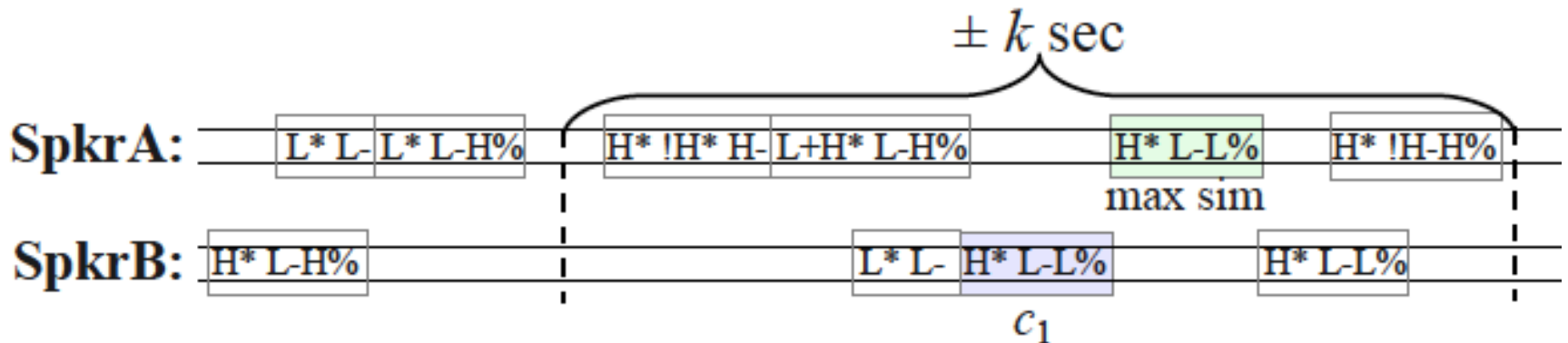
- Measures of contour similarity between speakers: when I use contour X are you more likely to do the same?
 - **Perplexity of language models** of sequences trained on Speaker A and used to model prosodic sequences of Speaker B: *low perplexity indicates greater similarity*
 - **Levenshtein distance** 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 are Speaker B's?

ϵ_1 Measure: N-Gram Perplexity

- How well does a prosodic model trained on A predict B's prosody?
 - SpkrA: L* L-H% H* H-L% ...
 - SpkrB: H* !H* H- L* H-H% ...
 - TRAIN TRIGRAM
 - MODEL TEST $\epsilon_1(A, B) = -(\text{Perplexity of A's model on B's productions})$
 - *Lower perplexity means more similar*

ϵ_2 Measure: Levenshtein Distance

How far away from B's production of contour c is A's production of the same contour?



$L \leftarrow$ new list

for each contour c_1 from B :

$C \leftarrow$ contours from A at most k seconds before/after c_1

append $\left(\max_{c_2 \in C} \text{sim}(c_1, c_2) \right)$ to L

$\mathcal{E}_2(A, B) \leftarrow \text{mean}(L)$

ϵ_3 Measure: Kullback-Leibler Divergence

$$D_{KL}(P \parallel Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)}$$

Assymmetric measure of the difference between two probability distributions P and Q .

$$\epsilon_3(A, B) = -D_{KL}(\text{contours}(B), \text{contours}(A))$$

Experiments

- Build a 24D vector with the value of ϵ_i for each member of each speaker pair
- Build similar vector for each social variable v (e.g., bored-with-game) where A_j, B_j are the two speakers from the same session j
- Run Pearson's correlation tests between entrainment vectors and social variable vectors

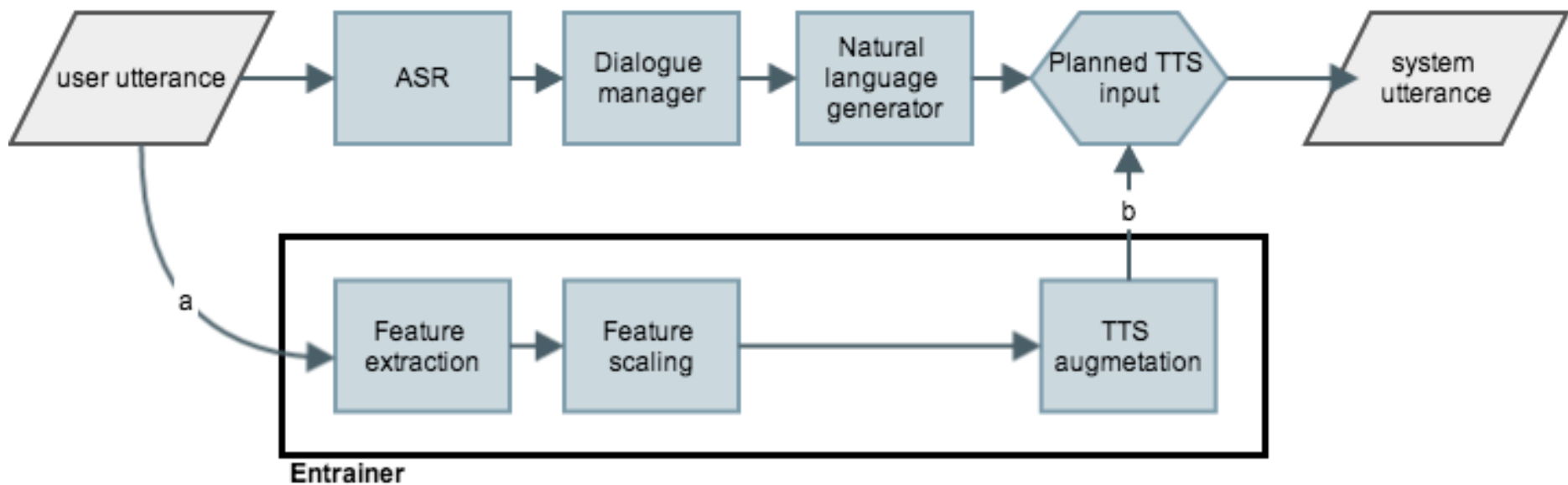
Correlations for Different Similarity Metrics with Social Variables

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

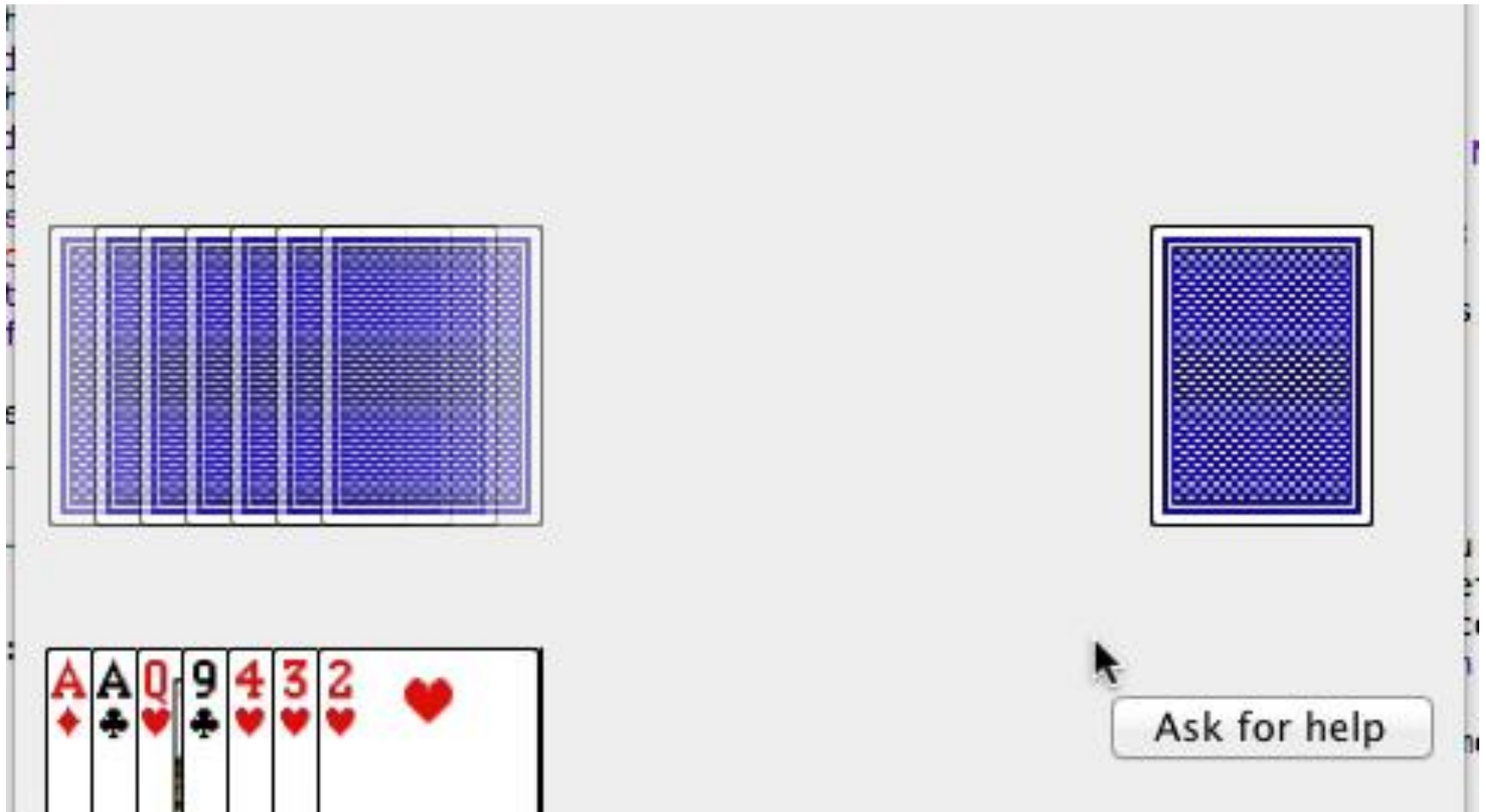
Conclusions and Future Research

- 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).

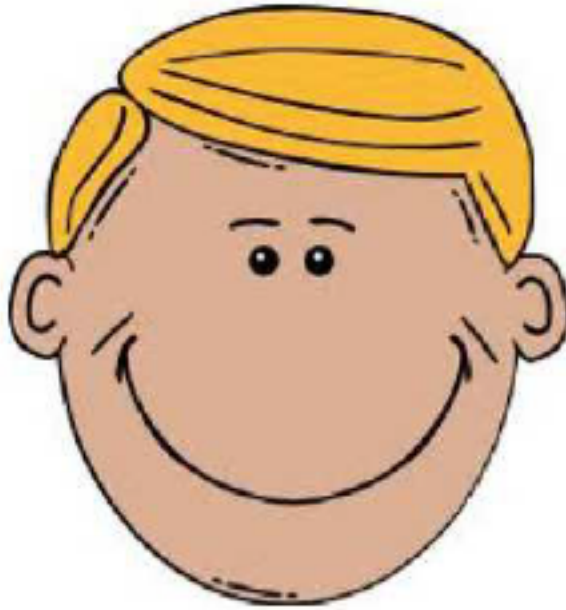
Entraining on Rate and Intensity to Users in Spoken Dialogue Systems (Levitan et al '16)



Go Fish: Do Users Prefer an Entraining System?



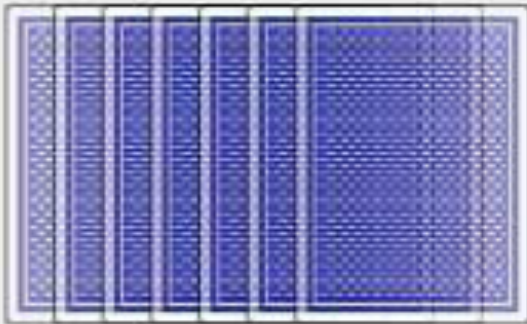
Go Fish Helpers



Bobby



Freddy



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 Porteño Spanish and Slovak GoFish-with-Helpers Games

- Differences from English:
 - System entrained only on speech rate – not intensity
 - Different TTS systems, avatar gender (female), varied pitch range
 - No effect for entraining avatar
- Why?
 - Issues with identifying speech rate accurately from ASR...

Current and Future Research

- **Entrainment and trust:**
 - *GoFish*, *NavGame* (Harry Potter like adventure game), *GuessWho* (aka *TwentyQuestions*) games being developed and tested for Slovak and Spanish
 - New research on entrainment and trust in tech
- **Entrainment in deceptive speech** (CxC corpus): significant evidence that speakers more similar to partner than to themselves: pitch, intensity, VQ, high-frequency words
- **Entrainment in *code-switching*** (*Miami Bangor Corpus*): significant evidence of entrainment

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

- 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, J. Hirschberg. Interspeech 2011.
- Entrainment in speech preceding backchannels. R. Levitan, A. Gravano, 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, J. Hirschberg. ACL/HTL 2008.