### Goals

**STOCHASTIC TURN-TAKING MODELS** predict a speaker’s future speech activity using that speaker’s past speech activity and their interlocutor’s past speech activity.

### Findings

1. Although there is “no data like more data”, stochastic turn-taking models are surprisingly speaker-specific.
2. Variability between person-specific models is correlated with person identity.
3. This variability is large enough to permit use of these models for:
   - speaker attribution: better than a task-specific baseline by 45%rel on unseen data
   - speaker detection: better than a task-specific baseline by 36%rel on unseen data

Note that task-specific baselines also relied on the unconditional probability of speaking, which these stochastic turn-taking models do not.

### Conclusions

1. People generally exhibit an interactional-timing self-similarity, even in longitudinal collections.
2. People generally exhibit an interactional-timing dissimilarity to other people.
3. Stochastic turn-taking models are able to quantify these similarities and dissimilarities.
4. Stochastic turn-taking models quantify aspects which appear to be independent of the unconditional probability of speaking, or talkativeness.

### Potential Impact

I. Stochastic turn-taking models appear to measure aspects of conversational social personality, based on a very simple representation of conversations.
II. They permit future study of behavior across different accents, languages, and topics, with no additional technical overhead.
III. They may be useful in the design of spoken dialogue system agents which exhibit a consistent personality.
IV. They may be suitable for speaker recognition applications in degraded-signal and privacy-sensitive environments.

### Two Diagnostic Tasks

- **Task 1: Speaker Attribution**
  - know the K identities of the K persons in a conversation
  - do not know which identity goes with each speech stream
  - random guessing: 80% classification error
- **Task 2: Speaker Detection**
  - do not know the identities of the K persons in a conversation
  - must draw from a larger population of N > K persons
  - random guessing: 93% classification error
- **Data**
  - 67 meetings from the ICSI Meeting Corpus
  - split into TrainSet, DevSet, and (unseen) TestSet
  - 14 persons occur frequently; all others pooled into Unk

### Task-Specific Baseline

- **Features** (per participant)
  - measure the unconditional probability of speaking
  - measure the conditional probability of speaking
  - conditioning context consists of one 100-ms frame back in time
  - conditioning context models a conversant’s behavior with respect to a specific other conversant’s behavior
- **Models**
  - one-Gaussian models of probability-of-speaking
  - Group Behavior model
  - product of individual hypothesized conversants
  - probability-of-speaking
  - Group Membership model
  - penalize hypothesized conversants which appear infrequently in the data
- **Search**
  - exhaustive search of K1 group assignments for Speaker Attribution
  - greedy search for Speaker Detection

### Evaluation of STT Models

- **Given a K-row chronogram representing the behavior of K conversants:**
  - form K chronograms with two rows
  - row one: specific to conversant k
  - row two: logical-OR of interlocutors of conversant k
- **Train an N-gram model for each of N identities in TrainSet**
  - over ten 100-ms frames from the modeled speaker and ten 100-ms frames from the interlocutor row → a 21-gram
- **For a test conversation, apply each of N models to each of K test conversants**
- **Greedy search over N × K matrix**

### Numerical Results

<table>
<thead>
<tr>
<th></th>
<th>BL</th>
<th>EXH</th>
<th>GZ</th>
<th>GRZ</th>
<th>GRZEXH</th>
<th>GRZT</th>
<th>BL</th>
<th>GZ</th>
<th>GRZ</th>
<th>GRZEXH</th>
<th>GRZT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Attribution</strong></td>
<td>52.9</td>
<td>41.5</td>
<td>46.1</td>
<td>50.0</td>
<td>48.0</td>
<td>49.2</td>
<td>53.8</td>
<td>48.1</td>
<td>49.2</td>
<td>53.8</td>
<td>48.1</td>
</tr>
<tr>
<td><strong>Detection</strong></td>
<td>52.9</td>
<td>41.5</td>
<td>46.1</td>
<td>50.0</td>
<td>48.0</td>
<td>49.2</td>
<td>53.8</td>
<td>48.1</td>
<td>49.2</td>
<td>53.8</td>
<td>48.1</td>
</tr>
</tbody>
</table>

*“BL”: baseline (2008)
“GRZ”/”GRZT”: final results of current work