1. **Goals**

   - Advance the state-of-the-art in stochastic turn-taking modeling:
     - history extension with model regularization
     - model re-estimation and/or adaptation
   - Enable quantitative social research into interactive conversational phenomena

2. **Conclusions**

   - With respect to turn-taking, people are:
     1. generally dissimilar
     2. but self-consistent
     3. and similar to their partner
   - Turn-taking is not 1st-order Markovian.

3. **Impact**

   - Natural turn timing now available to analytic and conversational agents.
   - Can now synthesize emergent behavior by composing dissimilar models.
   - Can now investigate how prosody may circumscribe chronogram cross entropy.

---

### Minimizing Model Cross-Entropy for Speech/Non-Speech Chronograms

1. A stochastic turn-taking model $\Theta$ is a model which accounts for the distribution of speech $\Box$ and non-speech $\Box$ in time and across both participants.

   $$P\left(\cdots \Box \Box \Box \Box \Box \Box \Box \Box \Box \Box \cdots\right) = \cdots \times P\left(\Box \Box \Box \Box \Box \Box \Box \Box \Box \Box \cdots\right) \times P\left(\Box \Box \Box \Box \Box \Box \Box \Box \Box \Box \cdots\right) \times P\left(\Box \Box \Box \Box \Box \Box \Box \Box \Box \Box \cdots\right) \times \cdots$$

2. Each factor can be further factored, by assuming that the $\Box/\Box$ behavior of both participants is independent:

   $$P\left(\Box \Box \Box \Box \Box \Box \Box \Box \Box \Box \cdots\right) = P\left(\Box \Box \Box \Box \Box \Box \Box \Box \Box \Box \cdots\right) \times P\left(\Box \Box \Box \Box \Box \Box \Box \Box \Box \Box \cdots\right)$$

   (conditional independence (CI))

   or

   $$P\left(\Box \Box \Box \Box \Box \Box \Box \Box \Box \Box \cdots\right) = \left(P\left(\Box \Box \Box \Box \Box \Box \Box \Box \Box \Box \cdots\right) \times P\left(\Box \Box \Box \Box \Box \Box \Box \Box \Box \Box \cdots\right)\right)$$

   (unconditional independence (UI))

3. Train $n$-gram models with recursive Jelinek-Mercer interpolation.

4. Score using normalized negative log-likelihood as conditional cross entropy.

---

### Questions and Implications

#### Question 1: Is one dialogue participant affected by the other?

If so, then for fixed history duration, conditioning on the other side should help: $\Theta_C$ should be better than $\Theta_U$.

#### Question 2: Is there turn-taking variation within the population?

If so, then it may be better to estimate $\Theta_C$ parameters using tiny but matched data than large but mismatched data.

#### Question 3: Are turn-taking systematics time-dependent?

If so, then incremental training of $\Theta_C$ should be accompanied by "forgetting" the least-recent past.

#### Question 4: Is one dialogue participant similar to the other?

If so, then it may be better to estimate $\Theta_C$ parameters using tiny data from the other participant than large but mismatched data.

---

Supported by the Riksbankens Jubileumsfond (RJ) project Samtalets Prosodi.

kornel@cs.cmu.edu, edlund@speech.kth.se, mattias@speech.kth.se

ISCA INTERSPEECH 2011, Firenze, Italia