INCREMENTAL LEARNING AND FORGETTING IN STOCHASTIC TURN-TAKING MODELS

Kornel Laskowski, Jens Edlund & Mattias Heldner

KTH Speech, Music and Hearing, Stockholm, Sweden

Goals

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

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.

Impact

1. **Natural** turn timing now available to analytic and conversational agents.
2. Can now synthesize **emergent** behavior by composing dissimilar models.
3. 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 Θ** is a model which accounts for the distribution of speech ■ and non-speech □ in time and across both participants.

$$\boldsymbol{P}\left(\cdots\ \begin{array}{cccccc} & t-3 & t-2 & t-1 & t & t+1 \\ A & \square & \square & \blacksquare & \blacksquare & \square \\ B & \blacksquare & \blacksquare & \blacksquare & \square & \square \end{array}\ \cdots\right) = \cdots \times \boldsymbol{P}\left(\begin{array}{cc} & t-3 \\ A & \square \\ B & \blacksquare \end{array} \middle| \cdots \right) \times \boldsymbol{P}\left(\begin{array}{cc} & t-2 \\ A & \square \\ B & \blacksquare \end{array} \middle| \cdots \begin{array}{cc} t-3 & \\ \square & A \\ \blacksquare & B \end{array}\right) \times \boldsymbol{P}\left(\begin{array}{cc} & t-1 \\ A & \blacksquare \\ B & \blacksquare \end{array} \middle| \cdots \begin{array}{ccc} t-3 & t-2 & \\ \square & \square & A \\ \blacksquare & \blacksquare & B \end{array}\right) \times \boldsymbol{P}\left(\begin{array}{cc} & t \\ A & \blacksquare \\ B & \square \end{array} \middle| \cdots \begin{array}{cccc} t-3 & t-2 & t-1 & \\ \square & \square & \blacksquare & A \\ \blacksquare & \blacksquare & \blacksquare & B \end{array}\right) \times \cdots$$

2. Each factor can be further factored, by assuming that the ■/□ behavior of both participants is **independent**:

$$\boldsymbol{P}\left(\begin{array}{cc} & t \\ A & \blacksquare \\ B & \square \end{array} \middle| \cdots \begin{array}{cccc} t-3 & t-2 & t-1 & \\ \square & \square & \blacksquare & A \\ \blacksquare & \blacksquare & \blacksquare & B \end{array}\right) \doteq \boldsymbol{P}\left(\begin{array}{cc} & t \\ A & \blacksquare \end{array} \middle| \cdots \begin{array}{cccc} t-3 & t-2 & t-1 & \\ \square & \square & \blacksquare & A \\ \blacksquare & \blacksquare & \blacksquare & B \end{array}\right) \times \boldsymbol{P}\left(\begin{array}{cc} & t \\ B & \square \end{array} \middle| \cdots \begin{array}{cccc} t-3 & t-2 & t-1 & \\ \blacksquare & \blacksquare & \blacksquare & B \\ \square & \square & \blacksquare & A \end{array}\right)$$

conditional independence (CI)

or

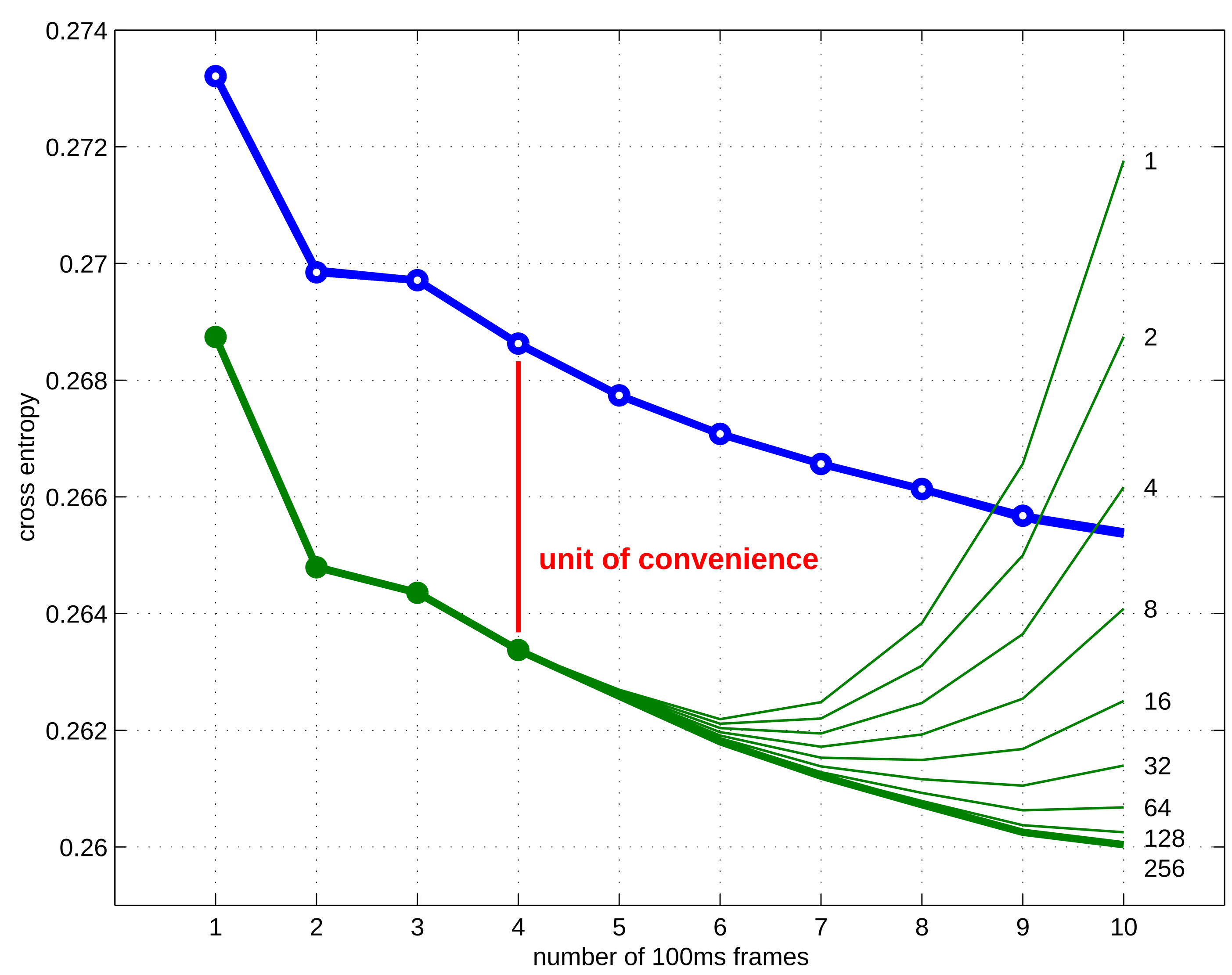
$$\doteq \underbrace{\boldsymbol{P}\left(\begin{array}{cc} & t \\ A & \blacksquare \end{array} \middle| \cdots \begin{array}{cccc} t-3 & t-2 & t-1 & \\ \square & \square & \blacksquare & A \end{array}\right)}_{\text{from A's point of view}} \times \underbrace{\boldsymbol{P}\left(\begin{array}{cc} & t \\ B & \square \end{array} \middle| \cdots \begin{array}{cccc} t-3 & t-2 & t-1 & \\ \blacksquare & \blacksquare & \blacksquare & B \end{array}\right)}_{\text{from B's point of view}}$$

unconditional independence (UI)

3. Train ***n***-gram models with recursive Jelinek-Mercer interpolation.
4. Score using **normalized negative log-likelihood** ≡ **conditional cross entropy**.

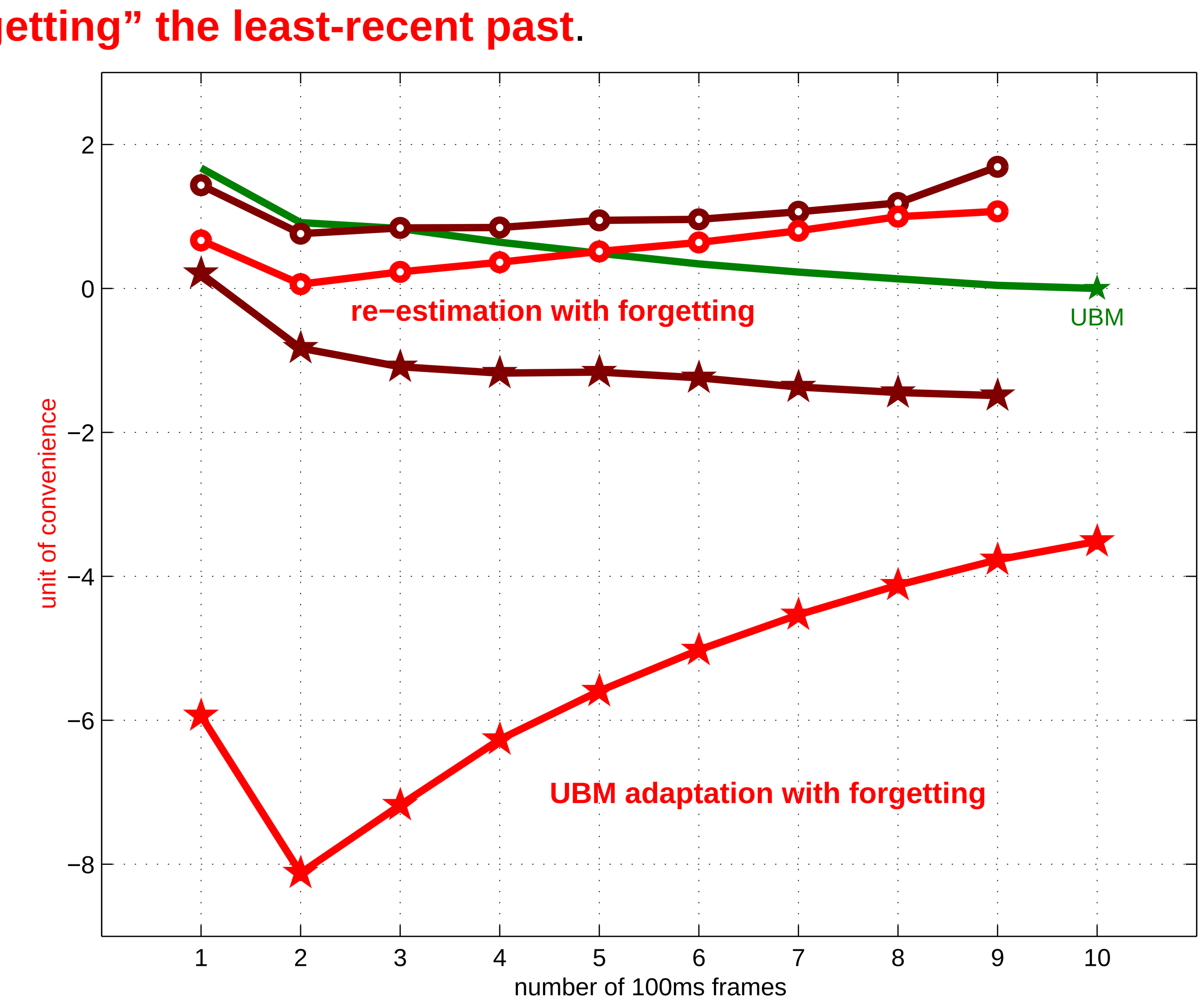
Question 1: Is one dialogue participant affected by the other?

If so, then for fixed history duration, conditioning on the other side should help: Θ_{CI} should be better than Θ_{UI}.



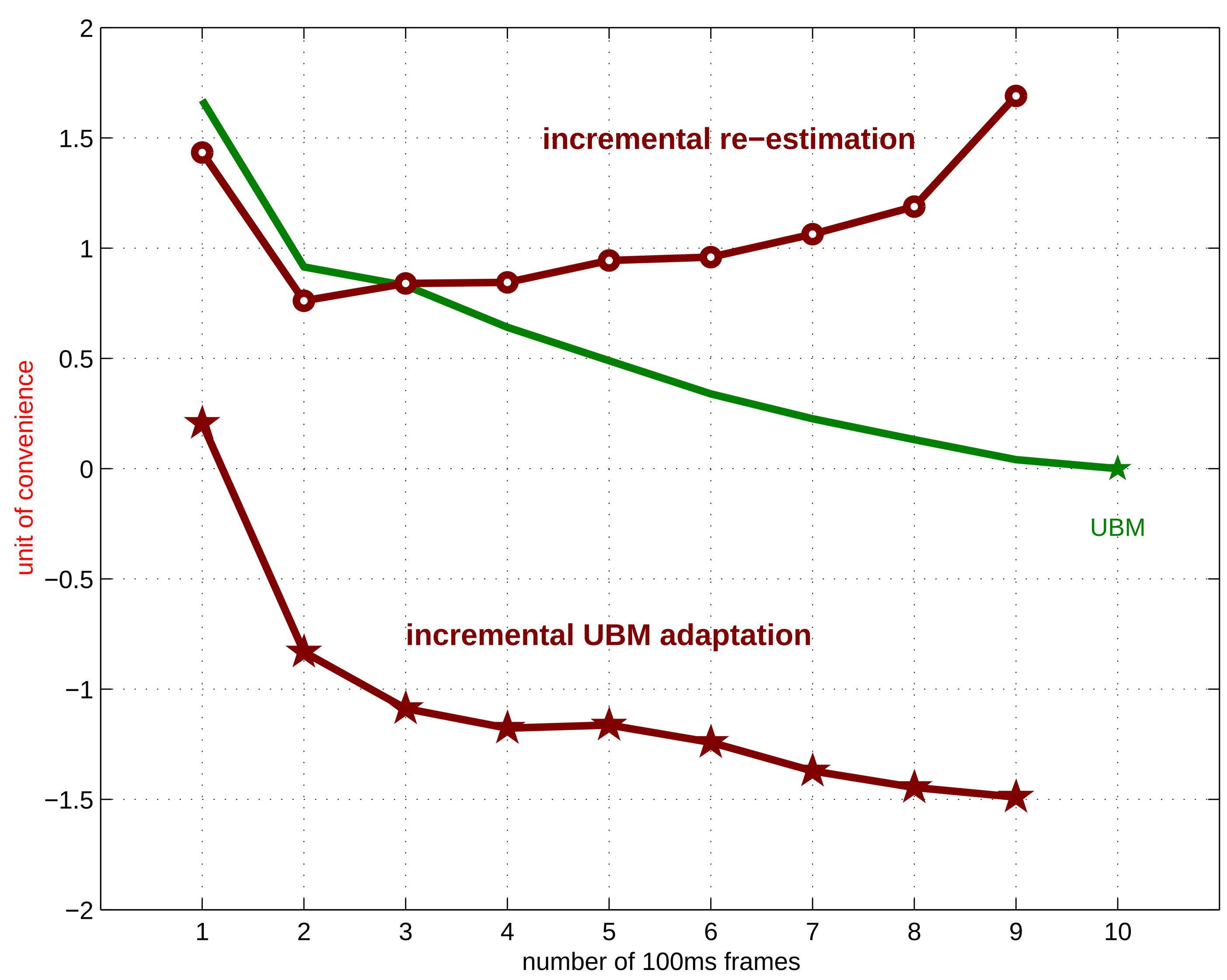
Question 3: Are turn-taking systematics time-dependent?

If so, then incremental training of Θ_{CI} should be accompanied by **"forgetting" the least-recent past**.



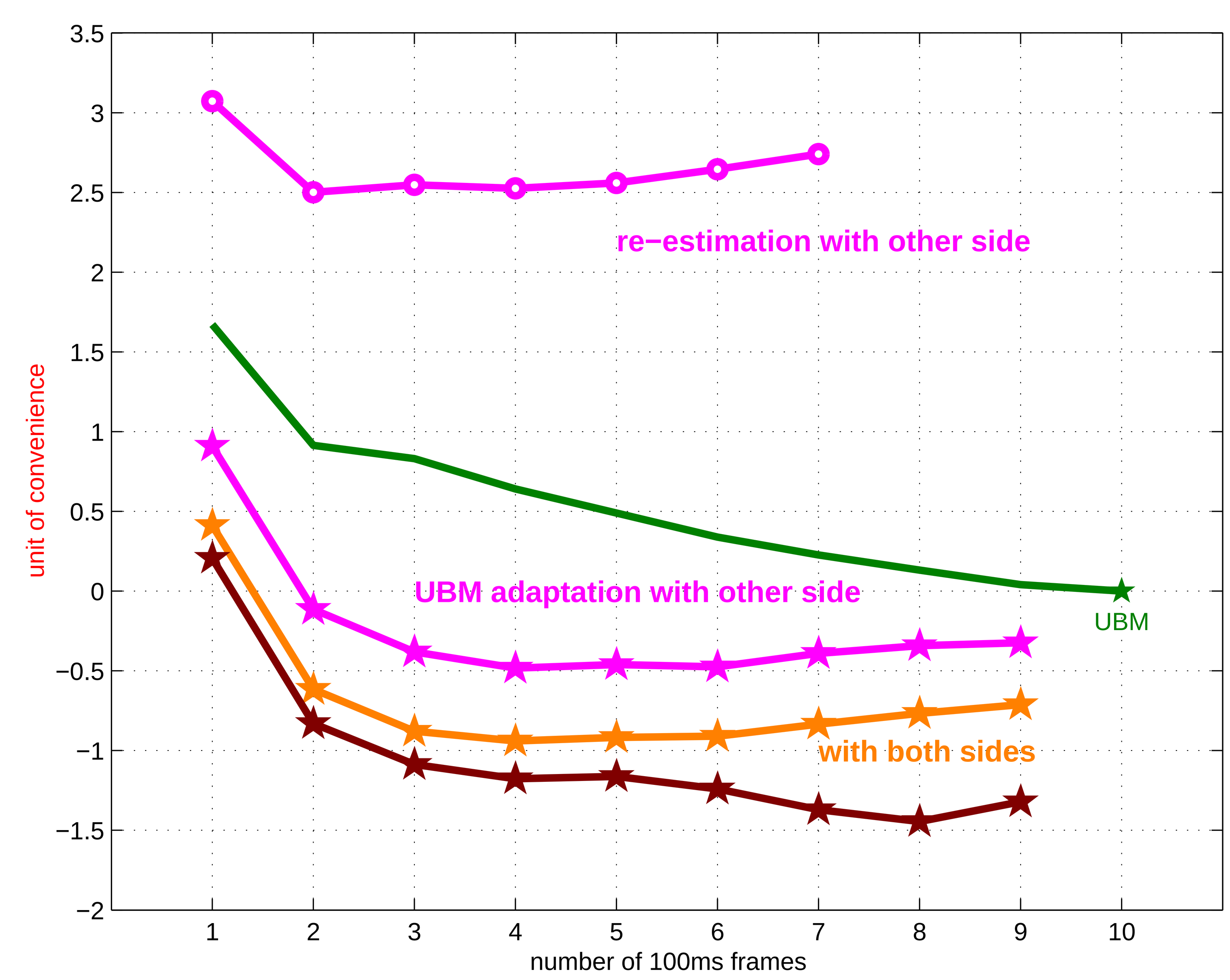
Question 2: Is there turn-taking variation within the population?

If so, then it may be better to estimate Θ_{CI} parameters using **tiny but matched data** than **large but mismatched data**.



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

If so, then it may be better to estimate Θ_{CI} parameters using **tiny data from the other participant** than **large but mismatched data**.



ISCA INTERSPEECH 2011, Firenze, Italia

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

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