A SINGLE-PORT NON-PARAMETRIC MODEL OF TURN-TAKING IN MULTI-PARTY CONVERSATION

Kornel Laskowski, Jens Edlund & Mattias Heldner
KTH Speech, Music and Hearing, Stockholm, Sweden

Goals

Develop a **conditionally independent (aka "single-port")** stochastic turn-taking model Θ for K-party conversation.

Θ provides the likelihood of a conversation's **speech/non-speech chronogram, Q**:

$$\mathbf{Q} = \left[\begin{array}{c} \square\square\square\square\square\blacksquare\blacksquare\square\blacksquare\blacksquare \\ \blacksquare\square\blacksquare\blacksquare\blacksquare\square\square\square\square\square \\ \cdots\ \square\square\square\square\square\square\square\square\square\square\ \cdots \\ \square\square\square\blacksquare\blacksquare\square\square\square\square \\ \square\square\square\square\square\square\square\square\square\square \end{array} \right]$$

Prior Model Limitations

Markov models of **1st order** which are in existence fall into two broad categories:

- Conditionally **independent**, for **two-party** conversation.
- Conditionally **dependent**, for K**-party** conversation.

As a result:

1. **Cannot** easily **predict/evaluate individual behavior** in the context of a given group.
2. **Do not know** how **less-recent history** conditions the future.
3. **Do not know** how history conditions the **more-distant future**.

Conclusions

1. The proposed conditionally independent SPEDO model is **at least as good** a K-party chronogram density estimator as the existing conditionally dependent MPEDO model.
2. The 1st-order form of the model is more compact than its MPEDO counterpart.
3. Overt evidence that participants in K-party conversation **influence one anothers' decisions** to speak.
4. Longer conditioning histories help: this **refutes 2-party findings in the literature**.
5. Distant futures are predictable, but only poorly.

Impact

I. Can now easily infer a model for any individual from the many conversations which that individual attended.

II. Can now easily compare the behavior of individuals in disparate conversational settings.

III. Can now easily synthesize anticipated emergent group behavior from models of individuals who never previously formed a group.

IV. Can now begin to evaluate the role of prosody in prediciting future individual behavior, with zero manual annotation of prosody.

"Extended-Degree-of-Overlap" (EDO) Models

1. Factor in time, Markovian assumption of order $m = 1$:

$$P\left(\cdots \right) = \cdots \times P\left(\begin{array}{c|c} \square & \square \\ \blacksquare & \blacksquare \\ \square & \square \\ \blacksquare & \square \\ \square & \square \end{array}\right) \times P\left(\begin{array}{c|c} \square & \square \\ \square & \blacksquare \\ \square & \square \\ \blacksquare & \square \\ \square & \square \end{array}\right) \times P\left(\begin{array}{c|c} \blacksquare & \square \\ \square & \square \\ \square & \square \\ \square & \blacksquare \\ \square & \square \end{array}\right) \times \cdots$$

in the multi-port (MPEDO) model

$$\doteq \cdots \times P\left(\begin{array}{c|c} 2 & 1 \\ 1 & \end{array}\right) \times P\left(\begin{array}{c|c} 1 & 2 \\ 1 & \end{array}\right) \times P\left(\begin{array}{c|c} 1 & 1 \\ 0 & \end{array}\right) \times \cdots$$

2. In the conditionally independent single-port (SPEDO) model, factor each factor further:

$$P\left(\begin{array}{c|c} \square & \square \\ \square & \blacksquare \\ \square & \square \\ \blacksquare & \square \\ \square & \square \end{array}\right) = P_1\left(\begin{array}{c|c} 1\ \square & \square \\ 2 & \blacksquare \\ 3 & \square \\ 4 & \blacksquare \\ 5 & \square \end{array}\right) \times P_2\left(\begin{array}{c|c} 1 & \square \\ 2\ \square & \blacksquare \\ 3 & \square \\ 4 & \blacksquare \\ 5 & \square \end{array}\right) \times P_3\left(\begin{array}{c|c} 1 & \square \\ 2 & \blacksquare \\ 3\ \square & \square \\ 4 & \blacksquare \\ 5 & \square \end{array}\right) \times P_4\left(\begin{array}{c|c} 1 & \square \\ 2 & \blacksquare \\ 3 & \square \\ 4\ \blacksquare & \square \\ 5 & \square \end{array}\right) \times P_5\left(\begin{array}{c|c} 1 & \square \\ 2 & \blacksquare \\ 3 & \square \\ 4 & \blacksquare \\ 5\ \square & \square \end{array}\right)$$

assume index independence

$$\doteq P\left(\begin{array}{c|c} 1\ \square & \square \\ 2 & \blacksquare \\ 3 & \square \\ 4 & \blacksquare \\ 5 & \square \end{array}\right) \times P\left(\begin{array}{c|c} 2\ \square & \blacksquare \\ 1 & \square \\ 3 & \square \\ 4 & \blacksquare \\ 5 & \square \end{array}\right) \times P\left(\begin{array}{c|c} 3\ \square & \square \\ 1 & \square \\ 2 & \blacksquare \\ 4 & \blacksquare \\ 5 & \square \end{array}\right) \times P\left(\begin{array}{c|c} 4\ \blacksquare & \blacksquare \\ 1 & \square \\ 2 & \blacksquare \\ 3 & \square \\ 5 & \square \end{array}\right) \times P\left(\begin{array}{c|c} 5\ \square & \square \\ 1 & \square \\ 2 & \blacksquare \\ 3 & \square \\ 4 & \blacksquare \end{array}\right)$$

***count** speaking interlocutors*

$$\doteq P\left(\begin{array}{c|c} 1\ \square & \square \\ oth & \mathbf{2} \end{array}\right) \times P\left(\begin{array}{c|c} 2\ \square & \blacksquare \\ oth & \mathbf{1} \end{array}\right) \times P\left(\begin{array}{c|c} 3\ \square & \square \\ oth & \mathbf{2} \end{array}\right) \times P\left(\begin{array}{c|c} 4\ \blacksquare & \blacksquare \\ oth & \mathbf{1} \end{array}\right) \times P\left(\begin{array}{c|c} 5\ \square & \square \\ oth & \mathbf{2} \end{array}\right)$$

3. The integer in the conditioning context is the *number (degree) of interlocutors speaking simultaneously* at instant $t - 1$. To reduce model size: cluster this number into K_{max} bins, with the K_{max}th bin containing all degrees $\geq K_{max}$.

Evaluation : Perplexity as a Function of K_{max}

Using the ICSI Meeting Corpus:

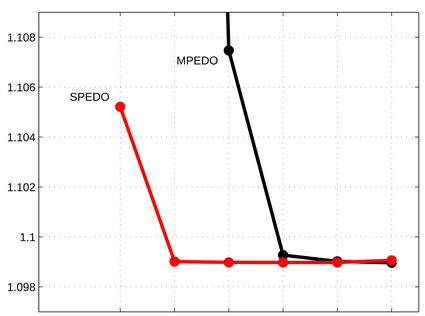
- 75 meetings, each $\approx$ **60** minutes long
- 3–9 participants per meeting
- round-robin: for each meeting, train on all others
- score with duration- and K- normalized perplexity

Model complexity controlled using:

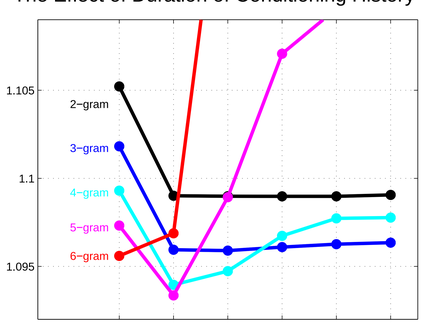
- n, the order of the Markov assumption
- K_{max}, the maximum allowed "degree-of-overlap"

  1: ignore interlocutors altogether
  2: distinguish among 0 or >**0** other talkers
  3: distinguish among 0, 1 or >**1** other talkers

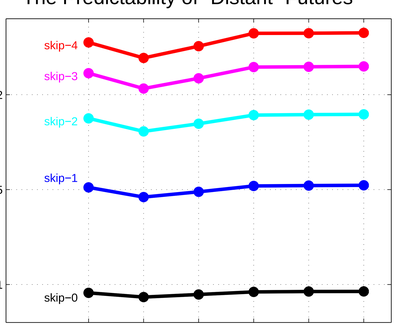
Single-Port (SPEDO) versus Multi-Port (MPEDO)



The Effect of Duration of Conditioning History



The Predictability of "Distant" Futures



Supported by the Riksbankens Jubileumsfond (RJ) project *Samtalets Prosodi*.
ktl@kth.se, edlund@speech.kth.se, mattias@speech.kth.se