Nature’s Inventive Mishmash of 
Turntaking 
Mechanisms & Stuff

Kristinn R. Thórisson, Ph.D. 
Associate Professor, Reykjavik University 
Co-director, Center for Analysis & Design of Intelligent Agents
Definitions

• “Turntaking”
  – Observable phenomenon
  – People take turns speaking in any sustained dialog
    • as measured at particular timescales & modes

• “Mechanism”
  – A causal chain of some significance
    • Recursive definition
Overview

• Introduction
• YTTM - Ymir Turntaking Model
  – Contexts, contextual alignment, coupling
  – Evaluation, extension: Other models
• Conclusions
F2F Dialog: Biggest Constraint

• Cognitive apparatus
  – Made for incremental interpretation

• Cognitive constraints
  – Interference between usage of key for production and understanding
    • E.g. working memory
Turntaking: What it is Good For

• Nature’s “workaround”
• Ensures alignment of content
  – By ensuring that understanding can progress incrementally
    • Without interference from e.g. planning and presentation processes
      – Taking advantage of various information-carrying systems like face, intonation, tone of voice
YTTM

• Generative model
  – From perception to action (to fully generate, you have to include both)

• Runnable
• Multimodal
• Realtime
Separation of dialog behaviors into:

- Envelope processes
  - Those which control the process of communication

- Content processes
  - Those responsible for the topic of discussion
YTTM: Participants

- Modeled as separate processes with independent perception-action loops
YTMM: Participants

• Modeled as separate processes with independent perception-action loops
  – Multiple loops at different timescales

\[ t_3 \geq t_2 \geq t_1 \]
YTMM: Participant

- Cognition = \{P, D, C_u, C_g, B, G, P\}
  - Set of perceptual feature processes, P
  - Set of decision-making processes, D
  - Content understanding mechanism, C_u
  - Content generation mechanism, C_g
  - Behavioral displays, B
  - Plans with goals, P, G
- P = \{p_1 \ldots p_n\}
- D = \{d_1 \ldots d_n\}
- B = \{b_1 \ldots b_n\}
- G = \{g_1 \ldots g_n\}
- P = \{p_1 \ldots p_n\}
YTTM: Contexts

• Indirect control of anticipation and prediction
  – A form of attentional control
• Guide perception, planning, action realization
YTTM: Contexts

I-have-turn
I-give-turn
I-want-turn
I-accept-turn
I-hold-turn

O-wants-turn
O-gives-turn
O-accepts-turn
O-has-turn
YTTM: Contexts

- I-have-turn
- O-has-turn
- I-give-turn
- O-gives-turn
- O-wants-turn
- I-accept-turn
- O-accepts-turn
- I-want-turn
- O-has-turn
- O-accepts-turn
- I-accept-turn
Contexts are:

• The basis for the coupling of communicating individuals
  – When the contexts of participants are aligned, envelope behaviors are synchronized
    • and content can be exchanged

• Cognitive processes related to content presentation and content interpretation can run efficiently
  – In all participants
Coupling

I-have-turn
I-give-turn
I-want-turn
I-accept-turn
I-hold-turn

O-has-turn
O-wants-turn
O-gives-turn
O-accepts-turn
Coupling

I-have-turn
I-accept-turn
I-want-turn
I-give-turn
I-hold-turn

O-accepts-turn
O-wants-turn
O-gives-turn
O-has-turn

Participant 1
Participant 2
Coupling

Participant 1

Participant 2

P1

P2
Coupling

I-give-turn

P1

O-gives-turn

P2
Coupling

Coupled contexts:

Associated with each context are perceptions and behaviors that the agents have learned over the years.
Coupling

\[ p^{P2}(b_e \subset B^{P2}) \]

\( b_e \) is behavior P2 expects from P1 that P2 has learned to be useful for alignment with P2 in the O-gives-turn context.
Coupling

$p^{P1}(b_e \subset B^{P1})$

Symmetrically

$p^{P2}(b_e \subset B^{P2})$

$b_e$ is behavior P2 expects from P1 that P2 has learned to be useful for alignment with P2 in the O-gives-turn context.
Coupling

\[ p^{P1}(b_e \subseteq B^{P1}) \]

\[ b_i \subseteq B^{P1} \quad p^{P2}(b_e \subseteq B^{P2}) \]

\( b_i \) has been learned by \( P1 \) to be effective to align the current contexts
Coupling

\[ p_{P1}^{P1}(b_e \subset B^{P1}) \]

I-give-turn

P1

\[ b_i \subset B^{P1} \]

P2

\[ p_{P2}^{P2}(b_e, b_i) \]

O-gives-turn
Coupling

\[ p^{P1} (b_e \subset B^{P1}) \]

P2’s decision mechanisms decide that the output of \( p_{P2} \) provides sufficient evidence that \( b_e \) and \( b_i \) match.
Coupling

\[ p^{P1}(b_e \subset B^{P1}) \]

P2 decides that to be aligned with P1, the best context is should be I-accept-turn.
Coupling

$p^P_1 \left(b_e \subset B^P_1\right)$
Coupling

$p^{P1} (b_e \subset B^{P1})$

Associated with I-accept-turn are cues $b_e$ that P2 expects from P1
Indicating that P2 has turn
As a new context becomes active, certain behaviors $b_j$ are exhibited.
Coupling

\[ d_{\Theta}^{P1} (p^{P1} (b_e, b_j)) \quad b_j \subset B^{P2} \]

I-give-turn

I-accept-turn

P1

P2

\[ p^{P2} (b_e \subset B^{P2}) \]
Coupling

\[ d_{\Theta}^{P1} (p^{P1} (b_e, b_j)) \]

\[ b_j \subset B^{P2} \]

\[ p^{P2} (b_e \subset B^{P2}) \]
Coupling

$p^{p1} (b_e \subset B^{p1})$

O-accepts-turn

P1

I-accept-turn

$p^{p2} (b_e \subset B^{p2})$

P2
The agents have aligned their new contexts again
Coupling

$p^{P1}(b_i \subset B^{P1})$

$P1$

$O$-accepts-turn

$P2$

$I$-accept-turn

$b_i \subset B$

$p^{P2}(b_e \subset B^{P2})$
Coupling

$p^{P1}(b_e \subset B^{P1})$

$D_{\Theta}^{P2}(p^{p2}(b_e, b_i))$
Coupling

$p^{P_1}(b_e \subset B^{P_1})$

$\Theta$

$d_x^{P_2}$

O-accepts-turn

P1

I-have-turn

I-accept-turn

$P_2$

$d_{\Theta}^{P_2}(p_2^{P_2}(b_e, b_i))$
Coupling

$p^{P_1}(b_e \subset B^{P_1})$

O-accepts-turn

P1

I-have-turn

P2
Coupling

\[ p^{P1}(b_e \subset B^{P1}) \quad \text{and} \quad b_j \subset B^{P2} \]

P1

O-accepts-turn

P2

I-have-turn
Coupling

\[ d_{\Theta}^{P1}(p^{P1}(b_e, b_j)) \]
Coupling

\[ d_\Theta^{P1} (p^{P1} (b_e, b_j)) \]
Coupling

P1

O-has-turn

I-have-turn

P2
Coupling

P1

O-has-turn

P2

I-have-turn
Coupling in Turntaking

• Happens as a
  – continuous alignment of predicted contexts
  – using predicted behavior displays
  – at multiple levels of detail and
  – multiple timescales

• Coupling ensures
  – Synchronized perceptual and planning mechanisms
YTTM Resulting Hypotheses

- Content production and interpretation can be separated from turntaking control
  - via a simple set of primitives
    - Topic-Knowledge-System-Received-Speech-Data
    - Speech-Data-Available-For-Analysis
    - Topic-Knowledge-System-Parsing-Speech-Data
    - Topic-Knowledge-System-Successful-Parse
    - Content-Layer-Action-Available
    - I-Have-Reply-Ready
    - Topic-Knowledge-System-Real-World-Action-Available
    - Im-Executing-Topic-Speech-Task
    - Im-Executing-Topic-Realworld-Task
    - Im-Executing-Topic-Multimodal-Act
    - Im-Executing-Topic-Communicative-Act
    - Im-Executing-Communicative-Act
YTTRM Resulting Hypotheses

• Features perceived during dialog are logically combined to determine appropriate behaviors at any point in time
• Decisions about multimodal behaviors are based on boolean combinations of perceptual data
Summary

• YTTM
  – accounts for many macro phenomena in realtime multimodal dialog
  – explains coupling between participants in multimodal dialog, as observed in turntaking
  – ready to be merged with related theories

• Theories from different levels of detail
  – Constrain and extend each other
  – Producing better theories