Prior Lessons of Incremental Dialogue and Robot Action Management for the Age of Language Models

DIALOGUE & DISCOURSE

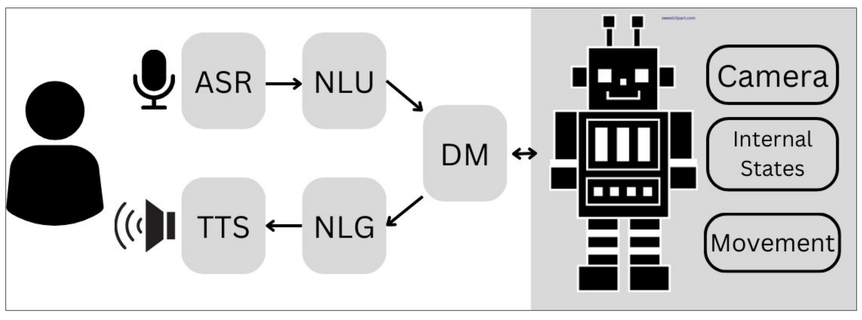
Casey Kennington
Boise State University

Pierre Lison
Norwegian Computing Center

David Schlangen
University of Potsdam

Introduction

- Efforts towards endowing robots with the ability to speak have benefited from recent advances in NLP, in particular large language models (LLMs).
- Current LMs are not fully **incremental**, as their processing is inherently **monotonic** and thus lack the ability to revise their interpretations or output in light of newer observations. Monotonicity has important implications for the development of dialogue systems for human–robot interaction.
- We r**eview the literature** on interactive systems that operate incrementally (i.e., at the word level or below it).
- We motivate the need for incremental systems, survey **incremental modeling** of key dialogue components
- Primary focus is on the part of the system that makes decisions, known as the **dialogue manager**.
- We find that there is very little research on incremental dialogue management, offer some **requirements for practical incremental dialogue management**, and implications of **incremental dialogue for embodied, robotic platforms** in the age of LLMs.



Dialogue Management on Robotic Platforms

The Dialogue Manager (or robot action manager) must determine not just *what* action to take, but *when* to take it, making timing critical for natural HRI

Requirements for Robotic, Multimodal DM:

- **Timing over Content:** must prioritize *when* to act, not only what to say.
- **Action on Incomplete Information:** must respond to incomplete information (e.g., early physical backchannels).
- **Fine-Grained, Concurrent Decisions:** Must support continuous, parallel updates beyond strict turn-taking.
- **Incremental LLM Efficiency:** Must adapt LLMs (which typically lack concurrent processing) to handle real-time, robotic updates.

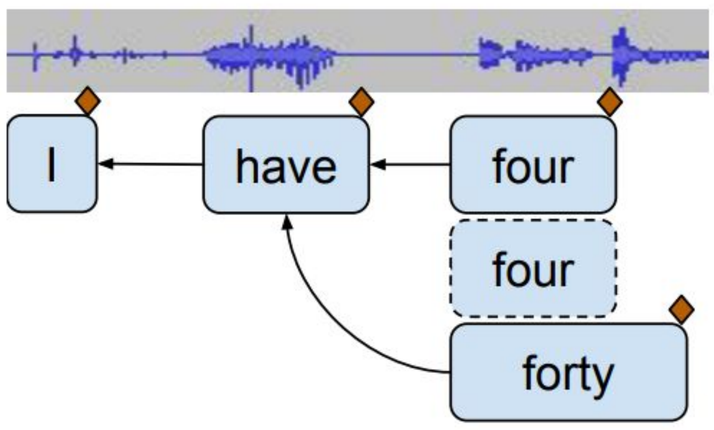
Some key challenges:

- **Managing Uncertainty:** Coping with both *sensory uncertainty* (noisy robot sensor data) and *conversational uncertainty* (inherent unpredictability of human dialogue).
- **Evaluation:** Both online execution and offline evaluation of a dialogue manager on physical robotic hardware remain incredibly difficult to standardize.
- **Non-monotonic processing:** systems must revise, update, or undo interpretations as new input arrives, rather than committing to a fixed output.
- **Beyond endpointing:** systems must act continuously on unfolding input, instead of waiting for pauses or utterance completion.

Incremental Processing in Spoken Dialogue Systems

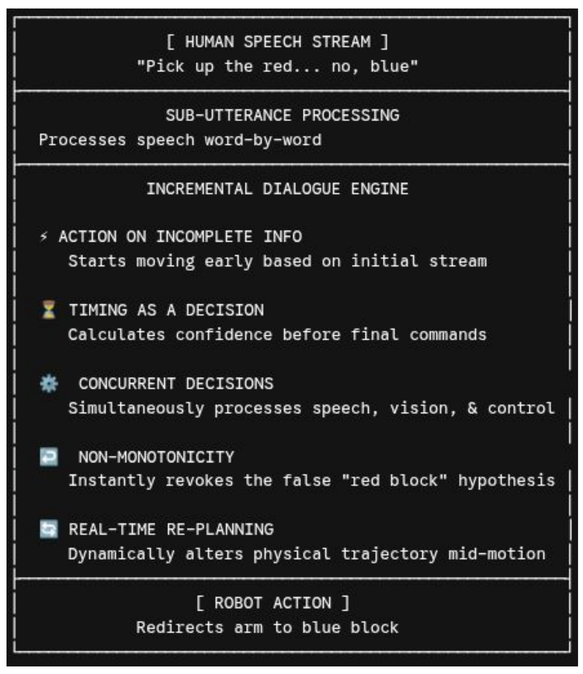
Unlike traditional systems that operate on complete utterances, incremental spoken dialogue systems (SDSs) process speech as it unfolds, at the word or sub-word level.

- **IU Framework (Schlangen & Skantze 2011):** represents all intermediate data as *Incremental Units* (IUs) in a shared network, enabling continuous updates and revisions.
- **Structural Links in the IU Network:** organizes units via *same-level links* (capturing temporal order within a module) and *grounded-in links* (dependencies on upstream data).

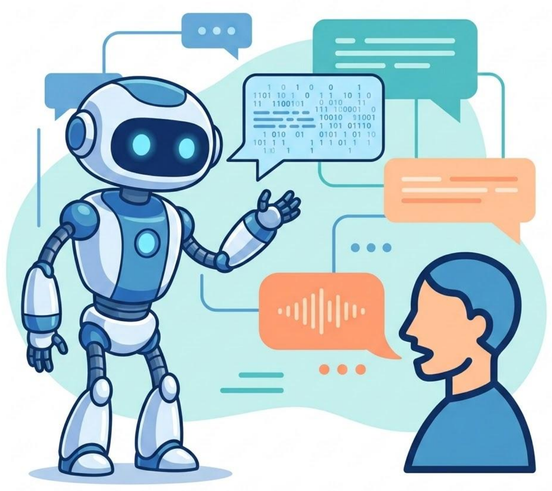


Incremental Dialogue Management

- **Timing as a decision variable**: systems must actively decide when to act, not just what to do.
- **Sub-utterance processing**: operate *word-by-word* to avoid delays from endpointing.
- **Non-monotonicity**: support mid-utterance revision as new information arrives.
- **Action on partial input**: enable early, incremental behaviors before user completion.
- **Continuous decisions**: make many small, *parallel updates* instead of one final decision.
- **Real-time re-planning**: continuously adapt dialogue and actions to user input and context.



Desiderata for Incremental DM on Robotic Platforms



To bridge the gap between language processing and physical execution, a robotic dialogue manager must support tightly coupled, real-time interaction between understanding and action:

1. Timing as a Choice (When, Not Just What). The system must actively decide when to act, balancing latency (fast responses) with confidence (waiting for more input).

2. Action on Incomplete Information. It must support early physical and conversational behaviors (e.g., gaze shifts, motion initiation) before the human interlocutor has completely finished speaking.

3. Fast, Concurrent, Fine-Grained Decisions. The system must continuously make many small, parallel updates to beliefs, speech planning, and motor control, rather than relying on single turn-level decisions.

4. Incremental LLMs (Transformer Efficiency). LLm architectures must enable efficient left-to-right updates, moving beyond sentence-level processing to handle real-time, sub-utterance interaction.

References. • **Timo Baumann, Casey Kennington, Julian Hough,** and **David Schlangen.** Recognising conversational speech: What an incremental ASR should do for a dialogue system and how to get there. In Proceedings of the International Workshop Series on Spoken Dialogue Systems Technology (IWSDS) 2016. • **Jens Edlund, Joakim Gustafson, Mattias Heldner,** and **Anna Hjalmarsson.** Towards human-like spoken dialogue systems. Speech Communication, 50(8):630–645, 2008 • **Julian Hough and David Schlangen.** Investigating fluidity for human-robot interaction with realtime, real-world grounding strategies. In Proceedings SigDial 2016.
• **Pierre Lison and Casey Kennington.** Who's in charge? roles and responsibilities of decision-making components in conversational robots. In Proceedings of the Workshop on Human-Robot Conversational Interaction, March 2023.
• **David Schlangen and Gabriel Skantze.** A general, abstract model of incremental dialogue processing. Dialogue & Discourse, 2(1):83–111, May 2011.