DEMONSTRATING THE INCREMENTAL INTERACTION MANAGER IN AN END-TO-END "LETS GO!" DIALOGUE SYSTEM

E.O. Selfridge, P.A. Heeman

Center for Spoken Language Understanding
Oregon Health & Science University
Portland, OR, USA

I. Arizmendi

AT&T Labs - Research
Florham Park, NJ, USA

J.D. Williams[*]

Microsoft Research
Redmond, WA, USA

ABSTRACT

The Incremental Interaction Manager supports the use of incremental speech recognition with conventional turn-based dialogue managers. We demonstrate its use in a complete statistical dialogue system in the "Lets Go" bus information domain. Our demonstration is two-fold. First, the user is able to call the system and have a complete dialogue. They are then able to inspect their own call using a visualizer or load logs of previous calls for a fuller picture of the system's capabilities.

Index Terms— turn-taking, incremental processing

1. INTRODUCTION

The primary goal is to the demonstrate the Incremental Interaction Manager [1] (IIM). The IIM, an intermediary layer between recognition/synthesis components and the dialogue manager, enables a standard turn-based dialogue manager to operate with some of the benefits of incremental processing, such as faster reaction times and better barge-in handling. It also supports the simple integration of incremental speech recognition results with POMDP-based dialogue managers, increasing the interpretation accuracy substantially [1].

2. SYSTEM DESCRIPTION

The dialogue system demonstrated here consists of three main elements: the recognition/synthesis components, which perform incremental speech recognition, scoring, and prompt production; the Incremental Interaction Manager, which handles the system turn-taking behavior; and the dialogue manager, which provides appropriate system prompts and responses for a given dialogue context.

Here, we use a subsequent version of the statistical dialogue manager described in Williams [2]. This manager tracks a probability distribution over each (hidden) state known as a belief state, which is intended to be updated every *turn*. In other words, we are using a dialogue manager and statistical models that are designed for conventional whole-phrase recognition with *incremental* recognition.

2.1. Incremental Speech Recognition and Synthesis

The system runs Lattice-Aware Incremental Speech Recognition [3] on the AT&T WATSON^{SM} speech recognizer [4]. In addition to pre-recorded prompts, it uses AT&T Natural Voices speech synthesis. As the recognizer produces partial results, they are filtered so only those that could be a complete utterance are retained. Partials are often unstable, in that they often change, and so we score them for stability. More details can be found in Selfridge et al.[3].

2.2. Incremental Interaction Manager

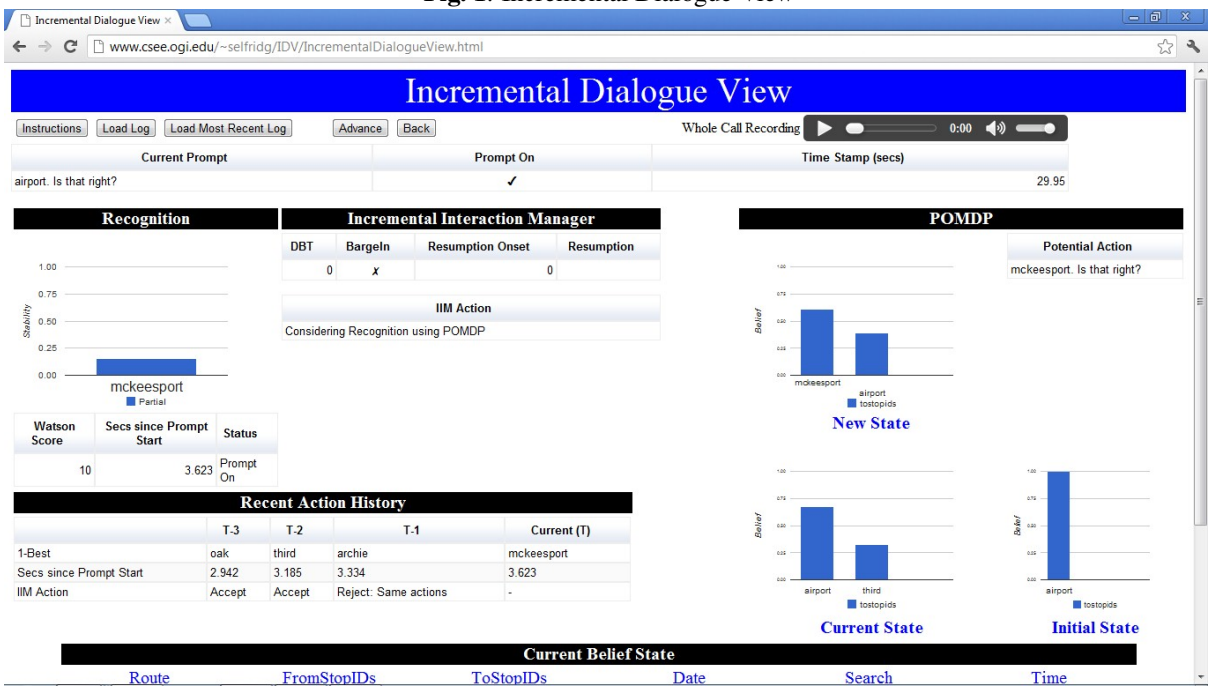
The Incremental Interaction Manager drives system turn-taking behavior. Using the stability score, it first decides whether to pass the partial to the dialogue manager or not. Currently, it passes all partials to the dialogue manager.

The IIM then evaluates a potential dialogue move by applying the partial result to a temporary instances of the DM. The IIM *copies* the current state of the DM, provides the copied DM with a recognition result, and considers the action that the copied DM would take. If the action will move the dialogue forward, the action is accepted by the IIM and the system begins playing the prompt. However, the IIM does not reset the recognizer yet, as more recognition may occur. If another partial (a revision) is recognized, it makes another copy of the initial DM and then evaluates another potential move while maintaining the current DM as well. If the move is accepted by the IIM, it stops the current prompt and begins playing the new one. More details of these procedures can be found in Selfridge et al. (2012).

Since it could be quite jarring for the user to hear multiple starts, we mask revisions by playing a short sound prior to every recognition-triggered prompt. While this *does* delay the prompt, it indicates to the user that the system is reacting and so any adverse effects should be minimal. Since partials are generally close together, the user hears multiple short sounds

[*]Work done while at AT&T Labs - Research

Fig. 1. Incremental Dialogue View



followed by a single prompt. When a phrase result is recognized (the final recognition of the utterance as determined by the recognizer), the IIM "refreshes" the decoder by loading the appropriate grammar and restarting the recognizer.

The IIM is also responsible for handling two complementary behaviors: barge-in detection and prompt resumption following a false barge-in. The IIM uses the partial's stability score to trigger barge-in, and the action decision to determine a false barge-in. Initially, any partial can trigger a barge-in. However, as the number of false barge-ins increases this *Dynamic Barge-In Threshold* increases as well, so that only more stable (and so more reliable) partials trigger barge-in. When barge-in is triggered, the IIM computes the appropriate place to resume the prompt (roughly two words from the barge-in time point) and then sets a timer (beginning at 1 second and then decreasing as the number of false barge-ins grow). If no action has been accepted before the timer lapses, the IIM will resume the prompt. Note that the number of false barge-ins is decremented by successful recognitions.

3. DEMONSTRATION DESCRIPTION

We provide a two-part demonstration of this incremental dialogue system. First, a live end-to-end system in the "Lets Go!" bus information domain. Second, a call inspection visualizer, Incremental Dialogue View, that allows the caller to step through the most recent system dialogue as well as those made previously. Incremental Dialogue View, shown in Figure 1, displays recognition results, IIM decisions, and the belief state. The visualizer (and a whole call recording) is available at *http://www.csee.ogi.edu/~selfridg/IDV/IDV.html*.

These two elements provides a full picture for users. They can first call the system as they normally would, with no visual feedback. Then, they can then step through their own call to see the behavior that was previously hidden from view and better understand the use of the IIM.

4. ACKNOWLEDGMENTS

We acknowledge funding from the NSF under grant IIS-0713698.

5. REFERENCES

[1] E.O. Selfridge, I. Arizmendi, P.A. Heeman, and J.D. Williams, "Integrating incremental speech recognition and pomdp-based dialogue systems," in *Proc. of the SIGdial*, 2012.

[2] J.D. Williams, "An empirical evaluation of a statistical dialog system in public use," in *Proc. of the SIGDIAL*, 2011.

[3] E.O. Selfridge, I. Arizmendi, P.A. Heeman, and J.D. Williams, "Stability and accuracy in incremental speech recognition," in *Proc. of the SIGdial*, 2011.

[4] V. Goffin, C. Allauzen, E. Bocchieri, D. Hakkani-Tur, A. Ljolje, S. Parthasarathy, M. Rahim, G. Riccardi, and M. Saraclar, "The AT&T WATSON speech recognizer," in *Proc. of ICASSP*, 2005, pp. 1033–1036.