MIMIC: An Adaptive Mixed Initiative Spoken Dialogue System for Information Queries

Jennifer Chu-Carroll
Lucent Technologies Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974, U.S.A.
jencc@research.bell-labs.com

Abstract

This paper describes MIMIC, an adaptive mixed initiative spoken dialogue system that provides movie showtime information. MIMIC improves upon previous dialogue systems in two respects. First, it employs initiative-oriented strategy adaptation to automatically adapt response generation strategies based on the cumulative effect of information dynamically extracted from user utterances during the dialogue. Second, MIMIC's dialogue management architecture decouples its initiative module from the goal and response strategy selection processes, providing a general framework for developing spoken dialogue systems with different adaptation behavior.

1 Introduction

In recent years, speech and natural language technologies have matured enough to enable the development of spoken dialogue systems in limited domains. Most existing systems employ dialogue strategies pre-specified during the design phase of the dialogue manager without taking into account characteristics of actual dialogue interactions. More specifically, mixed initiative systems typically employ rules that specify conditions (generally based on local dialogue context) under which initiative may shift from one agent to the other. Previous research, on the other hand, has shown that changes in initiative strategies in human-human dialogues can be dynamically modeled in terms of characteristics of the user and of the on-going dialogue (Chu-Carroll and Brown, 1998) and that adaptability of initiative strategies in dialogue systems leads to better system performance (Litman and Pan, 1999). However, no previous dialogue system takes into account these dialogue characteristics or allows for initiative-oriented adaptation of dialogue strategies.

In this paper, we describe MIMIC, a voice-enabled telephone-based dialogue system that provides movie showtime information, emphasizing its dialogue management aspects. MIMIC improves upon previous systems along two dimensions. First, MIMIC automatically adapts dialogue strategies based on participant roles, characteristics of the current utterance, and dialogue history. This automatic adaptation allows appropriate dialogue strategies to be employed based on both local dialogue context and dialogue history, and has been shown to result in significantly better performance than non-adaptive systems. Second, MIMIC employs an initiative module that is decoupled from the goal selection process in the dialogue manager, while allowing the outcome of both components to jointly determine the strategies chosen for response generation. As a result, MIMIC can exhibit drastically different dialogue behavior with very minor adjustments to parameters in the initiative module, allowing for rapid development and comparison of experimental prototypes and resulting in general and portable dialogue systems.

2 Adaptive Mixed Initiative Dialogue Management

2.1 Motivation

In naturally occurring human-human dialogues, speakers often adopt different dialogue strategies based on hearer characteristics, dialogue history, etc. For instance, the speaker may provide more guidance if the hearer is having difficulty making progress toward task completion, while taking a more passive approach when the hearer is an expert in the domain. Our main goal is to enable a spoken dialogue system to simulate such human behavior by dynamically adapting dialogue strategies during an interaction based on information that can be automatically detected from the dialogue. Figure 1 shows an excerpt from a dialogue between MIMIC and an actual user where the user is attempting to find the times at which the movie *Analyze This* playing at theaters in Montclair. S and U indicate system and user utterances, respectively, and the italicized utterances are the output of our automatic speech recognizer. In addition, each system turn is annotated with its *task* and *dialogue* initiative holders, where *task initiative* tracks the lead in the process toward achieving the dialogue participants' domain goal, while *dialogue initiative* models the lead in determining the current discourse focus (Chu-Carroll and Brown, 1998). In our information query application domain, the system has task (and thus dialogue) initiative if its utterances provide helpful guidance toward achieving the user's domain goal, as in utterances (6) and (7) where MIMIC provided valid response choices to its query intending to solicit a theater name, while the system has dialogue but not task initiative if its utterances only specify the current discourse goal, as in utterance (4).[1]

This dialogue illustrates several features of our adaptive mixed initiative dialogue manager. First, MIMIC automatically adapted the initiative distribution based on information extracted from user utterances and dialogue history. More specifically, MIMIC took over task initiative in utterance (6), after failing to obtain a valid answer to its query soliciting a missing theater name in (4). It retained task initiative until utterance (12), after the user implicitly indicated her intention to take over task initiative by providing a fully-specified query (utterance (11)) to a limited prompt (utterance (10)). Second, initiative distribution may affect the strategies MIMIC selects to achieve its goals. For instance, in the context of soliciting missing information, when MIMIC did not have task initiative, a simple information-seeking query was generated (utterance (4)). On the other hand, when MIMIC had task initiative, additional guidance was provided (in the form of valid response choices in utterance (6)), which helped the user successfully respond to the system's query. In the context of prompting the user for a new query, when MIMIC had task initiative, a limited prompt was selected to better constrain the user's response (utterance (10)), while an open-ended prompt was generated to allow the user to take control of the problem-solving process otherwise (utterances (1) and (13)).

In the next section, we briefly review a framework for dynamic initiative modeling. In Section 3, we discuss how this framework was incorporated into the dialogue management component of a spoken dialogue system to allow for automatic adaptation of dialogue strategies. Finally, we outline experiments evaluating the resulting system and show that MIMIC's automatic adaptation capabilities resulted in better system performance.

2.2 An Evidential Framework for Modeling Initiative

In previous work, we proposed a framework for modeling initiative during dialogue interaction (Chu-Carroll and Brown, 1998). This framework predicts task and dialogue initiative holders on a turn-by-turn basis in human-human dialogues based on participant roles (such as each dialogue agent's level of expertise and the role that she plays in the application domain), cues observed in the current dialogue turn, and dialogue history. More specifically, we utilize the Dempster-Shafer theory (Shafer, 1976; Gordon and Shortliffe, 1984), and represent the current initiative distribution as two *basic probability assignments (bpas)* which indicate the amount of support for each dialogue participant having the task and dialogue initiatives. For instance, the bpa $m_{t-cur}(\{S\}) = 0.3, m_{t-cur}(\{U\}) = 0.7$ indicates that, with all evidence taken into account, there is more support (to the degree 0.7) for the user having task initiative in the current turn than for the system. At the end of each turn, the bpas are updated based on the effects that cues observed during that turn have on changing them, and the new bpas are used to predict the next task and dialogue initiative holders.

In this framework, cues that affect initiative distribution include *NoNewInfo*, triggered when the speaker simply repeats or rephrases an earlier utterance, implicitly suggesting that the speaker may want to give up initiative, *AmbiguousActions*, triggered when the speaker proposes an action that is ambiguous in the application domain, potentially prompting the hearer to take over initiative to resolve the detected ambiguity, etc. The effects that each cue has on changing the current bpas are also represented as bpas, which were determined by an iterative training procedure using a corpus of transcribed dialogues where each turn was annotated with the task/dialogue initiative holders and the observed cues.

The bpas for the next turn are computed by combining the bpas representing the current initiative distribution and the bpas representing the effects of cues observed during the current turn, using Dempster's combination rule (Gordon and Shortliffe, 1984). The task and dialogue initiative holders are then predicted based on the new bpas. This framework was evaluated using annotated dialogues from four task-oriented domains, achieving, on average, a correct prediction rate of 97% and 88% for task and dialogue initiative holders, respectively. In Section 3.2, we discuss how this predictive model is converted into a generative model by enabling the system to automatically detect cues that were previously labelled manually. We further discuss how the model is used by the dialogue manager for dynamic dialogue strategy adaptation.

3 MIMIC: Mixed Initiative Movie Information Consultant

MIMIC is a telephone-based dialogue system that provides movie showtime information. It consists of the following main components, implemented on a distributed, client-server architecture (Zhou et al., 1997):

1. **Telephony server**: this component detects rings and hang-ups, and enables streaming of audio data on channels of a telephony board.
2. **Speech recognizer**: the recognizer receives audio data from the telephony server and generates the word string hypothesis that best matches the audio input. We used the Lucent Automatic Speech Recognizer (Reichl and Chou, 1998; Ortmanns et al., 1999), configured to use class-based probabilistic n-gram language models to allow for rapid updates of movie/theater/town names.

[1] Although the dialogues we collected in our experiments (Section 5) include cases in which MIMIC has neither initiative, such cases are rare in this application domain, and will not be discussed further in this paper.

		Utterance	Task Initiative	Dialogue Initiative
(1)	S:	Hello, this is MIMIC, the movie information system.	User	System
(2)		How can I help you?		
(3)	U:	What time is Analyze This playing in Montclair?		
		[what time is analyze this playing in montclair]		
(4)	S:	What theater would you like?	User	System
(5)	U:	Montclair.		
		[montclair]		
(6)	S:	Choices for theater are Wellmont Theater and Clearviews Screening Zone.	System	System
(7)		What theater would you like?		
(8)	U:	Clearviews Screening Zone.		
		[um movie screening zone]		
(9)	S:	Analyze This is playing at Clearviews Screening Zone in Montclair at 1:15pm, 4:10pm, and 6:30pm.	System	System
(10)		Please say the name of the movie or theater or town you would like information about.		
(11)	U:	What time is Analyze This playing at the Wellmont Theater?		
		[what time is analyze this playing at the wellmont theater]		
(12)	S:	Analyze This is playing at Wellmont Theater in Montclair at 3:15pm, 5:45pm, 8:20pm, and 10:10pm.	User	System
(13)		Can I help you with anything else?		

Figure 1: Excerpt of an Adaptive Mixed Initiative Dialogue

3. **NLP/Dialogue component**: this main application-dependent component receives a user utterance hypothesis from the speech recognizer, and generates system utterance(s) in response. Three major tasks are carried out by this component: 1) semantic interpretation, which constructs frame-based semantic representations from user utterances, 2) dialogue management, where response strategies are selected based on the semantic representation of the user's utterance, system's domain knowledge, and initiative distribution, and 3) utterance generation, where utterance templates are chosen and instantiated to realize the selected response strategies. These three tasks will be discussed in further detail in the rest of this section.

4. **Text-to-speech engine**: the TTS system receives the word string comprising the system's response from the dialogue component and converts the text into speech for output over the telephone. We used the Bell Labs TTS system (Sproat, 1998), which in addition to converting plain text into speech, accepts text strings annotated to override default pitch height, accent placement, speaking rate, etc.[2]

3.1 Semantic Interpretation

MIMIC utilizes a non-recursive frame-based semantic representation commonly used in spoken dialogue systems (e.g. (Seneff et al., 1991; Lamel, 1998)), which represents an utterance as a set of attribute-value pairs. Figure 2(a) shows the frame-based semantic representation for the utterance *"What time is Analyze This playing in Montclair?"*

Question-Type:	When
Movie:	Analyze This
Theater:	null
Town:	Montclair

(a) Semantic Representation

Question-Type:	When
Movie:	mandatory
Theater:	mandatory
Town:	optional

(b) Task Specification

Figure 2: Semantic Representation and Task Specification

MIMIC's semantic representation is constructed by first extracting, for each attribute, a set of keywords from the user utterance. Using a vector-based topic identification process (Salton, 1971; Chu-Carroll and Carpenter, 1999), these keywords are used to determine a set of likely values (including *null*) for that attribute. Next, the utterance is interpreted with respect to the dialogue history and the system's domain knowledge. This allows MIMIC to handle elliptical sentences and anaphoric references, as well as automatically infer missing values and detect inconsistencies in the current representation.

This semantic representation allows for decoupling of domain-dependent task specifications and domain-

[2] See (Nakatani and Chu-Carroll, 2000) for how MIMIC's dialogue-level knowledge is used to override default prosodic assignments for concept-to-speech generation.

independent dialogue management strategies. Each query type is specified by a template indicating, for each attribute, whether a value must, must not, or can optionally be provided in order for a query to be considered well-formed. Figure 2(b) shows that to solicit movie showtime information (question type *when*), a movie name and a theater name must be provided, whereas a town may optionally be provided. These specifications are determined based on domain semantics, and must be reconstructed when porting the system to a new domain.

3.2 Dialogue Management

Given a semantic representation, the dialogue history and the system's domain knowledge, the dialogue manager selects a set of strategies that guides MIMIC's response generation process. This task is carried out by three subprocesses: 1) *initiative modeling*, which determines the initiative distribution for the system's dialogue turn, 2) *goal selection*, which identifies a goal that MIMIC's response attempts to achieve, and 3) *strategy selection*, which chooses, based on the initiative distribution, a set of dialogue acts that MIMIC will adopt in its attempt to realize the selected goal.

3.2.1 Initiative Modeling

MIMIC's initiative module determines the task and dialogue initiative holders for each system turn in order to enable dynamic strategy adaptation. It automatically detects cues triggered during the current user turn, and combines the effects of these cues with the current initiative distribution to determine the initiative holders for the system's turn.

Cue Detection The cues and the bpas representing their effects are largely based on a subset of those described in (Chu-Carroll and Brown, 1998),[3] as shown in Figures 3(a) and 3(b). Figure 3(a) shows that observation of *TakeOverTask* supports a task initiative shift to the speaker to the degree .35. The remaining support is assigned to Θ, the set of all possible conclusions (i.e., $\{$speaker,hearer$\}$), indicating that to the degree .65, observation of this cue does not commit to identifying which dialogue participant should have task initiative in the next dialogue turn.

The cues used in MIMIC are classified into two categories, *discourse cues* and *analytical cues*, based on the types of knowledge needed to detect them:

1. Discourse cues, which can be detected by considering the semantic representation of the current utterance and dialogue history:
   - *TakeOverTask*, an implicit indication that the user wants to take control of the problem-solving process, triggered when the user provides more information than the discourse expectation.
   - *NoNewInfo*, an indication that the user is unable to make progress toward task completion, triggered when the semantic representations of two consecutive user turns are identical (a result of the user not knowing what to say or the speech recognizer failing to recognize the user utterances).
2. Analytical cues, which can only be detected by taking into account MIMIC's domain knowledge:
   - *InvalidAction*, an indication that the user made an invalid assumption about the domain, triggered when the system database lookup based on the user's query returns null.
   - *InvalidActionResolved*, triggered when the previous invalid assumption is corrected.
   - *AmbiguousAction*, an indication that the user query is not well-formed, triggered when a mandatory attribute is unspecified or when more than one value is specified for an attribute.
   - *AmbiguousActionResolved*, triggered when the attribute in question is uniquely instantiated.

Computing Initiative Distribution To determine the initiative distribution, the bpas representing the effects of cues detected in the current user utterance are instantiated (i.e., *speaker/hearer* in Figure 3 are instantiated as *system/user* accordingly). These effects are then interpreted with respect to the current initiative distribution by applying Dempster's combination rule (Gordon and Shortliffe, 1984) to the bpas representing the current initiative distribution and the instantiated bpas. This results in two new bpas representing the task and dialogue initiative distributions for the system's turn. The dialogue participant with the greater degree of support for having the task/dialogue initiative in these bpas is the task/dialogue initiative holder for the system's turn[4] (see Section 4 for an example).

3.2.2 Goal Selection

The goal selection module selects a goal that MIMIC attempts to achieve in its response by utilizing information from analytical cue detection as shown in Figure 4. MIMIC's goals focus on two aspects of cooperative dialogue interaction: 1) initiating subdialogues to resolve anomalies that occur during the dialogue by attempting to instantiate an unspecified attribute, constraining an attribute for which multiple values have been specified, or correcting an invalid assumption in the case of invalid

[3]We selected only those cues that can be automatically detected in a spoken dialogue system with speech recognition errors and limited semantic interpretation capabilities.

[4]In practice, this is the *preferred* initiative holder since practical reasons may prevent the dialogue participant from actually holding the initiative. For instance, if having task initiative dictates inclusion of additional helpful information, this can only be realized if MIMIC's knowledge base provides such information.

Cue Class	Cue	BPA
Discourse	TakeOverTask	$m_{t-tot}(\{speaker\}) = 0.35; m_{t-tot}(\Theta) = 0.65$
	NoNewInfo	$m_{t-nni}(\{hearer\}) = 0.35; m_{t-nni}(\Theta) = 0.65$
Analytical	InvalidAction	$m_{t-ia}(\{hearer\}) = 0.35; m_{t-ia}(\Theta) = 0.65$
	InvalidActionResolved	$m_{t-iar}(\{hearer\}) = 0.35; m_{t-iar}(\Theta) = 0.65$
	AmbiguousAction	$m_{t-aa}(\{hearer\}) = 0.35; m_{t-aa}(\Theta) = 0.65$
	AmbiguousActionResolved	$m_{t-aar}(\{speaker\}) = 0.35; m_{t-aar}(\Theta) = 0.65$

(a) Task Initiative

Cue Class	Cue	BPA
Discourse	TakeOverTask	$m_{d-tot}(\{speaker\}) = 0.35; m_{d-tot}(\Theta) = 0.65$
	NoNewInfo	$m_{d-nni}(\{hearer\}) = 0.35; m_{d-nni}(\Theta) = 0.65$
Analytical	InvalidAction	$m_{d-ia}(\{hearer\}) = 0.7; m_{d-ia}(\Theta) = 0.3$
	InvalidActionResolved	$m_{d-iar}(\{hearer\}) = 0.7; m_{d-iar}(\Theta) = 0.3$
	AmbiguousAction	$m_{d-aa}(\{hearer\}) = 0.7; m_{d-aa}(\Theta) = 0.3$
	AmbiguousActionResolved	$m_{d-aar}(\{speaker\}) = 0.7; m_{d-aar}(\Theta) = 0.3$

(b) Dialogue Initiative

Figure 3: Cues and BPAs for Modeling Initiative in MIMIC

Select-Goal(SemRep):
(1) If *AmbiguousAction* detected
(2) ambiguous-attr ← get-ambiguous(SemRep)
/* *get name of ambiguous attribute* */
(3) If (number-values(ambiguous-attr) == 0)
/* *attribute unspecified* */
(4) **Instantiate**(ambiguous-attr)
(5) Else /* *more than one value specified* */
(6) **Constrain**(ambiguous-attr)
(7) Else if *InvalidAction* detected
(8) **ProvideNegativeAnswer**(SemRep)
(9) Else /* *well-formed query* */
(10) answer ← database-query(SemRep)
(11) **ProvideAnswer**(answer)

Figure 4: Goal Selection Algorithm

user queries (steps 1-8)[5] (van Beek et al., 1993; Raskutti and Zukerman, 1993; Qu and Beale, 1999), and 2) providing answers to well-formed queries (steps 9-11).

3.2.3 Strategy Selection

Previous work has argued that initiative affects the degree of control an agent has in the dialogue interaction (Whittaker and Stenton, 1988; Walker and Whittaker, 1990; Chu-Carroll and Brown, 1998). Thus, a cooperative system may adopt different strategies to achieve the same goal depending on the initiative distribution. Since task initiative models contribution to domain/problem-solving goals, while dialogue initiative affects the current discourse goal, we developed alternative strategies for achieving the goals in Figure 4 based on initiative distribution, as shown in Table 1.

The strategies employed when MIMIC has only dialogue initiative are similar to the mixed initiative dialogue strategies employed by many existing spoken dialogue systems (e.g., (Bennacef et al., 1996; Stent et al., 1999)). To instantiate an attribute, MIMIC adopts the *InfoSeek* dialogue act to solicit the missing information. In contrast, when MIMIC has both initiatives, it plays a more active role by presenting the user with additional information comprising valid instantiations of the attribute (*GiveOptions*). Given an invalid query, MIMIC notifies the user of the failed query and provides an open-ended prompt when it only has dialogue initiative. When MIMIC has both initiatives, however, in addition to *NotifyFailure*, it suggests an alternative close to the user's original query and provides a limited prompt. Finally, when MIMIC has neither initiative, it simply adopts *NotifyFailure*, allowing the user to determine the next discourse goal.

3.3 Utterance Generation

MIMIC employs a simple template-driven utterance generation approach. Templates are associated with dialogue acts as shown in Table 2.[6] The generation component receives from the dialogue manager the selected dialogue acts and the parameters needed to instantiate the templates. It then generates the system response, which is sent to the TTS module for spoken output synthesis.

[5] An alternative strategy to step (4) is to perform a database lookup based on the ambiguous query and summarize the results (Litman et al., 1998), which we leave for future work.

[6] In most cases, there is a one-to-one-mapping between dialogue acts and templates. The exceptions are Answer, NotifyFailure, and SuggestAlternative, whose templates vary based on the question type.

	Task + Dialogue	**Dialogue**	**None**
Constrain	Clarify	Clarify	Acknowledge
Instantiate	GiveOptions InfoSeek	InfoSeek	Acknowledge
ProvideAnswer	Answer LimitedPrompt	Answer OpenPrompt	Answer
ProvideNegativeAnswer	NotifyFailure SuggestAlternative LimitedPrompt	NotifyFailure OpenPrompt	NotifyFailure

Table 1: Strategy Selection Based on Goal and Initiative Distribution

Dialogue Act	**Template**
Clarify	"Did you say $< value_1 > \ldots$, or $< value_n >$?"
Acknowledge	"Uh-huh."
GiveOptions	"Choices for $< attribute >$ are $< value_1 > \ldots < value_n >$."
InfoSeek	"What $< attribute >$ would you like?"
Answer	E.g., "$< movie >$ is playing at $< theater >$ at $< time_1 > \ldots < time_n >$."
OpenPrompt	"Can I help you with anything else?"
LimitedPrompt	"Please say the name of the movie or theater or town you would like information about."
NotifyFailure	E.g., "$< movie >$ is not playing at $< theater >$.
SuggestAlternative	E.g., "$< movie >$ is playing at $< alternativetheater >$ at $< time_1 > \ldots < time_n >$."

Table 2: Mappings Between Dialogue Acts and Utterance Templates

4 Examples

To illustrate MIMIC's adaptation capabilities, we return to the dialogue in Figure 1, which is repeated in Figure 5 and annotated with the cues detected in each user turn (in boldfaced italics) and the dialogue acts employed for response generation in each system turn (in boldface).

The bpas representing the initiative distribution for utterance (3) are the initial bpas, which, based on MIMIC's role as an information provider, are

$$m_{t-(3)}(\{S\}) = 0.3, \qquad m_{t-(3)}(\{U\}) = 0.7;$$
$$m_{d-(3)}(\{S\}) = 0.6, \qquad m_{d-(3)}(\{U\}) = 0.4.$$

The cue *AmbiguousAction* is detected in utterance (3) because the mandatory attribute *theater* was not specified and cannot be inferred (since the town of Montclair has multiple theaters). The bpas representing its effect are instantiated as follows (Figure 3):

$$m_{t-aa}(\{S\}) = 0.35, \qquad m_{t-aa}(\Theta) = 0.65;$$
$$m_{d-aa}(\{S\}) = 0.7, \qquad m_{d-aa}(\Theta) = 0.3.$$

Combining the current bpas with the effects of the observed cue, we obtain the following new bpas:

$$m_{t-(4)}(\{S\}) = 0.4, \qquad m_{t-(4)}(\{U\}) = 0.6;$$
$$m_{d-(4)}(\{S\}) = 0.83, \qquad m_{d-(4)}(\{U\}) = 0.17.$$

The updated bpas indicate that MIMIC should have dialogue but not task initiative when attempting to resolve the detected ambiguity in utterance (4).

MIMIC selects *Instantiate* as its goal to be achieved (Figure 4), which, based on the initiative distribution, leads it to select the *InfoSeek* action (Table 1) and generate the query *"What theater would you like?"*

The user's response in (5) again triggers *AmbiguousAction*, as well as *NoNewInfo* since the semantic representations of (3) and (5) are identical, given the dialogue context. When the effects of these cues are taken into account, we have the following initiative distribution for utterance (6):

$$m_{t-(6)}(\{S\}) = 0.62, \qquad m_{t-(6)}(\{U\}) = 0.38;$$
$$m_{d-(6)}(\{S\}) = 0.96, \qquad m_{d-(6)}(\{U\}) = 0.04.$$

Although *Instantiate* is again selected as the goal, MIMIC now has both task and dialogue initiatives; thus it selects both *GiveOptions* and *InfoSeek* to achieve this goal and generates utterances (6) and (7). The additional information, in the form of valid theater choices, helps the user provide the missing value in (8), allowing MIMIC to answer the query in (9) and prompt for the next query. However, despite the limited prompt, the user provides a well-formed query in (11), triggering *TakeOverTask*. Thus, MIMIC answers the query and switches to an open-ended prompt in (13), relinquishing task initiative to the user.

In addition to its automatic adaptation capabilities, another advantage of MIMIC is the ease of modifying its adaptation behavior, enabled by the decoupling of the initiative module from the goal and strategy selection processes. For instance, a system-initiative version of MIMIC can be achieved by setting the initial bpas as follows:

$$m_{t-initial}(\{S\}) = 1;$$
$$m_{d-initial}(\{S\}) = 1.$$

(1)	S:	Hello, this is MIMIC, the movie information system.	
(2)		How can I help you?	
(3)	U:	What time is Analyze This playing in Montclair?	
		[what time is analyze this playing in montclair]	***[AmbiguousAction]***
(4)	S:	What theater would you like?	**[InfoSeek]**
(5)	U:	Montclair.	
		[montclair]	***[AmbiguousAction,NoNewInfo]***
(6)	S:	Choices for theater are Wellmont theater and Clearviews Screening Zone	**[GiveOptions]**
(7)		What theater would you like?	**[InfoSeek]**
(8)	U:	Clearviews Screening Zone.	
		[um movie screening zone]	***[AmbiguousActionResolved]***
(9)	S:	Analyze This is playing at Clearviews Screening Zone in Montclair at 1:15pm, 4:10pm, and 6:30pm.	**[Answer]**
(10)		Please say the name of the movie or theater or town you would like information about.	**[LimitedPrompt]**
(11)	U:	What time is Analyze This playing at the Wellmont theater?	
		[what time is analyze this playing at the wellmont theater]	***[TakeOverTask]***
(12)	S:	Analyze this is playing at Wellmont theater in Montclair at 3:15pm, 5:45pm, 8:20pm, and 10:10pm.	**[Answer]**
(13)		Can I help you with anything else?	**[OpenPrompt]**

Figure 5: Annotated Dialogue Shown in Figure 1

This is because in the Dempster-Shafer theory, if the initial bpas or the bpas for a cue provide definite evidence for drawing a certain conclusion, then no subsequent cue has any effect on changing that conclusion. Thus, MIMIC will retain both initiatives throughout the dialogue. Alternatively, versions of MIMIC with different adaptation behavior can be achieved by tailoring the initial bpas and/or the bpas for each cue based on the application. For instance, for an electronic sales agent, the effect of *InvalidAction* can be increased so that when the user orders an out-of-stock item, the system will always take over task initiative and suggest an alternative item.

5 System Evaluation

We conducted two experiments to evaluate MIMIC's automatic adaptation capabilities. We compared MIMIC with two control systems: MIMIC-SI, a system-initiative version of MIMIC in which the system retains both initiatives throughout the dialogue, and MIMIC-MI, a non-adaptive mixed-initiative version of MIMIC that resembles the behavior of many existing dialogue systems. In this section we summarize these experiments and their results. A companion paper describes the evaluation process and results in further detail (Chu-Carroll and Nickerson, 2000).

Each experiment involved eight users interacting with MIMIC and MIMIC-SI or MIMIC-MI to perform a set of tasks, each requiring the user to obtain specific movie information. User satisfaction was assessed by asking the subjects to fill out a questionnaire after interacting with each version of the system. Furthermore, a number of performance features, largely based on the PARADISE dialogue evaluation scheme (Walker et al., 1997), were automatically logged, derived, or manually annotated. In addition, we logged the cues automatically detected in each user utterance, as well as the initiative distribution for each turn and the dialogue acts selected to generate each system response.

The features gathered from the dialogue interactions were analyzed along three dimensions: *system performance*, *discourse features* (in terms of characteristics of the resulting dialogues, such as the cues detected in user utterances), and *initiative distribution*. Our results show that MIMIC's adaptation capabilities 1) led to better system performance in terms of user satisfaction, dialogue efficiency (shorter dialogues), and dialogue quality (fewer ASR timeouts), and 2) better matched user expectations (by giving up task initiative when the user intends to have control of the dialogue interaction) and more efficiently resolved dialogue anomalies (by taking over task initiative to provide guidance when no progress is made in the dialogue, or to constrain user utterances when ASR performance is poor).

6 Conclusions

In this paper, we discussed MIMIC, an adaptive mixed-initiative spoken dialogue system. MIMIC's automatic adaptation capabilities allow it to employ appropriate strategies based on the cumulative effect of information dynamically extracted from user utterances during dialogue interactions, enabling MIMIC to provide more cooperative and satisfactory responses than existing non-adaptive systems. Furthermore, MIMIC was implemented as a general framework for information query systems by decoupling its initiative module from the goal selection process, while allowing the outcome of both processes to jointly determine the response strategies employed. This feature enables easy modification to MIMIC's adaptation behavior, thus allowing the framework to be used for rapid development and comparisons of experimental prototypes of spoken dialogue systems.

Acknowledgments

The author would like to thank Egbert Ammicht, Antoine Saad, Qiru Zhou, Wolfgang Reichl, and Stefan Ortmanns for their help on system integration and on ASR/telephony server development, Jill Nickerson for conducting the evaluation experiments, and Bob Carpenter, Diane Litman, Christine Nakatani, and Jill Nickerson for their comments on an earlier draft of this paper.

References

S. Bennacef, L. Devillers, S. Rosset, and L. Lamel. 1996. Dialog in the RAILTEL telephone-based system. In *Proceedings of the 4th International Conference on Spoken Language Processing*.

Jennifer Chu-Carroll and Michael K. Brown. 1998. An evidential model for tracking initiative in collaborative dialogue interactions. *User Modeling and User-Adapted Interaction*, 8(3-4):215–253.

Jennifer Chu-Carroll and Bob Carpenter. 1999. Vector-based natural language call routing. *Computational Linguistics*, 25(3):361–388.

Jennifer Chu-Carroll and Jill S. Nickerson. 2000. Evaluating automatic dialogue strategy adaptation for a spoken dialogue system. In *Proceedings of the 1st Conference of the North American Chapter of the Association for Computational Linguistics*. To appear.

Jean Gordon and Edward H. Shortliffe. 1984. The Dempster-Shafer theory of evidence. In Bruce Buchanan and Edward Shortliffe, editors, *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, chapter 13, pages 272–292. Addison-Wesley.

Lori Lamel. 1998. Spoken language dialog system development and evaluation at LIMSI. In *Proceedings of the International Symposium on Spoken Dialogue*, pages 9–17.

Diane J. Litman and Shimei Pan. 1999. Empirically evaluating an adaptable spoken dialogue system. In *Proceedings of the 7th International Conference on User Modeling*, pages 55–64.

Diane J. Litman, Shimei Pan, and Marilyn A. Walker. 1998. Evaluating response strategies in a web-based spoken dialogue agent. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics*, pages 780–786.

Christine H. Nakatani and Jennifer Chu-Carroll. 2000. Using dialogue representations for concept-to-speech generation. In *Proceedings of the ANLP-NAACL Workshop on Conversational Systems*.

Stefan Ortmanns, Wolfgang Reichl, and Wu Chou. 1999. An efficient decoding method for real time speech recognition. In *Proceedings of the 5th European Conference on Speech Communication and Technology*.

Yan Qu and Steve Beale. 1999. A constraint-based model for cooperative response generation in information dialogues. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence*.

Bhavani Raskutti and Ingrid Zukerman. 1993. Eliciting additional information during cooperative consultations. In *Proceedings of the 15th Annual Meeting of the Cognitive Science Society*.

Wolfgang Reichl and Wu Chou. 1998. Decision tree state tying based on segmental clustering for acoustic modeling. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*.

Gerald Salton. 1971. *The SMART Retrieval System*. Prentice Hall, Inc.

Stephanie Seneff, James Glass, David Goddeau, David Goodine, Lynette Hirschman, Hong Leung, Michael Phillips, Joseph Polifroni, and Victor Zue. 1991. Development and preliminary evaluation of the MIT ATIS system. In *Proceedings of the DARPA Speech and Natural Language Workshop*, pages 88–93.

Glenn Shafer. 1976. *A Mathematical Theory of Evidence*. Princeton University Press.

Richard Sproat, editor. 1998. *Multilingual Text-to-Speech Synthesis: The Bell Labs Approach*. Kluwer, Boston, MA.

Amanda Stent, John Dowding, Jean Mark Gawron, Elizabeth Owen Bratt, and Robert Moore. 1999. The CommandTalk spoken dialogue system. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 183–190.

Peter van Beek, Robin Cohen, and Ken Schmidt. 1993. From plan critiquing to clarification dialogue for cooperative response generation. *Computational Intelligence*, 9(2):132–154.

Marilyn Walker and Steve Whittaker. 1990. Mixed initiative in dialogue: An investigation into discourse segmentation. In *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, pages 70–78.

Marilyn A. Walker, Diane J. Litman, Candance A. Kamm, and Alicia Abella. 1997. PARADISE: A framework for evaluating spoken dialogue agents. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, pages 271–280.

Steve Whittaker and Phil Stenton. 1988. Cues and control in expert-client dialogues. In *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics*, pages 123–130.

Qiru Zhou, Chin-Hui Lee, Wu Chou, and Andrew Pargellis. 1997. Speech technology integration and research platform: A system study. In *Proceedings of the 5th European Conference on Speech Communication and Technology*.