Multimodal Dialogue Management - State of the art

Trung H. BUI
Human Media Interaction Department
University of Twente,
7500 AE Enschede, The Netherlands
buith@cs.utwente.nl

January 3, 2006

Version 1.0

Abstract

This report is about the state of the art in dialogue management. We first introduce an overview of a multimodal dialogue system and its components. Second, four main approaches to dialogue management are described (finite-state and frame-based, information-state based and probabilistic, plan-based, and collaborative agent-based approaches). Finally, the dialogue management in the recent dialogue systems is presented.

Contents

1 Introduction

Dialogue is a conversation between two or more agents, be they human or machine. Research on dialogue usually follows two main directions: *human-human dialogue* and *human-computer dialogue*. The later is involved in a *dialogue system*, a computer program that communicates with a human user in a natural way.

Previous research work has been focusing on *spoken dialogue systems*, which are defined as computer systems that human interact on a turn-by-turn basic and in which spoken natural language interface plays an important part in the communication [Fraser, 1997]. Recently, it has been extended to *multimodal dialogue systems*, which are dialogue systems that process two or more combined user input modes - such as speech, pen, touch, manual gestures, gaze, and head and body movements - in a coordinated manner with multimedia system output [Oviatt, 2002].

Both spoken dialogue system and multimodal dialogue system need a central management module called the *Dialogue Manager*. The Dialogue Manager (DM) is the program which coordinates the activity of several subcomponents in a dialogue system and its main goal is to maintain a representation of the current state of the ongoing dialogue.

This report describes the state of the art of the dialogue management research in a context of both spoken and multimodal dialogue systems. Section 2 describes an overview of a multimodal dialogue system and its components (readers who are only interested in spoken dialogue systems can consult [McTear, 2002]). Sections 3 and 4 presents approaches to dialogue management. Section 5 is about dialogue management in recent dialogue systems. Finally, the summary of the report is presented in section 6.

2 Overview of a multimodal dialogue system

A multimodal dialogue system normally consists of the following components (cf. Fig. 1): *Input*, *Fusion*, *Dialogue Manager* (DM) and *General Knowledge*, *Fission*, and *Output*.

2.1 Input

Inputs of a multimodal dialogue system are a subset of the various modalities such as: speech, pen, facial expressions, gestures, gazes, and so on. Two types of input modes are distinguished: *active input modes* and *passive input modes*. *Active input modes* are the modes that are deployed by the user intentionally as an explicit command to the computer such as speech. *Passive input modes* refer to naturally occurring user behavior or actions that are recognized by a computer (e.g., facial expressions, manual gestures). They involve user input that is unobtrusively and passively monitored, without requiring any explicit command to a computer [Oviatt, 2002].

A popular set of input modalities are: (1) speech and lips movement, (2) speech and gesture (including pen gesture, pointing gesture, human gesture), (3) speech, gesture, and facial expressions.

2.2 Fusion

Information from various input modalities is extracted, recognized and fused. Fusion processes the information and assigns a semantic representation which is eventually sent to DM.

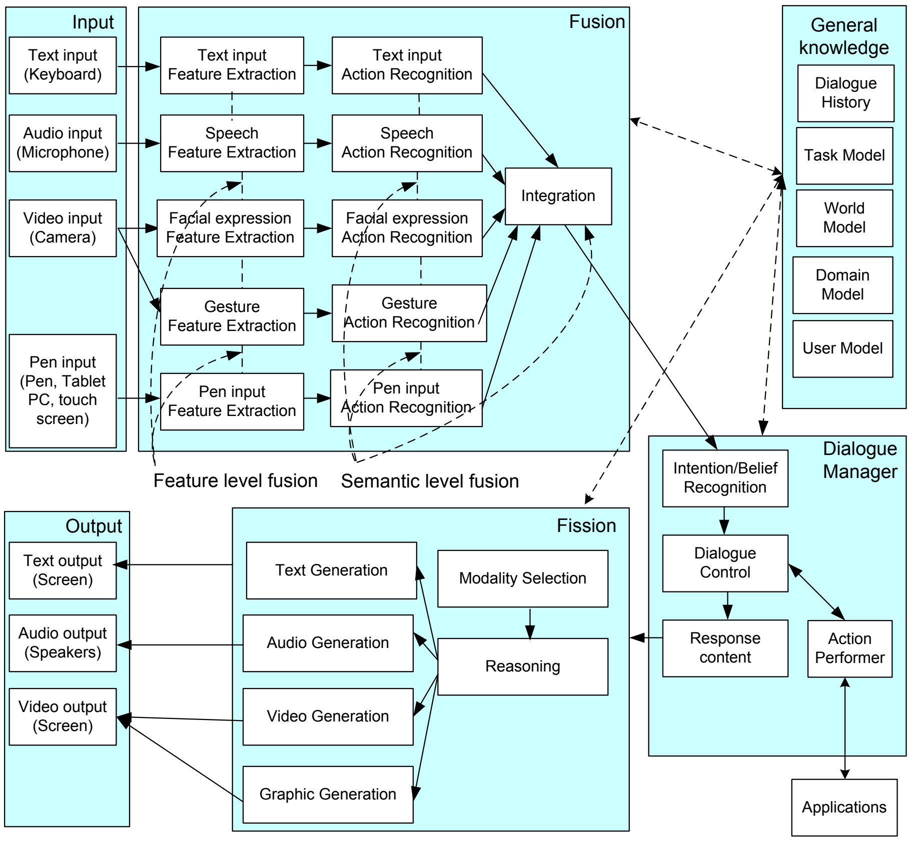

Figure 1: *General architecture of a multimodal dialogue system.*

In the context of multimodal dialogue systems, two main levels of fusion are often used: *feature-level fusion*, *semantic-level fusion*. The first one is a method for fusing low-level feature information from parallel input signals within a multimodal architecture (for example, in Fig. 1, feature-level fusion happens between input modality feature extraction modules). The second one is a method for integrating semantic information derived from parallel input modes in a multimodal architecture (for example, in Fig. 1, semantic-level fusion happens between input modality action recognizer [1] modules such as speech and gesture).

Another related work on low-level fusion is *sensor fusion*, which is the combining of sensory data from disparate sources such that the resulting information is in some sense better than would be possible when these sources were used individually [2].

Semantic-level fusion is usually involved in the DM and needs to consult the knowledge source from the DM. To date, three popular semantic fusion techniques are used:

- *Frame-based fusion* is a method for integrating semantic information derived from parallel input modes in a multimodal architecture, which has been used for processing speech and gesture input (e.g. [Vo and Wood, 1996]).
- *Unification-based fusion* is a logic-based method for integrating partial meaning fragments derived from two input modes into a common meaning representation during multimodal language processing. Compared with frame-based fusion, unification-based fusion derives from logic programming, and has been more precisely analyzed and widely adopted within computational linguistics (e.g. [Johnston, 1998]).
- *Hybrid symbolic/statistical fusion* is an approach to combine statistical processing techniques with a symbolic unification-based approach (e.g. Members-Teams-Committee (MTC) hierarchical recognition fusion [Wu *et al.*, 2002]).

2.3 Dialogue Manager and General Knowledge

Dialogue Manager is the core module of the system. The main tasks of DM are [Traum and Larsson, 2003]:

- updating the dialogue context on the basis of interpreted communication
- providing context-dependent expectations for interpretation of observed signals as communicative behavior
- interfacing with task/domain processing (e.g., database, planner, execution module, other back-end system), to coordinate dialogue and non-dialogue behavior and reasoning
- deciding what content to express next and when to express it

The term "dialogue context" can be viewed as *the totality of conditions that may influence the understanding and the generation of communicative behavior* [Bunt, 2000]. This definition is quite vague, and Bunt restricts to "local" aspect of the dialogue context (also called local context) which can be changed through communication. Local context

[1]this term is described in http://www.cs.berkeley.edu/ jfc/cs160/SP04/

[2]http://en.wikipedia.org/wiki/Sensor_fusion

factors can be grouped into five categories of conceptually different information dimensions: *linguistic, cognitive. physical, semantic, and social* as shortly described in table 1. More detail about these contexts are described in [Bunt, 2000].

Linguistic context	Surrounding linguistic material, 'raw' as well as analysed
Semantic context	state of the underlying task; facts in the task domain.
Cognitive context	participants' states of processing and models of each other's states.
Physical and perceptual context	availability of communicative and perceptual channels; partners' presence and attention.
Social context	communicative rights, obligations and constraints of each participant.

Table 1: Local dialogue context in the different dimensions

A number of general knowledge sources that is usually used by Dialogue Manager, Fusion, and Fission is as follows ([McTear, 2002], [Sharma *et al.*, 2003]):

- *Dialogue history*: A record of the dialogue so far in terms of the propositions that have been discussed and the entities that have been mentioned. This representation provides a basis for conceptual coherence and for the resolution of anaphora and ellipsis.
- *Task model*: A representation of the information to be gathered in the dialogue. This record, often referred to as a form, template, or status graph, is used to determine what information has not yet been acquired.
- *World model*: This model contains general background information that supports any commonsense reasoning required by the system, for example, that Christmas day is December 25.
- *Domain model*: A model with specific information about the domain in question, for example, flight information.
- *User model*: This model may contain relatively stable information about the user that may be relevant to the dialogue such as the users age, gender, and preferences (user preferences) as well as information that changes over the course of the dialogue, such as the users goals, beliefs, and intentions (user's mental states).

2.4 Fission

Fission is the process of realizing an abstract message through output on some combination of the available channels. The tasks of a fission module is composed of three categories [Foster, 2002]:

- *Content selection and structuring*: the presented content must be selected and arranged into an overall structure.
- *Modality selection*: the optimal modalities is determined based on the current situation of the environment, for example when the user device has a limited display and memory, the output can be presented as the graphic form such as a sequence of icons.

- *Output coordination*: the output on each of the channels should be coordinated so that the resulting output forms a coherent presentation.

2.5 Output

Various output modalities can be used to present the information content from the fission module such as: speech, text, 2D/3D graphics, avatar, haptics, and so on.

Popular combinations of the output modalities are: (1) graphics and avatar, (2) speech and graphics, (3) text and graphics, (4) speech and avatar, (5) speech, text, and graphics, (6) text, speech, graphics, and animation, (7) graphics and haptics, (8) speech and gesture.

3 Goal of the dialogue management

There is a distinction between dialogue models and dialogue management models or equivalently, between dialogue modeling and dialogue management modeling [Xu *et al.*, 2002]. The goal of dialogue modeling is to develop general theories of (usually, cooperative or collaborative task-oriented) dialogues and to uncover the universals in dialogues and to provide dialogue management with theoretical support. It takes an analyzer's point of view. Whereas, the goal of dialogue management modeling is to integrate dialogue model with task model in some specific domain to develop algorithms and procedures to support a machine's participation in a cooperative or collaborative dialogue. It takes the viewpoint of a dialogue system designer. In this report, we consider both theoretical and practical perspectives and group dialogue modeling and dialogue management modeling in a general terminology: *dialogue management*.

The purpose of studying dialogue management is to provide models allowing us to explore how language is used in different activities [Allwood, 1997]. Some of the questions that are addressed by theories of dialogue management are:

- What enables agents to participate in dialogue?
- What kind of information does a dialogue participant need to keep track of?
- How is this information used for interpreting and generating linguistic behavior?
- How is dialogue structured, and how can these structures be explained?

A part from these more theoretical motivations, there are also practical reasons for being interested in these fields. We are interested in creating *practical dialogue systems* [Allen *et al.*, 2001] to enable natural human-machine interaction. There is a widely held belief that interfaces using spoken dialogue and non-verbal modalities may be an important thing in the field of human-computer interaction. However, we believe that before this can happen, dialogue systems must become more flexible and more intelligent than currently available commercial systems. In order to achieve this, we need to base our implementations on reasonable theories of dialogue management. And of course, the implementation of dialogue systems can also feed back into the theoretical modeling of dialogue, provided the actual implementations are closely related to the underlying theory of dialogue [Larsson, 2002].

4 Approaches to dialogue management

According to how task model and dialogue model are used, there are several ways to classify dialogue management approaches. In [McTear, 2002], three strategies for dialogue control (i.e. dialogue management) are mentioned : finite state-based, frame-based, and agent-based. In [Xu *et al.*, 2002], four categories for dialogue management are distinguished: DITI (implicit dialogue model, implicit task model: like finite state-based models), DITE (implicit dialogue model, explicit task model: like frame-based models), DETI (explicit dialogue model, implicit task model), DETE (explicit dialogue model, explicit task model). In [Cohen, 1997] and [Catizone *et al.*, 2002], dialogue management approaches are classified into three categories: dialogue grammars, plan-based approaches, and cooperative approaches (i.e. agent-based approaches). These approaches are not mutually exclusive, and are often used together. For instance, plan-based approaches include features of dialogue grammars, and collaborative approaches include features of plan-based approaches.

Based on the recent development of the information state and the probabilistic approaches, we classify the approaches in four categories: (1) *Finite-state and frame-based approaches*, (2) *Information state and the probabilistic approaches*, (3)*Plan-based approaches*, and (4) *Collaborative agent-based approaches*.

4.1 Finite state-based and frame-based approaches

Finite state models are the simplest models used to develop a dialogue management system. The dialogue structure is represented in the form of state transition network in which the nodes represent the system's utterances (e.g. prompts) and the transitions between the nodes determine all the possible paths through the network. The dialogue control is system-driven and all the system's utterances are predetermined. In this approach, both task model and dialogue model are implicit and they are encoded by a dialogue designer. More detail about the theory of this approach is described in [Cohen, 1997].

Examples of the implemented dialogue management systems using this approach are the Nuance automatic banking system [McTear, 2002].

The major advantage of this approach is the simplicity. It is suitable for simple dialogue systems with well-structured task. However, the approach lacks of flexibility (i.e. only one state result from a transition), naturalness, and applicability to other domains.

An extension of finite state-based models, *frame-based models*, is developed to overcome the lack of flexibility of the finite state models. In the frame-based approach, rather than build a dialogue according to a predetermined sequence of system's utterances, they take the analogy of a form-filling (or slot-filling) task in which a predetermined set of information is to be gathered. The approach allows some degree of mixed-initiative and multiple slot fillings. The task model is represented explicitly and the dialogue model is (implicitly) encoded by a dialogue designer.

For example, [Hulstijn *et al.*, 1996], who developed a theatre booking system, arranged frame hierarchically to reflex the dependence of certain topics on others. In [van Zanten, 1996], a train timetable enquiry system, a frame structure relates the entities in the domain to on another, and this structure captures the meaning of all possible queries the user can make.

In [Goddeau *et al.*, 1996], they discussed a more complex type of form, the E-form (electronic form), which has been used in a spoken language interface to a database of classified advertisements for used cars. E-forms differ from the types of form and frame described so far, in that the slots may have different priorities for different users, etc.

Other variations of frame-based models allowing to deal with more complex dialogues include: *schemas, agenda* are used in the Carnegie Mellon Communicator system to model more complex tasks than the basic information retrieval tasks that use forms ([Constantinides *et al.*, 1998], [Rudnicky *et al.*, 1999], [Xu and Rudnicky, 2000], [Bohus and Rudnicky, 2003]), *task structure graphs* provide a similar semantic structure to E-form and are used to determine the behavior of the dialogue control as well as the language understanding module [Wright *et al.*, 1998], and *type hierarchies* are used to model the domain of a dialogue and as a basic for clarification questions [Denecke and Waibel, 1997], *blackboard* is used to managed contextual information relevant to dialogue manager such as history board, control board, presentation board, etc. [Rothkrantz *et al.*, 2000].

The frame-based approaches have several advantages over the finite state-based approaches: greater flexibility, the dialogue flow is more efficient and natural. However, the system context that contributes to the determination of the system's next action is fairly limited, more complex transactions cannot be modeled using these approaches.

Various rapid dialogue prototyping toolkits are available for the development and evaluation of dialogue systems using the finite-state based and frame-based approaches. Some of which will be presented hereafter.

In [Luz, 1999], a set of softwares/tools (e.g. CSLU's Rapid Dialogue Developer (RAD), UNISYS's Dialogue Design Assistant (DDA), GULAN, SpeechMania's HDDL-based toolkit, etc.) allowing to quickly develop dialogue management systems have been reviewed. These tools usually represent as a graphical based authoring environment (i.e. graphical editors) for designing and implementing spoken dialogue systems. For instance, RAD has been developed at the Center for Spoken Language Understanding (CSLU) at the Oregon Graduate Institute of Science and Technology to support speech-related research and development activities. A major advantage of the RAD interface is that users are shielded from many of the complex specification processes involved in the construction of a spoken dialogue system. Building a dialogue system involves selecting and linking graphical dialogue objects into a finite-state dialogue model, which may include branching decision, loops, jumps, and sub-dialogues ([Cole, 1999], [McTear, 1998], [Cole *et al.*, 1993]).

In [Denecke, 2002], Decnecke proposed a rapid prototyping for spoken dialogue systems which is an object-oriented approach. The architecture of the dialogue system was designed specifically to support rapid prototyping with three levels of abstraction: Abstract Dialogue Engine, Interaction Pattern Layer, and Dialogue Control Layer. Dialogues are rapidly created utilizing a class-based, hierarchical approach that does not require the creator to have a vast knowledge of dialogue systems.

In [Bui and Rajman, 2004], the authors have developed a Rapid Dialogue Prototyping Methodology (RDPM). The methodology is decomposed into 5 consecutive main steps: (1) producing the task model; (2) deriving the initial interaction model; (3) using a Wizard-of-Oz experiment to instantiate the initial interaction model; (4) using an internal field test to refine the interaction model; and (5) using an external field test to evaluate the final interaction model. Several extensions of this approach are being investigated [Bui *et al.*, 2005], [Pavel Cenek and Rajman, 2005].

4.2 Information state-based and the probabilistic approaches

Information state approach and its extensions are an effort to overcome limitations in finite-state based and frame-based approaches. An information state-based theory of dialogue consists of five main components [Traum and Larsson, 2003]:

- a description of *informational components* (e.g. participants, common ground, linguistic and intentional structure, obligations and commitments, beliefs, intentions, user models, etc.).
- *formal representations* of the above components (e.g., as lists, sets, typed feature structures, records, Discourse Representation Structures, propositions or modal operators within a logic, etc.)
- a set of *dialogue moves* that will trigger the update of the information state.
- a set of *update rules*, that govern the updating of the information state.
- an *update strategy* for deciding which rule(s) to apply at a given point from a set of applicable ones.

The theory has been implemented as a dialogue move engine toolkit called TrindiKit using Sicstus Prolog. A number of systems has been developed using this toolkit such as GoDiS [Larsson *et al.*, 2000], EDIS [Larsson and Traum, 2000].

The general idea of the information state approaches is being used for the development of multimodal dialogue systems such as Virtual Music Center [Hofs *et al.*, 2003], MATCH system for multimodal access to city help [Johnston *et al.*, 2002], Immersive Virtual Worlds [Traum and Rickel, 2002].

For example in [Traum and Rickel, 2002], Traum extends the original idea of the information state to develop a *multi-layer* dialogue model, each layer contains an information state representing the current status of that layer and a set of dialogue acts corresponding to the well-defined changes to the information state. A short description of this dialogue model is summarized as follow:

- contact: whether and how other individuals can be accessible for communication
- attention: the object or the process that agents attend to
- conversation: the separate dialogue episodes that go on during an interaction
  - participants: active speakers, addresses, overhearers
  - turn: indicating the participant with the right to communicate using the primary channel
  - initiative: indicating the participant who is controlling the direction of the conversation
  - grounding: tracking how information is added to the common ground of the participants
  - topic: governing the relevance
  - rhetorical: connecting between individual content units
- social commitments (obligations)

- negotiation: modeling how agents comme to agree on the commitments

Another extension the information state approaches is to use *probabilistic* techniques such as (fully observable) Markov Decision Process (MDP) or a Partially Observable Markov Decision Process (POMDP). The idea is to dynamically allow changing of the dialogue strategy and the actions of a dialogue systems based on optimizing some kinds of rewards or costs given the current state. Related research work focuses on a probabilistic modeling of dialogues in order to enable the system to statistically learn an optimal confirmation and initiative strategy. This is done by modeling the dialogue as a MDP ([Levin and Pieraccini, 1997],[Levin *et al.*, 2000],[Singh *et al.*, 2002]), or as a POMDP ([Roy *et al.*, 2000],[Zhang *et al.*, 2001],[Williams *et al.*, 2005]). Reinforcement learning is then used to learn an optimal strategy. The actions are the system's responses and question and the rewards are either defined by the designer (high reward for task completion, low punishment for confirmation and questions and so on) as in [Roy *et al.*, 2000, Zhang *et al.*, 2001], or they are provided by the user who is asked to rate the system at the end of each dialogue, as in [Singh *et al.*, 2002].

In [Young, 1999, Scheffler and Young, 2002], the system is optimizing a dialogue cost function either through Q-learning [Scheffler and Young, 2002] or using dynamic programming or sampling-based reinforcement learning.

In [Lecoeuche, 2001], inductive logic programming is used in order to extract rules from the result of reinforcement learning and in [Walker *et al.*, 1998] Q-learning is used to choose between strategies that are more specific than initiative or confirmation.

Other approaches using Bayesian Networks to recognize the dialogue act and/or control the dialogue strategy (e.g. [Wai *et al.*, 2001], [Keizer, 2003]) are recently studied.

4.3 Plan-based approaches

Plan-based approaches support a greater complexity to dialogue modeling than the approaches presented in previous sections. These approaches are based on the view that humans communicate to achieve goals, including changes to the mental state of the listener. The dialogue's input is not only considered as a sequence of words but as performing speech acts [Searle, 1969] and is used to achieve these goals. Usually, the task of the listener is to discover and appropriately respond to the speaker's underlying plan. In [Cohen, 1997], for example, in response to the customer's question of *"Where are the steaks you advertised?"*, a butcher's reply of *"How many do you want?"* is appropriate because the butcher understands the underlying plan of the customer.

The plan-based approaches are based on the plan-based theories of communicative action and dialogue (e.g. [Allen and Perault, 1980], [Appelt, 1985], [Cohen and Levesque, 1990]). The theories claim that the speaker's speech act is part of a plan and that it is the listener's job to identify and respond appropriately to this plan.

In the Verbmobil project, a number of approaches (finite state models of sequences of speech acts and plan-based operators) to modeling dialogue is deployed. The system represents the intentional structure of the dialogue on four levels: the lowest dialogue act level representing the speech acts of the speaker; a turns level which models more than one speech act within an actual turn; a dialogue phase level indicating the phase of the dialogue (e.g. opening, negotiation, closing); and the dialogue level at which individual dialogues take place. The dialogue is parsed from overall goal to sub-goals by a set of plan operators derived from an annotated corpus of example dialogues [Churcher *et al.*, 1997].

Other developed systems/projects based on these approaches include: SUNDIAL (Speech UNderstanding in DIALogue) project ([McGlashan *et al.*, 1992], [Eckert and McGlashan, 1993], [McGlashan, 1996]), TRAINS-96, etc.

Plan-based approaches have been criticised on practical and theoretical grounds. For example, the processes of plan-recognition and planning are combinatorically intractable in the worst case, and in some cases they are even undecidable. These approaches also lack of a sound theoretical basis. There is often no specification of what the system should do, for example, in terms of the kinds of dialogue phenomena and properties the framework can handle or what the various constructs like plans, goals, etc. are. In the example above of the customer and the butcher, the butcher interprets the intention of the customer by identifying the relevant plan. The actual interpretation is quite different from the illocutionary act and it is not clear how the butcher identified the plan from it.

An extension of the plan-based approaches is taken from the *conversational games theory* ([Carletta *et al.*, 1995], [Kowtko *et al.*, 1991]). The theory uses techniques from both dialogue grammars and plan-based approaches by including a goal or plan-oriented level in its structural approach. It can be used to model conversations between a human and a computer in a task-oriented dialogue [Williams, 1996].

A task-oriented dialogue consists of one or more transactions, each of which represents a subtask. A transaction comprises a number of conversational games, which in turn consist of an opening move, and end move (optional). For example, an INSTRUCTION game which consists of three nested games: an EXPLAINING game, a QUERY-YN game, and a CHECKING game. The CHECKING game can consist of a QUERY-YN and a REPLY-Y or a REPLY-N.

The approach deals with discourse phenomena such as side sequences, clarifications etc. by allowing games to be have another game embedded with in it. This technique provides a method of modeling mixed-initiative, complex natural language dialogue.

4.4 Collaborative agent-based approaches

Collaborative approaches or agent-based dialogue management approaches are based on viewing dialogues as a collaborative process between intelligent agents. Both agents work together to achieve a mutual understanding of the dialogue. The motivations that this joint activity places on both agents motivates discourse phenomena such as confirmation and clarification - which are also evident in human to human conversations [3].

Unlike the dialogue grammars and plan-based approaches which concentrate on the structure of the task, the collaborative approaches try to capture the motivations behind a dialogue and the mechanisms of dialogue itself. The beliefs of at least two participants will be explicitly modeled. A proposed goal, which is accepted by the other partner(s), will become part of the shared belief and the partners will work cooperatively to achieve this goal.

Several classes of these approaches have been developed using theorem proving, distributed architectures, and conversational agents. Some of which will be discussed below, and each approach uses a combination of techniques from dialogue grammar and plan-based approaches.

[3] we view the agent-based approaches broader that Blaylock [Blaylock, 2005b]. Blaylock views dialogue as collaborative problem solving between agents

In [Novick and Ward, 1993] and [Novick and Hansen, 1995], a model of collaboration (extended from [Clark and Schaefer, 1989]) has been proposed to establish the mutual information (i.e. shared knowledge). Agents collaborate to build a mutual model of conversation and shared belief using a set of domain dependent and independent speech acts. The belief system is similar to that proposed in [Allen, 1991], [Traum, 1991], [Traum and Hinkelman, 1992]. This theory has been implemented for a number of domains: the conversation between the Air Traffic Controller and a pilot conducting an Instrument Landing System landing [Novick and Ward, 1993]; and a letter-sequence completion task where the letters are distributed among the two participants [Novick and Hansen, 1995].

In the TRAINS-93 dialogue manager [Traum, 1996], Traum's model of conversation agency extended Bratman's Beliefs Desires Intentions (BDI [4]) agent architecture [Bratman *et al.*, 1991]. Traum states two major problems with the BDI model. Firstly, he argues that agent's perceptions not only influence its beliefs but also its desires and intentions. Secondly, the BDI model does not support more than one agent. Traum thus extended the BDI model by incorporating mutual beliefs, i.e. what both agents belief to be true and also let perceptions influence desires and intentions as well as beliefs.

Similar approaches to simulating collaborative dialogue have also been developed by Chu-Carroll [Chu-Carroll, 1996] who extends Sidners work ([Sidner, 1992], [Sidner, 1994]), and Beun [Beun, 1996].

In [Wilks and Ballim, 1991], *ViewGen* is a representational system for modeling agents and their beliefs and goals as part of a dialogue system. Two types of structures are used in Viewgen: the first one for agents that can have views of other agents and entities, and the second one for entities that have no points of view of their own. The model is based on a virtual machine that nests these entities to any depth required for analysis by nesting either type of object inside the first type (i.e. agents can have perspectives of entities and other agents). Specially, the nested beliefs are created only at need and not prestored in advance. The system was initially implemented in Prolog as a dialogue system which allows a doctor/system to interact separately with a couple who needed genetic counseling and could be supposed top have different levels of expertise on the subject matter. It was later reimplemented [Lee and Wilks, 1996] as part of an advisor on printing facilities, and embodied [Bontcheva, 2001] in a system that generated hypertext appropriate to a user's level of knowledge.

Various recent dialogue management frameworks have been following the collaborative approaches such as COLLAGEN, TRIPS, [Marsic and A. Flanagan, 2000], [Rothkrantz *et al.*, 2004], [Nguyen and Wobcke, 2005], [Cai *et al.*, 2005], [Blaylock, 2005a]. Some of these approaches will be presented in detail in section 5.

The advantages of the collaborative approaches are the ability to deal with more complex dialogues that involve collaborative problem solving, negotiation, and so on. But the approaches require much more complex resources and processing than the dialogue grammars and plan-based approaches.

[4] in the BDI model, actions in the world affect agent's beliefs and the agent can reason about its beliefs and thus formulate desires and intentions. Beliefs are how the agent perceives the world, desires are how the agent would like the world to be, and intentions are formulated plans of how to achieve these desires.

4.5 Summary

Which such a number of different approaches to dialogue management, it is reasonable to ask which approach is most appropriate for a particular application. Obviously the complexity of task model, dialogue model and domain-application increases from the finite state- to agent-based approaches. Conversational agents that incorporate principles of rationality and cooperation would seem to be the obvious choice, as they come closest to modeling human conversation competence. Certainly, for applications that involve co-operative problem solving with negotiated solutions, the simpler type of dialogue control are not sufficient. On the other hand, for simple applications and for constrained sub-tasks within some applications, more basic techniques such as finite state and frame-based control may be appropriate. The idea that the dialogue structure is often depend on the application domain, initially pushed people away from generic dialogue systems, dialogue grammars and plan-based approaches are reflect this direction.

5 Dialogue management in the recent systems

5.1 RDPM (Cooperative, Frame-based)

The general architecture of the dialogue system produced by the RDPM is represented in Fig. 2.

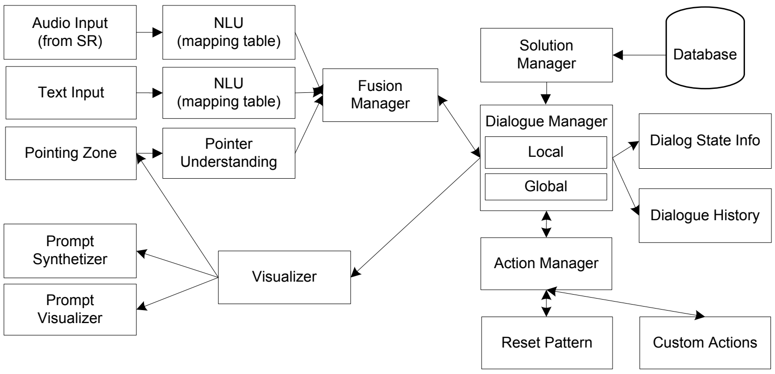

Figure 2: *Architecture of dialogue systems produced by RDPM.*

Three input modalities: voice, text and pointing can be used independently or simultaneously depending on the configuration of the current active mGDN [Lisowska *et al.*, 2004]. These inputs are pre-processed by the Natural Language Understanding (NLU) modules and the Pointer Understanding (PU) module. The outputs from NLU and PU modules are semantic triples (attribute, value, time-stamp). The fusion manager integrates the semantic triples receiving from the NLU and PU modules and sends a set of integrated semantic triples to the dialogue manager. In the current implemented version, the fusion manager simply collects the semantic triples based on their time-stamp relation and forwards them to the dialogue manager.

The dialogue manager encodes the local dialogue flow management strategy and global dialogue management strategy. Therefore, the input to the dialogue manager is first processed by the local dialogue management strategy in which we define five types of generic situations: *OK*, *Request for Repetition*, *Request for Help*, *NoInput*, and *NoMatch* [Bui *et al.*, 2004a].

In the case of the *OK* situation, control is handed back to the global dialogue manager which applies the global dialogue management strategy for the activation of the next mGDN. The dialogue state information (e.g. the current dialogue state, the active mGDN, etc.) and the recognized semantic triples are updated to the dialogue state info module and the dialogue history module respectively. When the dialogue manager gathers enough constraints [5], it sends the request to the action manager, the application connected with this module performs the task and sends the feedbacks to the action manager, the action manager then forwards these feedbacks to the dialogue manager. In addition, functions related with user modeling and system customization have been integrated such as *Reset Patterns* and *Custom Actions*. Reset Patterns allows the system to adapt to the behavior of a specific user or population of users by anticipating their next decisions. The idea is to develop an intelligent reset algorithm that estimates the most probable values for some mGDNs slots in a new dialogue session according to the previous interactions with the user. Custom Actions allows the users to dynamically associate sequences of solutions with a single new solution. The main goal of these two functions is to reduce the time to perform a task with the interface. The hypothesis is that these functions will indeed increase the quality of the interaction as perceived by the user. These two functions are described in detail in [Bui *et al.*, 2004b].

The outputs from the dialogue manager to the visualizer are multimedia prompts containing messages and a pointing zone update content. The messages are visualized in the user interface (Prompt Visualizer) and/or uttered by the mGDN during the interaction (Prompt Synthesizer). The messages are combined with the pointing zone update content (the content is a map, a calendar or a table depending on the nature of the mGDN) to allow the user to provide the desired values using keyboard, microphone or mouse click/touchscreen.

5.2 Smartkom (Cooperative, Information-State based & Plan-based)

Smartkom is a research project targeted to the development of multimodal task-oriented dialogue systems for two main domains: (1) information seeking (e.g. movie information and reservation, car and pedestrian navigation) and device control (e.g. telephony and biometric verification). It integrates 14 applications in three scenarios (public information booth, mobile information assistant, and home scenario). Four input modalities (speech, gesture, pen input, and facial expression) and three output modalities (speech, graphics, and gestures of the animated agent) are used in the system. The system is implemented using MULTIPLATFORM architecture [Herzog *et al.*, 2003]. XML schema-based messages are used to exchange data between sub-components in the systems. A common ontological knowledge representation is used for spoken utterances and other modalities, which facilitates disambiguation and resolution of cross-modal references [Löckelt, 2004].

Dialogue Manager in Smartkom has been developed based on a combination of the conversational games theory (cf. section 4.3) and information state based approaches (cf. section 4.2). In comparison with information state-based approaches, PRIVATE and SHARED information are split in two separate parts.

Dialogue Management in Smartkom tackles with two main challenges: a larger number

[5] this happens when the number of solutions (extracted from the solution manger) satisfying the current constraints is smaller than or equal to a pre-defined solution threshold.

of interacting modules as well as several concurrent applications and a set of nontrivial modalities. Modules connect directly with the dialogue manager are represented in Fig. 3. The most important connection is between the dialogue manager and the discourse modeler.

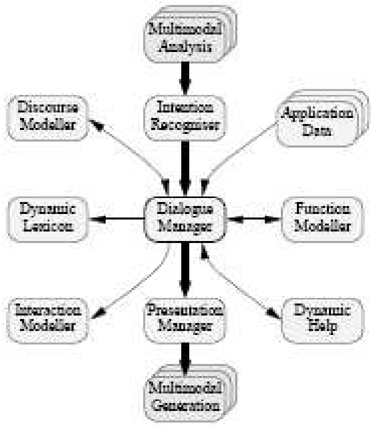

Figure 3: *Smartkom Dialogue Management Architecture.*

5.3 TRIPS (Collaborative, Agent-based)

TRIPS is the latest project in a series of prototype collaborative planning assistants developed at the University of Rochester. The project's goal is an intelligent planning assistant interacts with its human manager using a combination of natural language and graphical displays such as maps, charts, windows. The system and the human work in a collaborative way to solve harder problem faster than each member solves alone. The architecture of TRIPS is represented in Fig. 5. The physical of TRIPS is a hub and spoke message passing system, modules exchange the messages using KQML.

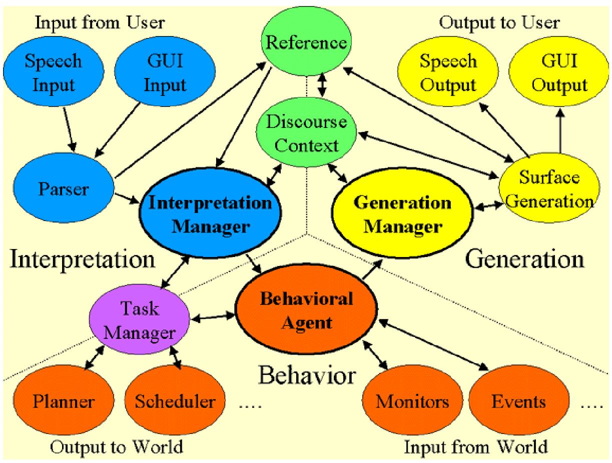

Figure 4: *TRIPS Architecture.*

There is no explicit dialogue manager module in the architecture. The dialogue management tasks are realized by three main modules: Interpretation Manager, Behavioral Agent, and Generation Manager.

5.4 COLLAGEN (Collaborative, Agent-based)

COLLAGEN (COLLaborative AGENt) is a Java middleware for building collaborative dialogue systems. This framework has been used to develop prototype systems for a range of applications such as air travel planning, email reading and responding, GUI design tool operation, car navigation system operation, airport landing path planning, gas turbine operator training, personal video recorder operation, programmable thermostat operation, multi-modal web-based form-filling.

The DM module (cf. Fig. 5.4) is developed based on the SharedPlan discourse theory [Grosz and Sidner, 1990] and the system has been implemented in Java.

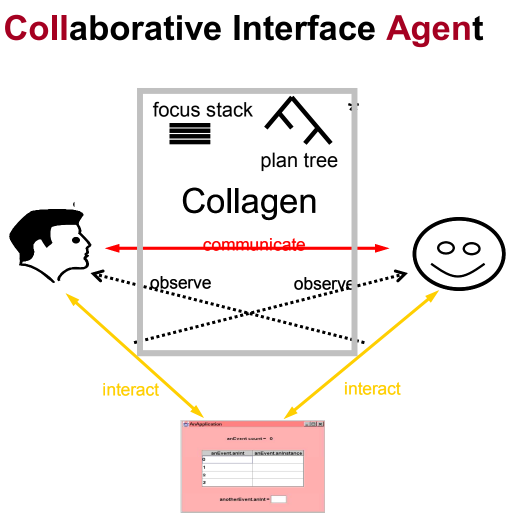

Figure 5: *COLLAGEN Architecture.*

6 Conclusions

We have presented the state of the art in the dialogue management research. In the ICIS domain, some of the applications are quite complex and require collaborative activities between human and machine as well as multi-modal, multi-party interaction. As mentioned, each approach has advantages and disadvantages points. The collaborative agent-based approach seams the appropriate choice, but it also takes a lot of efforts and time-consuming in the development of a workable dialogue systems. The complexity even increases in developing multimodal dialogue management systems for a number of reasons such as the user's multimodal input may be misinterpreted or misunderstood, ill-formed, incomplete [Sharma *et al.*, 2003].

We, therefore, are investigating a domain-independent dialogue management framework which is able to deal with complex, collaborative activities as well as allowing to facilitate the development of dialogue manager prototype.

Several issues required for this framework will be studied in detail: (1) methods for interpreting user's multimodal input in context, and (2) flexible and cooperative dialogue control strategies for dealing with dialogue phenomena such as repair, clarification, confirmation, etc.

7 Acknowledgements

We would like to thank Maarten van Someren and Léon Rothkrantz for their useful comments on the first version of this report. This work is part of the ICIS program (http://www.decis.nl/html/icis.html). ICIS is sponsored by the Dutch government under contract BSIK 03024.

References

[Allen and Perault, 1980] J.F. Allen and C.R. Perault. Analyzing Intentions in Dialogues. *Artificial Intelligence*, 15(3):143–178, 1980.

[Allen *et al.*, 2001] James Allen, Donna Byron, Myroslava Dzikovska, George Ferguson, Lucian Galescu, and Amanda Stent. Towards conversational human-computer interaction. *AI Magazine*, 2001.

[Allen, 1991] J.F. Allen. Discourse structure in the TRAINS project. In *Proceedings of the 4th DARPA Workshop on Speech and Natural Language*, 1991.

[Allwood, 1997] Jens Allwood. An Activity-Based Approach to Pragmatics. In David Traum, editor, *Working Notes: AAAI Fall Symposium on Communicative Action in Humans and Machines*, pages 6–11, Menlo Park, California, 1997. American Association for Artificial Intelligence.

[Appelt, 1985] D.E. Appelt. Planning English Sentences. *Cambridge, University Press*, 1985.

[Beun, 1996] R. J. Beun. Speech Act generation in cooperative dialogue. 1996.

[Blaylock, 2005a] Nate Blaylock. A collaborative problem-solving model of dialogue. In *SIGDIAL*, 2005.

[Blaylock, 2005b] Nate Blaylock. *Towards Tractable Agent-based Dialogue*. PhD thesis, University of Rochester, 2005.

[Bohus and Rudnicky, 2003] D. Bohus and A. Rudnicky. Ravenclaw: Dialog management using hierarchical task decomposition and an expectation agenda. In *Eurospech-2003*, Geneva, Switzerland, 2003.

[Bontcheva, 2001] K. Bontcheva. *Generation of Adaptive Hypertext*. PhD thesis, Department of Computer Science, University of Sheffield, 2001.

[Bratman *et al.*, 1991] Michael E. Bratman, David Israel, and Martha Pollack. *Plans and Resource-Bounded Practical Reasoning*. The MIT Press, Cambridge, Massachusetts, 1991.

[Bui and Rajman, 2004] Trung H. Bui and Martin Rajman. Rapid Dialogue Prototyping Methodology. Technical Report IC/2004/01, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland, January 2004.

[Bui *et al.*, 2004a] Trung H. Bui, Martin Rajman, and Miroslav Melichar. Rapid dialogue prototyping methodology. In *Proceedings of the 7th International Conference on Text, Speech & Dialogue (TSD 2004), P. Sojka, I. Kopecek & K. Pala (Eds.), Lecture Notes in Computer Science, Springer-Verlag, Berlin Heidelberg New York pp. 579-586*, Brno, Czech Republic, September 8-11 2004.

[Bui *et al.*, 2004b] Trung H. Bui, Martin Rajman, and Silvia Quarteroni. Extending the Rapid Dialogue Prototyping Methodology: User Modeling. Technical Report WP4 Deliverable 4.3, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland, August 2004.

[Bui *et al.*, 2005] Trung H. Bui, Job Zwiers, Anton Nijholt, and Mannes Poel. Generic dialogue modeling for multi-application dialogue systems. In *Proceedings of the 2nd Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms (MLMI05)*, Edinburgh, UK, July 2005.

[Bunt, 2000] H. Bunt. *Abduction, Belief and Context in Dialogue*, volume 1, chapter Dialogue pragmatics and context specification, pages 81–150. Benjamins, 2000.

[Cai *et al.*, 2005] Guoray Cai, Hongmei Wang, Alan M. MacEachren, and Sven Fuhrmann. Natural conversational interfaces to geospatial databases. *Transactions in GIS*, 9:199–221, 2005.

[Carletta *et al.*, 1995] J.H. Carletta, A. Isard, J. Kowtko, G. Doherty-Sneddon, and A. Anderson. The Coding of Dialogue Structure in a Corpus. *Andernach et al.*, 1995.

[Catizone *et al.*, 2002] R. Catizone, A. Setzer, and Y. Wilks. State of the Art in Dialogue Management. 2002.

[Chu-Carroll, 1996] J. Chu-Carroll. Response generation in collaborative dialogue interactions. *Luperfoy et al.*, 1996.

[Churcher *et al.*, 1997] Gavin E. Churcher, Eric S. Atwell, and Clive Souter. Dialogue management systems: a survey and overview. 1997.

[Clark and Schaefer, 1989] H. Clark and E. Schaefer. Contributing to Discourse. *Cognitive Science*, 13:259–294, 1989.

[Cohen and Levesque, 1990] P.R. Cohen and H.J. Levesque. Rational Interaction as the Basis for Communication. *In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, Intentions in Communication. M.I.T. Press, Cambridge, Massachusetts*, 1990.

[Cohen, 1997] P. Cohen. Dialogue Modeling in Survey the State of the art in Human Language Technology. 1997.

[Cole *et al.*, 1993] R. A. Cole, D.G. Novick, M. Fanty, S. Sutton, B. Hansen, and D. Burnett. Rapid prototyping of spoken language systems: The year 2000 census project. In *Proceedings of the International Symposium on Spoken Dialogue*, Tokyo, Japan, 1993.

[Cole, 1999] R. Cole. Tools for Research and Education in Speech Science. In *Proceedings of the International Conference of Phonetic Sciences*, San Francisco, USA, 1999.

[Constantinides *et al.*, 1998] P. Constantinides, S. Hansma, C. Tchou, and A. Rudnicky. A schema based approach to dialog control. In *Proceedings of the International Conference on Spoken Language Processing*, pages 409–412, Sidney, Australia, 1998.

[Denecke and Waibel, 1997] Matthias Denecke and Alex Waibel. Dialogue strategies guiding users to their communicative goals. In *Proc. Eurospeech '97*, pages 1339–1342, Rhodes, Greece, 1997.

[Denecke, 2002] M. Denecke. Rapid prototyping for spoken dialogue systems. In *Proceedings of the COLING '02*, August 2002.

[Eckert and McGlashan, 1993] W. Eckert and S. McGlashan. Managing Spoken Dialogues for Information Services. In *EUROSPEECH*, pages 1653–1656, Berlin, Germany, 1993.

[Foster, 2002] Mary Ellen Foster. State of the art review: Multimodal fission. 2002.

[Fraser, 1997] N. Fraser. Assessment of Interactive systems. *Handbook of Standards and Resources for Spoken Language Systems, D. Gibbon, R. Moore, and R. Winski, Eds. Mouton de Gruyter, New York, NY*, 1997.

[Goddeau *et al.*, 1996] D. Goddeau, H. Meng, J. Polifroni, S. Seneff, and S. Busayapongchai. A Form-Based Dialogue Manager for Spoken Language Applications. In *Proc. ICSLP '96*, volume 2, pages 701–704, Philadelphia, PA, 1996.

[Grosz and Sidner, 1990] B. J. Grosz and C. L. Sidner. Plans for discourse. 1990.

[Herzog *et al.*, 2003] G. Herzog, H. Kirchmann, S. Merten, A. Ndiaye, and P. Poller. MULTIPLATFORM Testbed: An Integration Platform for Multimodal Dialog Systems. In *Proceedings of the HLT-NAACL 2003 Workshop on Software Engineering and Architecture of Language Technology Systems (SEALTS)*, pages 75–82, Edmonton, Canada, 2003.

[Hofs *et al.*, 2003] D. Hofs, R. op den Akker, and A. Nijholt. A generic architecture and dialogue model for multimodal interaction. In *Proceedings 1st Nordic Symposium on Multimodal Communication*, 2003.

[Hulstijn *et al.*, 1996] H. Hulstijn, R. Steetskamp, H. Doest, S. Burgt, and A. Nijholt. Topics in schisma dialogues, 1996.

[Johnston *et al.*, 2002] Michael Johnston, Srinivas Bangalore, Gunaranjan Vasireddy, Amanda Stent, Patrick Ehlen, Marilyn A. Walker, Steve Whittaker, and Preetam Maloor. Match: An architecture for multimodal dialogue systems. In *ACL*, pages 376–383, 2002.

[Johnston, 1998] Michael Johnston. Unification-based multimodal parsing. In *Proceedings of the 36th annual meeting on Association for Computational Linguistics*, pages 624–630, Morristown, NJ, USA, 1998. Association for Computational Linguistics.

[Keizer, 2003] S. Keizer. *Reasoning under Uncertainty in Natural Language Dialogue using Bayesian Networks.* PhD thesis, University of Twente, Enschede, The Netherlands, September 2003.

[Kowtko *et al.*, 1991] J.C. Kowtko, S.D. Isard, and G.M. Doherty. Conversational Games within Dialogue. In *Proceedings of the ESPRIT Workshop on Discourse Coherence, University of Edinburgh*, 1991.

[Larsson and Traum, 2000] Staffan Larsson and David Traum. Information state and dialogue management in the TRINDI Dialogue Move Engine Toolkit. *In Natural Language Engineering Special Issue on Best Practice in Spoken Language Dialogue Systems Engineering, Cambridge University Press, U.K*, pages 323–340, 2000.

[Larsson *et al.*, 2000] Staffan Larsson, Peter Ljunglöf, Robin Cooper, Elisabet Engdahl, and Stina Ericsson. GoDiS - An Accommodating Dialogue System. In *Proceedings of ANLP/NAACL-2000 Workshop on Conversational Systems*, pages 7–10, 2000.

[Larsson, 2002] Staffan Larsson. *Issued-based Dialogue Management*. PhD thesis, Department of Linguistics, Göteborg University, Sweden, June 2002.

[Lecoeuche, 2001] Renaud Lecoeuche. Learning optimal dialogue management rules by using reinforcement learning and inductive logic programming, 2001.

[Lee and Wilks, 1996] M. Lee and Y. Wilks. An ascription-based approach to speech acts. In *Proceedings of the 16th Conference on Computational Linguistics (COLING-96)*, Copenhagen, 1996.

[Levin and Pieraccini, 1997] E. Levin and R. Pieraccini. A stochastic model of human-machine interaction for learning dialogue strategies. *Eurospeech*, 1997.

[Levin *et al.*, 2000] E. Levin, R. Pieraccini, and W. Eckert. A stochastic model of human-machine interaction for learning dialogue strategies. *IEEE Transactions on Speech and Audio Processing*, 8(1):11–23, january 2000.

[Lisowska *et al.*, 2004] Agnes Lisowska, Martin Rajman, and Trung H. Bui. Archivus: A system for accessing the content of recorded multimodal meetings. In Samy Bengio and Hervé Bourlard, editors, *Machine Learning for Multimodal Interaction, First International Workshop,MLMI 2004, Martigny, Switzerland, June 21-23, 2004, Revised Selected Papers*, volume 3361, pages 291 – 304. Springer, 2004.

[Löckelt, 2004] Markus Löckelt. Dialogue management in the smartkom system. *Kovens 2004*, pages 125–133, 2004.

[Luz, 1999] Saturnino Luz. State-of-the-art survey of dialogue management tools. 1999.

[Marsic and A. Flanagan, 2000] I. Medl Marsic and J. A. Flanagan. Natural communication with information systems. In *Proceedings of the IEEE*, volume 88, pages 1354–1366, August 2000.

[McGlashan *et al.*, 1992] S. McGlashan, N.M. Fraser, G.N. Gilbert, E. Bilange, P. Heisterkamp, and N.J. Youd. Dialogue Management for Telephone Information Systems. In *Proceedings of the International Conference on Applied Language Processing*, Trento, Italy, 1992.

[McGlashan, 1996] S. McGlashan. Towards multimodal dialogue management. In *Proceedings of Twente Workshop on Language Technology 11*, Enschede, The Netherlands, 1996.

[McTear, 1998] M. McTear. Modelling spoken dialogues with state transition diagrams: experiences with the cslu toolkit. In *Proc 5th International Conference on Spoken Language Processing*, pages 1223–1226, Sydney, Australia, Dec 1998.

[McTear, 2002] M. McTear. Spoken dialogue technology: Enabling the conversational user interface. *ACM Computing Survey, Volume 34, No 1*, 2002.

[Nguyen and Wobcke, 2005] Anh Nguyen and Wayne Wobcke. An agent-based approach to dialogue management in personal assistants. In *IUI '05: Proceedings of the 10th international conference on Intelligent user interfaces*, pages 137–144, New York, NY, USA, 2005. ACM Press.

[Novick and Hansen, 1995] D. G. Novick and B. Hansen. Mutuality strategies for reference in task-orientated dialogue. *Andernach et al. (eds.)*, 1995.

[Novick and Ward, 1993] D. G. Novick and K. Ward. Mutual Beliefs of Multiple Conversants: A computational model of collaboration in Air Traffic Control. In *Proceedings of AAAI93*, 1993.

[Oviatt, 2002] Sharon Oviatt. Multimodal interfaces. 2002.

[Pavel Cenek and Rajman, 2005] Miroslav Melichar Pavel Cenek and Martin Rajman. A framework for rapid multimodal application design. In *Proceedings of the 8th International Conference on Text, Speech Dialogue (TSD 2005)*, 2005.

[Rothkrantz *et al.*, 2000] Léon J. M. Rothkrantz, Robert J. van Vark, Alexandra Peters, and Niels A. Andeweg. Dialogue control in the alparon system. In Petr Sojka, Ivan Kopecek, and Karel Pala, editors, *TSD*, volume 1902 of *Lecture Notes in Computer Science*, pages 333–338. Springer, 2000.

[Rothkrantz *et al.*, 2004] Léon J. M. Rothkrantz, Pascal Wiggers, Frans Flippo, Dimitri Woei-A-Jin, and Robert J. van Vark. Multimodal dialogue management. In Petr Sojka, Ivan Kopecek, and Karel Pala, editors, *TSD*, volume 3206 of *Lecture Notes in Computer Science*, pages 621–628. Springer, 2004.

[Roy *et al.*, 2000] N. Roy, J. Pineau, and S. Thrun. Spoken dialogue management using probabilistic reasoning. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL-200)*, 2000.

[Rudnicky *et al.*, 1999] A. Rudnicky, E. Thayer, P. Constantinides, C. Tchou, R. Shern, K. Lenzo, W. Xu, and A. Oh. Creating natural dialogs in the Carnegie Mellon Communicator system. In *Proceedings of Eurospeech, volume 4*, pages 1531–1534, 1999.

[Scheffler and Young, 2002] K. Scheffler and S. J. Young. Automatic learning of dialogue strategy using dialogue simulation and reinforcement learning. In *Proc. Human Language Technology*, pages 12–18, 2002.

[Searle, 1969] J.R. Searle. Speech acts: An essay in the philosophy of language. *Cambridge, University Press*, 1969.

[Sharma *et al.*, 2003] R. Sharma, M. Yeasin, N. Krahnstoever, I. Rauschert, G. Cai I. Brewer, A. MacEachren, and K. Sengupta. Speech-gesture driven multimodal interfaces for crisis management. *PROCEEDINGS OF THE IEEE,*, 91(9):28, September 2003.

[Sidner, 1992] C. L. Sidner. Using discourse to negotiate in collaborative activity: an artificial language. pages 121–128, 1992.

[Sidner, 1994] C. L. Sidner. An artificial discourse language for collaborative negotiation. In *Proceedings of the AAAI*, pages 814–819, 1994.

[Singh *et al.*, 2002] S. Singh, D. Litman, and M. Kearns. Optimizing dialogue management with reinforcement learning: Experiments with the njfun system. *Journal of Artificial Intelligence Research*, 2002.

[Traum and Hinkelman, 1992] D. R. Traum and E. Hinkelman. Conversation acts in task-oriented spoken dialogue. 1992.

[Traum and Larsson, 2003] David Traum and Staffan Larsson. The information state approach to dialogue management. pages 325–353, 2003.

[Traum and Rickel, 2002] David R. Traum and Jeff Rickel. Embodied agents for multi-party dialogue in immersive virtual worlds. In *AAMAS*, pages 766–773. ACM, 2002.

[Traum, 1991] D. R. Traum. Towards a computational theory of grounding in natural language conversation. 1991.

[Traum, 1996] D. R. Traum. Conversational agency: The trains-93 dialogue manager. In *Proceedings of the Twente Workshop on Langauge Technology: Dialogue Management in Natural Language Systems (TWLT 11)*, pages 1–11, 1996.

[van Zanten, 1996] Gert Veldhuijzen van Zanten. Pragmatic interpretation and dialogue management in spoken-language systems. In *Proceedings of the 11th Twente Workshop on Language Technology, TWLT 11*, pages 81–88, June 1996.

[Vo and Wood, 1996] Minh Tue Vo and Cindy Wood. Building an application framework for speech and pen input integration in multimodal learning interfaces. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 1996.

[Wai *et al.*, 2001] Carmen Wai, Helen M. Meng, and Roberto Pieraccini. Scalability and portability of a belief network-based dialog model for different application domains. In *Proceedings of the 2001 Human Language Technology Conference (HLT-2001)*, 2001.

[Walker *et al.*, 1998] Marilyn A. Walker, Jeanne Frommer, and Shrikanth Narayanan. Learning optimal dialogue strategies: A case studcolorado dialoguy of a spoken dialogue agent for email. In *COLING-ACL*, pages 1345–1351, 1998.

[Wilks and Ballim, 1991] Y. Wilks and A. Ballim. Beliefs, stereotypes and dynamic agent modeling. 1991.

[Williams *et al.*, 2005] Jackson D. Williams, Pascal Poupart, and Steve Young. Partially Observable Markov Decision Processes with Continuous Observations for Dialogue Management. In *Proceedings of the 6th SigDial Workshop on Discourse and Dialogue*, Sept 2005.

[Williams, 1996] S. Williams. Dialogue management in a mixed-initiative, cooperative, spoken language system. 1996.

[Wright *et al.*, 1998] Jerry Wright, Allen Gorin, and Alicia Abella. Spoken language understanding within dialogs using a graphic model of task structure. *ICSLP98*, 1998.

[Wu *et al.*, 2002] L. Wu, S. Oviatt, and P. Cohen. From members to teams to committee - a robust approach to gestural and multimodal recognition. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, 13(4):11, july 2002.

[Xu and Rudnicky, 2000] W. Xu and A. Rudnicky. Task-based dialogue management using an agenda. In *Proc. ANLP/NAACL 2000 Workshop on Conversational Systems*, Seattle, WA, 2000.

[Xu *et al.*, 2002] Weiqun Xu, Bo Xu, Taiyi Huang, and Hairong Xia. Bridging the gap between dialogue management and dialogue models. *revue Traitement Automatique des Langues (TAL), 41(2), Paris*, 2002.

[Young, 1999] S. Young. Probabilistic methods in spoken dialogue systems. In *Proceedings of the Royal Society*, London, sept 1999.

[Zhang *et al.*, 2001] Bo Zhang, Qingsheng Cai, Jianfeng Mao, and Baining Guo. Spoken dialog management as planning and acting under uncertainty. In *Proceedings of Eurospeech*, 2001.