Technologies for Inclusive Education:

Beyond Traditional Integration Approaches

David Griol Barres
Carlos III University of Madrid, Spain

Zoraida Callejas Carrión
University of Granada, Spain

Ramón López-Cózar Delgado
University of Granada, Spain

A volume in the Advances in Educational Technologies and Instructional Design (AETID) Book Series

Information Science REFERENCE

Managing Director: Lindsay Johnston
Editorial Director: Joel Gamon
Book Production Manager: Jennifer Romanchak
Publishing Systems Analyst: Adrienne Freeland
Development Editor: Myla Merkel
Assistant Acquisitions Editor: Kayla Wolfe
Typesetter: Deanna Jo Zombro
Cover Design: Nick Newcomer

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Technologies for inclusive education : beyond traditional integration approaches / David Griol Barres, Zoraida Callejas Carrion and Ramon Lopez-Cozar Delgado, Editors
p. cm.
Includes bibliographical references and index.
Summary: “This book introduces the basic concepts, current research guidelines and future perspectives on the current state of inclusive education by highlighting technological advances in applied e-learning, cognitive learning and education multimedia”--Provided by publisher.
ISBN 978-1-4666-2530-3 (hardcover) -- ISBN 978-1-4666-2531-0 (ebook) -- ISBN 978-1-4666-2532-7 (print & perpetual access) 1. Inclusive education. 2. Special education. 3. Educational technology. I. Griol Barres, David, 1976- II. Callejas Carrion, Zoraida, 1982- III. Lspez-Cszar Delgado, Ramsn.
LC1201.T46 2013
371.33--dc23
2012026061

This book is published in the IGI Global book series Advances in Educational Technologies and Instructional Design (AE-TID) Book Series (ISSN: 2326-8905; eISSN: 2326-8913)

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter 1
Towards the Use of Dialog Systems to Facilitate Inclusive Education

David Griol Barres
Carlos III University of Madrid, Spain

José M. Molina López
Carlos III University of Madrid, Spain

Zoraida Callejas Carrión
University of Granada, Spain

Araceli Sanchis de Miguel
Carlos III University of Madrid, Spain

ABSTRACT

Continuous advances in the development of information technologies have currently led to the possibility of accessing learning contents from anywhere, at anytime, and almost instantaneously. However, accessibility is not always the main objective in the design of educative applications, specifically to facilitate their adoption by disabled people. Different technologies have recently emerged to foster the accessibility of computers and new mobile devices, favoring a more natural communication between the student and the developed educative systems. This chapter describes innovative uses of multimodal dialog systems in education, with special emphasis in the advantages that they provide for creating inclusive applications and learning activities.

1. INTRODUCTION

Technological advances currently reached by computers and mobile devices allow their use to access information and a number of services. In addition, users want to access these services anywhere and anytime in a natural, intuitive and efficient way. Speech-based interfaces have become one of the main options to facilitate this kind of communication as it is a good solution to the shrinking size of mobile devices, eases the communication in environments where this access is not possible using traditional input interfaces (e.g., keyboard and mouse), and facilitates information access for people with visual or motor disabilities.

With the advances of speech, image and video technology, human-computer interaction (HCI) has reached a new phase, in which multimodal

DOI: 10.4018/978-1-4666-2530-3.ch001

information is a key point to enhance the communication between humans and machines. Unlike traditional keyboard- and mouse-based interfaces, multimodal interfaces enable greater flexibility in the input and output, as they permit users to employ different input modalities as well as to obtain responses through different means, for example, speech, gestures, and facial expressions. This is especially important for users with special needs, for whom the traditional interfaces might not be suitable (McTear, 2004; López-Cózar and Araki, 2005; Wahlster, 2006).

In addition, the widespread use of mobile technology implementing wireless communications enables a new type of advanced applications to access information. As a result, users can effectively access huge amounts of information and services from almost everywhere and through different communication modalities.

There is a large variety of applications in which spoken dialog systems can be used. One of the most wide-spread is providing information on a specific topic, such as flight/railway and booking information, tourist and travel information, weather forecast, banking systems, or conference help (Glass et al., 1995; Zue et al., 2000; Bohus and Rudnicky, 2005; Andeani et al., 2006; Callejas and López-Cózar, 2008). In some cases, spoken interaction can be the only way to access information, as, for example when the screen is too small to display information (e.g. hand-held devices) or when the eyes of the user are busy in other tasks (e.g. driving) (Mattasoni et al., 2002; Jokinen et al., 2004; Weng et al., 2006). Spoken interaction is also useful for remote control devices and robots, especially in smart environments (Lemon et al., 2001; Montoro et al., 2006; Ábalos et al., 2006; Menezes et al., 2007; Augusto, 2009). Finally, one of the most demanding applications for fully natural and understandable dialogs are virtual agents and companions (Hubal et al., 2000; Catizone et al., 2003; Corradini et al., 2005).

With the growing maturity of speech technologies, the possibilities for integrating conversation and discourse in e-learning are receiving greater attention, including tutoring, question-answering, conversation practice for language learners, pedagogical agents and learning companions, and dialogs to promote reflection and metacognitive skills. This chapter focuses on some of the most important challenges that researchers have recently envisioned for future multimodal interfaces applied to educative purposes. It describes current efforts to develop intelligent, adaptive, proactive, portable and affective multimodal interfaces. All these concepts are not mutually exclusive, for example, the system's intelligence can be concerned with the system's adaptation enabling better portability to different environments.

To deal with all these important topics required for the design of educative multimodal interfaces, this chapter is organized as follows. Section 2 provides an overview of the main modules and functionalities required for the development of spoken dialog systems. Section 3 describes the main principles involved in the development of educative multimodal interfaces. This section also provides important examples showing the benefits of the integration of this kind of systems in educative applications. Section 4 describes our work related to the development of educative multimodal interfaces describing two systems developed to respectively facilitate the access to Internet and learn foreign languages. Finally, Section 5 presents the conclusions and outlines possibilities for future research directions.

2. DIALOG SYSTEMS: MODULAR ARCHITECTURE AND PROCESSES

The complexity of the interaction between the user and the dialog system can vary and some of the previously described components might not be used. For example, for a simple menu, semantic analysis is not necessary. However, for a conversational companion all the modules must be used in order to interpret the user input, take justified decisions on what the system will respond, and finally tailor the answer to user needs and expectations. This way, the implementation of multimodal dialog systems is a complex task in which a number of technologies are involved, including signal processing, phonetics, linguistics, natural language processing, affective computing, graphics and interface design, animation techniques, telecommunications, sociology and psychology. The complexity is usually addressed by dividing the implementation into simpler problems, each associated with a system's module that carries out specific functions. Usually, this division is based on the traditional architecture of spoken dialog systems: automatic speech recognition (ASR), spoken language understanding (SLU), dialog management (DM), natural language generation (NLG) and text-to-speech synthesis (TTS).

Speech recognition is the process of obtaining a sentence (text string) from a voice signal (Rabiner et al., 1996). It is a very complex task given the diversity of factors that can affect the input, basically concerned with the speaker, the interaction context and the transmission channel. Different applications demand different complexity on the speech recognizer. Cole et al. (1997) identified eight parameters that allow an optimal tailoring of the recognizer: speech mode, speech style, dependency, vocabulary, language model, perplexity, signal-to-noise ratio (SNR) and transduction. Nowadays, general-purpose ASR systems are usually based on Hidden Markov Models (HMMs) (Rabiner and Juang, 1993).

Spoken Language Understanding is the process of extracting the semantics from a text string (Minker, 1998). It generally involves employing morphological, lexical, syntactical, semantic, discourse and pragmatic knowledge. In a first stage, lexical and morphological knowledge allow dividing the words in their constituents distinguishing lexemes and morphemes. Syntactic analysis yields a hierarchical structure of the sentences, whereas the semantic analysis extracts the meaning of a complex syntactic structure from the meaning of its constituents. There are currently two major approaches to carry out SLU: rule-based (Mairesse et al., 2009) and statistical (Meza et al., 2008), including some hybrid methods (Liu et al., 2006).

Dialog Management is concerned with deciding the next action to be carried out by the dialog system. The simplest dialog model is implemented as a finite-state machine, in which machine states represent dialog states and the transitions between states are determined by the user's actions. Frame-based approaches have been developed to overcome the lack of flexibility of the state-based dialog models, and are used in most current commercial systems. For complex application domains, plan-based dialog models can be used. They rely on the fact that humans communicate to achieve goals, and during the interaction, the humans' mental state might change (Chu et al., 2005). Currently, the application of machine-learning approaches to model dialog strategies is a very active research area (Griol et al., 2008; Williams and Young, 2007; Cuayáhuitl et al., 2006; Lemon et al., 2006).

Natural language generation is the process of obtaining texts in natural language from a non-linguistic representation of information. It is usually carried out in five steps: content organization, content distribution in sentences, lexicalization, generation of referential expressions and linguistic realization. The simplest approach uses predefined text messages (e.g. error messages and warnings). Although intuitive, this approach is very inflexible (Reiter, 1995). The next level of sophistication is template-based generation, in which the same message structure can be produced with slight differences. This approach is used mainly for multi-sentence generation, particularly in applications where texts are fairly regular in structure, such as business reports (Reiter, 1995). Phrase-based systems employ what can be considered generalized templates at the sentence level (in which case the phrases resemble phrase structure grammar rules), or at the discourse level (in which case they are often called text plans) (Elhadad and

Robin, 1996). Finally, in feature-based systems, each possible minimal alternative of expression is represented by a single feature to obtain the maximum level of generalization and flexibility (Oh and Rudnicky, 2000).

Text-to-speech synthesizers transform text strings into acoustic signals. A TTS system is composed of two parts: front-end and back-end. The front-end carries out two major tasks. Firstly, it converts text strings containing symbols such as numbers and abbreviations into their equivalent words. This process is often called text normalization, pre-processing or tokenization. Secondly, it assigns a phonetic transcription to each word, which requires dividing and marking the text into prosodic units, i.e. phrases, clauses, and sentences. The back-end (often referred to as the synthesizer) converts the words in text format into sound. Concatenative synthesis employs pre-recorded units of human voice that are put together to obtain words. It generally produces the most natural synthesized speech; however, differences between variations in speech and in the nature of the automated techniques for segmenting the waveforms sometimes result in audible glitches.

3. SPOKEN DIALOG SYSTEMS AND EDUCATION

With the growing maturity of conversational technologies, the possibilities for integrating conversation and discourse in educative applications are receiving greater attention, including tutoring (Pon-Barry et al., 2006), question-answering (Wang et al., 2007), conversation practice for language learners (Fryer et al., 2006), pedagogical agents and learning companions (Cavazza et al., 2010), and dialogs to promote reflection and metacognitive skills (Kerly et al., 2008).

The design, implementation and strategies of dialog systems employed in e-learning applications vary widely, reflecting the diverse nature of the evolving speech technologies. The conversations are generally mediated through simple text based forms (Heffernan, 2002), with users typing responses and questions at a keyboard. Some systems use embodied dialog systems (Graesser et al., 2001) capable of displaying emotion and gesture, whereas others employ a simpler avatar (Kerly et al., 2008b). Speech output, using text to speech synthesis is used in some systems (Graesser et al., 2001), and speech input systems are increasingly viable (Litman and Silliman, 2004) (Bos et al., 2002).

According to Roda et al. (Roda et al., 2001), enhanced e-learning systems are expected to i) accelerate the learning process, ii) facilitate access, iii) personalize the learning process, and iv) supply a richer learning environment. In addition, three main emerging approaches are described to integrate dialog systems in learning environments at the individual as well as at the group level:

1. Advanced help and learning process facilitation tools;
2. Personal coaches equipped with specific domain knowledge;
3. Role-playing actors in simulated experiential learning environments.

Dialog systems as advanced help and learning process facilitation tools are designed to provide an advanced helping service. To do this, they integrate structured knowledge models about the application domain and the environment. Although this kind of systems can be very useful, they are usually considered to be annoying and not intelligent. Their main limitations are due the inability to contextualize the users' actions within the set of possible uses sequences of the computer application. This way, the dialog system does not have any knowledge related to the user and cannot be adapted according to their preferences and motivations. In addition, they do not usually include dialog functionalities and can only be used to solve isolated questions, providing information already present in the corresponding manual of the application. This way, these systems do not incorporate any learning model. Examples of this kind of systems are the Microsoft Conversational Agents integrated in the Office desktop applications, as well as similar agents that are available in database applications.

Dialog systems as personal coaches integrate information about the domain of the application. Systems of this kind are characterized by the possibility to represent and continuously update information that represents the cognitive and social users' state. The main objective is to guide and manage users in the learning process, providing suggestions and other interaction functionalities not only with the developed application but also with the rest of students. To do this, these applications usually integrate realistic and interactive interfaces.

K-Inca is an artificial conversational agent designed to help people to adopt knowledge management practices (Angehrn et. al 2001) (Roda et al., 2001). Users' profiles have been included to represent their similarity with a set of predefined hierarchic behavior profiles, providing a personalized helping service that takes into account tutoring (e.g., providing specific exercises), suggestions and stimuli that modify the way in which the system interacts with their users.

Dialog systems as role-playing actors in simulated experiential learning environments are systems which are able to carry out a specific function in a very realistic way inside a simulated environment that emulates the real learning environment. These systems integrate knowledge about this environment or domain application (tasks, behaviors, objects, relationships, etc.) and are able to maintain a dialog updating and adapting this knowledge by considering users' social and cognitive state.

The Change VIBE (C-VIBE) system (Angehrn, 2001) has been developed to interact within the EIS simulated environment (Manzoni and Angehrn, 1997), currently used in leading schools and universities to train managers in the theory and practice of managing change and organizational transformation facing the natural resistance to innovation and change latent in organizations. Students interact in this platform by means of avatars and with the main objective of managing (individually or in group) a specific mission that is proposed in a Virtual Board Room.

3.1 Tutoring Applications

Tutoring is one of the most substantially research areas for the use of natural language dialog in e-learning. In educational domains, Kumar et. al. (20011) have shown that agents playing the role of a tutor in a collaborative learning environment can lead to over one grade improvement. Additional works (Liu and Chee, 2004) have explored a variety of interaction patterns and tactics that could be used in multi-party educational situations.

Most of the existing research on interaction strategies for dialog systems used in various interactive settings has focused on task-related strategies. In the case of conversational tutors, the task (or work) related interaction include aspects like instructing students about the task, delivering appropriate interventions in suitable form (e.g. socratic dialog, hints), providing feedback and other such tactics (Graesser et. al., 2005). Some studies (Rosé et. al., 2001b; Wang and Johnson, 2008) have evaluated the effect of these task related conversational behavior in tutorial dialog scenarios. Work in the area of affective computing and its application to tutorial dialog has focused on identification of student's emotional states (D'Mello et. al., 2008) and using those to improve choice of task related behavior by tutors.

The AutoTutor project (Graesser et al., 1999) provides tutorial dialogs on subjects including university level computer literacy and physics. The tutoring tactics employed by this system assist students in actively constructing knowledge, and are based on extensive analysis of naturalistic tutoring sessions by human tutors. AutoTutor includes the use of a dialog manager, curriculum scripts and latent semantic analysis. This system was demonstrated to give an important improvement when compared to control conditions for gains in learning and memory.

Another tutoring system employing dialog is Ms Lindquist (Heffernan, 2003), which offers “coached practice” to high school students in algebra by scaffolding “learning by doing” rather than offering explicit instruction. Early work with the system found that students using Ms Lindquist did fewer problems, but that they learned equally well or better than students who were simply told the answer. The results also suggested that the dialog was beneficial in maintaining student motivation. The authors concluded that Ms Lindquist was a “Less is More” approach, where learners tackled fewer problems, but learnt more per problem when they were engaged in an intelligent dialog (Heffernan, 2003).

CycleTalk (Forbus et. al., 1999) is an intelligent tutoring system that helps students to learn principles of thermodynamic cycles in the context of a power plant design task. Teams of two students work on designing a Rankine cycle using a Thermodynamics simulation software package. As a part of the design lab during which this learning task is performed, students participated in a collaborative design interaction for 30-45 minutes using ConcertChat, a text based collaboration environment (Mühlpfordt and Wessner, 2005). ITSPOKE is a tutoring spoken dialog system which engages the students in a spoken dialog to provide feedback and correct misconceptions (Litman and Silliman, 2004). It is speech-based dialog system that uses a text-based system for tutoring conceptual physics. A list with additional projects developed at the University of Pittsburgh can be found at http://www.cs.pitt.edu/~litman/itspoke.html.

Another example of natural language tutoring is the Geometry Explanation Tutor (Aleven et al., 2004), where students explain their answers to geometry problems in their own words. The system uses a knowledge-based approach to recognize explanations as correct or partially correct, and a statistical text classifier when the knowledge-based method fails. Studies with this system found that students who explain in a dialog learn better to provide general explanations for problem-solving steps (in terms of geometry theorems and definitions) than those who explain by means of a menu

The Oscar conversational intelligent tutoring system (CITS) (Latham et al. 2012) aims to mimic a human tutor by implicitly modeling the learning style during tutoring, personalizing the tutorial to boost confidence and improving the effectiveness of the learning experience. The system uses natural language to provide communication about specific topics with the users and dynamically predicts and adapts to a student’s learning style. It is implemented using the Index of Learning Styles (ILS) model (Felder and Silverman, 1988) to deliver an SQL tutorial. The results of an evaluation carried out with real students that all learning styles in the ILS model were successfully predicted from a natural language tutoring conversation, with an accuracy of 61–100%.

An educational dialog system to support e-learning in the subject of geometry is described in (Kim, 2007). Knowledge in the system was created and represented by XML-based AIML (Artificial Intelligence Markup language). The system can answer the student’s questions by referring and saving the previous knowledge while having a conversation with a student. To do this, context information related to the student is considered using an overlay student model. An educational dialog system was evaluated to test the efficiency of the designed and implemented system with geometry learning.

An educational environment developed for a modular spoken dialog system is described in (Gustafson et al.,1998). The aim of the environment is to provide students, with different backgrounds, means to understand the behavior of spoken dialog systems. Dialog is recorded in a dialog tree whose nodes are dialog objects, which model the constituents of the dialog and consist of parameters for modeling dialog structure, focus structure and a process description describing the actions of the dialog system. The educational system has been used in a number of courses at various universities in Sweden.

Other question-answering systems have included a student discussion board (Feng et al., 2006) where the dialog system mines a corpus to retrieve an answer based on cosine similarities between the query post and the corpus passages, and the Intelligent Verilog Compiler Project (Taylor and Moore, 2006), which allows learners to ask questions in English that query the same ontology as is used to provide the system's 'help' texts. This style of use most closely mirrors the most common use in commercial environments where dialog systems are used for information retrieval.

3.2 Learning Companions and Embodied Dialog Systems

Developing more human-like systems seems to improve interaction by establishing a more engaging relation with this kind of systems (Dehn and van Mulken, 2000). Learning companions are simulated characters that act as a mate of the student, and take a non-authoritative role in a social learning environment (Chou et al., 2003). This way, a number of Embodied Dialog systems (ECA) have been developed to assist students during the learning process. According to research and evaluation studies in the field of intelligent interfaces, ECAs (Gratch et al., 2002) have shown to be a good interaction metaphor when acting in the role of counselors (Marsella et al., 2003), personal trainers (Bickmore, 2003), or healthy living advisors (de Rosis et al., 2003). Indeed, ECAs have the potential to involve users in a human-like conversation using verbal and non-verbal signals for providing feedback, showing empathy and emotions in their behavior (Cassell, 2001). Due to these features, ECAs can be successfully employed as interaction metaphor in the pedagogical domain (Johnson et al., 2004) and in other domains where it is important to settle long-term relations with the users (Bickmore and Picard, 2005).

These agents, which may employ gesture, synthesized speech and emotional facial displays, have been investigated in domains ranging from helping children to learn plant biology (Lester et al., 1999) to continuing medical education (Shaw et al., 1999) and naval training (Rickel and Johnson, 1999). Research into the roles which may be played by a pedagogical agent or learning companion has investigated agents as mentors, peers, experts or instructors (Baylor, and Kim, 2005). In some systems the student must teach the agent (Chan and Baskin, 1998), or interact with peer agents or co-learners (Dillenbourg and Self, 1992), who may even include trouble makers intended to provoke cognitive dissonance to prompt learning (Aimeur et al., 1997). Researchers have also investigated user preferences for the agent expertise. Findings suggest that in general, similarities in competency between an agent and learner have positive impacts on the learners' affective attainments, for example, academically strong students showed higher self-efficacy beliefs in a task after working with a high-competency agent, while academically weak students showed higher self-efficacy after working with a low-competency agent (Kim, 2007).

Moreover, even if ECAs have shown to have a good impact on settling an emphatic relation with the user (de Rosis et al., 2005; Cassell et al. 2000) (Ai et al., 2006) (Bailly et al., 2010) (Edlund et al., 2008), involving them in a deeper and intimate interaction, it is difficult to communicate with these agents whenever needed (i.e. when the user is not in front of a computer but he/she has the need to get suggestions and advices). For example, DESIA (Johnson et al., 2004) is a step in this direction. This agent, presented in Carmen's Bright IDEAS, has been adapted for running successfully on a handheld device in order to assist in a psychosocial intervention for acquiring problem solving skills (Marsella et al., 2003).

Another example is the VU-MAS architecture, a Virtual University Multi-Agent System (MAS), is described in (De Carolis et al., 2006). Each student can interact with VU-MAS using a personal agent, called MyCoach, represented as an ECA. The main goal of this agent is to monitor the student activities, following his/her learning improvements, but also to select useful material according to the recognized student's goals and needs. The agent is also capable to proactively provide the student with useful suggestions whenever it is needed. As it is designed to run on a smart phone or a PDA, this agent combines e-learning capabilities with mobile computing, thus realizing an m-learning experience where the student can feel always in touch with his advisor.

3.3 Other Applications

There may be possibilities to integrate dialog systems in Learning 2.0 communities, as assistants, moderators, guides or as virtual peers within the community. Dialog and anthropomorphic characteristics of pedagogical and dialog systems may help support the social dimension of e-learning activities, and the social context has been argued to catalyze the cultivation and motivation for knowledge (Chou et al., 2003).

For example, dialog systems have been proposed to offer conversation practice for language learners. Jia (Jia, 2002) found that users were dissatisfied with the responses provided by a basic ALICEbot (www.alicebot.org) implementation, and the pattern-matching mechanism was deemed insufficient for use as a foreign language practice environment. In contrast, Jabberwacky (Fryer and Carpenter, 2006) (www.jabberwacky.com) uses a very different technology to ALICEbots, learning from all its previous conversations. It has been suggested for providing language practice; Fryer and Carpenter note that agents are willing to repeat the same material as often as students require. They also argue that chatbots give students the opportunity to use varying language structures and vocabulary (for example slang and taboo words), which they may otherwise get little chance to experience (Fryer and Carpenter, 2006).

Spoken dialog systems have also been proposed to improve phonetic and linguistic skills. Vocaliza is a dialog application for computer-aided speech therapy in the Spanish language, which helps in the daily work of speech therapists who teach linguistic skills to Spanish speakers with different language pathologies (Vaquero et al., 2006). The Listem system (Literacy Innovation that Speech Technology Enables) is an automated Reading Tutor that displays stories on a computer screen, and listens to children read aloud (Mostow, 2012).

Finally, dialog is also used as a prompt for reflection. Grigoriadou et al. describe a system where the learner reads a text about a historical event before stating their position about the significance of an issue and their justification of this opinion (Grigoriadou et al., 2003). Answers are classified as scientific, towards-scientific or non-scientific, and a dialog generator produces "appropriate reflective diagnostic and learning dialog for the learner". CALMsystem (Kerly et al., 2008) promotes reflection of a different kind. Users answer questions on the domain, and state their confidence in their ability to answer correctly. The system infers a knowledge level for the student based on their answers, and encourages the learner to engage in a dialog to reflect on their self-assessment and any differences between their belief and that of the system about their knowledge levels. Studies have shown this dialog improved self-assessment accuracy significantly more than reflection based only on visual inspection of the system and learner beliefs (Kerly and Bull, 2008). Motivation and user engagement enhancements have also been frequently noted (Baylor and Kim, 2005), (Heffernan, 2003), (Fryer and Carpenter, 2006). In some cases motivation may be actively supported through deliberate motivational tutoring techniques; in others it may be a useful by-product of exposure to a novel technique.

If motivational benefits are to be retained, then this novelty cannot be relied upon, and further research into deliberate scaffolding of affect may be required. A key feature of dialog systems in educative applications is the use of a natural communication method. The user of natural language allows users' cognitive resources to be spent on the learning task, rather than stretched by the interface (Beun et al., 2003). Computer literacy, and familiarity with online chatting media, is becoming ubiquitous and a greater number of users are expected to find conversing with their learning tools a feasible option.

This section has demonstrated the wide variety in conversational systems in e-learning. Implementations may employ full embodied dialog systems with emotion or gesture display, synthetic voice output, simple text-based output, dialog with an accompanying avatar, and many variants or combinations of these. Developers have integrated conversational capabilities into systems for a range of reasons.

4. EXAMPLES OF EDUCATIVE DIALOG SYSTEMS

In this section we describe two interactive multimodal interfaces that we have developed covering some of the issues described in the previous sections.

4.1 The *VoiceApp* Multimodal Dialog System

The *VoiceApp* system (Griol et al, 2011) has been developed as a framework for the study of the XHTML+Voice technology to develop multimodal dialog systems that improve the accessibility to information on the Internet (http://www.w3.org/TR/xhtml+voice/). The XML, XHTML and VoiceXML (http://www.w3.org/TR/voicexml20/) programming languages respectively deal with the visual design of the application and allow spoken dialog with the user. This way, multimodal interaction capabilities have been integrated for both the input and output of the system. The use of additional programming languages, as PHP and JavaScript, as well as relational database management systems such as MySQL, facilitates the incorporation of adaptive features and the dynamic generation of contents for the application.

Accessibility has been defined as one of the most important design requisites of the system. This way, detailed instructions, help messages, and menus have been also incorporated to facilitate the interaction with the different applications in the system. Previous interactions of the users are also taken into account to adapt the system, considering users' most used application, recent topics searched using the application, or errors detected after each interaction with the system.

In order to interact with the XHTML+Voice documents that make up the system, a web search engine supporting speech interaction and the specifications of this language is required. There are different models for implementing this multimodal interaction on mobile devices. The fat client model employs embedded speech recognition on the specific device and allows conducting speech processing locally. The thin client model involves speech processing on a portal server and is suitable for mobile phones. The implementation of the *VoiceApp* multimodal application for both computers and mobile devices is based on the fat client model, including a multimodal browser and embedded speech recognition on the corresponding device, and a web application server in which the system is stored.

The development of oral interfaces with XHTML+Voice implies the definition of grammars, which delimit the speech communication with the system. The <*grammar*> element is used to provide a speech or DTMF grammar that specifies a set of utterances that a user may speak to perform an action or supply information, and for a matching utterance, returns a corresponding semantic interpretation. We have defined a specific strategy to cover the widest range of search criteria in *VoiceApp* by means of the definition of speech recognition grammars in the different applications. This strategy is based on different aspects such as the dynamic generation of the grammars built from the results generated by the interaction with a specific application, the definition of grammars that includes complete sentences to support the naturalness of the interaction with the system (e.g., to facilitate a more natural communication and cover more functionalities in *Voice Pronunciation*), and the use of the ICAO phonetic alphabet (http://www.icao.int/icao/en/trivia/alphabet.htm) in the cases in which spelling of the words is required in order not to restrict the contents of the search or in situations in which repetitive recognition errors are detected (e.g., in order not to delimit the topics to search using Voice Browser).

The system consists of several modules to access web information. Firstly, the *Voice Dictionary* application offers a single environment where users can search contents in the Wikipedia encyclopedia with the main feature that the access to the application and the results provided by the search are entirely facilitated to the user either through visual modalities or by means of speech. Once the result of an initial search is displayed on the screen and communicated to the user by means of speech, they can easily access any of the links included in the result of the search or visit the rest of applications in the system with the possibility of interrupting the system's speech in any case. This functionality is achieved by means of the dynamic generation of the corresponding grammars, in which the different links that are present in the result of a specific search are included in the dynamic XHTML+Voice page automatically generated by means of a PHP script that captures the different information sources to inform the user about them (headings, text, contents, formulas, links, etc.). Figure 1 shows the initial page of this application.

Secondly, Google is currently one of the most important companies for the management of information on the Internet due to its web search engine and a number of applications and services developed to access information on the net. This way, the *Voice Browser* application has been developed with the main objective of allowing the speech access to facilitate both the search and presentation of the results in the interaction with the Google search engine. The application interface receives the contents provided by the user and displays the results both visually and using synthesized speech. The application also allows the multimodal selection of any of the links included in the result of the search by numbering them and allowing using their titles as voice commands (Figure 1).

Thirdly, the *Voice Pronunciation* application has been developed with the main objective of implementing a web environment that facilitates second-language learning with two games that help to acquire new vocabulary and train the words pronunciation. The game *Words* shows on the screen and synthesizes orally the definition of one of the over one hundred thousand words stored in a database of the application and the user must guess the word. The game *Pictures* uses images stored in a database and annotated with different difficulties, whose exact name must be correctly uttered by the user to continue in the game and increase the score (Figure 2). The specific problems and errors detected during the previous interactions of the users with this application are taken into account for the selection of the different words and images and to consequently adapt both games to the specific evolution of each user during the learning process.

A number of tests and verifications have been carried out to maximize the functionalities and accessibility of the different applications included in the *VoiceApp* system. These tests have been very important to detect and correct programming errors and accessibility problems. One of

Figure 1. Main page of the Voice Dictionary application and screen showing the result of a search using the Voice Browser application

the main identified problems was related to the generation of inconsistencies when words with similar pronunciation were reserved to both interact with by the Opera search engine and the different applications in the system. These inconsistencies have been limited to the maximum so that the possible matches between selected words have been eliminated in the different applications.

Figure 2. Main page of the Voice Pronunciation application and the Words functionality

Current research lines include the adaptation of the system for its interaction using additional languages, a more detailed assessment of each specific application, and the incorporation of new features in each one of them. Another important research line consists of the adaptation of the different applications taking into account specific user profiles considering more detailed information about their preferences and evolution.

4.2 The LEGA System

The LEGA system (*Learning English? Go Ahead!*) (Griol et al., 2012) has been developed to facilitate self-learning English for foreign students. The design of the system considered the required capacities defined by the Association of Language Testers of Europe (ALTE), the Common European Framework of Reference for Languages (CEFR, http://www.coe.int/t/dg4/linguistic/CADRE_EN.asp), and the corresponding equivalence with Cambridge ESOL examinations (Figure 3). This way, the system is focused on the skills and kind of exercises included in two of the most representative ESOL exams: First Certificate in English (FCE) and Certificate in Advanced English (CAE).

Through the access to the main page of the web-based application, users find a friendly system to visualize and complete different exercises and tests. The application provides access to these exercises by means of traditional input interfaces (i.e., mouse and keyboard), using tactile devices such as tablet-PCs or mobile phones, by means of speech, or even alternating both visual and speech modalities.

The architecture of the system includes three main elements. Dynamic web pages are used to display the different exercises and interact with users dynamically. They have been developed using the VoiceXML and PHP programming languages. As can be observed in Figures 4 and 5, the application shows the type of examination and selected block, the randomly selected exercise by

Figure 3. Relationship between the Common European Framework of Reference for Languages and the exams of Cambridge EOSL

<table>
<tr><th colspan="2">CEFR level</th><th colspan="3">Skills</th><th>Equivalence Cambridge EOSL</th></tr>
<tr><td rowspan="2">Basic User</td><td>A1</td><td colspan="3">Can understand and elaborate basic spoken and written structures.</td><td></td></tr>
<tr><td>A2</td><td colspan="3">Can understand and communicate sentences and frequently used expressions related to areas of most immediate relevance.</td><td>Key English Test (KET)</td></tr>
<tr><td rowspan="2">Independent User</td><td>B1</td><td colspan="3">Can understand the main standard inputs, deal with most situations in an area where the language is spoken, and produce simple connected text on topics which are familiar or of personal interest.</td><td>English Self Study Platform (PET)</td></tr>
<tr><td>B2</td><td colspan="3">Can understand the main ideas of complex topics, interact with a degree of fluency and spontaneity with native speakers, and produce clear, detailed text on a wide range of subjects.</td><td>First Certificate in English (FCE)</td></tr>
<tr><td rowspan="2">Proficient User</td><td>C1</td><td colspan="3">Can understand and produce a wide range of well-structured texts, express him/herself fluently and spontaneously, and use language flexibly and effectively for social, academic and professional purposes.</td><td>Certificate in Advanced English (CAE)</td></tr>
<tr><td>C2</td><td colspan="3">Can understand with ease virtually everything heard or read, and express him/herself spontaneously, very fluently and precisely in the most complex situations.</td><td>Certificate of Proficiency in English (CPE)</td></tr>
<tr><td></td><td>Listening</td><td>Reading</td><td>Speaking</td><td>Writing</td><td></td></tr>
</table>

the application within the block and corresponding subsection, instructions, accumulated mark until the current and time remaining to complete the exercise. Each time students provide an erroneous answer in any of the questions, they receive a feedback explaining the reasons why it is not the right answer for the described situation. The different controls embedded in the page code allow the selection of the response by touching the screen, using the keyboard and mouse, playback audio files in the listening exercises, and interaction through speech for the assessment of oral expression.

The Application Database stores the different exercises, organized according to the kind of examination and corresponding block and sections. To incorporate a new exercise in the application it is required to introduce the statement, set of possible answers, valid answer, route of possible external files and feedback to the student for each of the options. This structure allows easy incorporation of new content in the application and modification of existing without advanced knowledge of databases using the phpMyAdmin tool (http://www.phpmyadmin.net). Each exercise is numbered with a unique code.

Finally, the Users Database stores the specific information related to each one of the users registered in the application. This feature allows including in this database information on the previous interactions of each user (exercises that have been tried and mistakes made). This information is taken into account by the application for the selection among the possible exercises of each block and the generation of recommendations that take into account students specific skills and relevant exercises that they should emphasize.

In summary, the LEGA system has been developed after a detailed study of Common European Framework of Reference for Languages and the ESOL program of the University of Cambridge. The main objective of the system is the integration of different technologies that make possible to emulate the different exercises so that the interaction with the system is similar to the real conditions that students are going to find during

Figure 4. Visual exercise included in a FCE exam

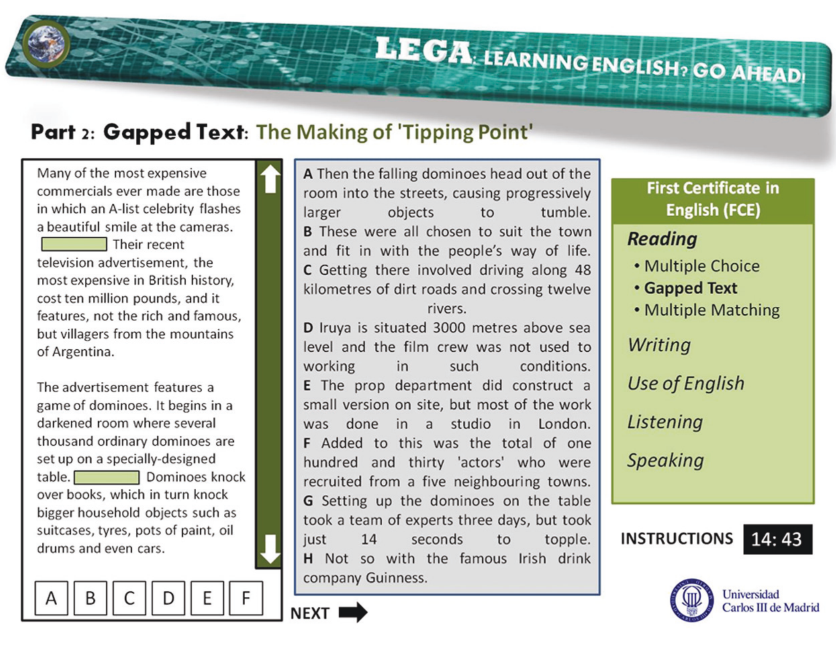

Figure 5. Oral exercise included in a CAE exam

the examination. Among these technologies, the VoiceXML language makes possible the spoken interaction with the system. In addition, it is possible to access the application by means of conventional computers or by means of mobile devices with the possibility of speech and/or tactile interaction. Finally, the system also considers the specific evolution of each student and the analysis of the errors to adapt the system by taking into account these important features. As future work we want to incorporate new contents and complete a detailed evaluation of each one of the modules of the system, as well as the extension of the number of functionalities related to the adaptation of the system.

5. CONCLUSION AND FUTURE RESEARCH DIRECTIONS

This chapter has explored the variety of conversational agent applications and techniques in the e-learning literature, identifying a variety of purposes and strategies including tutoring, language practice, learning companions, pedagogical agents, question answering and encouraging learner reflection. Reported benefits to learners include improvements in grades, motivation, engagement and metacognitive skills. Professors may also benefit from the ability of conversational systems to provide assessment, reporting and additional classroom tools. Through detailed examples we have described innovative uses for conversational agents in e-learning, and demonstrated their use as tools in a larger e-learning system and for the provision of support to parallel classroom activities. We have also explored a range of issues relating to the development of multimodal dialog systems, including questions regarding the design of conversational process and issues relating to technical implementation. We conclude that dialog systems have a valuable role to play in future e-learning and blended learning systems and we expect their use to become increasingly common and progressively more capable as this technology continues to develop. Finally, we have also described two systems developed using some of the innovative applications of multimodal dialogue systems described in the chapter.

REFERENCES

Ábalos, N., Espejo, G., López-Cózar, R., Callejas, Z., & Griol, D. (2010). A multimodal dialog system for an ambient intelligent application in home environments. *Lecture Notes in Computer Science*, *6231*, 491–498. doi:10.1007/978-3-642-15760-8_62

Ai, H., Litman, D., Forbes-Riley, K., Rotaru, M., Tetreault, J., & Purandare, A. (2006). Using systems and user performance features to improve emotion detection in spoken tutoring dialogs. *Proceedings of Interspeech'06-ICSLP*, (pp. 797–800).

Aimeur, E., Dufort, H., Leibu, D., & Frasson, C. (1997). Some Justifications for the Learning by Disturbing Strategy. *Proceedings of International Conference on Artificial Intelligence in Education*, (pp. 119-126).

Aleven, V., Ogan, A., Popescu, O., Torrey, C., & Koedinger, K. (2004). Evaluating the effectiveness of a tutorial dialog system for self-explanation. *Proceedings of International Conference on Intelligent Tutoring Systems*, (pp. 443-454).

Andeani, G., Fabbrizio, D. D., Gilbert, M., Gillick, D., Hakkani-Tur, D., & Lemon, O. (2006). Let's DISCOH: Collecting an annotated open corpus with dialogue acts and reward signals for natural language helpdesks. *Proceedings of IEEE Workshop on Spoken Language Technology (SLT'06)*, (pp. 218–221).

Angehrn, A. (2001). C-VIBE: A virtual interactive business environment addressing change management learning. *Proceedings of IEEE International Conference on Advanced Learning Technologies (ICALT'01)*, (pp. 174-177).

Angehrn, A., Nabeth, T., Razmerita, L., & Roda, C. (2001). K-InCA: Using artificial agents for helping people to learn new behaviours. *Proceedings of IEEE International Conference on Advanced Learning Technologies (ICALT'01)*, (pp. 225-226).

Augusto, J. (2009). Ambient intelligence: Opportunities and consequences of its use in smart classrooms. *Italics*, *8*(2), 53–63.

Bailly, G., Raidt, S., & Elisei, F. (2010). Gaze, dialog systems and face-to-face communication. *Speech Communication*, *52*(6), 598–612. doi:10.1016/j.specom.2010.02.015

Baylor, A. L., & Kim, Y. (2005). Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education*, *15*(2), 95–115.

Beun, R.-J., de Vos, E., & Witteman, C. (2003). Embodied dialog systems: Effects on memory performance and anthropomorphisation. *Proceedings of International Conference on Intelligent Virtual Agents*, (pp. 315-319).

Bickmore, T. (2003). *Relational agents: Effecting change through human-computer relationships*. PhD Thesis, Media Arts & Sciences, Massachusetts Institute of Technology.

Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction*, *12*(2), 293–327. doi:10.1145/1067860.1067867

Bohus, D., & Rudnicky, A. (2005). LARRI: A language-based maintenance and repair assistant. *Spoken Multimodal Human-Computer Dialogue in Mobile Environments*, *28*, 203–218. doi:10.1007/1-4020-3075-4_12

Bos, J., Klein, E., Lemon, O., & Oka, T. (1999). The verbmobil prototype system - A software engineering perspective. *Journal of Natural Language Engineering*, *5*(1), 95–112. doi:10.1017/S1351324999002132

Callejas, Z., & López-Cózar, R. (2008). Influence of contextual information in emotion annotation for spoken dialogue systems. *Speech Communication*, *50*(5), 416–433. doi:10.1016/j.specom.2008.01.001

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. F. (2001). *Embodied dialog systems*. The MIT Press.

Catizone, R., Setzer, A., & Wilks, Y. (2003). Multimodal dialog management in the COMIC Project. *Proceedings of EACL'03 Workshop on Dialog Systems*: *Interaction, Adaptation, and Styles of Management*, (pp. 25–34).

Cavazza, M., de la Camara, R. S., & Turunen, M. (2010). How was your day? A companion ECA. *Proceedings of AAMAS'10 Conference*, (pp. 1629-1630).

Chan, T.-W., & Baskin, A. B. (1988). Studying with the prince: The computer as a learning companion. *Proceedings of International Conference on Intelligent Tutoring Systems*, (pp. 194-200).

Chou, C.-Y., Chan, T.-W., & Lin, C.-J. (2003). Redefining the learning companion: The past, present and future of educational agents. *Computers & Education*, *40*, 255–269. doi:10.1016/S0360-1315(02)00130-6

Chu, S.-W., O'Neill, I., Hanna, P., & McTear, M. (2005). An approach to multistrategy dialog management. *Proceedings of Interspeech'05-Eurospeech*, Lisbon, Portugal, (pp. 865–868).

Cole, R., Mariani, J., Uszkoreit, H., Varile, G. B., Zaenen, A., Zampolli, A., & Zue, V. (Eds.). (1997). *Survey of the state of the art in human language technology*. Cambridge University Press.

Corradini, A., Mehta, M., Bernsen, N. O., & Charfuelán, M. (2005). Animating an interactive conversational character for an educational game system. *Proceedings of the International Conference on Intelligent User Interfaces*, (pp. 183–190).

Cuayáhuitl, H., Renals, S., Lemon, O., & Shimodaira, H. (2006). Reinforcement learning of dialog strategies with hierarchical abstract machines. *Proceedings of IEEE/ACL SLT'06 Workshop*, (pp. 182–186).

D'Mello, S. K., Craig, S. D., Gholson, B., Frankin, S., Picard, R., & Graesser, A. C. (2005). Integrating affect sensors in an intelligent tutoring system. *Proceedings of Workshop on Affective Interactions: The Computer in the Affective Loop at IUI*, (pp. 7-13).

De Carolis, B., Pelachaud, C., Poggi, I., & Steedman, M. (2003). APML, a markup language for believable behavior generation. In Prendinger, H., & Ishizuka, M. (Eds.), *Life-like characters: Tools, affective functions and applications*. Berlin, Germany: Springer.

de Rosis, F., Cavalluzzi, A., Mazzotta, I., & Novielli, N. (2005). Can embodied dialog systems induce empathy in users? *Proceedings of AISB'05 Virtual Social Characters Symposium*, (pp. 1-8).

Dehn, D. M., & van Mulken, S. (2000). The impact of animated interface agents: A review of empirical research. *International Journal of Human-Computer Studies*, *52*(1), 1–22. doi:10.1006/ijhc.1999.0325

Dillenbourg, P., & Self, J. (1992). *People power: A human-computer collaborative learning system in intelligent tutoring systems* (pp. 651–660). Berlin, Germany: Springer-Verlag.

Edlund, J., Gustafson, J., Heldner, M., & Hjalmarsson, A. (2008). Towards human-like spoken dialog systems. *Speech Communication*, *50*(8-9), 630–645. doi:10.1016/j.specom.2008.04.002

Elhadad, M., & Robin, J. (1996). An overview of SURGE: A reusable comprehensive syntactic realization component. *Proceedings of the Eighth International Natural Language Generation Workshop*, (pp. 1–4).

Felder, R. M., & Silverman, L. K. (1988). Learning and teaching styles in engineering education. *English Education*, *78*(7), 674–681.

Feng, D., Shaw, E., Kim, J., & Hovy, E. (2006). An intelligent discussion-bot for answering student queries in threaded discussions. *Proceedings of International Conference on Intelligent User Interfaces*, (pp. 171-177).

Forbus, K. D., Whalley, P. B., Evrett, J. O., Ureel, L., Brokowski, M., Baher, J., & Kuehne, S. E. (1999). CyclePad: An articulate virtual laboratory for engineering thermodynamics. *Artificial Intelligence*, *114*(1-2), 297–347. doi:10.1016/S0004-3702(99)00080-6

Fryer, L., & Carpenter, R. (2006). Bots as language learning tools. Language learning and technology. *Language Learning & Technology*, *10*(3), 8–14.

Glass, J., Flammia, G., Goodine, D., Phillips, M., Polifroni, J., & Sakai, S. (1995). Multilingual spoken-language understanding in the MIT Voyager system. *Speech Communication*, *17*, 1–18. doi:10.1016/0167-6393(95)00008-C

Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialog. *IEEE Transactions on Education*, *48*, 612–618. doi:10.1109/TE.2005.856149

Graesser, A. C., Person, N. K., & Harter, D. (2001). Teaching tactics and dialog in AutoTutor. *International Journal of Artificial Intelligence in Education*, *12*, 23–39.

Graesser, A. C., Wiemer-Hastings, K., Wiemer-Hastings, P., & Kreuz, R.TRG. (1999). AutoTutor: A simulation of a human tutor. *Journal of Cognitive Systems Research*, *1*, 35–51. doi:10.1016/S1389-0417(99)00005-4

Gratch, J., Rickel, J., Andre, J., Badler, N., Cassell, J., & Petajan, E. (2002). Creating interactive virtual humans: Some assembly required. *IEEE Intelligent Systems*, *17*(4), 54–63. doi:10.1109/MIS.2002.1024753

Grigoriadou, M., Tsaganou, G., & Cavoura, T. (2003). Dialog-based reflective system for historical text comprehension. *Proceedings of Workshop on Learner Modelling for Reflection at Artificial Intelligence in Education*.

Griol, D., Hurtado, L. F., Segarra, E., & Sanchis, E. (2008). A statistical approach to spoken dialog systems design and evaluation. *Speech Communication*, *50*(8-9), 666–682. doi:10.1016/j.specom.2008.04.001

Griol, D., Molina, J. M., Callejas, Z., & López-Cózar, R. (2012). (in press). Desarrollo de actividades de evaluación para un sistema on-line de aprendizaje de idiomas. *Revista Relada*, *6*.

Griol, D., Molina, J. M., & Corrales, V. (2011). Lecture Notes in Computer Science: *Vol. 7023*. *The VoiceApp System: Speech technologies to access the Semantic Web* (pp. 393–402). CAEPIA. doi:10.1007/978-3-642-25274-7_40

Gustafson, J., Elmberg, P., Carlson, R., & Jönsson, A. (1998). An educational dialogue system with a user controllable dialogue manager. *Proceedings of ICSLP*, *98*, 33–37.

Heffernan, N. T. (2003). Web-based evaluations showing both cognitive and motivational benefits of the Ms. Lindquist tutor. In Hoppe, U., Verdejo, F., & Kay, J. (Eds.), *Artificial intelligence in education* (pp. 115–122). Amsterdam, The Netherlands: IOS Press.

Hubal, R. C., Frank, G. A., & Guinn, C. I. (2000). AVATALK virtual humans for training with computer generated forces. *Proceedings of 9th Conference on Computer Generated Forces and Behavioral Representation*, Orlando, (pp. 617–623).

Jia, J. (2002). *The study of the application of a keywords-based chatbot system on the teaching of foreign languages*. University of Augsburg.

Johnson, W. L., LaBore, L., & Chiu, Y. C. (2004). A pedagogical agent for psychosocial intervention on a handheld computer. *Proceedings of AAAI Fall Symposium on Dialogue Systems for Health Communication*, (pp. 22-24).

Jokinen, K., Kanto, K., & Rissanen, J. (2004). Adaptative user modelling in AthosMail. *Lecture Notes in Computer Science*, *3196*, 149–158. doi:10.1007/978-3-540-30111-0_12

Kerly, A., Ellis, R., & Bull, S. (2008). Dialog systems in e-learning. *Proceedings of*, *AI-08*, 169–182.

Kerly, A., Ellis, R., & Bull, S. (2008). Children's interactions with inspectable and negotiated learner models. *Proceedings of International Conference on Intelligent Tutoring Systems*, (pp. 132-141).

Kerly, A., Ellis, R., & Bull, S. (2008). CALM-system: A dialog system for learner modelling. *Knowledge-Based Systems*, *21*(3), 238–246. doi:10.1016/j.knosys.2007.11.015

Kim, Y. (2007). Desirable characteristics of learning companions. *International Journal of Artificial Intelligence in Education*, *17*(4), 371–388.

Kumar, R., & Rose, C. P. (2011). Architecture for building dialog systems that support collaborative learning. *IEEE Transactions in Learning Technology*, *4*(1), 21–34. doi:10.1109/TLT.2010.41

Latham, A., Crockett, K. A., McLean, D., & Edmonds, B. A. (2012). Conversational intelligent tutoring system to automatically predict learning styles. *Computers & Education*, *59*(1), 95–109. doi:10.1016/j.compedu.2011.11.001

Lemon, O., Georgila, K., & Henderson, J. (2006). Evaluating Effectiveness and portability of reinforcement learned dialog strategies with real users: The TALK TownInfo evaluation. *Proceedings of IEEE-ACL SLT'06,* Palm Beach, Aruba, (pp. 178–181).

Lester, J. C., Stone, B. A., & Stelling, G. D. (1999). Lifelike pedagogical agents for mixed-initiative problem solving in constructivist learning environments. *User Modeling and User-Adapted Interaction*, *9*, 1–44. doi:10.1023/A:1008374607830

Litman, D. J., & Silliman, S. (2004). ITSPOKE: An intelligent tutoring spoken dialog system. *Proceedings of Human Language Technology Conference: North American Chapter of the Association for Computational Linguistics*, (pp. 5-8).

Liu, Y., & Chee, Y. S. (2004). Designing interaction models in a multiparty 3D learning environment. *Proceedings of 12th International Conference on Computers in Education (ICCE 2004)*, (pp. 293–302).

López-Cózar, R., & Araki, M. (2005). *Spoken, multilingual and multimodal dialog systems: Development and assessment*. John Wiley and Sons.

Mairesse, F., Gasic, M., Jurcícek, F., Keizer, S., Thomson, B., Yu, K., & Young, S. (2009). Spoken language understanding from unaligned data using discriminative classification models. *Proceedings of ICASSP*, *09*, 4749–4752.

Manzoni, J. F., & Angehrn, A. (1997). Understanding organizational dynamics of IT-enabled change: A multimedia simulation approach. *Journal of Management Information Systems*, *14*(3), 109–140.

Marsella, S. C., Johnson, W. L., & LaBore, C. M. (2003). Interactive pedagogical drama for health interventions. In Hoppe, I. U. (Eds.), *Artificial intelligence in education: Shaping the future of learning through intelligent technologies* (pp. 341–348). Amsterdam, The Netherlands: IOS Press.

Mattasoni, M., Omologo, M., Santarelli, A., & Svaizer, P. (2002). On the joint use of noise reduction and MLLR adaptation for in-car hands-free speech recognition. *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP'02)*, (pp. 289–292).

McTear, M. F. (2004). *Spoken dialog technology*. Springer. doi:10.1007/978-0-85729-414-2

Menezes, P., Lerasle, F., Dias, J., & Germa, T. (2007). Towards an interactive humanoid companion with visual tracking modalities. *International Journal of Advanced Robotic Systems*, 48–78.

Meza, I., Riedel, S., & Lemon, O. (2008). Accurate statistical spoken language understanding from limited development resources. *Proceedings of ICASSP*, *08*, 5021–5024.

Minker, W. (1998). Stochastic versus rule-based speech understanding for information retrieval. *Speech Communication*, *25*(4), 223–247. doi:10.1016/S0167-6393(98)00038-7

Montoro, G., Haya, P. A., Alamán, X., López-Cózar, R., & Callejas, Z. (2006). A proposal for an XML definition of a dynamic spoken interface for ambient intelligence. *Proceedings of International Conference on Intelligent Computing (ICIC'06)*, (pp. 711–716).

Mostow, J. (2012). Why and how our automated reading tutor listens. *Proceedings of International Symposium on Automatic Detection of Errors in Pronunciation Training (ISADEPT)*, (pp. 43-52).

Mühlpfordt, M., & Wessner, M. (2005). Explicit Referencing in chat supports collaborative learning. *Proceedings of 6th International Conference on Computer-Supported Collaborative Learning* (CSCL'05), (pp. 662-671).

Oh, A., & Rudnicky, A. (2000). Stochastic language generation for spoken dialog systems. *Proceedings of ANLP/NAACL Workshop on Conversational Systems*, (pp. 27–32).

Pon-Barry, H., Schultz, K., Bratt, E. O., Clark, B., & Peters, S. (2006). Responding to student uncertainty in spoken tutorial dialog systems. *International Journal of Artificial Intelligence in Education*, *16*, 171–194.

Rabiner, L. R., & Juang, B. H. (1993). *Fundamentals of speech recognition*. Prentice-Hall.

Rabiner, L. R., Juang, B. H., & Lee, C. H. (1996). An overview of automatic speech recognition. In Lee, C. H., Soong, F. K., & Paliwal, K. K. (Eds.), *Automatic speech and speaker recognition: Advanced topics* (pp. 1–30). Kluwer Academic Publishers. doi:10.1007/978-1-4613-1367-0_1

Reiter, E. (1995). NLG vs. templates. *Proceedings of the Fifth European Workshop in Natural Language Generation*, (pp. 95–105).

Rickel, J., & Johnson, W. L. (1999). Animated agents for procedural training in virtual reality: Perception, cognition, and motor control. *Applied Artificial Intelligence*, *13*, 343–382. doi:10.1080/088395199117315

Roda, C., Angehrn, A., & Nabeth, T. (2001). Dialog systems for advanced learning: Applications and research. *Proceedings of BotShow'01 Conference*, (pp. 1-7).

Roda, C., Angehrn, A., & Nabeth, T. (2001). Matching competencies to enhance organizational knowledge sharing: An intelligent agents approach. *Proceedings of 7th International Netties Conference*, (pp. 931-937).

Rosé, C. P., Moore, J. D., VanLehn, K., & Allbritton, D. (2001). A comparative evaluation of Socratic versus didactic tutoring. *Proceedings of Cognitive Sciences Society*.

Shaw, E., Johnson, W. L., & Ganeshan, R. (1999). Pedagogical agents on the Web. *Proceedings of International Conference on Autonomous Agents,* (pp. 283-290). ACM Press.

Taylor, K., & Moore, S. (2006). Adding question answering to an e-tutor for programming languages. *Proceedings of 26th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence*, (pp. 193-206). Cambridge, UK: Springer.

Vaquero, C., Saz, O., Lleida, E., Marcos, J., & Canalís, C. (2006). VOCALIZA: An application for computer-aided speech therapy in Spanish language. *Proceedings of IV Jornadas en Tecnología del Habla*, (pp. 321–326).

Wahlster, W. (Ed.). (2006). *SmartKom: Foundations of multimodal dialog systems*. Springer. doi:10.1007/3-540-36678-4

Wang, N., & Johnson, L. W. (2008). The politeness effect in an intelligent foreign language tutoring system. *Proceedings of 9th International Conference Intelligent Tutoring Systems,* (pp. 70-280).

Wang, Y., Wang, W., & Huang, C. (2007). Enhanced semantic question answering system for e-learning environment. *Proceedings of AINAW'07 Conference*, (pp. 1023-1028).

Weng, F., Varges, S., Raghunathan, B., Ratiu, F., Pon-Barry, H., & Lathrop, B. … Shriberg, L. (2006). CHAT: A conversational helper for automotive tasks. *Proceedings of the 9th International Conference on Spoken Language Processing (Interspeech-ICSLP)*, (pp. 1061–1064).

Williams, J., & Young, S. (2007). Partially observable Markov decision processes for spoken dialog systems. *Computer Speech & Language*, *21*(2), 393–422. doi:10.1016/j.csl.2006.06.008

Zue, V. W., & Glass, J. R. (2000). Conversational interfaces: Advances and challenges. *Proceedings of the IEEE*, *88*, 1166–1180. doi:10.1109/5.880078

ADDITIONAL READING

Bezold, M., & Minker, W. (2011). *Adaptive multimodal interactive systems*. Springer. doi:10.1007/978-1-4419-9710-4

Clark, B., Lemon, O., Gruenstein, A., Bratt, E., Fry, J., & Peters, S. … Treeratpituk, P. (2005), A general purpose architecture for intelligent tutoring systems. *Advances in Natural Multimodal Dialogue Systems*, (pp. 287–305).

Graesser, A. C., VanLehn, K., Rosé, C., Jordan, P. W., & Harter, D. (2001). Intelligent tutoring systems with conversational dialogue. *AI Magazine*, *22*(4), 39–51.

Grifoni, P. (Ed.). (2009). *Multimodal human-computer interaction and pervasive services*. Hershey, PA: IGI Global. doi:10.4018/978-1-60566-386-9

Jokinen, K. (2009). *Constructive dialog modelling: Speech interaction and rational agents*. Wiley.

Kurkovsky, S. (Ed.). (2009). *Multimodality in mobile computing and mobile devices: Methods for adaptable usability*. Hershey, PA: IGI Global. doi:10.4018/978-1-60566-978-6

Macías, J. A., Granollers, A., & Latorre, P. M. (Eds.). (2009). *New trends on human-computer interaction*. Springer. doi:10.1007/978-1-84882-352-5

Maragos, P., Potamianos, A., & Graos, P. (Eds.). (2010). *Multimodal processing and interaction: Audio, video, text*. Springer.

Tzovaras, D. (Ed.). (2010). *Multimodal user interfaces: From signals to interaction*. Springer.

Veletsianos, G., Heller, R., Overmyer, S., & Procter, M. (2010). Conversational agents in virtual worlds: Bridging disciplines. *British Journal of Educational Technology*, *41*(1), 123–140. doi:10.1111/j.1467-8535.2009.01027.x

KEY TERMS AND DEFINITIONS

Automatic Speech Recognition (ASR): Technique to determine the word sequence in a speech signal. To do this, this technology first detects basic units in the signal, e.g. phonemes, which are then combined to determine words.

Dialog Management (DM): Implementation of the "intelligent" behaviour of the conversational system. It receives some sort of internal representation obtained from the user input and decides the next action the system must carry out.

Multimodal Dialog System: Computer program that emulates a dialog between two human beings and process two or more combined user input modes such as speech, pen, touch, manual gestures, gaze, and head and body movement in a coordinated manner with multimedia system output.

Natural Language Generation (NLG): Creation of messages in text mode, grammatical and semantically correct, which will be either displayed on screen or converted into speech by means of text-to-speech synthesis.

Spoken Language Understanding (SLU): Technique to obtain the semantic content of the sequence of words provided by the ASR module. It must face a variety of phenomena, for example, ellipsis, anaphora and ungrammatical structures typical of spontaneous speech.

Speech Synthesis: Artificial generation of human-like speech. A particular kind of speech synthesis technique is called Text-To-Speech synthesis (TTS), the goal of which is to transform into speech of any input sentence in text format.

VoiceXML: Standard XML-based language to access web applications by means of speech.

XHTML+Voice (X+V): XML-based language that combines traditional web access using XHTML and speech-based access to web pages using VoiceXML.